The statistical errors in the estimated correlation function matrix for operational modal analysis

Tarpø, Marius; Friis, Tobias; Georgakis, Christos T.; Brincker, Rune

Published in:
Journal of Sound and Vibration

Link to article, DOI:
[10.1016/j.jsv.2019.115013](https://doi.org/10.1016/j.jsv.2019.115013)

Publication date:
2020

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Tarpø, M., Friis, T., Georgakis, C. T., & Brincker, R. (2020). The statistical errors in the estimated correlation function matrix for operational modal analysis. *Journal of Sound and Vibration*, 466, [115013].
<https://doi.org/10.1016/j.jsv.2019.115013>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The statistical errors in the estimated correlation function matrix for operational modal analysis

Marius Tarpø^{a,*}, Tobias Friis^b, Christos Georgakis^a, Rune Brincker^b

^a*Aarhus University, Department of Engineering, Inge Lehmanns Gade 10, Aarhus, Denmark*

^b*Technical University of Denmark, Department of Civil Engineering, Brovej B.118, Kgs. Lyngby, Denmark*

Abstract

Given the random vibration of a linear and time-invariant system, the correlation function matrix is equivalent to free decays when the system is excited by Gaussian white noise. Correlation-driven Operational Modal Analysis utilises these properties to identify modal parameters from systems in operation based on the response only. Due to the finite length of the system response, the correlation function matrix must be estimated and this introduces statistical errors. This article focuses on the statistical errors due to this estimation process and the effect it has on the envelope and zero crossings of the estimated correlation function matrix. It is proven that the estimated correlation function matrix is a Gaussian stochastic process. Furthermore, it is proven that the envelope of the modal correlation function matrix is Rice distributed. This causes the tail region of the correlation function to become erroneous - called the noise tail. The zero crossings are unbiased, but the random error related to the crossings increases fast in the noise tail. The theory is tested on a simulated case and there is a high agreement between theory and simulation. A new expression for the minimal time length is introduced based on the bias error on the envelope.

Keywords: Uncertainty, operational modal analysis, envelope, zero crossings, Rice distribution

1. Introduction

In operational modal analysis or output-only modal analysis, we use the random response of a system in its ambient environment to identify the modal parameters [1, 2]. A common practice is the use of a correlation function matrix in a two-stage modal identification process - the correlation-driven operational modal analysis. Firstly, we calculate the correlation function matrix of the measured random system response. Secondly, we apply an identification technique to identify the modal parameters since we interpret the correlation function matrix as transposed free decays of the linear and time-invariant system, assuming white Gaussian noise as excitation [3, 4]. This is the idealised and exact statistical properties of the random system response, and it is based on a system response with an infinite time length. Exact properties of random data, however, are inaccessible from sampled data and they must be estimated instead. Thus, in operational modal analysis, we calculate an estimated correlation function matrix based on a measured system response with a finite time length. This introduces statistical errors into the estimated correlation function matrix [5, 6]. Hence, in practice, the statistical properties of the estimated correlation function are a stochastic process.

In the literature, the statistical errors in the correlation function matrix are system and time-dependent [5, 6]. Furthermore, it is well-known that the statistical errors result in an erroneous and erratic tail region of the correlation function matrix - the noise tail [6–8]. Whereas the beginning of the estimated correlation matrix is more persistent despite the statistical errors, generally, the length of the "persistent" properties

*Corresponding author

Email address: martar@eng.au.dk (Marius Tarpø)

is called the "*correlation time*" of the correlation function. The noise tail causes biased errors in the identification of the modal parameters [8] and it is common practice to disregard this region and utilise the correlation time of the correlation function matrix in an identification process. To the best of the authors' knowledge, however, the theory behind this erratic tail region is scarcely studied in the literature. Bendat and Piersol [5] show the introduction of statistical errors in the estimation of a correlation function. Orlowitz and Brandt [9] study the random errors in the estimated correlation function in operational modal analysis numerically. Giampellegrini [6] derives an analytic expression of the random error in the auto-correlation function of a single-of-degree (SDOF) system that illustrates the system and time dependency of the random error. Tarpø et al. [7] introduce an algorithm for the automatic detection of the correlation time or the noise tail. Furthermore, Tarpø et al. [10] introduce an algorithm for reduction of statistical errors by applying windows to the noise tail of each modal auto-correlation function. They found that the estimated damping ratios were biased due to the statistical errors and a reduction of the statistical errors reduces the biased and random errors.

To account for the statistical errors, different studies exist on the optimal or required time length of measured system response [8, 11] since the time length relates to the statistical errors in the identification of modal parameters where an increase of time length decreases the statistical errors [5]. Damping estimates are often biased, but an increase in the time length results in a decrease in both random and biased errors in the damping estimates. Generally, different recommended time lengths exist for operational modal analysis [1, 2, 5, 12–14]. Most of these are based on calculation of the spectral density function [2, 5] while others are based on experience. [14] sets up a rule of thumb saying the time length should be 1,000 to 2,000 times the fundamental period.

Another approach is uncertainty quantifications that set up uncertainty bonds for the estimated modal parameters [15, 16]. For Bayesian operational modal analysis, the analytic expressions of the uncertainty for modal parameters are calculated for a frequency approach [12, 17].

This article will study the statistical errors in the estimated correlation function matrix used in correlation-driven operational modal analysis, and it adds novel information to the field. To illustrate the effect of statistical errors in this article, we will use two different recommended time lengths from Brincker and Ventura [1] & ANSI 2.47 [13]. This article will show that the length of the correlation function matrix - the correlation time - is an equally important parameter as the time length. To access the statistical errors of the estimated correlation function matrix, we will base this article on the following assumptions:

- The system is linear and time-invariant
- The excitation is stationary white Gaussian noise
- The system is lightly damped with orthogonal modes
- The time length of the measured system response is long compared to the time lags of the estimated correlation function matrix

In this article, the term "*statistical errors*" includes all random and biased errors, which affect the estimated correlation function matrix due to the estimation process. Thus, we exclude the effect of measurement noise on the correlation function matrix in this study, see [18] for more on this subject. We assess the random errors by the variance and any biased errors by the difference between the analytic and the expected value [5]. The term "*statistical errors*", however, exclude all errors in the identified modal parameter caused by any identification techniques. Likewise, we disregard eventual errors caused by violations of the core assumptions of operational modal analysis - linear and time-invariant system excited by white Gaussian noise.

We organise the remainder of this article in the following manner. In Section 2, we derive an expression for the random errors in the estimated correlation function matrix. Section 3 proves that the modal auto-correlation function has a biased envelope by applying the theory of peak probability density functions. Section 5 presents a new expression of the required time length based on the statistical errors in the correlation function matrix. Finally, in Section 6, we validate the analytic expressions of errors in the correlation function matrix from the previous sections by a simulation study.

2. Statistical errors in the correlation function matrix

In this section, we will go through the general statistical errors in the correlation function matrix. We will show that we can express the statistical errors as system dependent errors in each modal auto-correlation function.

2.1. Correlation function matrix

In operational modal analysis, we consider random vibrations from linear systems that are excited by white Gaussian noise. We measure the spatial limited system response, $\mathbf{y}(t)$, at a finite number degrees-of-freedom. In the correlation-driven operational modal analysis, the correlation function matrix, $\mathbf{R}(\tau)$, for the system response is equivalent to free decays corresponding to the modal parameters of the system [3, 4].

$$\mathbf{R}(\tau) \triangleq \mathbb{E}[\mathbf{y}(t)\mathbf{y}^\top(t+\tau)] \quad (1)$$

where $(\cdot)^\top$ denotes the transpose and τ is the time lag.

We can use the modal decomposition to transform the system response into the modal domain.

$$\begin{aligned} \mathbf{y}(t) &= \sum_{i=1}^N \boldsymbol{\Phi}_i q_i(t) \\ &= \boldsymbol{\Phi} \mathbf{q}(t) \end{aligned} \quad (2)$$

where $\boldsymbol{\Phi}_i$ is the mode shape for the i^{th} mode, $q_i(t)$ is the modal coordinates for the i^{th} mode, N is the total number of modes in the system, $\boldsymbol{\Phi}$ is the modal matrix containing mode shapes as columns, and $\mathbf{q}(t)$ is the modal coordinate vector with the modal coordinates in the rows.

We use the modal decomposition to decorrelate the correlation function matrix into the modal correlation function matrix by inserting Eq. (2) in Eq. (1) [1].

$$\mathbf{R}(\tau) = \boldsymbol{\Phi} \mathbf{R}_q(\tau) \boldsymbol{\Phi}^\top \quad (3)$$

where $\mathbf{R}_q(\tau)$ is the modal correlation function matrix. Each modal auto-correlation function corresponds to a single-degree-of-freedom (SDOF) system and a free decay of the given mode [19]. When white Gaussian noise excites a linear and time-invariant system with lightly damped and orthogonal modes, the modal coordinates are independent and the modal correlation function matrix is diagonal [1, 19]. Thus, we rewrite the equation of the correlation function matrix as a summation of the modal auto-correlation function.

$$\mathbf{R}(\tau) \approx \sum_{i=1}^N \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^\top R_{q_i}(\tau) \quad (4)$$

where $R_{q_i}(\tau)$ is the modal auto-correlation function for the i^{th} mode.

Clough and Penzien [19] made an expression for the correlation function for an SDOF system excited by white Gaussian noise that we will use to describe the modal auto-correlation functions analytically.

$$R_{q_i}(\tau) = \frac{\pi \omega_0 S_0}{2k^2 \zeta} e^{-\omega_0 \zeta |\tau|} \left(\cos(\omega_d \tau) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\omega_d |\tau|) \right) \quad (5)$$

where ω_0 is natural cyclic frequency, ω_d is natural damped cyclic frequency, ζ is the damping ratio, S_0 is the loading-constant from the white Gaussian noise in the frequency domain, and k is the modal stiffness of the i^{th} mode. We omit the index for the given mode on all modal parameters to improve readability of the equations. Based on this equation, we have an expression for the variance of the modal coordinates.

$$\sigma_{q_i}^2 = \frac{\pi \omega_0 S_0}{2k^2 \zeta} \quad (6)$$

The Hilbert transformation is often used to find the (Hilbert) envelope by the absolute value of the analytic signal [20].

Figure 1: Correlogram with two realisations of the same estimated correlation function (*black solid line*) and the corresponding theoretical correlation function, Eq. (5), (*red dashed line*)

$$\text{env}(X(t)) = \sqrt{X(t)^2 + \mathcal{H}[X(t)]^2} \quad (7)$$

where $X(t)$ is an arbitrary function and $\mathcal{H}$ denotes the Hilbert transformation.

We calculate the envelope of the modal auto-correlation function, Eq. (5), using the Hilbert transformation.

$$e_{q_i}(\tau) = \text{env}(R_{q_i}(\tau)) = \sigma_{q_i}^2 e^{-\omega_0 \zeta |\tau|} \sqrt{1 + \frac{\zeta^2}{1 - \zeta^2}} \quad (8)$$

For simplicity, we will assume that the system has low damping.

$$e_{q_i}(\tau) = \sigma_{q_i}^2 e^{-\omega_0 \zeta |\tau|} \quad (9)$$

2.2. Estimated correlation function matrix

Since the system response is ergodic, we use *time averaging* instead of *ensemble averaging*. The length of the signal, however, must tend towards infinite if the *time averaging* should equal the expected value [5].

$$\mathbf{R}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{y}(t) \mathbf{y}^\top(t + \tau) dt \quad (10)$$

where T is the time length of the measured system response. The measured system response has a finite time length, meaning, we have to estimate the correlation function matrix and this introduces statistical errors. This is illustrated in Fig. 1 where two estimated correlation functions are plotted next to the theoretical correlation function. Therefore, the estimated correlation function matrix is a stochastic process.

Different estimators exist for the correlation function [5]. The unbiased estimator has a higher variance as we increase the time lag, τ , whereas the biased error of the biased estimator increases with increasing time lags. The biased estimator has a biased envelope and this biased error transfers into the identified damping ratios. In operational modal analysis, however, the time length is often longer than the maximum time lag, $T \gg \tau$, thus the difference between the estimators is small. Since we use the correlation function matrix to identify modal parameters, then an unbiased estimator is preferable.

$$\tilde{\mathbf{R}}(\tau) = \frac{1}{T - \tau} \int_0^{T - \tau} \mathbf{y}(t) \mathbf{y}^\top(t + \tau) dt, \quad 0 \leq \tau < T \quad (11)$$

By increasing the time length of the recording, the variance of the estimated correlation function matrix decreases.

$$\lim_{T \rightarrow \infty} \tilde{\mathbf{R}}(\tau) \rightarrow \mathbf{R}(\tau) \quad (12)$$

2.3. Probability density function of the estimated correlation function matrix

Next, we will find the probability density function of the estimated correlation function matrix. We consider a linear and stationary multi-degree-of-freedom system excited by white Gaussian noise. In linear structural dynamics, we apply the modal decorrelation/modal superposition, Eq. (2), to decorrelate this system into multiple SDOF systems describing each mode [19]. The response of each SDOF system is a convolution of the excitation and the impulse response function of the given mode [19]. The (convolution) integral is defined as the limit of a summation and - according to the Central Limit Theorem - the summation of many random variables tends toward a normal distribution [1, 5]. Thus, the modal coordinates are normal distributed, $q_i(t) \sim \mathcal{N}(0, \sigma_{q_i}^2)$, if the time length, T , is large, $T \gg 0$, see Appendix A. The modal decomposition is a linear transformation of normal random vectors [21] so the system response is a multivariate Gaussian variable, $\mathbf{y}(t) \sim \mathcal{N}(\mathbf{0}, \Phi[\sigma_{q_i}^2]\Phi^\top)$ where $\mathbf{0}$ is a zero vector [1, 5].

The estimated modal auto-correlation functions are approximately Gaussian distributed, $\tilde{R}_{q_i}(\tau) \sim \mathcal{N}(R_{q_i}(\tau), \text{Var}[\tilde{R}_{q_i}(\tau)])$, see Appendix A. Therefore, the estimated correlation function matrix follows a multivariate Gaussian distribution, $\tilde{\mathbf{R}}(\tau) \sim \mathcal{N}(\mathbf{R}(\tau), \text{Cov}[\tilde{\mathbf{R}}(\tau)])$, see Appendix B. Thus, the expected value of the estimated correlation function matrix is given by

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{R}}(\tau)] &= \mathbf{R}(\tau) \\ &\approx \sum_{i=1}^N \Phi_i \Phi_i^\top R_{q_i}(\tau) \end{aligned} \quad (13)$$

Let us turn to the variance - or statistical error - of each entry of the correlation function matrix using Eq. (4).

$$\text{Var}[\tilde{\mathbf{R}}_{j,k}(\tau)] \approx \text{Var}\left[\sum_{i=1}^N \Phi_{j,i} \Phi_{k,i} \tilde{R}_{q_i}(\tau)\right] \quad (14)$$

We can express the variance as a summation of the covariances of all products in the summation.

$$\begin{aligned} \text{Var}[\tilde{\mathbf{R}}_{j,k}(\tau)] &= \sum_{i=1}^N (\Phi_{j,i} \Phi_{k,i})^2 \text{Var}[\tilde{R}_{q_i}(\tau)] + \sum_{i=1}^N \sum_{\substack{m=1 \\ m \neq i}}^N \Phi_{j,i} \Phi_{k,i} \Phi_{j,m} \Phi_{k,m} \text{Cov}[\tilde{R}_{q_i}(\tau), \tilde{R}_{q_m}(\tau)] \end{aligned} \quad (15)$$

When we are dealing with orthogonal modes, the covariance between modal auto-correlation functions is much smaller than the variance of each modal auto-correlation function. This enables us to approximate the statistical errors in the entire matrix as a summation of the variance of each modal auto-correlation function (a similar expression is derived in [6]).

$$\text{Var}[\tilde{\mathbf{R}}_{j,k}(\tau)] \approx \sum_{i=1}^N (\Phi_{j,i} \Phi_{k,i})^2 \text{Var}[\tilde{R}_{q_i}(\tau)] \quad (16)$$

Each product of mode shapes, $(\Phi_{j,i} \Phi_{k,i})^2$, is a linear transformation from the variance of the modal auto-correlation function to the contribution of the i^{th} modes' contribution of variance of the estimated correlation function matrix. Thus, the mode shapes distribute the statistical errors from the modal auto-correlation functions to the entire correlation function matrix. The statistical errors exist within the vector space spanned by the mode shapes of the system and this vector space of the estimated correlation function matrix is independent of the statistical errors, at least for systems with low damping and orthogonal modes. Thus, in correlation-driven operational modal analysis, an identification of the mode shapes is almost independent of the statistical errors.

Figure 2: Correlogram with probability density function of the estimated correlation function for an arbitrary system plotted for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13], the expected value (*black solid line*) and the interquartile range (*grey*)

To express the statistical errors in the estimated correlation function matrix, we have to derive an expression for the variance of the estimated modal auto-correlation functions. Bendat and Piersol [5] derived an expression for the variance of the cross-correlation function by assuming Gaussian distributed data as input in the correlation function matrix - here written for the unbiased auto-correlation function for the modal coordinates.

$$\text{Var} [\widetilde{R}_{q_i}(\tau)] = \frac{1}{T - \tau} \int_{-T+\tau}^{T-\tau} \left(1 - \frac{|r|}{T - \tau}\right) (R_{q_i}(r)^2 + R_{q_i}(r + \tau)R_{q_i}(r - \tau)) dr \quad (17)$$

Bendat and Piersol [5] further simplified the variance of the estimated correlation function by assuming a large record length, $T \gg 0$.

$$\text{Var} [\widetilde{R}_{q_i}(\tau)] \approx \frac{1}{T - \tau} \int_{-\infty}^{\infty} (R_{q_i}^2(r) + R_{q_i}(r + \tau)R_{q_i}(r - \tau)) dr \quad (18)$$

Giampellegrini [6] calculated the variance of the estimated correlation functions using this integral, Eq. (18), by assuming small damping.

$$\text{Var} [\widetilde{R}_{q_i}(\tau)] \approx \frac{\sigma_{q_i}^4}{2(T - \tau)\omega_0\zeta} \left(1 + e^{-2\omega_0\zeta\tau} \left(\cos(2\omega_d\tau)(1 + 2\zeta\omega_0\tau) - \sin(2\omega_d\tau)\frac{4\omega_0\zeta^2\tau}{\sqrt{1 - \zeta^2}}\right)\right) \quad (19)$$

Thus, the statistical errors are non-stationary while they are system and time-dependent. By increasing the time length, T , the statistical errors decrease. Figs. 2 and 3 illustrate this time dependence. The statistical errors tend to become approximately stationary as the time lags increase, see Fig. 3.

The minima of the variance of the correlation function give the variance at zero crossings or axis crossings and the maxima give the variance at the peaks [6], see Fig. 4. We want to find the envelopes of the variance, Eq. (18), since this converts the variance to either the zero crossings or the peaks of the modal auto-correlation function, similarly to an interpolation. By assuming small damping and removing all terms with ζ to the power of minimum 2, the variance at the peaks and zero crossings of the estimated correlation function is given by

$$\text{Var} [\widetilde{R}_{q_i}(\tau)]^{\pm} \approx \sigma_{q_i}^4 \frac{1 \pm e^{-2\omega_0\zeta\tau}(2\omega_0\zeta\tau + 1)}{2(T - \tau)\omega_0\zeta} \quad (20)$$

where $\pm$ denotes the upper (+) or lower (-) envelope of the variance from Eq. (18).

Figure 3: Variance of the estimated correlation function, Eq. (19), for an arbitrary system plotted for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13]

Figure 4: The relationship between the analytic correlation function, Eq. (5), (*top*) and the variance of the estimated correlation function, Eq. (19), (*bottom*)

Figure 5: The variance of the estimated correlation function for two arbitrary systems with the upper and lower envelopes, the variance, Eq. (19), (black solid line), upper envelope, Eq. (20), (red dashed line), and lower envelope, Eq. (20), (blue solid line)

Fig. 5 illustrates the variance and its envelopes for two SDOF systems. For the second SDOF system, we double the frequency-damping product, $f_0\zeta$, of the first system. Generally, the variance drops for the second system and it decays faster. Hence, for multi-degree-of-freedom systems, the mode with the lowest frequency-damping product has the highest level of statistical errors.

On their own, these expressions of variance do not explain the erratic behaviour in the tail region of the estimated correlation function, as seen in Fig. 7. We have to apply the theory of Peak Probability Density Functions and Level Crossing to obtain an expression of the density function of the envelope, see section 3, and the zero crossings, see section 4.

2.4. Error of the estimated modal auto-correlation functions

In this section, we will study the random error of the modal auto-correlation function, Eq. (20). Two special cases of this equation are worth mentioning. First, for variance at time lag zero, $\tau = 0$, the random error becomes

$$\text{Var} [\widetilde{R}_{qi}(0)]^+ \approx \sigma_{qi}^4 \frac{1}{T\omega_0\zeta} \quad , \quad \text{Var} [\widetilde{R}_{qi}(0)]^- \approx 0 \quad (21)$$

While for large time lags, $\tau \gg 0$, and time length much larger than the time lag, $T \gg \tau$, the random error is

$$\text{Var} [\widetilde{R}_{qi}(\tau)]^\pm \approx \sigma_{qi}^4 \frac{1}{2(T-\tau)\omega_0\zeta}, \quad \tau \gg 0, \quad T \gg \tau \quad (22)$$

For both cases, we decrease the random error by increasing the frequency-damping product or the time length, see Fig. 5.

In the literature, the normalised random error, $\frac{\sigma}{\mu}$, is often used. This is inversely proportional to the signal-to-noise ratio, $\frac{\mu}{\sigma}$.

$$\begin{aligned} \epsilon(\tau)^\pm &= \frac{\sqrt{\text{Var} [\widetilde{R}_{qi}(\tau)]^\pm}}{e_{qi}(\tau)} \\ &= \sqrt{\frac{1}{T-\tau} \frac{e^{2\omega_0\zeta\tau} \pm (2\omega_0\zeta\tau + 1)}{2\omega_0\zeta}} \end{aligned} \quad (23)$$

This is identical to the results in [6], who looked at the normalised random error by converting the envelope of the variance function for the biased auto-correlation function, $\tau = n \frac{\pi}{\omega_d}$, where n is the discrete number of each peak of the modal auto-correlation function.

Figure 6: The normalised random error of the estimated correlation function, Eq. (23), for two arbitrary systems, normalised random error for the upper envelope, $\epsilon(\tau)^+$, (red dashed line) and normalised random error for the lower envelope, $\epsilon(\tau)^-$, (blue solid line)

Let us have a look at Eq. (23). Let us say that we increase the frequency-damping product, then we increase the slope of the normalised random error. Thus, the normalised random error of modes with high frequency-damping products increases rapidly compared to modes with low frequency-damping products, see Fig. 6. This is, however, due to the normalisation process. The random errors are inversely proportional to the frequency-damping product and the time length, see Eqs. (20), (21), and (22). Thus, modes with low frequency-damping products have higher random errors than modes with higher frequency-damping products, as seen in Fig. 5.

This concludes this section. So far we have shown that the finite time length of measurement introduces statistical errors to the correlation function matrix. It was shown that the non-stationary statistical errors are system and time-dependent, and they are introduced in each modal auto-correlation function (for lightly damped and orthogonal modes). The auto-correlation function for the statistical errors and expression for the variance (statistical errors) are presented in the next section.

3. Statistical errors in the envelope of the modal auto-correlation function

In this section, we will go into more detail regarding the statistical errors in the envelope and the correlation time. Let us focus on the errors regarding the envelope of each modal auto-correlation function by applying the theory of Peak Probability Density Functions [5, 22].

3.1. Noise tail

The noise tail is a non-persistent feature in the estimated correlation function matrix where we have erratic behaviour in the tail region, see Fig. 7. This happens when the statistical errors dominate the correlation function and introduce a biased error in the envelope of the estimated correlation function. As seen in Fig. 7, the noise tail looks different for each realisation of the same system. We must neglect the notion that the estimated correlation function is a free decay in this tail region. This section will prove this phenomenon analytically by studying the (Hilbert) envelope of the estimated modal auto-correlation function and compare it with the theoretical envelope, Eq. (9).

$$\text{env}\left(\widetilde{R}_{q_i}(\tau)\right) = \sqrt{\widetilde{R}_{q_i}(\tau)^2 + \mathcal{H}\left[\widetilde{R}_{q_i}(\tau)\right]^2} \quad (24)$$

3.2. Probability density function of the Hilbert envelope

In this section, we will derive the general probability density function for the Hilbert envelope of a normal distributed random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$. The Hilbert envelope of X is given by: $\text{env}(X) = \sqrt{X^2 + \mathcal{H}[X]^2}$. We look at the probability density of the Hilbert transformed variable, $\mathcal{H}[X]$. By a linear

Figure 7: Correlogram with two example of the noise tail in a correlation function based on two realisations of the same estimated correlation function

Figure 8: The probability density function of the Rice distribution, Eq. (26), for different values of μ while σ is constant

combination of the standard normal distribution, $M \sim \mathcal{N}(0, 1)$, we write the normal distributed random variable as: $X = \sigma M + \mu$. Then we apply the Hilbert transformation to this linear combination and use the fact that the Hilbert transformation of a constant is zero.

$$\begin{aligned} \mathcal{H}[X] &= \mathcal{H}[\sigma M + \mu] \\ &= \sigma \mathcal{H}[M] \end{aligned} \quad (25)$$

Thus, the Hilbert transformation creates a modified normal distribution with an expected value of zero and the same variance, $\mathcal{H}[X] \sim \mathcal{N}(0, \sigma^2)$. So we have two normal variables with the same variance but different expected values. Hence, the Hilbert envelope has a Rice distribution [22], which is a special case of the non-central chi-square distribution with two degrees of freedom and non-centrality parameter $(\frac{\mu}{\sigma})^2$. A variable, $R \sim \text{Rice}(|\mu_1^2 + \mu_2^2|, \sigma)$, is Rice distributed when it is comprised of two normal variables, $Z \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, in the following way, $R = \sqrt{Z^2 + Y^2}$ [23]. Thus, the general Hilbert envelope of a normal distributed random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, is Rice distributed, $\text{env}(X) \sim \text{Rice}(\mu, \sigma)$.

The probability density function of the Rice distributed variable is given by

$$f_e(x, \mu, \sigma) \triangleq \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \mu^2}{2\sigma^2}\right) I_0\left(\frac{\mu x}{\sigma^2}\right) \quad (26)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind with order zero [24].

Then we find the expected value of the Hilbert envelope as a function of the original expected value, μ , and variance, σ^2 , of the normal distributed variable X .

$$\mu_m(\mu, \sigma) = \sqrt{\frac{\pi}{2}} \sigma L_{1/2} \left(-\frac{\mu^2}{2\sigma^2} \right) \quad (27)$$

where $L_{1/2}(\cdot)$ is a Laguerre polynomial for case $q = \frac{1}{2}$ [22, 24]. Likewise, we find the variance of the Hilbert envelope.

$$\begin{aligned} \sigma_m^2(\mu, \sigma) &= 2\sigma^2 + \mu^2 - \frac{\pi\sigma^2}{2} L_{1/2} \left(-\frac{\mu^2}{2\sigma^2} \right)^2 \\ &= 2\sigma^2 + \mu^2 - \mu_m(\mu, \sigma)^2 \end{aligned} \quad (28)$$

The Rice distribution changes shape depending on the signal-to-noise ratio, $\frac{\mu}{\sigma}$, see Fig. 8. When this ratio tends towards infinite, the envelope becomes normal distributed, and the envelope gets a Rayleigh distribution when the ratio tends towards zero [22, 23]. In practice, the distribution is approximately normal distributed when the ratio is above 3 [25].

3.3. Probability density function of the envelope for the estimated correlation function

We insert the expected value and variance of the underlying normal distribution into the general Rice distribution for the envelope, Eq. (26). Here we use the analytic envelope of the modal auto-correlation function, Eq. (9), and the upper envelope of the variance, Eq. (20).

$$f_i(x, \tau) = f_e \left(x, e_i(\tau), \text{Var} \left[\widetilde{R}_{q_i}(\tau) \right]^+ \right) \quad (29)$$

The shape of the Rice distribution depends on the signal-to-noise ratio or the relationship between the analytic envelope, $e_i(\tau)$, and the statistical errors at the peaks, $\text{Var}[\widetilde{R}_{q_i}(\tau)]^+$.

We find the expected value of the Hilbert envelope using Eq. 27.

$$\mu_{m_i}(\tau) = \sqrt{\frac{\pi}{2}} \sigma_{q_i}^2 \sqrt{\frac{1 + e^{-2\omega_0\zeta\tau}(2\omega_0\zeta\tau + 1)}{2(T - \tau)\omega_0\zeta}} L_{1/2} \left(\frac{(\tau - T)\omega_0\zeta}{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1} \right) \quad (30)$$

Likewise, we find the variance of the Hilbert envelope using Eq. 28.

$$\sigma_m^2(\tau) = \sigma_{q_i}^4 \frac{1 + e^{-2\omega_0\zeta\tau}(2\omega_0\zeta\tau + 1)}{2(T - \tau)\omega_0\zeta} \left(2 - 2 \frac{(\tau - T)\omega_0\zeta}{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1} - \frac{\pi}{2} L_{1/2} \left(\frac{(\tau - T)\omega_0\zeta}{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1} \right)^2 \right) \quad (31)$$

Fig. 9 shows the expected value and the interquartile range of the envelope for the modal auto-correlation function. In the beginning, the expected value follows the analytic envelope but it diverts as the signal-to-noise ratio decreases in the tail region, see Eq. (30). Hence, the expected envelope is biased in this region of the modal auto-correlation function since it tends toward a constant value due to the statistical errors. This explains the erratic noise tail in the estimated correlation function matrix.

The variance of the envelope is displayed in Fig. 10 for two different time lengths, and the shape and amplitude of the variance depend on the time length and frequency-damping product, see Eq. (31). The excitation level is related to the amplitude of the variance.

3.4. First part of the envelope - free decay

We want to look at the first part of the modal auto-correlation function that has low levels of statistical errors. In this region, the analytic envelope is much bigger than the variance, $e_i(\tau) \gg \text{Var}[\widetilde{R}_{q_i}(\tau)]$, and the length of this region is the correlation time. Then we have a limit of the Rice distribution where it is an approximated normal distribution with the expected value of $\sqrt{\mu^2 + \sigma^2}$ and variance of σ^2 [25]. The expected value of the envelope of a modal auto-correlation function is

Figure 9: Expected value and variance of the envelope of the estimated correlation function for an arbitrary system plotted for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13], the expected value, Eq. (30), (black solid line), the interquartile range (grey fill), and the analytic envelope, Eq. (9), (red dashed line)

Figure 10: Variance of the envelope of the estimated correlation function for an arbitrary system plotted for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13], the variance of the envelope, Eq. (31), (black solid line) and the general variance of modal auto-correlation function, Eq. (20), (red dashed line)

Figure 11: (a) The normalised biased error as a function of $\frac{\mu}{\sigma}$, (b) the normalised random error as a function of $\frac{\mu}{\sigma}$

$$\mathbb{E}[\tilde{e}_i(\tau)] = \sqrt{e_i(\tau)^2 + \text{Var}[\widetilde{R}_{q_i}(\tau)]^+}, \quad e_i(\tau) \gg \text{Var}[\widetilde{R}_{q_i}(\tau)] \quad (32)$$

The variance of this region of the envelope yields

$$\text{Var}[\tilde{e}_i(\tau)] = \text{Var}[\widetilde{R}_{q_i}(\tau)]^+, \quad e_i(\tau) \gg \text{Var}[\widetilde{R}_{q_i}(\tau)] \quad (33)$$

3.5. Last part of envelope - noise tail

We will look at the part of the correlation function where the noise tail appears. As the analytic envelope tends towards zero, $e_i(\tau) \approx 0$, the density function of the Hilbert envelope changes to another special case of the Rice distribution; the Rayleigh distribution [22, 23]. Then we calculate the expected value of the envelope in the noise tail region of the correlation function.

$$\mathbb{E}[\tilde{e}_i(\tau)] = \sqrt{\frac{\pi}{2}} \sqrt{\text{Var}[\widetilde{R}_{q_i}(\tau)]^+}, \quad e_i(\tau) \approx 0 \quad (34)$$

We calculate the variance of the envelope in this region of the modal auto-correlation function.

$$\text{Var}[\tilde{e}_i(\tau)] = \left(2 - \frac{\pi}{2}\right) \text{Var}[\widetilde{R}_{q_i}(\tau)]^+, \quad e_i(\tau) \approx 0 \quad (35)$$

The envelope depends only on the statistical errors in the noise tail. Therefore, the estimated modal auto-correlation function is no longer a free decay since it is independent of the analytic envelope of the modal auto-correlation function in this region. The noise tail violates the core assumption in the correlation-driven operational modal analysis: that the correlation function matrix consists of free decays. Therefore, it is crucial that we exclude the noise tail from an identification process.

3.6. Errors of the estimated envelope

Let us look at the biased and random errors to understand the statistical errors and their effect on the envelope. We start with the normalised biased error as a function of the signal-to-noise ratio, see Fig. 11 A).

$$\begin{aligned} \varepsilon_m\left(\frac{\mu}{\sigma}\right) &= \frac{\mu_m}{\mu} - 1 \\ &= \sqrt{\frac{\pi}{2}} \frac{\sigma}{\mu} L_{1/2}\left(-\frac{\mu^2}{2\sigma^2}\right) - 1 \end{aligned} \quad (36)$$

Figure 12: The normalised biased error of the envelope, Eq. (37), for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13]

We insert the equation for the envelope, Eq. (9) and the equation for the variance related to the peaks of the correlation function, Eq. (20), into Eq. (36).

$$\varepsilon_m(\tau) = \sqrt{\frac{\pi}{4}} \sqrt{\frac{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1}{(T - \tau)\omega_0\zeta}} L_{1/2} \left(\frac{(\tau - T)\omega_0\zeta}{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1} \right) - 1 \quad (37)$$

Fig. 12 shows the normalised biased error of the envelope of an estimated modal auto-correlation function for two different time lengths.

We turn to the normalised random error, sometimes called the coefficient of variation, see Fig. 11 B).

$$\begin{aligned} \epsilon_m \left(\frac{\mu}{\sigma} \right) &= \frac{\sigma_m}{\mu_m} \\ &= \frac{\sqrt{2 + \frac{\mu^2}{\sigma^2} - \frac{\pi}{2} L_{1/2} \left(-\frac{\mu^2}{2\sigma^2} \right)^2}}{\sqrt{\frac{\pi}{2} L_{1/2} \left(-\frac{\mu^2}{2\sigma^2} \right)}} \end{aligned} \quad (38)$$

The normalised random error is illustrated in Fig. 13. In the physical part of the envelope (high value of signal-to-noise ratio) the normalised random error of the envelope is approximately equal to the normalised random error of the estimated correlation function, Eq. (23). In the noise tail region, we calculate the normalised random error by tending $\frac{\mu}{\sigma}$ toward zero.

$$\lim_{\frac{\mu}{\sigma} \rightarrow 0} \epsilon_m \left(\frac{\mu}{\sigma} \right) = \sqrt{\frac{4}{\pi} - 1} \quad (39)$$

Both expressions of the error for each mode are independent of the excitation level. Therefore, the effect of the statistical errors depends on the modal parameters, time lag, and the total time length. Both the modal auto-correlation function and its variance have the same relative contribution from the excitation.

In this section, we have proven that the envelope of the estimated correlation function matrix is Rice distributed. Therefore, the envelope becomes increasingly biased as the signal-to-noise ratio decreases to a point where the expected envelope becomes almost constant, thus explaining the phenomenon of the noise tail. Hence, in operational modal analysis, this biased envelope introduces a biased error in the identified damping ratio.

Figure 13: The normalised random error of the envelope, Eq. (38), for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13]

Figure 14: Correlogram with two realisations of the same estimated correlation function and the zero crossings

4. Statistical errors in zero crossings of the modal auto-correlation function

In this section, we will study the zero crossings from the modal auto-correlation function, see Fig. 14. The statistical properties of zero crossings for a stochastic process are, however, largely unsolved and there is no closed-form solution for the probability density function for the level crossing of a stochastic process [5, 22, 26–30]. Level and zero crossings are studied in [5, 22, 27–29] for a sine wave with ergodic white Gaussian noise. The estimated modal auto-correlation function is an exponential decaying cosine with non-stationary Gaussian non-white noise, thus, we simplify the problem due to this complexity. Therefore, we will study the zero crossing in the noise tail region where the statistical errors dominate and where they are approximately stationary when the time length is much larger than the time lag, $T \gg \tau$.

4.1. Expected zero crossing

Rice [22] states that level crossing is given by the joint probability function, $p(\alpha, \beta)$, of a stationary and narrow banded random variable, α , and its derivative, β . The expected zero crossing is then given by

$$N_0 = \int_{-\infty}^{\infty} |\beta| p(0, \beta) d\beta \quad (40)$$

For a Gaussian process with zero means, it simplifies to [5]

$$N_0 = \frac{\sqrt{1 - \rho_{\alpha, \beta}^2}}{\pi} \frac{\sigma_\beta}{\sigma_\alpha} \quad (41)$$

where σ_α and σ_β are the standard deviation of α and β respectively, and $\rho_{\alpha,\beta}$ is the population correlation coefficient between α and β .

4.2. Variance of the derivative of the modal auto-correlation function

In this section, we will find the variance of the derivative of the estimated modal auto-correlation function with respect to the time lag, τ .

The correlation function has the following properties regarding the first and second derivative with respect to time lag [1, 5, 19]

$$-R_{\dot{x},x}(\tau) = R_{x,\dot{x}}(\tau) = \frac{d}{d\tau}R_{x,x}(\tau) = \dot{R}_{x,x}(\tau) \quad (42)$$

$$R_{\dot{x},\dot{x}}(\tau) = -\frac{d^2}{d\tau^2}R_{x,x}(\tau) = -\ddot{R}_{x,x}(\tau) \quad (43)$$

Thus, the derivative, with respect to time lag, of the modal auto-correlation function equals the cross-correlation function between the modal coordinate and its derivative in time, $\frac{d}{d\tau}R_{q_i}(\tau) = R_{q_i,\dot{q}_i}(\tau)$. We assume a large time length, T , and use the variance expression for a cross-correlation function [5] and we apply the properties of the correlation function from Eqs. (42) and (43) to rewrite it.

$$\begin{aligned} \text{Var} \left[\tilde{R}_{q_i,\dot{q}_i}(\tau) \right] &\approx \frac{1}{T-\tau} \int_{-\infty}^{\infty} (R_{q_i,q_i}(r)R_{\dot{q}_i,\dot{q}_i}(r) + R_{q_i,\dot{q}_i}(r+\tau)R_{\dot{q}_i,q_i}(r-\tau)) dr \\ &\approx \frac{1}{T-\tau} \int_{-\infty}^{\infty} \left(-R_{q_i}(r)\ddot{R}_{q_i}(r) - \dot{R}_{q_i}(r+\tau+)\dot{R}_{q_i}(r-\tau) \right) dr \end{aligned} \quad (44)$$

We calculate the first derivative of the modal auto-correlation function, Eq. (5), with respect to τ .

$$\dot{R}_{q_i}(\tau) = -\frac{\omega_0}{\sqrt{1-\zeta^2}}\sigma_{q_i}^2 e^{-\omega_0\zeta|\tau|} \sin(\omega_d\tau) \quad (45)$$

Clough and Penzien [19] calculated the second derivative of the auto-correlation function of an SDOF system with respect to τ .

$$\ddot{R}_{q_i}(\tau) = -\omega_0^2\sigma_{q_i}^2 e^{-\omega_0\zeta|\tau|} \left(\cos(\omega_d\tau) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin(\omega_d|\tau|) \right) \quad (46)$$

We can see that $\omega_0^2 R_{q_i}(\tau) = -\ddot{R}_{q_i}(\tau)$. So we simplify again.

$$\text{Var} \left[\frac{d}{d\tau} \tilde{R}_{q_i}(\tau) \right] \approx \frac{1}{T-\tau} \int_{-\infty}^{\infty} \left(\omega_0^2 R_{q_i}(r)^2 - \dot{R}_{q_i}(r+\tau+)\dot{R}_{q_i}(r-\tau) \right) dr \quad (47)$$

This is similar to Eq. (18) but the first term, $\omega_0^2 R_{q_i}(r)^2$, is multiplied by ω_0^2 . We insert Eqs. (5) and (46) into Eq. (47) and calculate the integral by assuming small damping.

$$\text{Var} \left[\frac{d}{d\tau} \tilde{R}_{q_i}(\tau) \right] \approx \omega_0^2 \frac{\sigma_{q_i}^4}{2(T-\tau)\omega_0\zeta} (1 - e^{-2\omega_0\zeta\tau} \cos(2\omega_d\tau)(1 + 2\zeta\omega_0\tau)) \quad (48)$$

4.3. Auto-correlation of statistical errors

The statistical dependency of the correlation function matrix is another important statistical property of the correlation function. Since we calculate the estimated correlation function matrix for all time lags from the same realisation of the system response, then the matrix - at different time lags - is statistically dependent. Hence, the statistical errors in the estimated correlation function matrix are correlated. In this subsection, we will study this correlation since it relates to the zero crossings.

The estimated modal auto-correlation function is Gaussian distributed so the statistical errors are additive Gaussian noise, $s(\tau) \sim \mathcal{N}(0, \text{Var}[\tilde{R}_{q_i}(\tau)])$, in each estimated modal auto-correlation function.

$$\widetilde{R}_{q_i}(\tau) = R_{q_i}(\tau) + s(\tau) \quad (49)$$

We will look at the non-stationary and auto-correlation function of the statistical errors in each estimated modal auto-correlation function. This is equivalent to the covariance function of the estimated modal auto-correlation function.

$$\begin{aligned} R_{s,s}(\tau, v) &= \mathbb{E}[s(\tau)s(\tau+v)] \\ &= \mathbb{E}\left[\left(\widetilde{R}_{q_i}(\tau) - R_{q_i}(\tau)\right)\left(\widetilde{R}_{q_i}(\tau+v) - R_{q_i}(\tau+v)\right)\right] \\ &= \text{Cov}[\widetilde{R}_{q_i}(\tau), \widetilde{R}_{q_i}(\tau+v)] \end{aligned} \quad (50)$$

Priestley [31] derives an expression of the covariance function of the estimated correlation function, see Appendix C for a similar derivation of the expression of the auto-correlation function of the statistical errors.

$$R_{s,s}(\tau, v) \approx \frac{1}{T-\tau} \int_{-\infty}^{\infty} (R_{q_i}(r)R_{q_i}(r+v) + R_{q_i}(r+v+\tau)R_{q_i}(r-\tau)) dr \quad (51)$$

Giampellegrini [6] calculates the covariance function of the estimated correlation functions using this integral, Eq. (51), by assuming small damping. In this article, we calculate the auto-correlation function of the statistical errors based on the work in [6].

$$\begin{aligned} R_{s,s}(\tau, v) \approx & \frac{\sigma_{q_i}^4}{2(T-\tau)\omega_0\zeta} \cdot \left(e^{-\omega_0\zeta|v|} \left(\cos(\omega_d v)(1 + |v|\omega_0\zeta) + \sin(\omega_d |v|) \frac{2\omega_0\zeta^2|v|}{\sqrt{1-\zeta^2}} \right) + \right. \\ & \left. e^{-\omega_0\zeta(2\tau+v)} \left(\cos(\omega_d(2\tau+v))(1 + (2\tau+v)\omega_0\zeta) + \sin(\omega_d(2\tau+v)) \frac{2\omega_0\zeta^2(2\tau+v)}{\sqrt{1-\zeta^2}} \right) \right) \end{aligned} \quad (52)$$

For $v = 0$, the covariance function, Eq. (52), simplifies to the variance of the modal correlation function, Eq. (19). Fig. 15 shows the correlation function of the statistical errors. The statistical errors are a narrow banded process since it - for any value of τ - oscillates with the damped frequency, ω_d , of the given mode.

We will study two special instances: at time lag zero, $\tau = 0$, and high time lags, $\tau \gg 0$.

$$R_{s,s}(0, v) \approx \frac{\sigma_{q_i}^4}{T\omega_0\zeta} e^{-\omega_0\zeta|v|} \left(\cos(\omega_d v)(1 + |v|\omega_0\zeta) + \sin(\omega_d |v|) \frac{2\omega_0\zeta^2|v|}{\sqrt{1-\zeta^2}} \right) \quad (53)$$

Fig. 7 shows this phenomenon with two examples of the estimated modal auto-correlation function.

$$R_{s,s}(\tau, v) \approx \frac{\sigma_{q_i}^4}{2(T-\tau)\omega_0\zeta} e^{-\omega_0\zeta|v|} \left(\cos(\omega_d v)(1 + |v|\omega_0\zeta) + \sin(\omega_d |v|) \frac{2\omega_0\zeta^2|v|}{\sqrt{1-\zeta^2}} \right), \quad \tau \gg 0 \quad (54)$$

This case studies the statistical errors in the noise tail where each estimated modal auto-correlation function oscillates with its damped natural frequency in the tail region of the correlation function. Furthermore, given a large time length compared to the time lag, $T \gg \tau$, the statistical errors are approximately stationary.

4.4. Expected zero crossing for the modal auto-correlation function

We focus on the noise tail where the expected value, $\widetilde{R}_{q_i}$, is approximated zero and the variance is approximately stationary given that the time length is much larger than the time lag, $T \gg \tau$. Therefore, the variance of the modal auto-correlation function and its derivative obtain the following relationship.

Figure 15: Correlogram of statistical errors, Eq. (52), for time lag zero, $\tau = 0$, plotted for two different time lengths, (a) Brincker and Ventura [1], (b) ANSI [13]

$$\text{Var} \left[\frac{d}{d\tau} \widetilde{R}_{q_i}(\tau) \right] \approx \omega_0^2 \text{Var} \left[\widetilde{R}_{q_i}(\tau) \right], \quad \tau \gg 0 \quad (55)$$

We have the estimated modal auto-correlation function, $\widetilde{R}_{q_i} \sim \mathcal{N}(0, \text{Var}[\widetilde{R}_{q_i}(\tau)])$, and its derivative, $\frac{d}{d\tau} \widetilde{R}_{q_i}(\tau) \sim \mathcal{N}(0, \omega_0^2 \text{Var}[\frac{d}{d\tau} \widetilde{R}_{q_i}(\tau)])$. Since the statistical errors dominate, $\widetilde{R}_{q_i}(\tau) = s(\tau)$, we can write the correlation coefficient between the estimated modal auto-correlation function and its derivative as

$$\rho_{s,\dot{s}}(\tau) = \frac{R_{s,\dot{s}}(\tau, 0)}{\sqrt{\text{Var}[\widetilde{R}_{q_i}(\tau)]} \sqrt{\text{Var}[\frac{d}{d\tau} \widetilde{R}_{q_i}(\tau)]}}, \quad \tau \gg 0 \quad (56)$$

For stationary data, the cross-correlation function between a random process and its derivative is zero at time lags zero, $R_{x,\dot{x}}(0) = 0$, [5]. Hence, the cross-correlation, $R_{s,\dot{s}}(\tau, v)$, between the statistical errors and its derivative approach zero in the noise tail for $v = 0$. Thus, the correlation coefficient is approximately zero in the noise tail.

$$\rho_{s,\dot{s}}(\tau) \approx 0, \quad \tau \gg 0 \quad (57)$$

We apply Eq. (41) and calculate the expected zero crossing in the noise tail region.

$$N_0(\tau) \approx \frac{\omega_0}{\pi}, \quad \tau \gg 0 \quad (58)$$

This corresponds well with the findings in Section 4.3, however, here the oscillating frequency is not the damped frequency. This is due to the assumptions of narrow banded in Rice [22], Eq.(41), for zero crossings. Since the statistical errors are approximately stationary in the noise tail region and they are narrow banded with the same frequency, ω_0 , as the given mode, then the expected zero crossing should be twice this frequency. Hence, the zero crossings of the estimated modal auto-correlation function are an unbiased process. In the noise tail, the expected number of zero crossings as a function of τ is given by

$$\mathbb{E}[N(\tau)] = \int_0^\tau N_0 d\tau = \frac{\omega_0}{\pi} \tau + \frac{1}{2}, \quad \tau \gg 0 \quad (59)$$

4.5. Variance of zero crossings for the modal auto-correlation function

As previously stated, no closed-form solution exists for the probability density function of the periods of zero crossings. Thus, it is difficult to express the exact variance of each zero crossing.

Cobb [29] shows that the periods of zero crossings from a sine wave with additive Gaussian noise are normal distributed for high signal-to-noise ratios.

$$\text{Var}[N(\tau)] = \frac{2(1 + \rho_1)}{\pi^2 a^2}, \quad \mu \ll \sigma \quad (60)$$

where $a = \frac{1}{\epsilon(\tau)^-}$, Eq. (23), and ρ_1 is the normalised correlation function for noise alone at the value $T = 1$.

It illustrates that the variance of the zero crossings depends on the signal-to-noise ratio at the beginning of the modal auto-correlation function.

Raninal [28] proves that the variance of the number of zero crossings in a time length, t , increase linearly for low signal-to-noise ratios and high time lags.

$$\text{Var}[N(\tau)] = N_0 \tau, \quad \mu \gg \sigma \quad (61)$$

This indicates that the variance of zero crossings is much higher and increases linearly in the noise tail.

This concludes this section on zero crossings. We have shown that the statistical errors on each modal auto-correlation function are a non-stationary process but they are narrow banded with the same frequency as the damped natural frequency of each mode. We found that the expected zero crossings are unbiased but the variance of the zero crossings increase linearly in the noise tail. Thus in operational modal analysis, the identified natural frequencies are unbiased regardless of the noise tail but we should disregard the noise tail to reduce the uncertainty of the identification process.

5. Noise tail relationship / time length and correlation time relationship

In this section, we will study the relationship between the time length of data and the correlation time of the modal auto-correlation functions. The correlation time is, however, a stochastic process that describes the gradual transition from a persistent physical part to the erratic noise tail. So we will have to express the relationship between the time length and correlation time with some stochastic expression of error. We will set up a simplified relationship between the time length, correlation time, and the modal parameters based on the biased error since Eq. (37) is non-trivial to solve. Both the biased and random error of the envelope depend on the signal-to-noise ratio or the normalised random error, Eq. (23), of the peaks of the estimated correlation function.

We know from Sections 2.3 and 2.4 that the mode with the lowest frequency-damping product has the highest level of statistical errors. Thus, we should use the mode with the lowest frequency-damping product to find the required time length or correlation time.

In operational modal analysis, the excitation is unknown so we must use the relative expression of the statistical errors since these are unrelated to the excitation levels.

We want to solve the equation for the biased error of the envelope, Eq. (36), to find an expression for the signal-to-noise ratio as a function of this biased error. We use the asymptotic expansion of the Laguerre polynomial for the case $\frac{1}{2}$ for negative values [23].

$$L_{1/2}(-x) = \frac{2\sqrt{x}}{\sqrt{\pi}} + \frac{1}{2\sqrt{\pi x}} + \frac{1}{16\sqrt{\pi x^{\frac{3}{2}}}} + \mathcal{O}\left(x^{-\frac{5}{2}}\right) \quad (62)$$

We insert this into the normalised biased error, Eq. (36), and solve it for the signal-to-noise ratio, $\frac{\mu}{\sigma}$.

$$SNR(\varepsilon_m) = \sqrt{\frac{\sqrt{2\varepsilon_m + 1} + 1}{4\varepsilon_m}} \quad (63)$$

where $SNR(\varepsilon_m)$ is the approximated signal-to-noise ratio, $\frac{\mu}{\sigma}$, as a function of the normalised biased error. Since we based it on an asymptotic expansion, it only works for a small normalised biased error.

We know that the signal-to-noise ratio is inversely proportional to the normalised random error. Thus, we can equal the approximated signal-to-noise ratio, Eq. (63), to the normalised random error, Eq. (23), that we derived in Section 2.4.

$$\frac{1}{SNR(\varepsilon_m)} = \sqrt{\frac{1}{T-\tau} \frac{e^{2\omega_0\zeta\tau} + 2\omega_0\zeta\tau + 1}{2\omega_0\zeta}} \quad (64)$$

Since the normalised biased error is based on the biased error relative to the envelope, we will express the correlation time relative to the decay of the correlation function, $a = e^{-\omega_0\zeta|\tau_0|}$. We simplify Eq. (64) by removing the smallest terms of the normalised random error and we insert Eq. (63).

$$T_{min}(\varepsilon_m, f_0\zeta, a) = \frac{\sqrt{2\varepsilon_m + 1} + 1}{4\varepsilon_m} \frac{a^{-2}}{4\pi f_0\zeta} \quad (65)$$

where τ_0 is the correlation time for the given mode.

This new expression adds novel information on the relationship between the normalised biased error of the envelope, ε_m , the frequency-damping product, $\omega_0\zeta$, the time length of the recording, T , and the correlation time of the correlation function matrix, τ_0 . For this expression, we should use the mode with the lowest frequency-damping product since it has the highest level of statistical errors as previously stated. We can use this expression to find an appropriate time length, T_{min} , based on a given time lag, biased error, and correlation time. Or we can find the correlation time, τ_{max} , for a given time length if we have the desired level of normalised biased error.

We will show the application of Eq. (65) with an example. Let us say that we want to find the time length of a measurement setup meant for operational modal analysis. We have an initial estimate of the modal parameters of the first fundamental mode. Now we must specify how much correlation time, a , we need for the identification process, e.g. $a = 10\%$. In practice, the authors find that the noise tail seems to begin at 10% normalised biased error of the envelope. Thus, we need a time length of approximately $T_{ex} \approx \frac{42}{f_0\zeta}$.

5.1. Comparison to other required time lengths

By applying Eq. (65), we will study two minimum required time lengths, Brincker and Ventura [1] and ANSI 2.47 [13]. These two expressions recommend minimum time lengths with a difference of 2,000%.

$$T_{Brincker} = \frac{10}{f_0\zeta} \quad (66)$$

$$T_{ANSI} = \frac{200}{f_0\zeta} \quad (67)$$

The required time length from Brincker and Ventura, Eq. (66), assumes that the lowest frequency of the system defines the longest correlation time. It uses spectral density functions as the basis for this time length requirement [5]. The correlation function for the lowest mode must decay to an amplitude of 4% and we need 20 data segments with 50% overlap and 40 frequency averages [1]. We want to compare this time length requirement with Eq. (65). We know from Eq. (66) that we must have $a = 4\%$. We equal Eq. (65) to Eq. (66), $T_{min}(\varepsilon_m, f_0\zeta, 0.04) = T_{Brincker}$, and solve for the normalised biased error, ε_m , which is approximately 440%. Since ε_m was based on an asymptotic expansion of the Laguerre polynomial for negative values, it is inaccurate for large values of normalised biased error. The normalised biased error is, however, still so big that we should decrease the correlation time for the given time length. We set the normalised biased error to 10%, so the correlation time ends at an amplitude of approximately 20%. This contradicts the 4% decay of the original derivation.

We turn to the other time length requirement from ANSI 2.47, Eq. (67). It uses the first fundamental mode of a system and has a biased error of 4% and a variance error of 10% [13]. We insert a normalised biased error, ε_m , of 4% into Eq. (65) and equal it to Eq. (67), $T_{min}(0.04, f_0\zeta, a) = T_{ANSI}$. Then the correlation time should end at an amplitude of 7.1%. If we accept a normalised biased error of 10% then the correlation time ends at an amplitude of approximately 5%.

A normalised biased error at 10% often occurs at the end of the correlation time and the beginning of the noise tail. So, for Brincker and Ventura the noise tail begins when the amplitude drops to $\sim 20\%$ while

Figure 16: The system used in the case study

it begins at an amplitude of $\sim 5\%$ for ANSI 2.47. The increased correlation time from ANSI 2.47, however, comes with a 2,000% increase in the time length. The two recommended time lengths result in widely different levels of error, but Brincker and Ventura provide the more practical of the two recommended time lengths, though the correlation time is shorter.

In this section, we showed that the normalised random error of the estimated correlation function can be used to set up a simple relationship between the normalised biased error, the time length, correlation time, and the system parameters.

6. Case study

In this section, we will verify the statistical properties derived in this article of the estimated correlation function matrix based on a simulation and an experimental study. For the discrete estimation of the correlation function matrix, we apply the direct method in the time domain where we compute average products of the sampled data using the toolbox from Brincker and Ventura [1].

$$\hat{\mathbf{R}}(k) = \frac{1}{N-k} \sum_{n=0}^{N-k} \mathbf{y}(n) \mathbf{y}^\top(n+k), \quad 0 \leq k < N \quad (68)$$

where $\hat{\mathbf{R}}(k)$ is the discrete estimated correlation function matrix, N is the total number of samples, $T = N\Delta t$, k is the discrete time lag, $\tau = k\Delta t$, and Δt is the sampling interval. For more information on the discrete estimation of the correlation function matrix, we refer to [1, 5].

6.1. Simulation study

We simulate the response of a three degree-of-freedom system corresponding to masses in series supported in one end, see the system in Fig. 16. To model a system with closely spaced and orthogonal modes, we modify the system so the third mode has the similar natural frequency as the second mode but shifted upwards with 0.05 Hz, see the modal parameters in Table 1. The time length of each simulation was 13 minutes. Thus, the correlation time corresponds to a decay of 6.7% of the original amplitude while the normalised error is below 25% according to Eq. (65). We run 100,000 simulations of the system response by generating white Gaussian noise as excitation and use the Fourier transformed modal superposition to simulate the linear system [1]. For each simulation of the response, we calculate the estimated correlation function matrix and use the analytic mode shapes of the system to decorrelate each estimated correlation function matrix into the modal auto-correlation functions, similar to Eq. (4). We estimate the envelope of each modal auto-correlation function using the Hilbert transformation [20].

Table 1: Modal parameters of the numerical system

	Mode 1	Mode 2	Mode 3
Frequency (Hz)	3.39	10.01	10.06
Damping Ratio (%)	1.49	1.34	5.00

Figure 17: Simulation study: Analytic expectation, Eq. (13), (red dashed line) against sample mean of the Estimated Correlation Function Matrix, $\tilde{\mathbf{R}}(\tau)$, based on 100,000 Monte Carlo Simulations (black solid line)

Figure 18: Simulation study: Analytic variance, Eq. (16), (*red dashed line*) against sample variance of the Estimated Correlation Function Matrix, $\tilde{\mathbf{R}}(\tau)$, based on 100,000 Monte Carlo Simulations (*black solid line*)

Figure 19: Simulation study: Analytic expected value, Eq. (5), and analytic variance, Eq. (19), (red dashed line) against sample mean and variance of the Estimated Modal Auto-Correlation Function, $\widehat{R}_{q_i}(\tau)$, for each mode based on 100,000 Monte Carlo Simulations (black solid line)

Figure 20: Simulation study: Analytic expected value, Eq. (30), and variance, Eq. (31), of envelope (red dashed line) against sample mean and variance of envelope based on 100,000 Monte Carlo simulations (black solid line)

Figure 21: Simulation study: The normalised biased error, Eq. (37), and normalised random error, Eq. (38), (red dashed line) against the sampled versions from 100,000 Monte Carlo simulations (black solid line)

Figure 22: Simulation study: Analytic correlation of statistical errors, Eq. (52), (*red dashed line*) against estimated correlation of statistical errors based on 100,000 Monte Carlo simulations (*black solid line*)

Figure 23: Simulation study: Analytic expected zero crossings, Eq. (59), (*red dashed line*) against sample mean and variance of zero crossings based on 100,000 Monte Carlo simulations (*black solid line*)

Figs. 17 and 18 show the expected value, Eq. (13), and variance, Eq. (16), respectively, of the correlation function matrix next to sampled values of all estimated correlation function matrices from the simulation. Fig. 19 compares the expected value, Eq. (5), and the variance, Eq. (19), of the estimated modal auto-correlation functions, $\widetilde{R}_{q_i}(\tau)$, to the sample mean and sample variance of the decorrelated estimated correlation function matrix from the Monte Carlo simulations. There is a high correlation between the theoretical and the simulated statistical properties.

In Fig. 20, we move to the envelope and compare the analytic expressions of the expected value, Eq. (30), and the variance, Eq. (31), of the envelope to sample mean and variance from the Monte Carlo simulation. The sample mean and analytic expression of the expected envelope are identical while the sample variance is almost equivalent to the analytic variance of the envelope. Fig. 21 illustrates the normalised biased and random error of the Hilbert envelope.

Fig. 22 illustrates the correlation function of the statistical errors in each modal correlation function at two time lags, τ , by comparing the analytic expression, Eq. (52), with the results from the simulation. The analytic correlation function of the statistical errors matches with the simulation study. Fig. 23 shows the zero crossings for each modal auto-correlation function, and the analytic expression is similar to the sample mean zero crossing. The variance of the zero crossings increases rapidly and linearly in the noise tail region while the variance is small for zero crossings located in the correlation time. By comparing it with Fig. 21, this growth of variance happens as the biased error exceeds 10%.

The assumption of a sufficient time length is system dependent. For the second and third mode, there is a high correlation between theory and the simulation but the first mode deviates a bit. Thus, the time length must be sufficiently long for the individual modes, see sections 2.4 and 3.6. Thus, modes with low frequency-damping products require longer time lengths compared to modes with higher products. Therefore in this case study, we should expect the third mode to behave more as the analytic equations from this article than the first mode would.

Thus, based on these findings, we will expect relatively large random and biased errors in the estimated damping ratios, some random errors in the estimated frequency, and only identification errors in the estimated mode shapes in an identification process using correlation-driven operational modal analysis.

The assumptions in this article are mainly based on lightly damped and orthogonal modes. Very closely spaced modes, however, have correlated modal coordinates, and the cross-correlation between the modes is influential. In such a case, the expectation and variance of the estimated correlation function matrix differ from the summation of the expectations and variances of the modal auto-correlation functions. In this case, these summations are more a rule of thumb.

Based on this case study, the statistical properties of the estimated correlation function matrix match well with the analytic expression.

6.2. Experimental study

In this section, the analytic expressions are verified by an experimental study of a T shaped steel structure with a hollow cross-section, RHS40×40×2, in a laboratory facility, see Fig. 24. To verify the statistical properties, we need many data sets measured over a longer period. The fluctuation of the modal parameters is, however, near inevitable due to environmental effects during the recording. Even in a laboratory, it is difficult to obtain complete stationarity and completely remove environmental effects over many data sets, see Fig. 26. The variances in temperature and humidity cause slightly non-stationary in the modal parameters over all the data sets. Thus for the experimental study, we have a variance from both the statistical errors and non-stationary modal parameters. This additional variance obstructs the verification of any analytic expression regarding variance; however, we can still verify the analytic expression regarding expected values using the experimental study.

For the measurement-setup, we applied 10 Brüel & Kjær 4508-B-002 uniaxial accelerometers to the structure as seen in Fig. 25. The data sets were recorded over the course of four days, primarily at night. We sampled the data at a frequency of 1651 Hz and we used 386 data sets with a time length of 238 seconds. Thus, the correlation time corresponds to a decay of 14.6% given a normalised error of 25% according to Eq. (65) using the lowest frequency-damping product. We bandpass filter the data with cut-off frequencies

Figure 24: Experimental study: Photo of the test specimen in the laboratory

Figure 25: Experimental study: Position of sensors on the test specimen, arrows denotes uniaxial accelerometers

(a)

(b)

Figure 26: Experimental study: The environmental effects during the experimental tests, variance in (a) temperature and (b) humidity

Figure 27: Experimental study: Analytic expected value, Eq. (5), (*red dashed line*) against sample mean of the Estimated Modal Auto-Correlation Function, $\widehat{R}_{q_i}(\tau)$, for each mode based on experimental study (*black solid line*)

Figure 28: Experimental study: Analytic expected value, Eq. (30), of envelope (*red dashed line*) against sample mean of envelope based on experimental study (*black solid line*)

at 5 and 100 Hz to remove noise and higher modes. The Ibrahim time domain technique [32] identified the modal parameters from an arbitrary data set using the correlation function matrix. The identified mode shapes were used to decorrelate the estimated correlation function matrix from each data set. Finally, we identified the modal parameters based on the sample mean of each modal auto-correlation functions using the Ibrahim time domain, see Table 2. In this study, only the first three modes are included.

Table 2: Modal parameters of the experimental system			
	Mode 1	Mode 2	Mode 3
Frequency (Hz)	7.30	8.91	22.8
Damping ratio (%)	2.47	0.392	0.237

Fig. 27 compares the expected value of the estimated modal auto-correlation functions, Eq. (5), to the sample mean of the decorrelated estimated correlation function matrix from the experimental study. For the envelope, Fig. 28 compares the analytic expressions of the expected value, Eq. (30), of the envelope to the sample mean. While Fig. 29 shows the expected and sample mean of zero crossings for each modal auto-correlation function. Generally, the analytic expression correlates well with this experimental study, taking into account, that the uncertainty of the actual modal parameters is a source of error.

Based on this experimental case study, the statistical properties of the estimated correlation function matrix match well with the analytic expression for all expected values.

Figure 29: Experimental study: Analytic expected zero crossings, Eq. (59), (*red dashed line*) against sample mean of zero crossings based on experimental study (*black solid line*)

7. Conclusion

In this article, we have studied the estimation of the correlation function matrix in operational modal analysis and the associated statistical errors, which relates to the finite time length of the measured system response. To verify the statistical properties derived in this article, we performed a Monte Carlo simulation of a three DOF system, and generally, these simulations agree well with the analytical expressions from this article. We showed that the statistical errors are non-stationary and additive Gaussian noise in each modal auto-correlation function - for orthogonal modes. The relative statistical errors depend on the system parameters, the time length of the measurements, and time lags (correlation time) in the estimated correlation function matrix. The statistical errors are narrow banded with the same frequency as the damped frequency of the mode. The zero crossing of each modal auto-correlation function is unbiased.

We have shown that the envelope of the modal auto-correlation function is Rice distributed, meaning it becomes biased as the statistical errors increase in the tail region. We derived an expression for the expected envelope of the modal auto-correlation function. In this article, we call this biased region the *noise tails* of the modal auto-correlation functions, and the correlation function matrix contains a noise tail for each mode in the system. The biased error of the envelope transfers to an identification process and creates a biased error in the estimated modal parameters. Since damping ratios are primarily identified based on the envelope, the estimation of damping is highly affected by the noise tail.

These statistical errors affect identification of modal parameters in operational modal analysis. The mode shapes span the vector space of the estimated correlation function matrix and the statistical errors are located within this space. Thus, for orthogonal modes, there are no statistical errors in the subspace of the estimated correlation function matrix. The frequency estimates are unbiased since the zero crossings of the correlation function matrix are unbiased but the variance increases rapidly in the noise tail. The damping estimates are biased when the correlation function matrix includes parts of the noise tail. Generally, there is a higher variance on the envelopes than the zero crossings, thus, a higher variance is expected for damping estimates than the frequency estimates.

We derived a relationship between the modal parameters, time lags, and the total time length of the recording. The expression is based on the assumption of white Gaussian noise as excitation, a linear and ergodic system, low damping, and a large time length.

For further research, the effect of statistical errors on the identification of modal parameters in the correlation-driven operational modal analysis should be studied. The theory and analytic expressions provided within this article provide the theoretical framework for such a study.

Acknowledgements

The authors acknowledge the funding received from the Centre for Oil and Gas - DTU/Danish Hydrocarbon Research and Technology Centre (DHRTC).

The authors thank Michael Vigsø, Thomas Kabel, Peter Olsen, and Martin Juul from Aarhus University for their helpful discussions and suggestions. Furthermore, the authors thank Sebastian Glavind from Aalborg University for his contribution to the experimental case study.

Appendix A. Probability density function of the estimated modal auto-correlation function

We will look at the probability density function of the modal auto-correlation function and we will use the Transformation of Probability Densities, the Central Limit Theorem, and the Moment Generating Function to obtain it [5, 21]. We want to find the distribution of the modal auto-correlation function.

$$R_{q_i}(\tau) = \mathbb{E}[q_i(t)q_i(t+\tau)] \quad (\text{A.1})$$

First, we go back to the equation of motion for a stationary system excited with white Gaussian noise.

$$\mathbf{M}\ddot{\mathbf{y}}(t) + \mathbf{C}\dot{\mathbf{y}}(t) + \mathbf{K}\mathbf{y}(t) = \mathbf{x}(t) \quad (\text{A.2})$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix, and $\mathbf{x}(t)$ is the excitation vector. The excitation vector is multivariate normal distributed, $\mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is the covariance matrix.

We apply the modal decomposition, Eq. (2), for mode i and pre-multiply with the transposed mode shape for mode i to decorrelate the equation of motion into an SDOF system for mode i .

$$\boldsymbol{\phi}_i^\top \mathbf{M} \boldsymbol{\phi}_i \ddot{q}_i(t) + \boldsymbol{\phi}_i^\top \mathbf{C} \boldsymbol{\phi}_i \dot{q}_i(t) + \boldsymbol{\phi}_i^\top \mathbf{K} \boldsymbol{\phi}_i q_i(t) = \boldsymbol{\phi}_i^\top \mathbf{x}(t) \quad (\text{A.3})$$

The response of an SDOF system is a convolution of the modal excitation, $\boldsymbol{\phi}_i^\top \mathbf{x}(t)$, and the impulse response function, $h_i(t)$, [19].

$$q_i(t) = \int_{-\infty}^{\infty} \boldsymbol{\phi}_i^\top \mathbf{x}(t-\tau) h_i(t) d\tau \quad (\text{A.4})$$

The integral is defined as the limit of a sum and the Central Limit Theorem states that the sum of many random variables tends to be normal distributed [1]. Thus, the modal coordinates, $q_i(t)$, are approximated normal distributed, $q_i(t) \sim \mathcal{N}(0, \sigma_{q_i}^2)$, where $\sigma_{q_i}^2$ is the variance of the modal coordinates. Any time lag of the response is still a normal variable with the same variance due to the time-invariance of the system and the excitation, $q_i(t+\tau) \sim \mathcal{N}(0, \sigma_{q_i}^2)$.

For simplicity, we deal with a single discrete variable, $q_i(t) \rightarrow X \sim \mathcal{N}(0, \sigma^2)$, and the distribution of the estimated auto-correlation coefficient. The auto-correlation function will have similar proof but the number of averaging changes from N to $N-j$ where N is the discrete time length, $T = N\Delta t$, and j is the discrete time lag, $\tau = j\Delta t$. We leave that proof to the reader.

The density function for a normal distribution with zero mean is given by

$$f_x(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (\text{A.5})$$

Then the correlation coefficient (auto-correlation function at time lag zero) of this normal variable is given by

$$\begin{aligned} \rho_x(\tau) &= \mathbb{E}[X^2] \\ &= \frac{1}{N} \sum_{i=1}^N X_i^2 \end{aligned} \quad (\text{A.6})$$

Firstly, we transform the normal distribution by squaring the random variables, $Y = X^2$, using the Transformation of Probability Densities.

$$g(x) = x^2 \quad , \quad g^{-1}(y) = \pm\sqrt{y} \quad (\text{A.7})$$

We calculate the transformed probability density function for Y :

$$\begin{aligned} f_y(y, \sigma^2) &= f_x(g^{-1}(y), \sigma^2) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= f(\sqrt{y}, \sigma^2) \left| \frac{1}{2\sqrt{y}} \right| + f(-\sqrt{y}, \sigma^2) \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp\left(-\frac{y}{2\sigma^2}\right) \end{aligned} \quad (\text{A.8})$$

We rewrite the probability density function to a Gamma distribution with shape factor $k = \frac{1}{2}$ and scale factor $\theta = 2\sigma^2$.

$$f_y(y, k, \theta) = \frac{y^{k-1}}{\Gamma(k)(\theta)^k} \exp\left(-\frac{y}{\theta}\right) \quad (\text{A.9})$$

Therefore, we can write the squared normal variable as a Gamma distributed variable, $Y \sim \Gamma(k, \theta) = \Gamma(\frac{1}{2}, 2\sigma^2)$.

Now, we will transform this Gamma distribution by finding the sum of N Gamma variables with the same shape and scale factor, $Z = \sum_{i=1}^N Y_i$. We will use the Moment Generating Function [5] to find the new probability density function. The Moment Generating Function for a Gamma distribution is given by

$$\begin{aligned} M(t) &= \mathbb{E}[\exp(tX)] \\ &= \frac{1}{(1 - t\theta)^k} \end{aligned} \quad (\text{A.10})$$

The Moment Generating Function for a summation of random variables equals the product of all the Moment Generating Functions.

$$\begin{aligned} M_{\sum_{i=1}^N Y_i}(t) &= \prod_{i=1}^N M_{Y_i}(t) \\ &= \prod_{i=1}^N \frac{1}{(1 - t\theta)^k} \\ &= \frac{1}{(1 - t\theta)^{Nk}} \end{aligned} \quad (\text{A.11})$$

The Moment Generating Function for the summation is still valid for a Gamma distribution with a shape factor of Nk . Therefore, we have $Z \sim \Gamma(Nk, \theta) = \Gamma(\frac{N}{2}, 2\sigma^2)$.

Finally, we need to divide with the number of variables, $\rho_x = \frac{Z}{N}$, which is a scaling of a Gamma distribution. We use the Transformation of probability densities.

$$g(z) = \frac{z}{N} \quad , \quad g^{-1}(r) = Nr \quad (\text{A.12})$$

The transformed probability density function for ρ_x is given by

$$\begin{aligned}
f_r(r, \sigma^2) &= f_z(g^{-1}(r), \sigma^2) \left| \frac{d}{dr} g^{-1}(r) \right| \\
&= \frac{(Nr)^{Nk-1}}{\Gamma(Nk)(\theta)^{Nk}} \exp\left(-\frac{Nr}{\theta}\right) N \\
&= \frac{r^{Nk-1}}{\Gamma(Nk)(\frac{\theta}{N})^{Nk}} \exp\left(-\frac{Nr}{\theta}\right)
\end{aligned} \tag{A.13}$$

Thus, proving that a scaling of a Gamma distribution is another Gamma distribution with a new scale parameter, $\rho_x \sim \Gamma(Nk, \frac{\theta}{N}) = \Gamma(\frac{N}{2}, \frac{2\sigma^2}{N})$.

We rewrite the Gamma distribution as a summation, $\rho_x = \sum_{i=1}^{Nk} \Omega_i$, using the Moment Generating Functions.

$$\begin{aligned}
M_{\rho_x}(r) &= \frac{1}{\left(1 - r \frac{\theta}{N}\right)^{Nk}} \\
&= \prod_{i=1}^{Nk} \frac{1}{1 - r \frac{\theta}{N}}
\end{aligned} \tag{A.14}$$

Thus, we have $\Omega_i \sim \Gamma(1, \frac{\theta}{N})$ [21]. The Central Limit Theorem states that a large sum of random variables is approximately normal distributed [5]. For a long discrete time length, N , the modal auto-correlation function is approximately normal distributed.

Appendix B. Probability density function of the estimated correlation function matrix

We will look at the probability density function of the correlation function matrix.

$$\mathbf{R}(\tau) = \mathbb{E}[\mathbf{y}(t)\mathbf{y}(t+\tau)^\top] \tag{B.1}$$

We know from Appendix A that the modal coordinates are uncorrelated normal variables, $q_i(t) \sim \mathcal{N}(0, \sigma_{q_i}^2)$. We stack the modal coordinates in a multivariate normal variable, $\mathbf{q}(t) \sim \mathcal{N}(\mathbf{0}, [\sigma_{q_i}^2])$ where $\mathbf{0}$ is a zero vector and $[\sigma_{q_i}^2]$ is the covariance matrix where the non-diagonal elements are zero.

Let us look at the modal decomposition. It is a linear transformation of Gaussian random vectors or a linear transformation of a multivariate Gaussian variable [21].

$$\mathbf{y}(t) = \Phi \mathbf{q}(t) \tag{B.2}$$

If we assume that the mode shapes are time invariant, we have the multivariate normal variable, $\mathbf{y}(t) \sim \mathcal{N}(\mathbf{0}, \Phi[\sigma_{q_i}^2]\Phi^\top)$ [21]. As for the modal coordinates, any time lag of the response is a multivariate normal variable with the same covariance $\mathbf{y}(t+\tau) \sim \mathcal{N}(\mathbf{0}, \Phi[\sigma_{q_i}^2]\Phi^\top)$.

We turn to the correlation function matrix and assume uncorrelated modal coordinates [1, 19].

$$\begin{aligned}
\mathbf{R}(\tau) &= \Phi \mathbb{E}[\mathbf{q}(t)\mathbf{q}(t+\tau)^\top] \Phi^\top \\
&= \sum_{i=1}^N \Phi_i \Phi_i^\top R_{q_i}(\tau)
\end{aligned} \tag{B.3}$$

Thus, the correlation function matrix is a summation of the product, $\Phi_i \Phi_i^\top R_{q_i}(\tau)$. We know from Appendix A that each modal correlation function is approximately normal distributed. Therefore, the correlation function matrix is a summation of linear transformations of multivariate Gaussian variables. Hence, the correlation function matrix is also approximately normal distributed due to the Central Limit Theorem. Thus, the correlation function matrix, $\mathbf{R}(\tau)$, is approximated multivariate normal distributed for a long time series in Operational Modal Analysis.

Appendix C. Auto-correlation of statistical errors

The estimated modal auto-correlation function has non-stationary and additive Gaussian noise, $s(\tau) \sim \mathcal{N}(0, \text{Var}[\widetilde{R}_{q_i}(\tau)])$.

$$\widetilde{R}_{q_i}(\tau) = R_{q_i}(\tau) + s(\tau) \quad (\text{C.1})$$

The auto-correlation function of these statistical errors is equivalent to the covariance function of the estimated modal auto-correlation function, $\text{cov}[\widetilde{R}_{q_i}(\tau), \widetilde{R}_{q_i}(\tau + v)]$, and given by

$$\begin{aligned} R_s(\tau, v) &= \mathbb{E}[s(\tau)s(\tau + v)] \\ &= \mathbb{E}\left[\left(\widetilde{R}_{q_i}(\tau) - R_{q_i}(\tau)\right)\left(\widetilde{R}_{q_i}(\tau + v) - R_{q_i}(\tau + v)\right)\right] \\ &= \mathbb{E}\left[\widetilde{R}_{q_i}(\tau)\widetilde{R}_{q_i}(\tau + v)\right] + \mathbb{E}\left[R_{q_i}(\tau)R_{q_i}(\tau + v)\right] \\ &\quad - \mathbb{E}\left[\widetilde{R}_{q_i}(\tau)R_{q_i}(\tau + v)\right] - \mathbb{E}\left[R_{q_i}(\tau)\widetilde{R}_{q_i}(\tau + v)\right] \end{aligned} \quad (\text{C.2})$$

We arrive at the following equation.

$$R_s(v) = \mathbb{E}\left[\widetilde{R}_{q_i}(\tau)\widetilde{R}_{q_i}(\tau + v)\right] - R_{q_i}(\tau)R_{q_i}(\tau + v) \quad (\text{C.3})$$

The rest of the derivation follows a similar derivation of the variance of the auto-correlation function from [5]. The estimated modal auto-correlation function is defined as (assuming data exist for $T - \tau$).

$$\widetilde{R}_{q_i}(\tau) = \frac{1}{T - \tau} \int_0^{T - \tau} q_i(t)q_i(t + \tau) dt \quad (\text{C.4})$$

We insert the definition of the estimated modal auto-correlation function into Eq. (C.3).

$$R_s(v) = \mathbb{E}\left[\frac{1}{(T - \tau)^2} \int_0^{T - \tau} \int_0^{T - \tau} q_i(t)q_i(t + \tau)q_i(u)q_i(u + \tau + v) dt du\right] - R_{q_i}(\tau)R_{q_i}(\tau + v) \quad (\text{C.5})$$

We use the fourth-order statistical expression of stationary Gaussian processes on the modal coordinates inside the integral [5].

$$\begin{aligned} \mathbb{E}[q_i(t)q_i(t + \tau)q_i(u)q_i(u + \tau + v)] &= R_{q_i}(\tau)R_{q_i}(\tau + v) + R_{q_i}(u - t)R_{q_i}(u + v - t) + \\ &\quad R_{q_i}(u + v - t + \tau)R_{q_i}(u - t - \tau) \end{aligned} \quad (\text{C.6})$$

We insert this back into the integral Eq. (C.5).

$$R_s(\tau, v) = \frac{1}{(T - \tau)^2} \int_0^{T - \tau} \int_0^{T - \tau} (R_{q_i}(u - t)R_{q_i}(u + v - t) + R_{q_i}(u + v - t + \tau)R_{q_i}(u - t - \tau)) dt du \quad (\text{C.7})$$

We rewrite this expression by $r = u - t$.

$$R_s(\tau, v) = \frac{1}{T - \tau} \int_{-T + \tau}^{T - \tau} \left(1 - \frac{|r|}{T - \tau}\right) (R_{q_i}(r)R_{q_i}(r + v) + R_{q_i}(r + v + \tau)R_{q_i}(r - \tau)) dr \quad (\text{C.8})$$

We assume a long time length, T , so we simplify the equation.

$$R_s(\tau, v) \approx \frac{1}{T - \tau} \int_{-\infty}^{\infty} (R_{q_i}(r)R_{q_i}(r + v) + R_{q_i}(r + v + \tau)R_{q_i}(r - \tau)) dr \quad (\text{C.9})$$

References

- [1] R. Brincker, C. Ventura, Introduction to Operational Modal Analysis, John Wiley and Sons, Inc., Chichester, West Sussex, UK, 2015.
- [2] C. Rainieri, G. Fabbro, Operational Modal Analysis of Civil Engineering: An Introduction and Guide for Applications, Springer Science+Business, New York, USA, 2014.
- [3] G. C. James, T. G. Carne, J. P. Lauffer, The Natural Excitation Technique (NExT) for modal parameter extraction, The International Journal of Analytical and Experimental Modal Analysis 10 (4) (1995) 260–277.
- [4] R. Brincker, On the application of correlation function matrices in OMA, Mechanical Systems and Signal Processing 87 (2017) 17–22. doi:10.1016/j.ymssp.2016.08.009.
- [5] J. S. Bendat, A. G. Piersol, Random Data: Analysis and Measurement Procedures, 4th Edition, Wiley-Blackwell, Hoboken, N.J., USA, 2010.
- [6] L. Giampellegrini, Uncertainty in correlation-driven operational modal analysis, Ph.D. thesis, University College London, UK, 64-74 (2007).
- [7] M. Tarpø, P. Olsen, S. Amador, M. Juul, R. Brincker, On minimizing the influence of the noise tail of correlation functions in operational modal analysis, Procedia Engineering 199 (2017) 1038 – 1043, x International Conference on Structural Dynamics, EURODYN 2017. doi:https://doi.org/10.1016/j.proeng.2017.09.282.
URL <http://www.sciencedirect.com/science/article/pii/S1877705817337633>
- [8] B. A. Pridham, J. C. Wilson, A study of damping errors in correlation-driven stochastic realizations using short data sets, Probabilistic Engineering Mechanics 18 (1) (2003) 61 – 77. doi:https://doi.org/10.1016/S0266-8920(02)00042-5.
URL <http://www.sciencedirect.com/science/article/pii/S0266892002000425>
- [9] E. Orlovitz, A. Brandt, Influence of correlation estimation methods on damping estimates, 5th International Operational Modal Analysis Conference, Guimarães, Portugal (2013) 6.
- [10] M. Tarpø, T. Friis, P. Olsen, M. Juul, C. Georgakis, R. Brincker, Automated reduction of statistical errors in the estimated correlation function matrix for operational modal analysis, Mechanical Systems and Signal Processing 132 (2019) 790 – 805. doi:https://doi.org/10.1016/j.ymssp.2019.07.024.
URL <http://www.sciencedirect.com/science/article/pii/S0888327019304522>
- [11] E. Cheynet, J. B. Jakobsen, J. Snæbjörnsson, Damping estimation of large wind-sensitive structures, Procedia Engineering 199 (2017) 2047 – 2053, x International Conference on Structural Dynamics, EURODYN 2017. doi:https://doi.org/10.1016/j.proeng.2017.09.471.
URL <http://www.sciencedirect.com/science/article/pii/S1877705817339656>
- [12] S.-K. Au, Uncertainty law in ambient modal identification—part ii: Implication and field verification, Mechanical Systems and Signal Processing 48 (1) (2014) 34 – 48. doi:https://doi.org/10.1016/j.ymssp.2013.07.017.
URL <http://www.sciencedirect.com/science/article/pii/S0888327013003695>
- [13] ANSI S2.47, Vibration of buildings-guidelines for the measurement of vibrations and evaluation of their effects on buildings, Standard, Acoustical Society of America (Jan. 1991).
- [14] R. Cantieni, Experimental methods used in system identification of civil engineering structures, Proceedings of Workshop Problemi di vibrazioni nelle strutture civili e nelle costruzioni meccaniche.
- [15] R. Pintelon, P. Guillaume, J. Schoukens, Uncertainty calculation in (operational) modal analysis, Mechanical Systems and Signal Processing 21 (6) (2007) 2359 – 2373. doi:https://doi.org/10.1016/j.ymssp.2006.11.007.
URL <http://www.sciencedirect.com/science/article/pii/S0888327006002561>
- [16] E. Reynders, K. Maes, G. Lombaert, G. D. Roeck, Uncertainty quantification in operational modal analysis with stochastic subspace identification: Validation and applications, Mechanical Systems and Signal Processing 66-67 (2016) 13 – 30. doi:https://doi.org/10.1016/j.ymssp.2015.04.018.
URL <http://www.sciencedirect.com/science/article/pii/S0888327015002009>
- [17] S.-K. Au, Uncertainty law in ambient modal identification—part i: Theory, Mechanical Systems and Signal Processing 48 (1) (2014) 15 – 33. doi:https://doi.org/10.1016/j.ymssp.2013.07.016.
URL <http://www.sciencedirect.com/science/article/pii/S0888327013003683>
- [18] E. Orlovitz, A. Brandt, Influence of noise in correlation function estimates for operational modal analysis, Topics in Modal Analysis & Testing, Proceedings of the 36th IMAC 9 (2019) 55–64. doi:https://doi.org/10.1007/978-3-319-74700-2.
- [19] R. W. Clough, J. Penzien, Dynamics of Structures, 2nd Edition, CSI Computers & Structures, Inc., Berkeley, Calif., USA, 2003.
- [20] R. G. Lyons, Understanding Digital Signal Processing, 3rd Edition, Prentice Hall, Upper Saddle River, N.J., USA, 2011.
- [21] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, Cambridge, Mass., USA, 2012.
- [22] S. O. Rice, Mathematical analysis of random noise, Bell System Technical Journal 24 (1) (1945) 46–156.
arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1945.tb00453.x, doi:10.1002/j.1538-7305.1945.tb00453.x.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1945.tb00453.x>
- [23] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, Statistical Analysis of Noise in MRI - Appendix A, Springer International Publishing, Cham, Schweiz, 2016. doi:10.1007/978-3-319-39934-8.
- [24] N. N. Lebedev, Special Functions & Their Applications, Dover Publications, Mineola, NY, USA, 1972.
- [25] H. Gudbjartsson, S. Patz, The rician distribution of noisy mri data, Magnetic Resonance in Medicine 34 (6) 910–914. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.1910340618, doi:10.1002/mrm.1910340618.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910340618>

- [26] R. Greco, G. C. Marano, A. Fiore, I. Vanzi, Nonstationary first threshold crossing reliability for linear system excited by modulated gaussian process, *Shock and Vibration* 2018 (2018) 17. doi:<https://doi.org/10.1155/2018/3685091>.
- [27] S. O. Rice, Statistical properties of a sine wave plus random noise, *The Bell System Technical Journal* 27 (1) (1948) 109–157. doi:[10.1002/j.1538-7305.1948.tb01334.x](https://doi.org/10.1002/j.1538-7305.1948.tb01334.x).
- [28] J. Rainal, Axis-crossings of the phase of the sine wave plus noise, *The Bell System Technical Journal* (1967) 737 – 754.
- [29] S. Cobb, The distribution of intervals between zero crossings of sine wave plus random noise and allied topics, *IEEE Transactions on Information Theory* 11 (2) (1965) 220–233. doi:[10.1109/TIT.1965.1053743](https://doi.org/10.1109/TIT.1965.1053743).
- [30] A. Firouzi, W. Yang, C.-Q. Li, Efficient solution for calculation of upcrossing rate of nonstationary gaussian process, *Journal of Engineering Mechanics* 144 (4) (2018) 04018015. doi:[10.1061/\(ASCE\)EM.1943-7889.0001420](https://doi.org/10.1061/(ASCE)EM.1943-7889.0001420).
- [31] M. Priestley, *Spectral Analysis and Time Series*, Academic Press Inc., Fift Avenue, New York, NY, USA, 1982.
- [32] S. R. Ibrahim, E. C. Mikulcik, A method for the direct identification of vibration parameters from the free response, *The Shock and Vibration Inform. Ctr. Shock and Vibration Bull.* (1977) 183–198.