Uncertainty quantification of power spectrum and spectral moments estimates subject to missing data

Yuanjin Zhang[1],Liam Comerford[2],Ioannis A. Kougioumtzoglou[3]
Edoardo Patelli[1],Michael Beer[124]

[1]Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, U.K.
[2]Institute for Risk and Reliability, Leibniz University of Hannover, Hannover, Germany.
[3]The Fu Foundation School of Engineering and Applied Science, Dept. of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027.
[4]School of Civil Engineering & Shanghai Institute of Disaster Prevention and Relief, Tongji University, China.

ABSTRACT

In this paper, the challenge of quantifying the uncertainty in stochastic process spectral estimates based on realizations with missing data is addressed. Specifically, relying on relatively relaxed assumptions for the missing data and on a Kriging modeling scheme, utilizing fundamental concepts from probability theory, and resorting to a Fourier based representation of stationary stochastic processes, a closed-form expression for the probability density function (PDF) of the power spectrum value corresponding to a specific frequency is derived. Next, the approach is extended for determining the PDF of spectral moments estimates as well. Clearly, this is of significant importance to various reliability assessment methodologies that rely on knowledge of the system response spectral moments for evaluating its survival probability. Further, it is shown that utilizing a Cholesky kind decomposition for the PDF related integrals the computational cost is kept at a minimal level. Several numerical examples are included and compared against pertinent Monte Carlo simulations for demonstrating

the validity of the approach.

Keywords: Uncertainty quantification; Survival probability; Spectral moments; Missing data; Kriging; Spectral estimation

INTRODUCTION

In research fields such as stochastic structural dynamics, stochastic processes are most often described by statistical quantities such as the power spectrum. In this regard, several approaches exist in the literature for stochastic process power spectrum estimation. For instance, a Fourier basis is typically utilized in the spectral estimation of stationary processes (Newland 1993). Further, similar to the stationary case, the evolutionary power spectrum related to non-stationary processes can be estimated by employing wavelet (e.g. (Spanos and Failla 2004); (Kougioumtzoglou et al. 2012)) or chirplet bases (Politis et al. 2007) among other alternatives; see also (Qian 2002) for a detailed presentation of joint time-frequency analysis techniques.

It is noted that the above spectral estimation approaches often require a large number of complete data samples for attaining a predefined adequate degree of accuracy. However, missing data in measurements is frequently an unavoidable situation. In fact, missing data are possible in almost any situation where data are collected and stored. Indicative reasons in engineering dynamics measurement applications include failure and/or restricted use of equipment, as well as data corruption and cost/bandwidth limitations.Thus, standard spectral analysis techniques that inherently assume the existence of full sets of data, such as those based on Fourier, wavelet and chirplet transforms, cannot be used in a straightforward manner.

To address this challenge, a number of signal reconstruction techniques subject to missing/incomplete data (e.g. Lomb-Scargle periodogram, iterative deconvolution method CLEAN, ARMA-model based techniques, etc) have been developed with various degrees of accuracy; see (Wang et al. 2005) for a review. Indicatively, (Comerford et al. 2016) developed recently a compressive sensing approach (e.g. (Eldar and Kutyniok 2012)) based on L1-norm minimization for stationary and non-stationary stochastic process/field (evolutionary) power spectrum estimation subject to highly incomplete data, which has already been applied to practical engineering problems (Comerford et al. 2017; Kougioumtzoglou et al. 2017). The approach has been shown to be particularly advantageous for cases where multiple records/realizations compatible with a stochastic process are available. In such cases, a re-weighting procedure can be introduced to improve the result to a large degree (Comerford et al. 2014). Further, an artificial neural network based approach was also developed recently having the advantage that no prior knowledge of the underlying process is required (Comerford et al. 2015a).

Although all of the above methodologies can, depending on the setting, potentially provide a relatively accurate stochastic process power spectrum estimate, they will also propagate inaccuracies from missing data predictions in the time domain through to the final spectral estimates. Most of the aforementioned techniques estimate the power spectrum by reconstructing missing parts of the data, and based on these reconstructed full data, standard spectral analysis methods are applied. Nevertheless, reconstructing the available records, and thus, deterministically estimating/predicting missing values, rarely accounts for the inherent uncertainty associated with the missing data. Hence, there is merit in developing a methodology for quantifying the uncertainty in a given spectral estimate as a result of the uncertainty related to the missing data in the time/space domain.

In this manner, to quantify the uncertainty of spectral estimates subject to missing data, a stochastic model accounting for the uncertainty in the missing data in the time/space domain can be considered based on any available prior knowledge (e.g. an appropriately estimated probability density function (PDF)). Further, the uncertainty in the missing data can be propagated and the PDF for each individual power spectrum point can be determined in the frequency domain. In this regard, (Comerford et al. 2015b) proposed a methodology and determined a closed form expression for the power spectrum estimate PDF under the assumption that the (missing data) variables in the time domain are independent Gaussian random variables. Note, however, that this approach does not consider the correlation between the missing points, and thus, can be largely unrepresentative, for instance, of a signal with harmonic features. Further, by virtue of the central limit theorem (Billingsley 2008), it is reasonable for many cases (e.g. environmental processes such as earthquakes, winds, sea waves and, for linear systems, the structural responses subject to these effects) to consider the missing points following a multi-variate Gaussian PDF.

In this paper, the approach developed in (Comerford et al. 2015b) is extended to account for the correlation between the missing data. Although determining the exact correlation between points is practically a quite challenging task, an estimate can be obtained by relying on existing available data and employing various modeling schemes such as Kriging (Stein 1999). Further, an additional significant contribution of the herein proposed methodology is that it is generalized to evaluate not only the power spectrum points PDFs, but also the PDFs of the corresponding spectral moments. Clearly, this is of considerable importance to various engineering dynamics applications such as to structural system reliability assessment, where the survival probability (or equivalently, the first-passage time) can be estimated approximately based on knowledge of spectral moments (Vanmarke 1975). Several numerical examples are included and compared against pertinent Monte Carlo simulations for demonstrating the validity of the approach.

MATHEMATICAL FORMULATION

Uncertainty quantification of the power spectrum estimate under missing data

Consider a zero mean stationary process represented as

$$f(t) = \int_{-\infty}^{+\infty} A(\omega) e^{i\omega t} dZ(\omega), \tag{1}$$

(Priestley 1982; Cramer and Leadbetter 1967), where $A(\omega)$ is a deterministic function and $dZ(\omega)$ is a zero mean orthonormal increment stochastic process. The two-sided power spectrum $S_f(\omega)$ of process $f(t)$ is then defined as $S_f(\omega) = |A(\omega)|^2$. In general, realizations of a stochastic process that are compatible with a given spectrum can be generated by a spectral representation methodology (Shinozuka and Deodatis 1991) in the form

$$f(t) = 2\sum_{n=0}^{N-1} \sqrt{S_f(\omega)\Delta\omega}\cos(\omega_n t + \phi_n), \tag{2}$$

where ϕ_n is the independent random phase angle distributed uniformly over the interval $[0, 2\pi]$. The realizations generated by Eq.(2) exhibit the property of ergodicity (Shinozuka and Deodatis 1991); hence, the power spectrum $S_f(\omega)$ of the underlying process can be estimated by utilizing a single realization only. In this regard, and employing the discrete Fourier transform (DFT) yields

$$S_f(\omega_k) = \lim_{N\longrightarrow\infty} \frac{T}{2\pi N^2} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N} \right|^2, \tag{3}$$

where N is the number of data points, t and k are the time and frequency indices respectively, and T is the time duration. In the following, the condition $N \longrightarrow \infty$ is omitted, for convenience, under the assumption that the length is long enough to provide with an accurate spectrum estimate. Following the notation of (Comerford et al. 2015b), the data points are divided into 2 parts: the known points x_α and missing points x_β, where α and β are indices of the known and unknown points, respectively; thus, Eq.(3) can be further cast in the form

$$S_f(\omega_k) = \frac{T}{2\pi N^2} \left| M_1 + M_2 - i(M_3 + M_4) \right|^2 = \frac{T}{2\pi N^2} \left[(M_1 + M_2)^2 + (M_3 + M_4)^2 \right] \tag{4}$$

where $M_1 = \sum_\alpha x_\alpha \cos\left(\frac{2\pi k\alpha}{N}\right)$, $M_2 = \sum_\beta x_\beta \cos\left(\frac{2\pi k\beta}{N}\right)$, $M_3 = \sum_\alpha x_\alpha \sin\left(\frac{2\pi k\alpha}{N}\right)$, and $M_4 = \sum_\alpha x_\alpha \sin\left(\frac{2\pi k\alpha}{N}\right)$. Next, $S_f(\omega_k)$ is rewritten into the simpler form

$$S_f(\omega_k) = (c_1 + a' X_\beta)^2 + (c_2 + b' X_\beta)^2 \tag{5}$$

where ($'$) denotes the transpose,

$$c_1 = \sqrt{\frac{T}{2\pi N^2}} \sum_{\alpha} x_\alpha \cos\left(\frac{2\pi k \alpha}{N}\right) \tag{6}$$

$$c_2 = \sqrt{\frac{T}{2\pi N^2}} \sum_{\alpha} x_\alpha \sin\left(\frac{2\pi k \alpha}{N}\right) \tag{7}$$

$$a = \sqrt{\frac{T}{2\pi N^2}} \left(\cos\left(\frac{2\pi k \beta_1}{N}\right), \cos\left(\frac{2\pi k \beta_2}{N}\right), ..., \cos\left(\frac{2\pi k \beta_u}{N}\right)\right)' \tag{8}$$

$$b = \sqrt{\frac{T}{2\pi N^2}} \left(\sin\left(\frac{2\pi k \beta_1}{N}\right), \sin\left(\frac{2\pi k \beta_2}{N}\right), ..., \sin\left(\frac{2\pi k \beta_u}{N}\right)\right)' \tag{9}$$

and

$$X_\beta = (x_{\beta 1}, x_{\beta 2}, ..., x_{\beta u})' \tag{10}$$

where u is the number of missing points.

By virtue of the central limit theorem (Billingsley 2008), it is reasonable in many cases to make the approximation that missing points follow a multi-variate Gaussian PDF. In this regard, the various statistical quantities such as the mean and variance for each missing point as well as the correlation between missing points are taken into consideration. In the ensuing analysis, it is assumed that the mean and correlation matrix of the missing data following a Gaussian distribution, i.e. $X_\beta \sim N(\mu, \Sigma)$, are obtained by some available estimation scheme, such as the Kriging model; see following section for more details.

Next, Eq.(5) is rearranged (see also (Papoulis and Pillai, 2002)) as a function of two variables in the form

$$S_f(\omega_k) = (c_1 + a'X_\beta)^2 + (c_2 + b'X_\beta)^2 = X_1^2 + X_2^2 \tag{11}$$

It is readily seen that $X_1 = c_1 + a'X_\beta \sim N(c_1 + a'\mu, a'\Sigma a)$ and $X_2 = c_2 + b'X_\beta \sim N(c_2 + b'\mu, b'\Sigma b)$. Because both X_1 and X_2 are related to the same set of random variables X_β, it is obvious that they exhibit some degree of correlation. In this regard, the correlation matrix $C_{X_1X_2}$ of joint Gaussian variables X_1 and X_2 is given by

$$C_{X_1X_2} = \begin{pmatrix} a'\Sigma a & \sum_i \sum_j a_i b_j (\Sigma_{ij} + \mu_1\mu_2) - b'\mu a'\mu \\ \sum_i \sum_j a_i b_j (\Sigma_{ij} + \mu_1\mu_2) - b'\mu a'\mu & b'\Sigma b \end{pmatrix} \tag{12}$$

and the mean vector of joint Gaussian variables X_1 and X_2 takes the form

$$\mu_{X_1X_2} = (c_1 + \mu, c_2 + \mu)' \tag{13}$$

Further, to determine the PDF of the variable $S_f(\omega_k)$ in Eq.(11), the celebrated input-output PDF relationship (Papoulis and Pillai 2002) is applied, and the cumulative distribution function (CDF) of $S_f(\omega_k)$ is defined as

$$F(S_f) = P(S_f \leq s) = P[(X_1, X_2) \in D_s] = \iint_{(X_1,X_2)\in D_s} f_{X_1,X_2}(X_1, X_2) dX_1 dX_2 \tag{14}$$

where D_s is the region such that $X_1^2 + X_2^2 \leq s$ is satisfied, $f_{X_1,X_2}(X_1, X_2)$ is the joint PDF of the variables X_1 and X_2; the PDF of $S_f(\omega_k)$ is given by

$$f_s(s) = \frac{dF(S_f)}{ds} \tag{15}$$

Thus, taking into account Eqs. (11-15), an analytical expression for the power spectrum PDF at a given frequency ω_k is derived in the form

$$\begin{aligned} p_{S_f(\omega_k)}(s) =& \frac{d}{ds} \iint_{X_1^2+X_2^2\leq s} \frac{1}{2\pi\sqrt{|C_{X_1X_2}|}} \\ & \exp\left[-\frac{1}{2}(X - \mu_{X_1X_2})' C_{X_1X_2}^{-1} (X - \mu_{X_1X_2})\right] dX_1 dX_2 \end{aligned} \tag{16}$$

In this section an approach has been developed for quantifying the uncertainty in a stochastic process power spectrum estimate subject to missing data. Specifically, a closed form analytical expression has been derived in Eq.(16) for the power spectrum estimate PDF corresponding to a given frequency. In comparison with the methodology in (Comerford et al. 2015b), which adopts the assumption that missing data in a given realization are independent and identically distributed Gaussian random variables, the rather strict assumption of independence is abandoned herein. In this manner, the correlation between the missing data is taken into account in estimating the power spectrum PDF.

Kriging model for estimating correlations between missing data

Clearly, the approach developed in the previous section relies on prior knowledge of the correlation between the missing data. Among the various available techniques in the literature for estimating data correlation relationships a Kriging based scheme (e.g. (Stein 1999); (Gaspar et al. 2014) and (Jia and Taflanidis 2013)) is considered in the ensuing analysis.

Specifically, let $f(t)$ be a sample of a stationary stochastic process with a power spectrum $S_f(\omega)$. Given the n known points $t_i, i = 1, 2, ..., n$, an estimate of $f(t_j)$ at the missing point t_j, can be obtained as a weighted linear combination of the available known points (Stein 1999), i.e.,

$$f(t_j) = \sum_{i=1}^{n} \lambda_i f(t_i) + z(t) \tag{17}$$

where λ_i is the weight of each known point, and $z(t)$ is a stationary Gaussian process with zero mean and covariance

$$C = cov\left(z(t_i), z(t_i - t_j)\right) = \gamma\left(|t_i - t_j|\right) = \sigma_z^2 R\left(|t_i - t_j|\right) \tag{18}$$

where σ_z^2 is the constant variance of the process and R is the correlation function. Several types of correlation functions, such as exponential, linear and Gaussian, have been proposed in the literature (Kaymaz 2005). Herein, a correlation function of exponential form is adopted due to its applicability in a wide range of engineering processes (Spanos et al. 2007), i.e.

$$\gamma(h) = \sigma_z^2 e^{-\theta_1|h|}\cos(\theta_2 h)(1+\theta_1|h|) \tag{19}$$

where $h = t_i - t_j$ is the interval between two time instants, and θ_1, θ_2 are constant values to be determined. Next, σ_z^2, θ_1 and θ_2 are obtained by least-squares fitting of eq.19 to the available data, i.e.,

$$\min_{\sigma_z^2,\theta_1,\theta_2} |\gamma(h) - \gamma_e(h)|_2 \tag{20}$$

where $|\cdot|_2$ denotes the L-2 norm, $\gamma_e(h) = \frac{1}{n}\sum_{i=1}^{n}[f(t_i+h)f(t_i)]$, and $f(t_i+h), f(t_i)$ are the known points.

Further, utilizing the Kriging model of Eq.(17) the estimate error variance is given by

$$V = Var[f^*(t_j) - f(t_j)] = 2\sum_{i=1}^{n}\lambda_i\gamma(|t_i - t_j|) - \sum_{i=1}^{n}\sum_{k=1}^{n}\lambda_i\lambda_k\gamma(|t_i - t_k|) - \sigma_z^2 \tag{21}$$

Next, to minimize the error variance V, a Lagrange multipliers approach is applied yielding the equations

$$\begin{cases} \sum_{i=1}^{n}\lambda_i\gamma(|t_i - t_k|) + \kappa = \gamma(|t_i - t_j|), (j = 1, ..., n) \\ \sum_{i=1}^{n}\lambda_i = 1 \end{cases} \tag{22}$$

to be solved for the weights λ_i and Lagrange multiplier κ . Further, an estimate of the missing point is given by Eq.(17). Then, the covariance matrix C of the sample could be easily obtained through Eq.(18).

Note that, denoting the time history vector x as $x = (x_\beta, x_\alpha)$, the covariance matrix C can be expressed as $C = \begin{pmatrix} C_{\beta\beta} & C_{\beta\alpha} \\ C_{\alpha\beta} & C_{\alpha\alpha} \end{pmatrix}$, where $C_{\beta\beta}$ is the matrix whose rows and columns correspond to the missing points x_β, while $C_{\alpha\alpha}$ corresponds to the known points x_α. In this regard, the conditional covariance matrix Σ of the missing points is calculated as (Papoulis and Pillai 2002)

$$\Sigma = C\{x_\beta|x_\alpha\} = C_{\beta\beta} - C_{\beta\alpha}C_{\alpha\alpha}^{-1}C_{\alpha\beta} \tag{23}$$

Overall, adopting a Kriging modeling approach in this section, the mean and covariance of missing data are estimated, and can be used as an input to the approach developed in the previous section.

Stochastic process spectral moment estimate uncertainty quantification under missing data

For stationary random processes, the spectral moments are defined as

$$\lambda_i = \int_{-\infty}^{+\infty} \omega^i S(\omega) d\omega \tag{24}$$

where $S(\omega)$ is the two-sided power spectrum (e.g. (Lutes and Sarkani 2004)). Considering next the case of a zero mean process, the zero spectral moment λ_0 is equal to the mean square $E[X^2]$ of the process X (also equal to the squared standard deviation σ_X^2 in this case), and the second spectral moment λ_2 is the mean square $E[\dot{X}^2]$ of the derivative process X.In a similar manner as the moments of a random variable are used to describe certain features of the related PDF, spectral moments are indispensable in a variety of applications such as determining approximately the survival probability (or equivalently, the first-passage time) and assessing the reliability of structural systems (e.g. (Vanmarke 1972); (Vanmarke 1975); (Lutes and Sarkani 2004)).

Further, Eq.(24) can be recast into a discrete form in the frequency domain, i.e.

$$\lambda_i = \sum_n \omega_n^i S(\omega_n) \Delta\omega \tag{25}$$

Clearly, based on Eq.(25) the spectral moment can be viewed as a linear combination of individual power spectrum points. Note that although the PDFs of the power spectrum points $S(\omega_n)$ can be obtained by the methodology developed in the previous sections, a straightforward determination of the PDF of the spectral moment λ_i can be quite daunting due to the following reasons. First, the various power spectrum points $S(\omega_n)$ do not, in general, follow the same PDF for different frequency values ω_n. Second, the variables $S(\omega_n)$ exhibit correlation as they are defined by utilizing the same set of random variables.

Next, to address these challenges, a methodology based on characteristic functions is proposed. The characteristic function of a random variable is defined as (Papoulis and Pillai 2002)

$$\Phi_X(\omega) = E[e^{i\omega x}] = \int_{-\infty}^{+\infty} f_X(x) e^{i\omega x} dx \tag{26}$$

where $f_X(x)$ is the probability density function of X. Clearly, the characteristic function and the PDF of a random variable form a Fourier transform pair. Further, the spectral moment Eq.(25) can be construed as a quadratic transformation of the missing points X_β. The correlated variables $X_\beta \sim N(\mu, \mathit{\Sigma})$, where $\mathit{\Sigma}$ can be cast into the Cholesky factorization form $\mathit{\Sigma} = AA'$ (A being a lower triangular matrix), are replaced by a new set of independent standard Gaussian variables $X_g \sim N(\mathit{0}, \mathit{I})$ as

$$X_\beta = \mu + AX_g \tag{27}$$

Next, employing Eqs.(25-27), Eq.(5) can be cast in the matrix form

$$S_f(\omega_k) = (c_{1,k} + a_k'\mu + a_k'X_g)^2 + (c_{2,k} + b_k'\mu + b_k'X_g)^2 = X_{gn}' B_k X_{gn} \tag{28}$$

where $c_{1,k}$, $c_{2,k}$, a_k, and b_k are defined by Eq.(6-9),

$$X_{gn} = [X_g', 1]' = [x_{g1}, x_{g2}, ..., x_{gu}, 1]', \tag{29}$$

and

$$B_{k,ij} = \begin{cases} a_{k,i}a_{k,i} + b_{k,i}b_{k,i}, & i,j \leq u \\ (c_{1,k} + a'_k\mu)a_{k,i} + (c_{2,k} + b'_k\mu)b_{k,i}, & j = u+1, i \neq u+1 \\ (c_{1,k} + a'_k\mu)a_{k,j} + (c_{2,k} + b'_k\mu)b_{k,j}, & i = u+1, j \neq u+1 \\ (c_{1,k} + a'_k\mu)^2 + (c_{2,k} + b'_k\mu)^2, & i = j = u+1 \end{cases} \tag{30}$$

Combining Eqs.(25) and (29), the spectral moments are given, alternatively, in the form

$$\lambda_i = X'_{gn}\left(\sum_k \omega_k^i \Delta\omega B_k\right) X_{gn} \tag{31}$$

whereas utilizing Eq.(31) the characteristic function of the spectral moments becomes (Papoulis and Pillai 2002)

$$\Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] = \int_{-\infty}^{+\infty} (2\pi)^{-\frac{u}{2}} \exp\left(-\frac{1}{2}\left[X'_g X_g - i\omega X'_{gn}\left(\sum_k \omega_k^i \Delta\omega B_k\right) X_{gn}\right]\right) dx_g \tag{32}$$

Note that, the evaluation of Eq.(32) can be simplified based on the following steps. Specifically,

1) Let

$$Y = \frac{1}{2}\left[X'_g X_g - i\omega X'_{gn}\left(\sum_k \omega_k^i \Delta\omega B_k\right) X_{gn}\right] \tag{33}$$

Eq.(33) can be divided into two parts, i.e., $Y = Y_1 + Y_2$. The first includes the second order terms, i.e. $Y_1 = \sum_{i,j} c_{ij} x_{gi} x_{gj}$,while the second includes the first order terms plus the constant term, i.e. $Y_2 = \sum_i c_i x_{gi} + c_{cons}$. Thus, Eq.(32) can be rewritten as

$$\Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] = \int_{-\infty}^{+\infty} (2\pi)^{-\frac{u}{2}} e^{-Y_1 - Y_2} dx_g \tag{34}$$

2) Similar to Eq.(31), Y_1 can be expressed as $Y_1 = X_g' B_{Y_1} X_g$ where B_{Y_1} is given by

$$B_{Y_1} = A_{Y_1}' A_{Y_1} \tag{35}$$

In Eq.(35) A_{Y_1} is a complex upper triangular matrix. Here, A_{Y_1}' indicates the non-conjugate transpose of A_{Y_1}, similarly in Eq.36. The factorization in Eq.(35) is numerically implemented via a Cholesky factorization kind algorithm (Golub and Van Loan 1996) with the note that the diagonal elements in B_{Y_1} are complex values.

3) After obtaining the upper triangular matrix A_{Y_1}, Y may be expressed in a similar form to Y_1 (after accounting for first order terms and the constant); thus simplifying the solution of the integral in Eq.(34). Hence

$$Y = (A_Y X_{gn})'(A_Y X_{gn}) + c_Y \tag{36}$$

where $A_Y = (A_{Y_1}, a_{u\times 1})$, and $a_{u\times 1}$ are the coefficients to account for the first order terms $\sum_i X_{gi}$ in Y_2 (with u being the number of missing data); and c_Y is a constant. A worked 2-variable example is shown in detail in Appendix.

4) Finally, substituting Eq.(36) into Eq.(32), the integral in Eq.(32) may be simplified significantly to a function of B_{Y_1}, and the constant term c_Y in the form

$$\Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] = 2^{-\frac{u}{2}}(\det(B_{Y_1}))^{-\frac{1}{2}} e^{-c_Y} \tag{37}$$

whereas the spectral moments PDFs are estimated via the inverse Fourier transform of Eq.(32), i.e.

$$p_{\lambda_i}(s) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_{\lambda_i}(\omega) e^{i\omega s} d\omega \tag{38}$$

In this section an efficient approach has been developed for quantifying the uncertainty in the spectral moments estimates of an underlying stochastic process based on available realizations with missing data. Specifically, a closed form expression has been derived in Eq.(32) for the spectral moment characteristic function. The rather daunting brute force numerical evaluation of the integral appearing in the derived expression has been conveniently circumvented via a Cholesky kind decomposition of the integrand function. Clearly, the development in this section is of considerable importance (as illustrated in the following section) to various engineering dynamics applications such as to structural system reliability assessment (Vanmarke 1975).

Survival probability estimate uncertainty quantification under missing data

A persistent challenge in the field of stochastic dynamics has been the determination of the system survival probability, i.e. the probability that the structural system response will stay below a certain threshold over a given period of time. Many research efforts for addressing the aforementioned challenge exist in the literature ranging from semi-analytical to purely numerical approaches (e.g. (Spanos and Kougioumtzoglou 2014); (Bucher 2001); (Au and Beck 2001)). One of the first semi-analytical approximate approaches proposed by Vanmarke (Vanmarke 1975) that relies on the knowledge of the system response spectral moments (Vanmarke 1972) is considered next.

Specifically, consider a linear single-degree-of-freedom (SDOF) oscillator, whose motion is governed by the stochastic differential equation

$$\ddot{x} + 2\zeta_0\omega_0\dot{x} + \omega_0^2 x = w(t) \tag{39}$$

where x is the response displacement, a dot over a variable denotes differentiation with respect to time t; ζ_0 is the ratio of critical damping; ω_0 is the oscillator natural frequency and $w(t)$ represents a Gaussian, zero-mean stationary stochastic process possessing a broad-band power spectrum $S(\omega)$. Focusing next on the stationary response of the oscillator, the response displacement and velocity power spectra are given by (Newland 1993)

$$S_X(\omega) = |H(\omega)|^2 S(\omega) \tag{40}$$

and

$$S_{\dot{X}}(\omega) = \omega^2 S_X(\omega) = \omega^2 |H(\omega)|^2 S(\omega) \tag{41}$$

respectively; and the frequency response function $H(\omega)$ is given by

$$H(\omega) = \frac{1}{\omega_0^2 - \omega^2 + 2i\zeta_0\omega_0\omega} \tag{42}$$

According to (Vanmarke 1975) and (Crandall 1970), the time-dependent survival probability $L_D(t)$ of a linear oscillator given a barrier level D can be approximated by

$$L_D(t) = \exp\left[-\frac{1}{\pi}\sqrt{\frac{\lambda_{X,2}}{\lambda_{X,0}}}t \ \exp\left(-\frac{D^2}{2\lambda_{X,0}}\right)\right] \tag{43}$$

where $\lambda_{X,i}$ is the i-th order spectral moment of the displacement x. Note that for the specific case of the linear oscillator of Eq.(39), and considering a low value for the damping ratio, i.e. $\zeta_0 \leq 0.05$, its response exhibits a narrow-band feature in the frequency domain due to the form of the frequency response function (see Eq.(40)). In particular, it can be seen that $|H(\omega)|^2$ is a function with a sharp peak around the oscillator natural frequency $\omega = \omega_0$, and decays quickly for $\omega \neq \omega_0$. Thus, it is reasonable to assume that the response of the linear oscillator exhibits a pseudo-harmonic behavior (Spanos 1978), and the response displacement and velocity can be represented, respectively, as

$$x = a\cos(\omega_0 t + \varphi) \tag{44}$$

and

$$\dot{x} = -a\omega_0 \sin(\omega_0 t + \varphi) \tag{45}$$

In Eq.(44), a and φ represent the response amplitude and phase processes, respectively; see also (Spanos 1978) and (Kougioumtzoglou and Spanos 2012) for more details. Considering next Eqs.(44-45), the independence of a with φ and taking into account that $E(\cos^2(\omega_0 t + \varphi)) = E(\sin^2(\omega_0 t + \varphi))$ yields

$$E(\dot{x}^2) = \omega_0^2 E(x^2) \tag{46}$$

or in other words

$$\lambda_{X,2} = \omega_0^2 \lambda_{X,0} \tag{47}$$

Substituting Eq.(47) into Eq.(43) yields an approximate expression for the oscillator survival probability that depends only on $\lambda_{X,0}$, i.e.

$$L_D(t) = \exp\left[-\frac{\omega_0}{\pi} t \ \exp\left(-\frac{D^2}{2\lambda_{X,0}}\right)\right] \tag{48}$$

In Eq.(48), the analytical expression for the PDF of $\lambda_{X,0}$ in the case of missing data can be derived by the methodology described in the previous sections. After determining the PDF $p_{\lambda_{X,0}}$, the system survival probability characteristic function can be obtained as

$$\Phi_{L_D}(\omega_k) = E[e^{i\omega_k L_D}] = \int_{-\infty}^{+\infty} e^{i\omega_k L_D} p_{\lambda_{X,0}} d\lambda_{X,0} \tag{49}$$

whereas, an inverse Fourier transform can applied to Eq.(49) for numerically evaluating the survival probability PDF.

NUMERICAL EXAMPLES

Excitation records with missing data

To demonstrate the validity of the developed uncertainty quantification approach, stationary stochastic process time histories compatible with the Kanai-Tajimi-like earthquake engineering power spectrum of the form

$$S(\omega) = S_0 \frac{\omega_g^4 + 4\zeta_g^2\omega_g^2\omega^2}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2\omega_g^2\omega^2} \tag{50}$$

where $\omega_g = 5\pi rad/s$ and $\zeta_g = 0.63$, are generated via Eq.(2) with a time duration of 8.64 seconds and time step of 0.039 seconds. To compare with the method described in (Comerford et al. 2015b), a factor $S_0 = 0.011$ is introduced to make the standard deviation equal to 1. Next, uniformly randomly distributed missing data are artificially induced to provide a Monte-Carlo simulation comparison; $10,000$ samples are used in the following results.

Figure 1 shows the estimated power spectrum PDFs and confidence ranges determined via the herein developed approach for 10% missing data. For comparison purposes Figure 2 is the result of applying the methodology in (Comerford et al. 2015b), where correlations between missing data are not taken into consideration and the missing points follow independent identical Gaussian distributions $X_\beta \sim N(0, I)$. Compared with Figure 2, the method developed herein provides with a smaller range, and the mean spectrum fits the original spectrum better. Figure 3 shows the PDFs corresponding to frequencies 10.9 and 30.5 rad/s with 10% missing data replaced both by correlated and by independent identically distributed Gaussian random variables. The vertical lines correspond to the spectral values without missing data. Figure 4 shows the spectral moment λ_0 of the excitation spectrum, compared with pertinent Monte Carlo simulations. It can be readily seen that in all cases accounting for the correlation of the missing data, as estimated via the Kriging model, yields spectral estimates PDFs that are much closer to the true value.

Structural response records with missing data

In the second example, consider a linear oscillator with $\omega_0 = 10.9 rad/s$, and $\zeta_0 = 0.05$. Further, the missing data are introduced into the stationary records of the oscillator response, which are generated by utilizing the same excitation spectrum as in the first example, and by numerically solving the equation of motion. Similarly, the artificially induced missing data in the response records are uniformly randomly distributed, this time with $100,000$ Monte-Carlo samples utilized for increased accuracy in the spectral moment comparison.

Figure 5 shows the power spectrum PDF and confidence ranges of the oscillator response with 70% missing data determined by the herein developed methodology. For comparison purposes Figure 6 is the result of applying the methodology in (Comerford et al. 2015b), where correlations between missing data are not taken into consideration and the missing points follow independent identical Gaussian distributions. As anticipated, it can be readily seen that neglecting the correlation structure in the missing data has a bigger negative effect when considering narrow-band signals (see Figures 5 and 6) rather than broad-band ones (see Figures 1 and 2). In fact, for the highly correlated oscillator response process disregarding the correlation structure yields an almost constant power spectrum estimate value. Figure 7 shows the PDF of the response spectral moment λ_0, compared with pertinent Monte Carlo simulations. In Figure 8 the PDF of the oscillator survival probability Eq.(48) with 70% missing data and a barrier level $a = 0.05$ is plotted and compared with pertinent Monte Carlo simulations of Eq.(43).

CONCLUSION

In this paper, an analytical approach for quantifying the uncertainty in stochastic process power spectrum estimates based on samples with missing data has been developed. Specifically, the correlations between the missing data are considered by employing a Kriging model, while utilizing fundamental concepts from probability theory, and resorting to a Fourier based representation of stationary stochastic processes, a closed form expression has been derived for the power spectrum estimate PDF at each frequency. Next, the approach has been extended for determining the PDF of spectral moments estimates as well. This is of considerable significance to reliability assessment methodologies as well, where spectral moments are used for evaluating the survival probability of the system. Further, it has been shown that utilizing a Cholesky kind decomposition for the PDF related integrals the computational cost is kept at a minimal level. Several numerical examples have been presented and compared against pertinent Monte Carlo simulations for demonstrating the validity of the approach.

ACKNOWLEDGEMENTS

The first author is grateful for the financial support from the China Scholarship Council.

APPENDIX

By factorizing part of the integrand of Eq.(32) (given as Y in Eq.(33), the solution of Eq.(32) may be greatly simplified. In the following, a 2-variable case is given as an example.

For a 2-variable case, Eq. (31) becomes

$$\lambda_i = ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f \tag{51}$$

where a, b, c, d, e, f are real constant with $a > 0, c > 0, f > 0$. Eq.(51) can be also recast into a matrix form as

$$\lambda_i = \begin{pmatrix} x_1 & x_2 & 1 \end{pmatrix} \begin{pmatrix} a & 0.5b & 0.5d \\ 0.5b & c & 0.5e \\ 0.5d & 0.5e & f \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \tag{52}$$

Further, according to Eq.(33), Y has the form

$$Y = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - i\omega(ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f) \tag{53}$$

The object of step 3 is to recast Eq.(53) into the form given by Eq.(36). To achieve this goal, second order terms of Y are separated and then factorized as follows,

$$\begin{aligned} Y_1 &= \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - i\omega(ax_1^2 + bx_1x_2 + cx_2^2) \\ &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0.5 - i\omega a & -0.5i\omega b \\ -0.5i\omega b & 0.5 - i\omega c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} A'_{Y_1} A_{Y_1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned} \tag{54}$$

where $A_{Y_1} = \begin{pmatrix} \sqrt{0.5 - i\omega a} & -\frac{i\omega b}{2\sqrt{0.5 - i\omega a}} \\ 0 & \sqrt{\frac{\omega^2 b^2}{2 - 4i\omega a} + 0.5 - i\omega c} \end{pmatrix}$, and A'_{Y_1} is the non-conjugate transpose of A_{Y_1}, i.e., $A'_{Y_1} A_{Y_1} = \begin{pmatrix} 0.5 - i\omega a & -0.5i\omega b \\ -0.5i\omega b & 0.5 - i\omega c \end{pmatrix}$. This calculation can use follow the same numerical implementation steps as a Cholesky factorization algorithm with the note that $\begin{pmatrix} 0.5 - i\omega a & -0.5i\omega b \\ -0.5i\omega b & 0.5 - i\omega c \end{pmatrix}$ is not a Hermitian positive-definite matrix. Then, extending Y_1 to account for the first order terms in Eq.(53), Y may be written as,

$$\begin{aligned} Y &= \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - i\omega(ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f) \\ &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} A'_{Y_1} A_{Y_1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - i\omega(dx_1 + ex_2 + f) \\ &= (A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix})'(A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}) + c_Y \end{aligned} \tag{55}$$

where $A_Y = \begin{pmatrix} \sqrt{0.5 - i\omega a} & -\frac{i\omega b}{2\sqrt{0.5-i\omega a}} & -\frac{i\omega d}{2\sqrt{0.5-i\omega a}} \\ 0 & \sqrt{\frac{\omega^2 b^2}{2-4i\omega a} + 0.5 - i\omega c} & \frac{\frac{bd\omega^2}{1-2i\omega a} - i\omega e}{2\sqrt{\frac{\omega^2 b^2}{2-4i\omega a} + 0.5 - i\omega c}} \\ 0 & 0 & 0 \end{pmatrix}$, $c_Y = -(-\frac{i\omega d}{2\sqrt{0.5-i\omega a}})^2 - (\frac{\frac{bd\omega^2}{1-2i\omega a} - i\omega e}{2\sqrt{\frac{\omega^2 b^2}{2-4i\omega a} + 0.5 - i\omega c}})^2 - i\omega f$.

Calculating the first term in Eq.(55), it can be seen that $(A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix})'(A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix})$ takes the form

$$(A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix})'(A_Y \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}) = (m_1 x_1 + m_2 x_2 + m_3)^2 + (m_4 x_2 + m_5)^2 \tag{56}$$

where the constants m_1, m_2, m_3, m_4, m_5 are calculated by A_Y. Hence, Y may be written as

$$Y = (m_1 x_1 + m_2 x_2 + m_3)^2 + (m_4 x_2 + m_5)^2 + c_Y \tag{57}$$

The form Eq.(57) is particularly useful in calculating the integral in Eq.(32), allowing it to be simplified as shown

$$\begin{aligned} \Phi_{\lambda_i}(\omega) = E[e^{i\omega\lambda_i}] &= \int_{-\infty}^{+\infty} (2\pi)^{-\frac{u}{2}} exp(-Y) dx_g \\ &= (2\pi)^{-1} \iint_{-\infty}^{+\infty} exp[-(m_1 x_1 + m_2 x_2 + m_3)^2 - (m_4 x_2 + m_5)^2 - c_Y] dx_1 dx_2 \\ &= (2\pi)^{-1} \frac{\sqrt{\pi}}{m_1} \int_{-\infty}^{+\infty} exp[-(m_4 x_2 + m_5)^2 - c_Y] dx_2 \\ &= \frac{1}{2 m_1 m_4} exp(-c_Y) \end{aligned} \tag{58}$$

For the general multi-variable case, the above steps are the same.

REFERENCES

Au, S.-K. and Beck, J. (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." *Prob. Eng. Mech.*, 16(4), 263–275.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Bucher, C. (2001). "Simulation methods in structural reliability." *Marine Technology and Engineering*, 2, 1071–1086.

Comerford, L., Jensen, H., Mayorga, F., Beer, M., and Kougioumtzoglou, I. (2017). "Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data." *Computers & Structures*, 182, 26–40.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2014). "Compressive sensing based power spectrum estimation from incomplete records by utilizing an adaptive basis." *Proceedings of IEEE SSCI 2014*, 117–124. doi:10.1109/CIES.2014.7011840.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2015a). "An artificial neural network approach for stochastic process power spectrum estimation subject to missing data." *Structural Safety*, 52, 150–160.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2015b). "On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data." *International Journal of Sustainable Materials and Structural Systems*, 2(1-2), 185–206.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2016). "Compressive sensing based stochastic process power spectrum estimation subject to missing data." *Probabilistic Engineering Mechanics*. doi:10.1016/j.probengmech.2015.09.015.

Cramer, H. and Leadbetter, R. (1967). *Stationary and related stochastic processes: sample function properties and their applications*. Wiley.

Crandall, S. H. (1970). "First-crossing probabilities of the linear oscillator." *Journal of Sound Vibration*, 12(3), 285.

Eldar, Y. and Kutyniok, G. (2012). *Compressed Sensing: Theory and Apps*. Cambridge Uni Press.

Gaspar, B., Teixeira, A. P., and Guedes, S. C. (2014). "Assessment of the efficiency of kriging surrogate models for structural reliability analysis." *Probabilistic Engineering Mechanics*, 37, 24–34.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. Johns Hopkins, Baltimore.

Jia, G. and Taflanidis, A. (2013). "Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment." *Computer Methods in Applied Mechanics and Engineering*, 261-262, 24–38.

Kaymaz, I. (2005). "Application of kriging method to structural reliability problems." *Struct. Saf.*, 27(2), 133–151.

Kougioumtzoglou, I. A., dos Santos, K. R., and Comerford, L. (2017). "Incomplete data based parameter identification of nonlinear and time-variant oscillators with fractional derivative elements." *Mechanical Systems and Signal Processing*, 94, 279 – 296.

Kougioumtzoglou, I. A., Kong, F., Spanos, P. D., and Li, J. (2012). "Some observations on wavelets based evolutionary power spectrum estimation." *Proceedings of the Stochastic Mechanics Conference (SM12)*, 37–44. ISSN: 2035-679X.

Kougioumtzoglou, I. A. and Spanos, P. D. (2012). "An analytical wiener path integral technique for non-stationary response determination of nonlinear oscillators." *Probabilistic Engineering Mechanics*, 22, 125–131.

Lutes, L. D. and Sarkani, S. (2004). *Random vibrations: analysis of structural and mechanical systems*. Elsevier.

Newland, D. E. (1993). *An introduction to random vibrations, spectral and wavelet analysis*. Dover Publications.

Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill.

Politis, N. P., Giaralis, A., and Spanos, P. D. (2007). "Joint time-frequency representation of simulated earthquake accelerograms via the adaptive chirplet transform." *Proc. Computational stochastic Mechanics*, 549–557.

Priestley, B. (1982). *Spectral analysis and time series*. Academic Press.

Qian, S. (2002). *Introduction to Time-Frequency and Wavelet Transforms*. Prentice Hall.

Shinozuka, M. and Deodatis, G. (1991). "Simulation of stochastic processes by spectral representation." *Appl. Mech. Rev.*, 44(4), 192–203.

Spanos, P. D. (1978). "Non-stationary random vibration of a linear structure." *International Journal of Solids and Structures*, 14, 861–867.

Spanos, P. D., Beer, M., and Red-Horse, J. (2007). "Karhunen–loève expansion of stochastic processes with a modified exponential covariance kernel." *Journal of Engineering Mechanics*, 133(7), 773–779.

Spanos, P. D. and Failla, G. (2004). "Evolutionary spectra estimation using wavelets." *J. Eng. Mech.*, 130, 952–960.

Spanos, P. D. and Kougioumtzoglou, I. A. (2014). "Survival probability determination of nonlinear oscillators subject to evolutionary stochastic excitation." *Journal of Applied Mechanics*, 81, 051016–1 051016–9.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.

Vanmarke, E. H. (1972). "Properties of spectral moments with applications to random vibration." *Journal of Engineering Mechanics Division*, 98, 425.

Vanmarke, E. H. (1975). "On the distribution of the first passage time for normal stationary random process." *Journal of Applied Mechanics*, 215–220.

Wang, Y., Li, J., and Stoica, P. (2005). *Spectral Analysis of Signals, the Mising Data Case*. Morgan & Caypool.

List of Figures

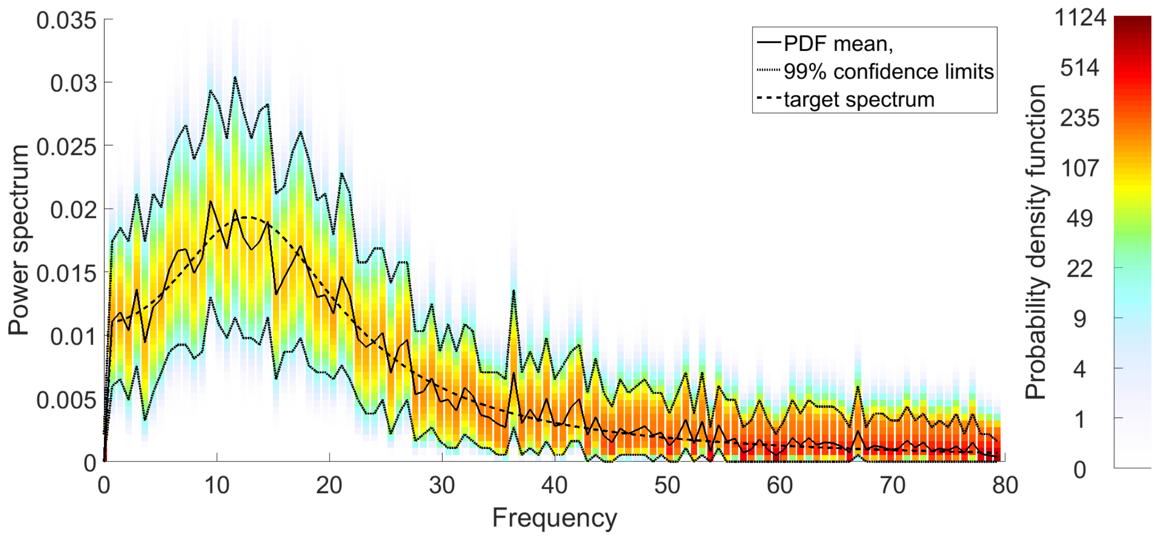

FIG. 1. Power spectrum probability densities with 10% missing data replaced by correlated Gaussian random variables

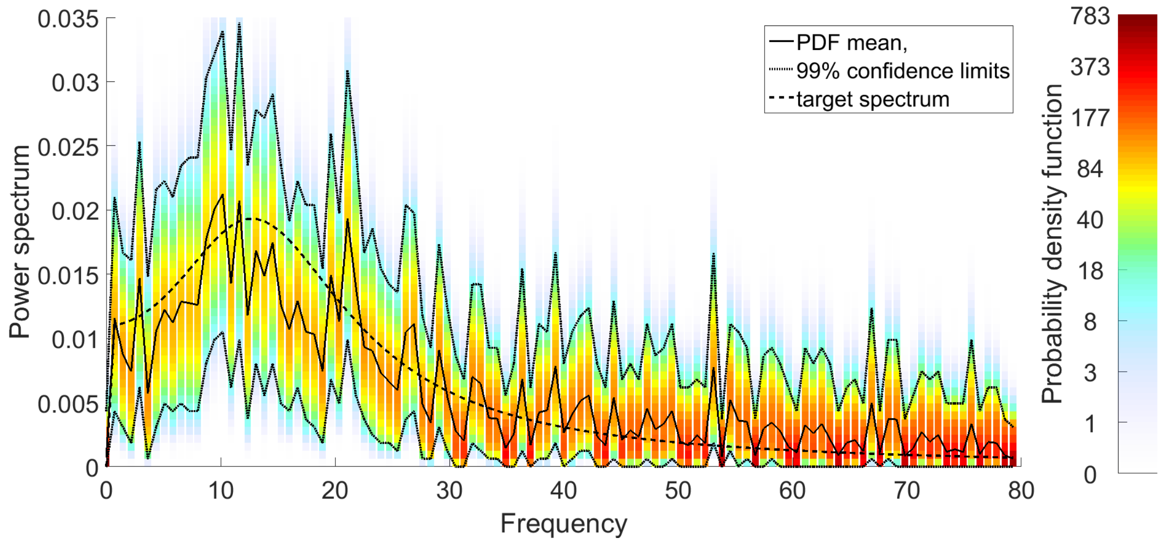

FIG. 2. Power spectrum probability densities with 10% missing data replaced by independent identically distributed Gaussian random variables

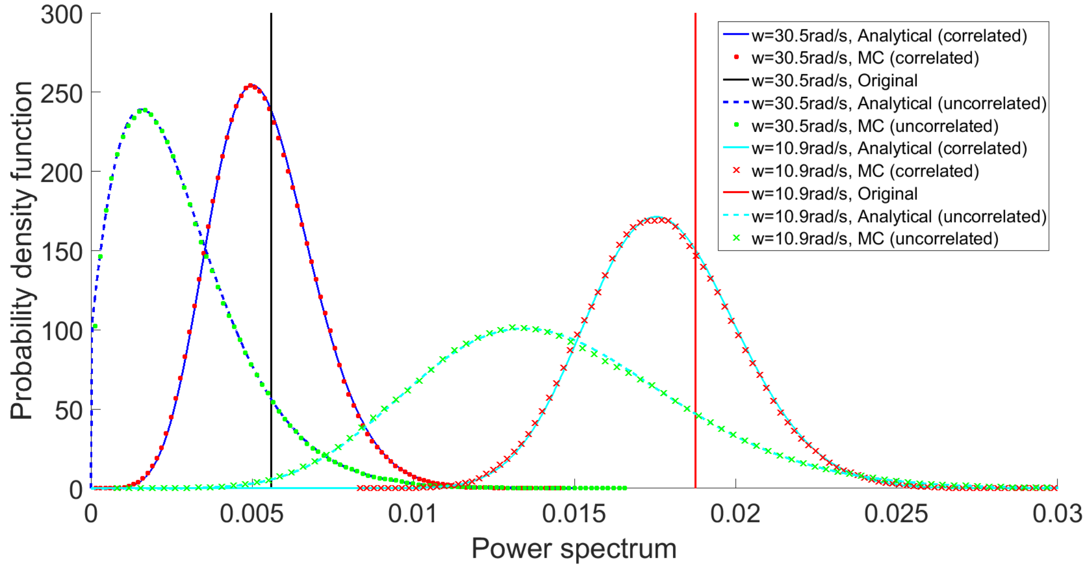

FIG. 3. PDFs at 10.9 and 30.5 rad/s with 10% missing data replaced by both correlated and independent identically distributed Gaussian random variables. Monte-Carlo estimated PDFs (MC) are shown for validation of the procedure. The vertical line shows the spectral value without missing data

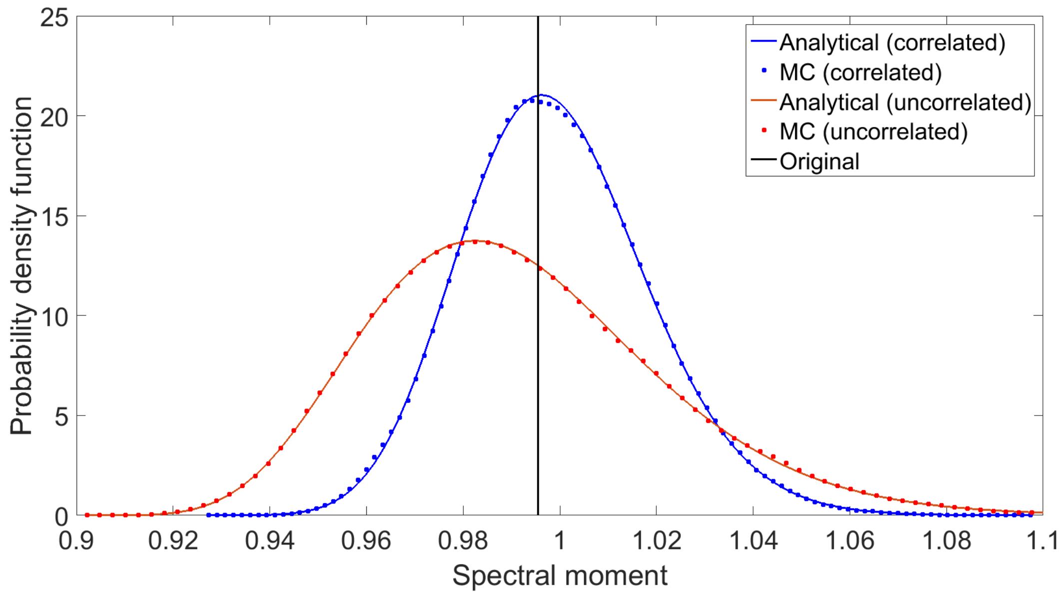

FIG. 4. PDF of spectral moment λ_0 with 10% missing data replaced by both correlated and independent identically distributed Gaussian random variables. Monte-Carlo estimated PDFs (MC) are shown for validation of the procedure. The vertical line shows the spectral moment λ_0 value without missing data

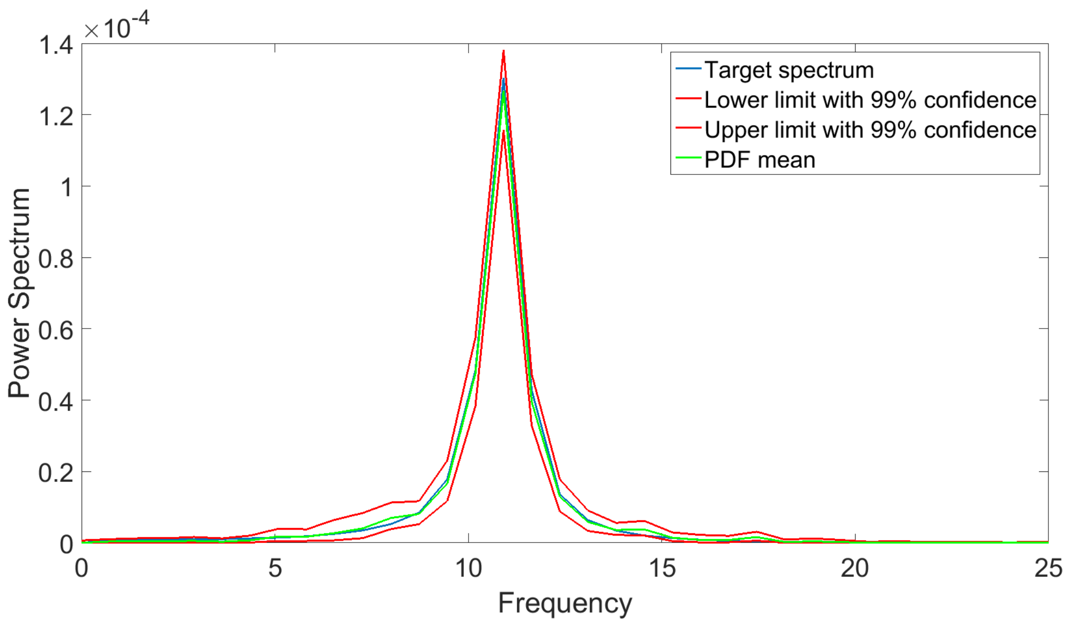

FIG. 5. Oscillator response power spectrum PDF with 70% missing data replaced by correlated Gaussian random variables

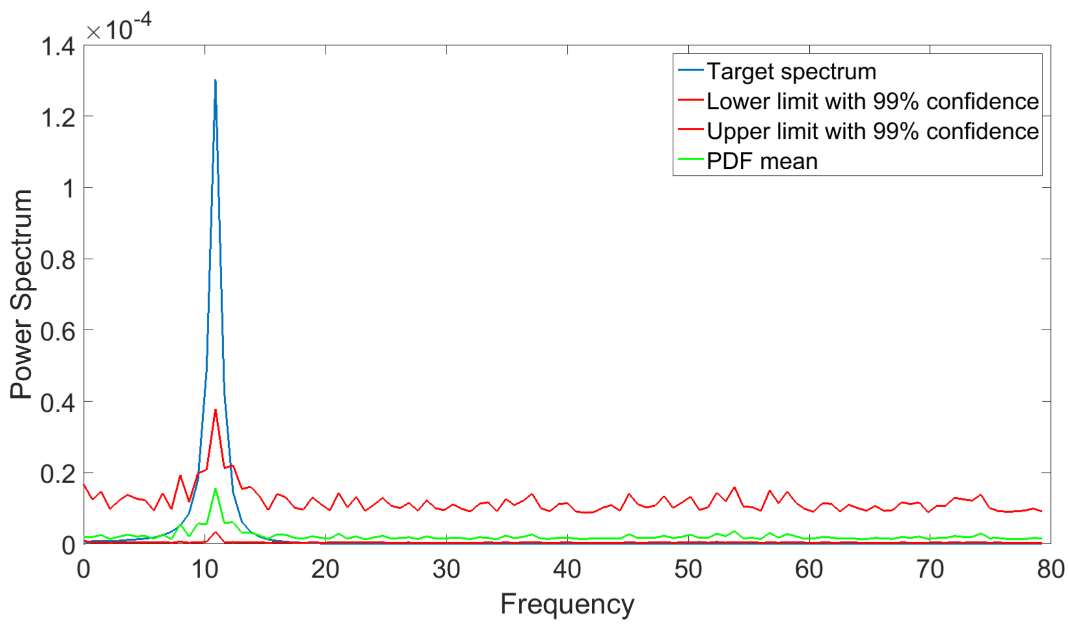

FIG. 6. Oscillator response power spectrum PDF with 70% missing data replaced by independent identically distributed Gaussian random variables

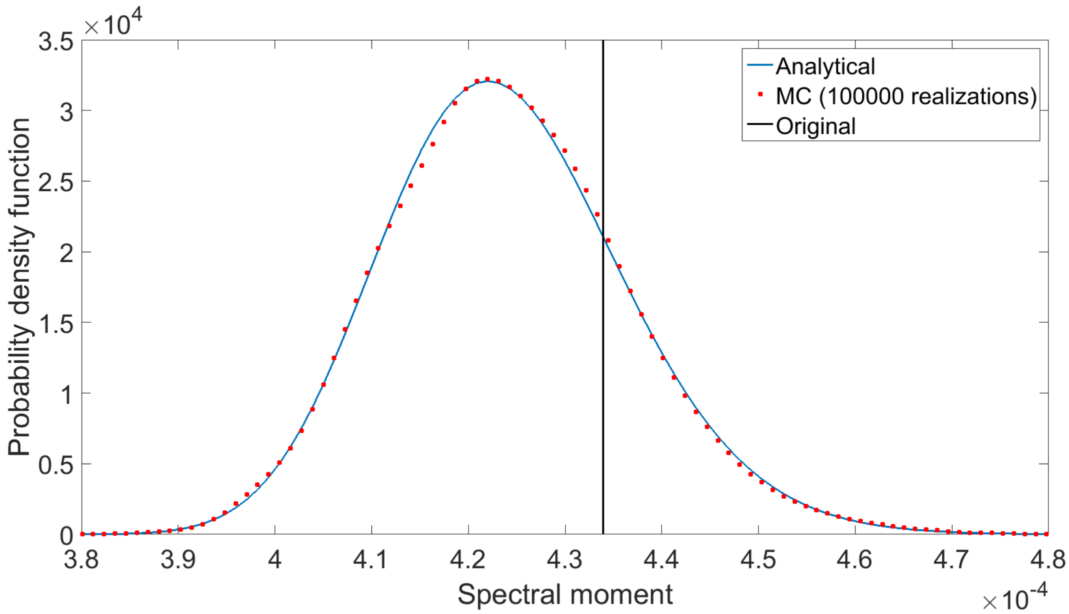

FIG. 7. PDF of response spectral moment λ_0 with 70% missing data. The Monte-Carlo estimated PDF (MC) is shown for validation of the procedure. The vertical line shows the spectral moment without missing data

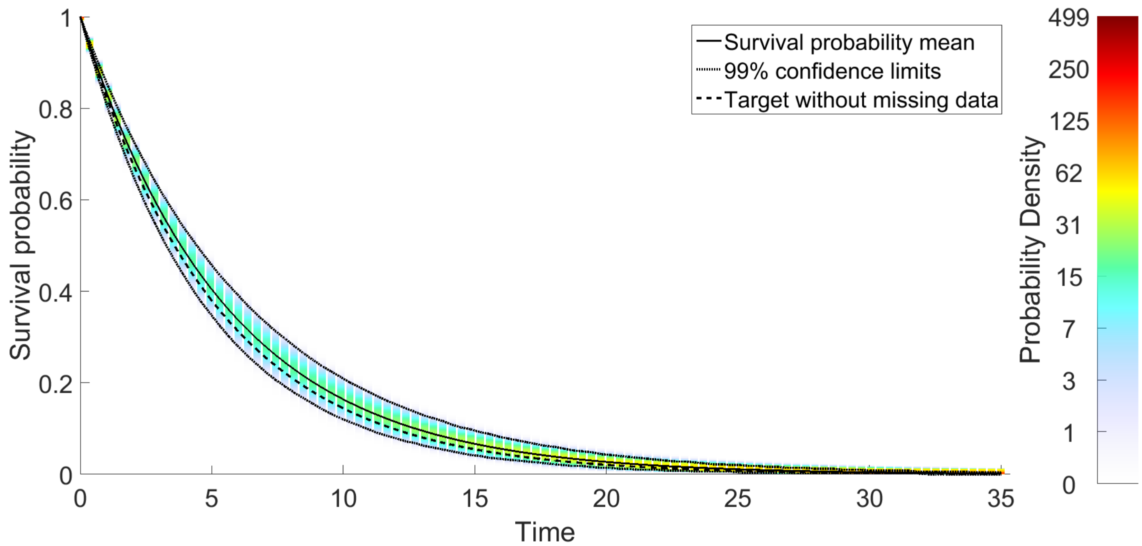

FIG. 8. Survival probability of oscillator response with 70% missing data and barrier $a = 0.05$ **via Eq.(48); comparisons with pertinent Monte Carlo simulations of Eq.(43)**