Generation of Stationary Gaussian Processes and Extreme Value Distributions for High-Cycle Fatigue Models -Application to Tidal Stream Turbines

M Suptille, E Pagnacco, L Khalij, J E Souza de Cursi, J. Brossard

M. Suptille

mickael.suptille@insa-rouen.fr

E. Pagnacco

Emmanuel.Pagnacco@insa-rouen.fr

L. Khalij

Leila.Khalij@insa-rouen.fr

J. E. Souza de Cursi

souza@insa-rouen.fr
LMR, EA 3828, INSA Rouen BP8, 76801
Saint-Etienne du Rouvray Cedex, France

J. Brossard

Jerome.brossard@univ-lehavre.fr
LOMC, FRE 3102 CNRS, BP540,76058
Le Havre Cedex, France

Generation of Stationary Gaussian Processes and Extreme Value Distributions for High-Cycle Fatigue Models – Application to Tidal Stream Turbines

The operating environment of tidal stream turbines is random due to the variability of the sea flow (turbulence, wake, tide, streams, among others). This yields complex time-varying random loadings, making it necessary to deal with high cycle multiaxial fatigue when designing such structures. It is thus required to apprehend extreme value distributions of stress states, assuming they are stationary multivariate Gaussian processes. This work focus on such distributions, addressing their numerical simulation with an analytical description. For that, we first focused on generating one-dimensional Gaussian processes, considering a band-limited white noise in both the narrow-band and the wide-band cases. We then fitted the resulting extreme value distributions with GEV distributions. We secondly extended the generation method to the correlated two-dimensional case, in which the joint extreme value distribution can be obtained from the associated margins. Finally, an example of application related to tidal stream turbines introduces a Bretschneider spectrum, whose shape is commonly encountered in the field of hydrology. Comparing the empirical calculations with the GEV fits for the extreme value distributions shows a very well agreement between the results.
Keywords: stochastic processes generation, correlated Gaussian processes, multivariate extreme value distribution, high cycle fatigue, tidal stream turbines

Introduction

Nowadays, it has become necessary to make the most of renewable energies. Tidal stream turbines allow this by harnessing the kinetic energy of sea streams (see Fig. **??**). But in order to come through they have to be economically competitive, which implies reduced maintenance. However, they are immersed in flows having not mastered characteristics resulting from many contributions such as turbulence, wake, swell, current, etc. These structures are then submitted to random time-varying multiaxial stress states, which can be described from probability distributions associated to a time evolution. Numerically, these stress states can be estimated from a statistical description of the sea load that acts on these submerged structures; see for example Theodorsen (1934), Gmür (1997), Guéna et al. (2006) and Thresher et al. (2009). Such stresses imply a risk of fatigue failure which must be taken into account during the design stage to ensure structural integrity. In this study, we are concerned with the infinite fatigue life, that is based on the greatest stress values reached during the service life (Lambert et al., 2010; Khalij et al., 2010). By taking into account their randomness, this leads us to study their multivariate extreme value distributions. We assume that the stress states are stationary multivariate Gaussian processes, described by their means and power spectral densities. More specifically, we focus in this paper on the way to get, from numerical simulations, bivariate extreme value distributions of components of the stress tensor, with analytical descriptions of the results. Since the mechanical stresses reach their maximum value on the skin of the structure, we shall consider the plane stress case only.

The paper is divided in three parts. The first part deals with generating one-dimensional Gaussian processes in order to evaluate its empirical extreme value distribution. Then, we perform a fit with a GEV distribution. Secondly, these concepts are adapted to multi-dimensional Gaussian processes and applied to a bivariate correlated case. A bivariate extreme value distribution, evaluated from the associated margins, is found to be a well-adapted model. Finally, we apply the previous considerations to a realistic example related to tidal stream turbines.
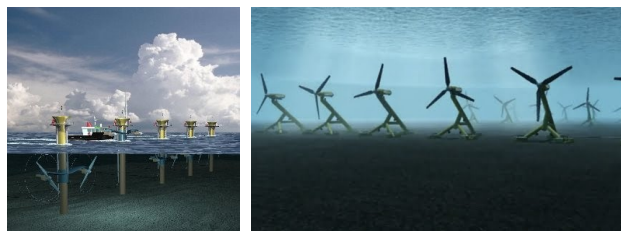
Figure 1: Projects of tidal stream turbines: Skerries Islands, Wales, 2015 (left) and Islay strait, Scotland, 2015 (right).

Nomenclature

Acronyms

AR	= Average Random
EVD	= Extreme Value Distribution
GEV	= Generalized Extreme Value
IFFT	= Inverse Fast Fourier Transform
MA	= Moving Average
PSD	= Power Spectrum Density

Operators

$\arg[\bullet]$ = argument of the complex number $\bullet$
$\mathrm{E}[\bullet]$ = statistical expectation of $\bullet$
$\mathrm{Prob}[\bullet]$ = probability of the event $\bullet$

Re $[\bullet]$ = real part of $\bullet$
V $[\bullet]$ = statistical variance of $\bullet$
$|\bullet|$ = modulus of the complex number $\bullet$
$\langle \bullet \rangle$ = temporal mean of $\bullet$

Latin Symbols

f = frequency (Hz)
j = imaginary unit
m = association parameter between the maxima of two processes
m_i = i-th spectral moment
n = number of stochastic processes
n_f = number of frequencies
s_p = thresholds associated to the maximum of the p-th process (MPa)
t = time (s)
F = cumulative distribution function
F_{pq} = bivariate distribution function between the p-th and the q-th processes
F_p = marginal distribution function of the p-th process
N_0 = number of cycles
R   = auto-correlation function (MPa²)
R_{pz} = component of correlation function matrix **R** (MPa²)
S   = PSD of a single process (MPa²/Hz)
S   = PSD matrix (MPa²/Hz)
S_{xx} = traction component of the stress tensor (MPa)
S_{xy} = shear component of the stress tensor (MPa)
T = samples duration (s)
X   = stochastic process (MPa)

Calligraphic Symbols

$\mathcal{U}$ = uniform distribution
$\mathcal{R}$ = Rayleigh distribution
$\mathcal{N}$ = normal distribution

Greek Symbols

μ_{X_p} = statistical mean of the process X_p (MPa)
$\mu_{X_p}^{(r)}$ = temporal mean of the r-th realization of the process X_p (MPa)
ρ = auto-correlation coefficient
ρ_m = correlation coefficient between the maxima of two processes
σ = standard deviation (MPa)
ω = angular frequency (rad/s)

The One-Dimensional Case

Let us consider a stationary process characterized by its one-sided PSD, measured in MPa²/Hz and denoted by S, and its statistical distribution: both these quantities provide the frequency distribution of the process. Classically, two extreme situations may be considered: wide-band and narrow-band cases. Let us introduce the frequency f, measured in Hz. Both the cases may be illustrated by considering a band-limited white noise:

$$S(f) = \begin{cases} S_0 & \text{if } f \in (f_1, f_2] \\ 0 & \text{else} \end{cases}. \qquad (1)$$

For instance:

- the narrow-band case corresponds to $|f_2 - f_1|$ "small": in the sequel, we take $f_1 = 9.5$ Hz, $f_2 = 10.5$ Hz, $S_0 = 7$ MPa²/Hz, the frequency f_1 being included in the spectral definition domain;

- the wide-band case corresponds to $|f_2 - f_1|$ "large": in the sequel, we take $f_1 = 0$ Hz, $f_2 = 20$ Hz, $S_0 = 0.35$ MPa²/Hz, the frequency f_1 being excluded from the spectral definition domain, since it is the zero value.

It should be noted that in either case, S_0 is such that $V[X(t)] = 7$ MPa², $V[X(t)]$ denoting the statistical variance of the process $X(t)$. These two extreme situations are represented in Fig. 2: the narrow-band PSD corresponds to the first case and the wide-band PSD represents the second one.
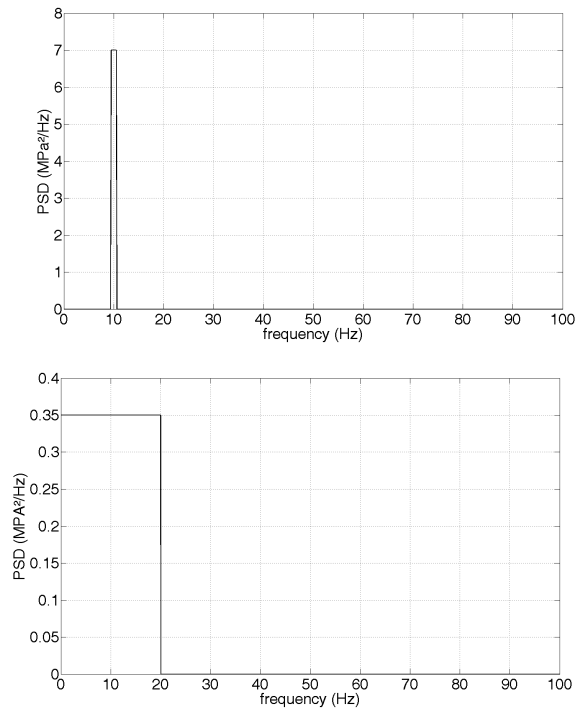
Figure 2: Narrow-band (top) and wide-band (bottom) PSDs of the investigated one-dimensional case.

As announced in the introduction, we consider the generation of these processes for the analysis of the extreme value distribution. In either case (wide and narrow) we proceed as follows: at first, 10^4 realizations are generated. Afterwards, the extreme value distribution is constructed from these empirical data and fitted with a GEV distribution. In the following, the results regarding both cases are presented and illustrated together whenever it is possible.

Theory of Gaussian process generation

Let $X(t)$ be a temporal zero-mean Gaussian process. Let $S(f)$ be the associated PSD, uniformly discretized in n_f points:

$$\begin{cases} (f_i, S(f_i)) = (f_i, S_i) \\ \Delta f = f_{i+1} - f_i, & 1 \le i \le n_f - 1 \end{cases}$$

where Δf is the frequency step and n_f denotes the number of frequencies f_i whose associated PSD magnitude S_i is such that $S_i \neq 0$ (i.e. the number of frequencies f_i such that $f_i \in [f_1, f_2]$). Two main

classes of generation methods exist in the literature (Sun, 2006). The first one includes time series models: Average Random (AR), Moving Average (MA) and ARMA processes. The other one is about spectral decomposition methods. We focus on the last category in this document. Each realization of $X(t)$ is made of a sum of harmonic functions of time, whose amplitude and frequency are determined from the PSD:

$$X(t) = \sum_{i=1}^{n_f} X_i(t). \qquad (2)$$

One way of computing $X(t)$ is to regard each harmonic contribution $X_i(t)$ as a cosine function with deterministic modulus M_i and random uniformly distributed phase ϕ_i:

$$X_i(t) = M_i \cos(2\pi f_i t + \phi_i) \qquad (3)$$

where:

$$\begin{cases} M_i = \sqrt{2 S_i \Delta f} \\ \phi_i \sim \mathcal{U}(0, 2\pi) \end{cases}.$$

Another possibility is to use the same expression as in Eq. (3) with M_i following a Rayleigh distribution, i.e.:

$$\begin{cases} M_i \sim \mathcal{R}\left(\sqrt{S_i \Delta f}\right) \\ \phi_i \sim \mathcal{U}(0, 2\pi) \end{cases}.$$

Finally, we can write $X_i(t)$ as a sum of a sine and a cosine functions with Gaussian amplitudes:

$$X_i(t) = A_i \cos(2\pi f_i t) + B_i \sin(2\pi f_i t) \qquad (4)$$

with:

$$\begin{cases} A_i, B_i \sim \mathcal{N}(0, \sigma_i) \\ \sigma_i = \sqrt{S_i \Delta f} \end{cases}.$$
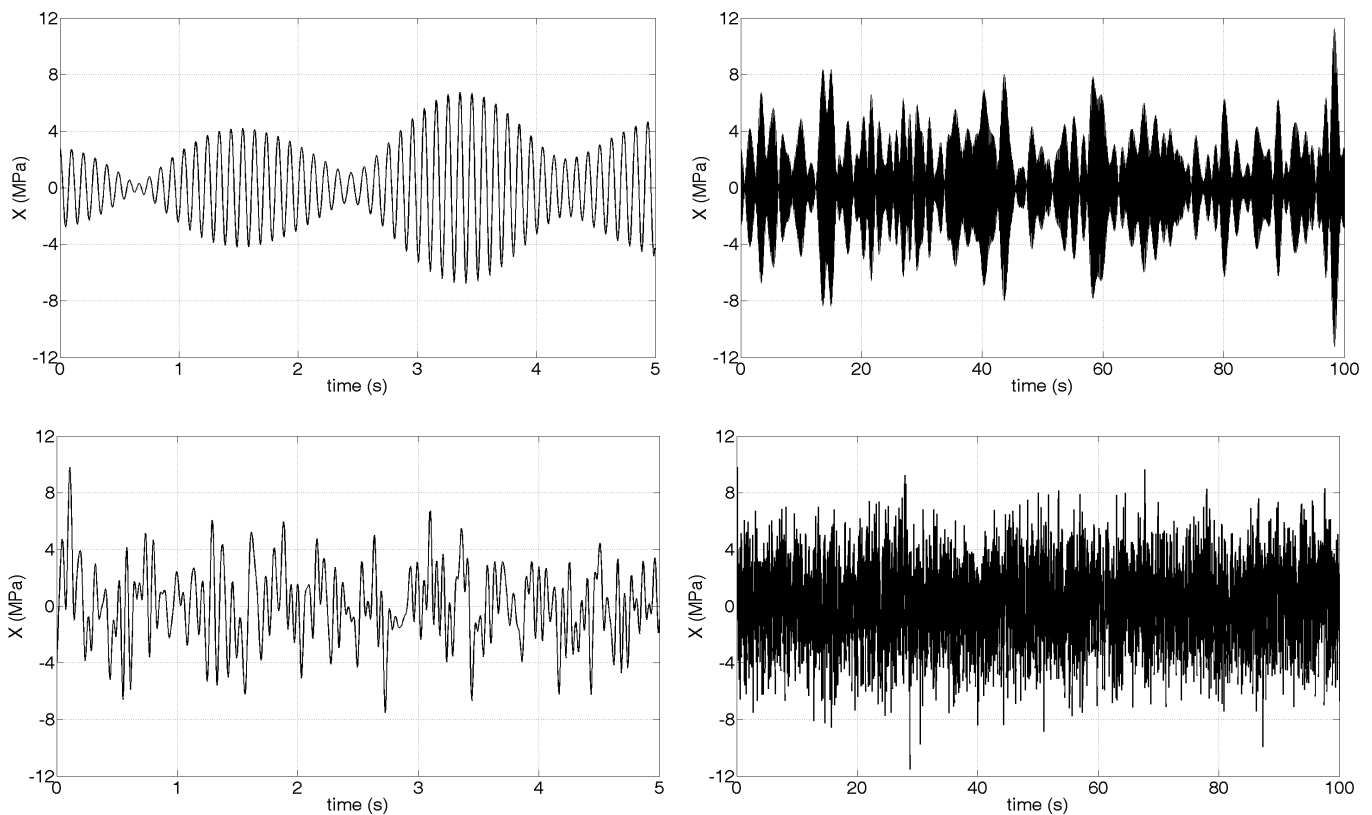

Figure 3: Process samples during a short (left) and a longer (right) time, narrow-band (top) and wide-band (bottom) PSDs.

It should be noted that processes generated from Eq. (3) considering deterministic moduli are asymptotically Gaussian when the number of frequency points n_f tends to infinity, in virtue of the Central Limit Theorem. Besides, the use of Eq. (3) with Rayleigh distributed moduli is equivalent to the use of Eq. (4) with Gaussian amplitudes (Tucker, 1984). In the sequel, computations are performed using Eq. (4). One realization of the processes under consideration is shown in Fig. 3 to illustrate the narrow-band and wide-band concepts.

Numerical aspects

The random amplitudes A_i and B_i involved in Eq. (4) are computed using the Matlab generator of normally distributed numbers (*randn* function). The expression proposed for their standard deviation arises from the statistical interpretation of the auto-PSD. As a matter of fact, the integral of S gives the variance of X. The definition of σ_i then ensures units coherence: with this choice, A_i and B_i have the same dimensional units as X. The total variance of the

process is the sum of the variances of each harmonic contribution:

$$V[X(t)] = \sum_{i=1}^{n_f} V[X_i(t)] = \sum_{i=1}^{n_f} \sigma_i^2 = \sum_{i=1}^{n_f} V[X_i] = V[X].$$

The temporal variable vanishes due to the expected stationarity of the process. However, the reader must keep in mind that empirical processes cannot be stationary if the number of frequencies n_f is not large enough. On the contrary, they are Gaussian for any number of frequencies since they are computed as a linear combination of Gaussian random variables. In addition, it can be shown that the processes under consideration are ergodic for the mean but not for the correlation function (see Appendix). Besides, each PSD is characterized by its central frequency f_0 such as:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{m_2}{m_0}}$$

where m_k is the k-th order spectral moment, defined from the PSD by:

$$m_k = \int_0^{+\infty} (2\pi f)^k S(f) \, df.$$

We recall that the PSD $S(f)$ is one-sided and involves the frequency variable. For PSDs derived in terms of the circular frequency variable and/or two-sided, the previous expression is to be adapted. Let N_0 denote the number of up-crossings of the zero threshold (i.e. the number of cycles) and T the time duration of the signals. We have the following relation:

$$N_0 = f_0 T.$$

Thus the knowledge of both the number of cycles and the PSD yields the samples duration. Moreover, it should be noted that the computed processes are periodic, due to the use of harmonic functions. So we make sure to adjust the time period T to the samples length in order to avoid any duplication of the signals in the observation window. In practice, the value of T must be large enough to cover 10^6 cycles considering high cycle fatigue applications. Then a large number of frequencies are needed, which makes the frequency step Δf being very small and the generation time-consuming. One way to solve this problem is to take advantage of the IFFT algorithm. As matter of fact, adapting Eq. (4), we have:

$$
\begin{aligned}
X(t) &= \sum_{i=1}^{n_f} A_i \cos(2\pi f_i t) + B_i \sin(2\pi f_i t) \\
&= \sum_{i=1}^{n_f} A_i \cos(2\pi f_i t) + B_i \cos\left(-\frac{\pi}{2} + 2\pi f_i t\right) \\
&= \sum_{i=1}^{n_f} A_i \text{Re}\left[e^{j2\pi f_i t}\right] + B_i \text{Re}\left[e^{-j\frac{\pi}{2}} e^{j2\pi f_i t}\right] \\
&= \text{Re}\left[\sum_{i=1}^{n_f} (A_i - B_i) e^{j2\pi f_i t}\right] \\
&= \sum_{i=1}^{n_f} (A_i - B_i) e^{j2\pi f_i t}.
\end{aligned}
$$

The real operator vanishes since the calculations yield an inverse Fourier transform whose result is real-valued. This transform can be rapidly computed using the *ifft* function implemented in Matlab. On top of that, computing extreme value distributions requires accurate peaks in the temporal signals, which is conditioned by a refined sampling, as shown in Fig. 4. To get enough time points, we

perform a zero-padding up to 100 Hz in the frequency domain. Indeed, due to the use of the IFFT algorithm, the number of sampling points is the same in both the time and the frequency domains. $n_f = 10^5$ frequencies are used in the narrow-band case and $n_f = 2.10^6$ frequencies are used in the wide-band case. To get accurate distributions, 10^4 realizations of the processes are required. Complying with the last remarks leads to dealing with a huge amount of data. Generating 10^4 narrow-band samples requires 7h15 user time using Matlab 7.11 without any specific launch options on a Linux server having 24 processors Intel(R) Xeon(R) CPU X5680 cadenced at 3.33 GHz with 12 MB cached size and 192 GB total memory. In comparison, nearly 115h are needed when using the *filter* function implemented in Matlab to generate the same Gaussian processes via AR series.
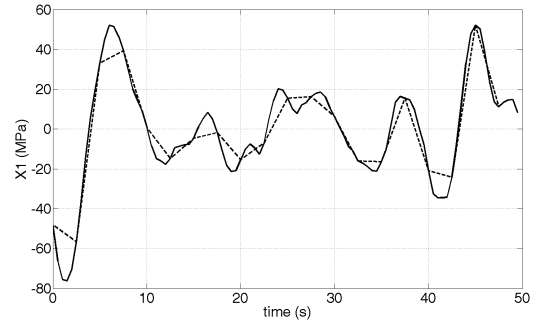

Figure 4: Importance of temporal discretization.

Results analysis

To verify the adequateness of the computed samples, the empirical auto-correlation coefficient is calculated and compared to the analytical one in Fig. 5. For stationary processes, the auto-correlation coefficient is defined as:

$$\rho(\tau) = \frac{R(\tau)}{V[X]}. \tag{5}$$

$R(\tau)$ being the auto-correlation function defined as:

$$R(\tau) = E[X(t)X(t+\tau)], \ t \in [0, T] \tag{6}$$

where τ is the time shift between the two instants t and $t + \tau$ involved in Eq. (6), $E[\bullet]$ being the mean operator. Note that $\tau \in [-T, T]$. Due to Eq. (1), in both the low-pass and the band-pass cases, the auto-correlation function is given by:

$$R(\tau) = \frac{S_0}{2\pi} \frac{\sin(2\pi f_2 \tau) - \sin(2\pi f_1 \tau)}{\tau}. \tag{7}$$

In order to compare the empirical and analytical auto-correlation coefficients in a proper way, both of the empirical and analytical auto-correlation functions are normalized using the theoretical value of the variance. To quantify the precision of the simulation, we shall compare in detail the theoretical and empirical values of the variances and auto-correlation coefficients. Figure 5 and Table 1 show the results, using here only $n_f = 10^2$ and 2.10^3 frequencies for the narrow-band and the wide-band case respectively. The empirical auto-correlation coefficient and variance agree well with the analytical ones, what implies that the simulated processes actually correspond to their respective target PSDs.

Table 1: Comparison between the theoretical and the empirical variances, narrow-band (top) and wide-band (bottom) cases.
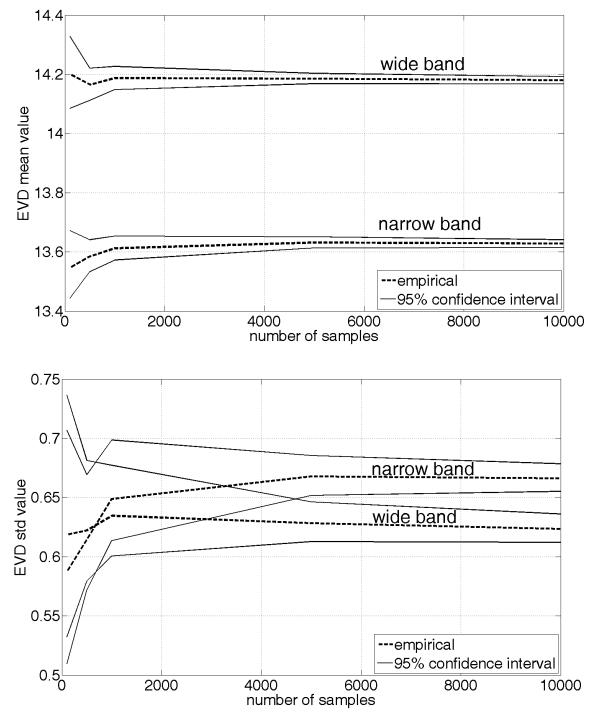
Empirical (target) variance	Relative gap
7.0082 (7)	0.12 %

Empirical (target) variance	Relative gap
7.0037 (7)	0.05 %
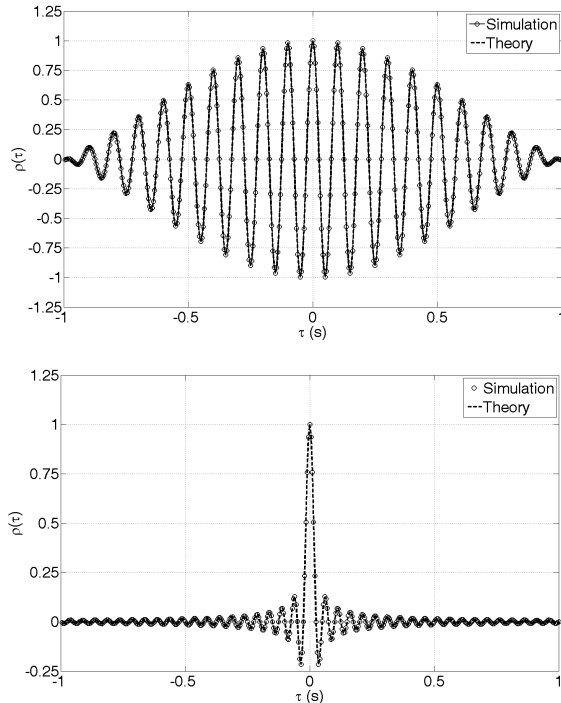
Figure 5: Comparison between empirical and analytical auto-correlation coefficients, narrow-band (top) and wide-band (bottom).

The greatest value of each realization is stored in order to compute the associated distribution. In the sequel, EVD stands for "extreme value distribution". Different verifications are made to evaluate the quality of the empirical EVDs. First, its mean and standard deviation are computed for an increasing number of samples. The values are provided in Fig. 6 with the related confidence intervals (95%), obtained from the bootstrap method. When generating 10^4 samples of each process, the mean and standard deviation have converged, as well as the related confidence intervals. Secondly, the empirical EVD is plotted together with its 95% confidence interval, computed from the Greenwood formula. It must be underlined that this interval is extremely reduced, due to the large number of realizations. Finally, it is interesting to check (see Fig. 7) if the empirical EVDs actually "look like" usual EVDs. These are then fitted with 3 parameters Generalized Extreme Value (GEV) distributions:

$$F\left(x; \xi, \mu, \beta\right) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\beta}\right)^{-\frac{1}{\xi}} \right], \ 1 + \xi \frac{x - \mu}{\beta} > 0, \ \xi \neq 0 \tag{8}$$

Figure 6: Convergence of the EVDs mean (top) and standard deviation (bottom) value.

In probability theory and statistics, the GEV distribution is a family of continuous probability distributions developed under the extreme value theory in order to combine the Gumbel, Fréchet and Weibull families. The parametrization involved in Eq. (8) results from the work of von Mises (1936) and Jenkinson (1955). The GEV distribution arises from the extreme value theorem (Fisher and Tippet, 1928) as the limiting distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. It has been shown that the framework can be extended to Gaussian processes; see for example De Haan and Ferreira (2006). The above-mentioned GEV distribution is computed from the simulation results by means of the maximum likelihood method (*gevfit* function in Matlab). It should be noted that the resulting curve lies in the confidence interval. A particular attention is paid to the distribution tails since these may be involved in reliability calculations. We end by evaluating the relative error $err\left(p\right)$ between the empirical CDF and the related GEV distribution for any probability level p, $0 \leq p \leq 1$. Let $s_{\text{th}}\left(p\right)$ be the threshold such that $\text{Prob}\left[\max_{t \in [0,T]}\left(\text{X}\left(t\right)\right) < s_{\text{th}}\right] = p$ in the theoretical model (GEV, here). Let $s_{\text{MC}}\left(p\right)$ be the threshold such that $\text{Prob}\left[\max_{t \in [0,T]}\left(\text{X}\left(t\right)\right) < s_{\text{MC}}\right] = p$ in the MC simulations. Then the relative error is computed as follows:

$$err\left(p\right) = \frac{s_{\text{th}}\left(p\right) - s_{\text{MC}}\left(p\right)}{s_{\text{MC}}\left(p\right)}. \tag{9}$$

It must be underlined that the relative error between the empirical EVD and its GEV fit is low. These observations allow us to state that the amount of data (in particular the number of realizations) is large enough to yield accurate EVDs.
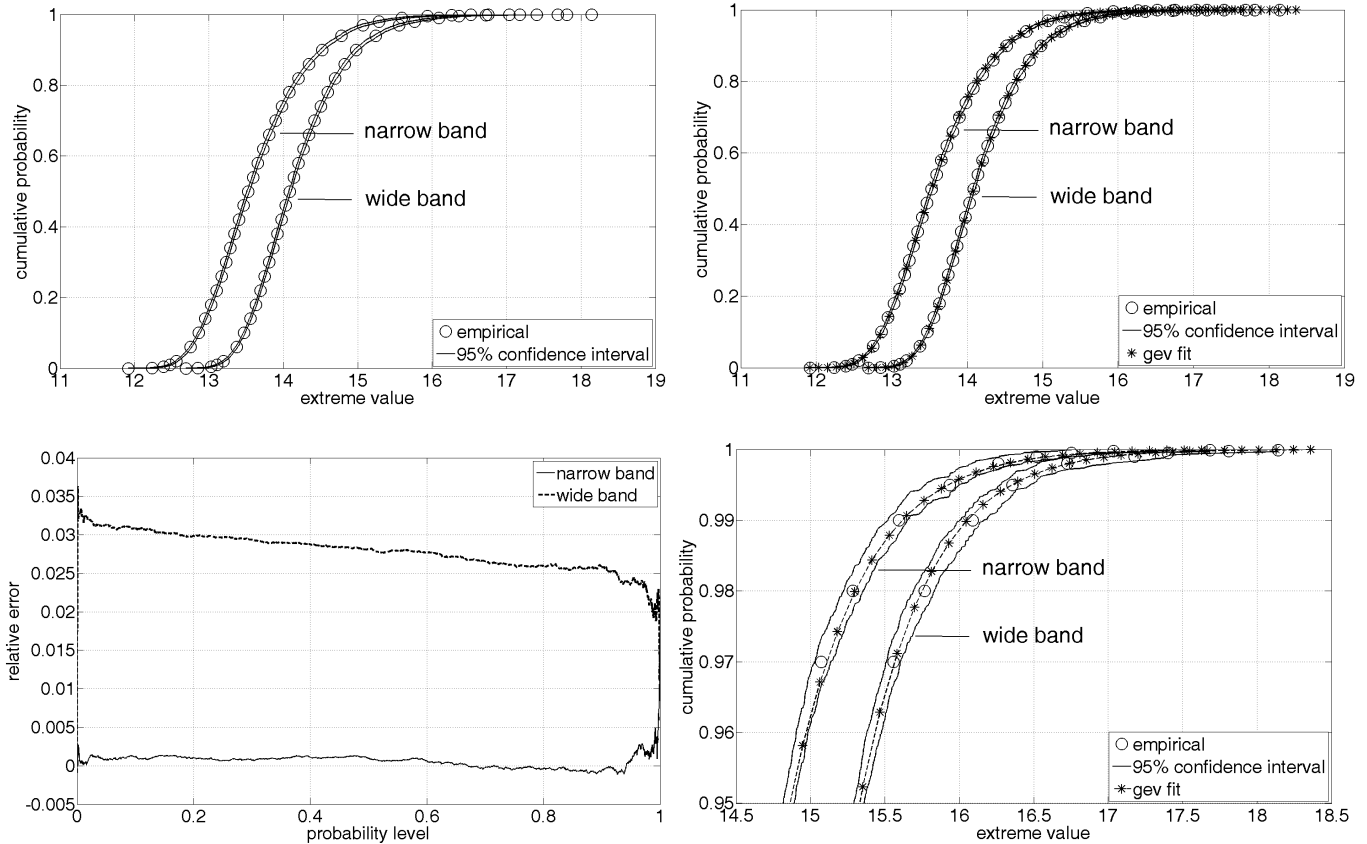
Figure 7: Confidence interval of the EVDs and fit with GEV distributions.

Extension to the Multi-Dimensional Case

Both cases introduced in the previous paragraph are now generalized to several processes. Using the PSDs relative to the n processes to be considered, we have the spectral matrix:

$$
\mathbf{S}(f) = \begin{bmatrix} \mathrm{S}_{11}(f) & \mathrm{S}_{12}(f) & \cdots & \mathrm{S}_{1n}(f) \\ \mathrm{S}_{21}(f) & \mathrm{S}_{22}(f) & \cdots & \mathrm{S}_{2n}(f) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{S}_{n1}(f) & \mathrm{S}_{n2}(f) & \cdots & \mathrm{S}_{nn}(f) \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathrm{S}_{11}(f) & \mathrm{S}_{12}(f) & \cdots & \mathrm{S}_{1n}(f) \\ \mathrm{S}_{12}^*(f) & \mathrm{S}_{22}(f) & \cdots & \mathrm{S}_{2n}(f) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{S}_{1n}^*(f) & \mathrm{S}_{2n}^*(f) & \cdots & \mathrm{S}_{nn}(f) \end{bmatrix}.
$$

$\mathrm{S}_{pp}(f)$, $p = 1, \ldots, n$ being the process p auto-PSD and $\mathrm{S}_{pq}(f)$, $p, q = 1, \ldots, n$ being the cross-PSD between processes p and q. Each auto-PSD shows the variance frequential repartition of the related process. Each cross-PSD provides the covariance frequential repartition of the associated processes. In particular, we now consider the two-dimensional correlated case. Unlike the general case, cross-PSDs are real-valued in the following. As in the first section, the academic case of band-limited white noise is considered:

$$
\mathrm{S}_{11}(f) = \begin{cases} S_{01} & \text{if } f \in (f_1, f_2] \\ 0 & \text{else} \end{cases}
$$

$$
\mathrm{S}_{22}(f) = \begin{cases} S_{02} & \text{if } f \in (f_1, f_2] \\ 0 & \text{else} \end{cases} \quad . \tag{10}
$$

$$
\mathrm{S}_{12}(f) = \begin{cases} S_{012} & \text{if } f \in (f_1, f_2] \\ 0 & \text{else} \end{cases}
$$

Analogously to the one-dimensional case:

- in the narrow-band case, we take $f_1 = 9.5$ Hz, $f_2 = 10.5$ Hz, $S_{01} = 3$ MPa^2/Hz, $S_{02} = 5$ MPa^2/Hz, $S_{012} = 0.9\sqrt{S_{01}S_{02}}$ MPa^2/Hz;

- the wide-band case, we take $f_1 = 0$ Hz, $f_2 = 20$ Hz, $S_{01} = 0.15$ MPa^2/Hz, $S_{02} = 0.25$ MPa^2/Hz, $S_{012} = 0.9\sqrt{S_{01}S_{02}}$ MPa^2/Hz.

It should be noted that in either case, S_{01} is such that $\mathrm{V}[\mathrm{X}_1(t)] = 3$ MPa^2, S_{02} is such that $\mathrm{V}[\mathrm{X}_2(t)] = 5$ MPa^2 and S_{012} is such that $\frac{\mathrm{cov}[\mathrm{X}_1(t), \mathrm{X}_2(t)]}{\sqrt{\mathrm{V}[\mathrm{X}_1(t)]\mathrm{V}[\mathrm{X}_2(t)]}} = 0.9$, $\mathrm{cov}[\mathrm{X}_1(t), \mathrm{X}_2(t)]$ denoting the statistical covariance between the processes $\mathrm{X}_1(t)$ and $\mathrm{X}_2(t)$. These two situations are represented in Fig. 8: the narrow-band PSD corresponds to the first case and the wide-band PSD represents the second one.
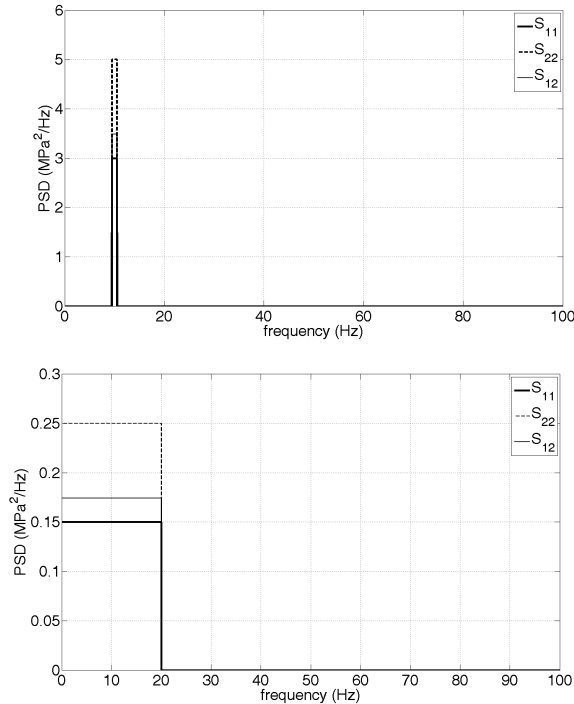
Figure 8: Narrow-band (top) and wide-band (bottom) PSDs of the investigated multi-dimensional case.

Analogously to the one-dimensional situation, for either case (wide and narrow) we proceed as follows: at first, 10^4 realizations are generated by assuming that the processes are Gaussian. Afterwards, the bivariate extreme value distribution is estimated from these empirical data as well as the associated margins and empirical cross-correlation coefficient. Like in the previous Results analysis subsection, the marginal EVDs are fitted with GEV distributions. The latter are then used together with the cross-correlation coefficient in order to compute a theoretical bivariate distribution.

Theory of Gaussian processes generation

Here again, each realization of a Gaussian process results from a sum of harmonic functions of time with random amplitudes computed from the PSD. Compared to the one-dimensional case, the difference lies in the fact that random quantities have to be appropriately correlated such that processes are correlated according to their associated target cross-PSD. Then, Eq. (4) is generalized as follows:

$$X_p(t) = \sum_{i=1}^{n_f} \sum_{q=1}^{n=2} \left\{ A_{ipq} \cos\left(2\pi f_i t + \theta_{ipq}\right) + B_{ipq} \sin\left(2\pi f_i t + \theta_{ipq}\right) \right\}$$

(11)

where:

$$\begin{cases} A_{ipq} = A_{iq}(f_i)\,\sigma_{ipq} \\ B_{ipq} = B_{iq}(f_i)\,\sigma_{ipq} \\ A_{iq}, B_{iq} \sim \mathcal{N}(0,1) \\ \sigma_{ipq} = \left|\mathrm{H}_{pq}(f_i)\right|\sqrt{\Delta f} \\ \theta_{ipq} = \arg\left[\mathrm{H}_{pq}(f_i)\right] \\ \mathbf{H}(f_i) = \mathbf{S}(f_i)^{\frac{1}{2}} \end{cases}.$$

i being the frequency index, p being the current process index and q being the summation index (p being fixed) on all the processes (to introduce correlation between X_p and X_q).

Results analysis

In order to verify the adequateness of the computed samples, the empirical auto and cross-correlation functions are calculated and compared to the analytical ones in Fig. 9. Analogously to Eq. (5), for a stationary process, the (auto and cross) correlation coefficients are defined as:

$$\rho_{pq}(\tau) = \frac{\mathrm{R}_{pq}(\tau)}{\mathrm{V}[X_p]\,\mathrm{V}[X_q]}.$$

(12)

$\mathrm{R}_{pq}(\tau)$ being the related correlation function defined as:

$$\mathrm{R}_{pq}(\tau) = \mathrm{E}\left[X_p(t)\,X_q(t+\tau)\right],\ t \in [0,T]$$

(13)

where τ is the time shift between the two instants t and $t+\tau$ involved in Eq. (13). As mentioned earlier, $\tau \in [-T,T]$. Using Eq. (11) in (13) yields:

$$\mathrm{R}_{pq}(\tau) = \sum_{i=1}^{n_f} \sum_{z=1}^{n} \left|\mathrm{H}_{pz}(f_i)\mathrm{H}_{qz}(f_i)\right|\Delta f \cos\left(2\pi f_i t\tau + \theta_{iqz} - \theta_{ipz}\right).$$

(14)

Due to Eq. (10), in both the low-pass and the band-pass cases, the correlation functions are given by:

$$\begin{aligned}
\mathrm{R}_{11}(\tau) &= \frac{S_{01}}{2\pi}\frac{\sin(2\pi f_2\tau)-\sin(2\pi f_1\tau)}{\tau} \\
\mathrm{R}_{22}(\tau) &= \frac{S_{02}}{2\pi}\frac{\sin(2\pi f_2\tau)-\sin(2\pi f_1\tau)}{\tau} \\
\mathrm{R}_{12}(\tau) &= \frac{\mathrm{Re}[S_{012}]}{2\pi}\frac{\sin(2\pi f_2\tau)-\sin(2\pi f_1\tau)}{\tau} \\
&+ \frac{\mathrm{Im}[S_{012}]}{2\pi}\frac{\cos(2\pi f_2\tau)-\cos(2\pi f_1\tau)}{\tau}.
\end{aligned}$$

(15)

Table 2: Comparison between the theoretical and the empirical variances and cross-correlation coefficients, narrow-band (top) and wide-band (bottom) cases.

	Empirical (target) value	Relative gap
$\mathrm{V}[X_1]$	3.0267 (3)	0.89%
$\mathrm{V}[X_2]$	5.0462 (5)	0.92%
ρ_{12}	0.9082 (0.9)	0.91 %
	Empirical (target) value	Relative gap
$\mathrm{V}[X_1]$	2.9999 (3)	-0.004%
$\mathrm{V}[X_2]$	5.0005 (5)	0.011%
ρ_{12}	0.9001 (0.9)	0.008 %

In order to compare the empirical and analytical correlation coefficients in a proper way, both the empirical and analytical correlation functions are normalized using the theoretical values of the variances. To quantify the precision of the simulation, we are interested in comparing in detail the theoretical and empirical values of the variances and cross-correlation coefficients. $n_f = 10^2$ frequencies are used in the narrow-band case and $n_f = 2.10^3$ frequencies are used in the wide-band case. We can see from Fig. 9 that the empirical auto and cross-correlation coefficients agree well with the analytical ones. Moreover, Table 2 shows that the empirical variances and cross-correlation coefficients of the computed processes are in total agreement with the theoretical ones. This implies that the simulated processes actually correspond to their respective target PSDs and are accurately correlated.

Here again, storing the greatest value of each realization enables us to compute the bivariate EVD as well as the associated margins. The latter are fitted by GEV distributions in Fig. 11, as we did for

the one-dimensional case. These can be used to find the bivariate distribution. In the literature, Raynal-Villasenor and Salas (1987) and Elshamy (1992) present two differentiable models that are well suited for our study. From them, we consider the more flexible and widely applicable model:

$$F_{m12}(s_1, s_2) = \exp\left\{ -\left[\{-\ln[F_{m1}(s_1)]\}^m + \{-\ln[F_{m2}(s_2)]\}^m \right]^{\frac{1}{m}} \right\} \tag{16}$$

where $F_{mi}(s_i)$ is the marginal EVD of the process $X_i(t)$, $F_{m12}(s_1, s_2)$ denotes the related bivariate EVD and m is the association parameter. It is such that $m \in [1, +\infty[$, where $m = 1$ gives the independence case and $m \to +\infty$ corresponds to a complete dependence between the margins. This parameter is a function of the correlation coefficient ρ_m between the marginal EVD:

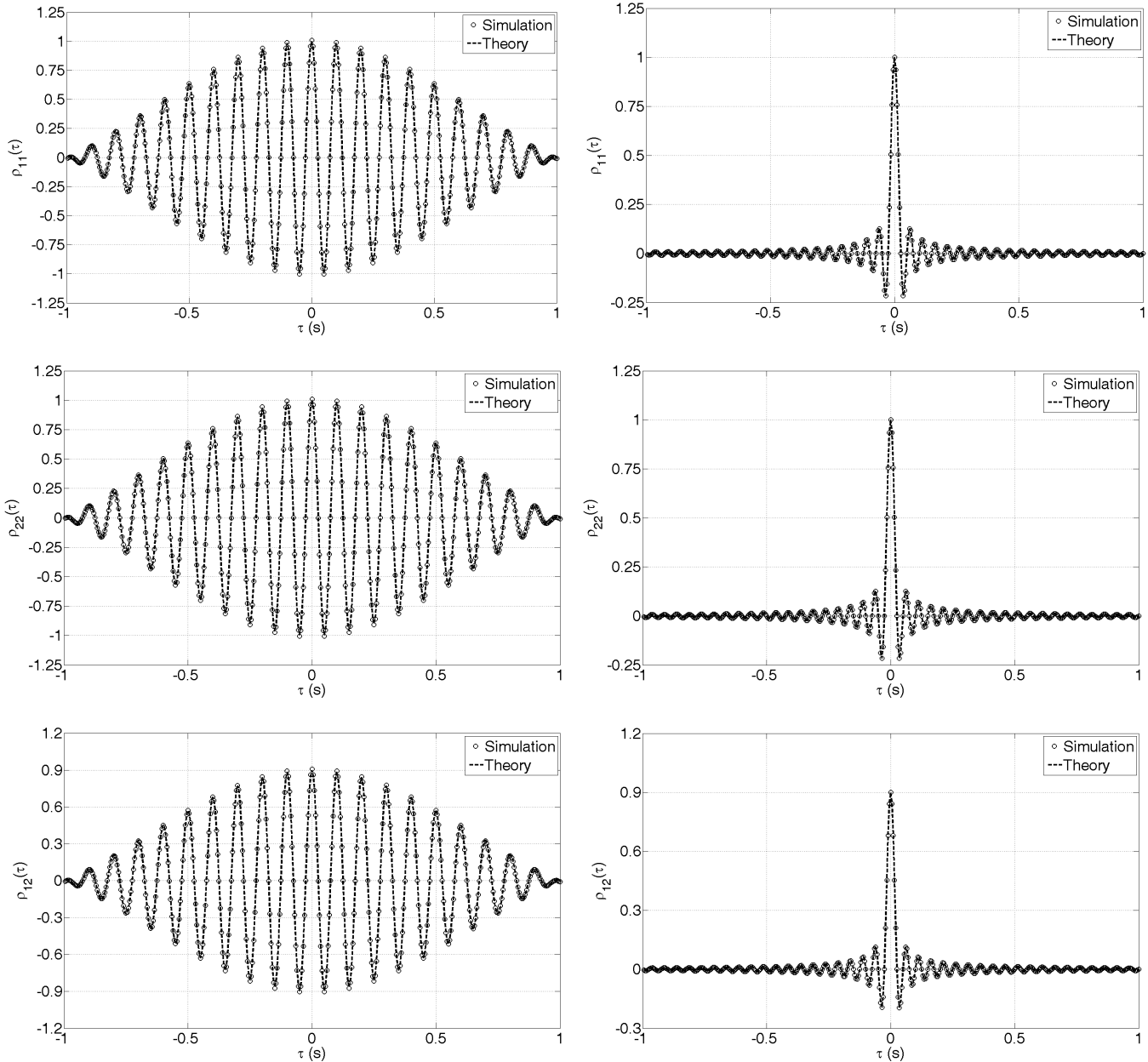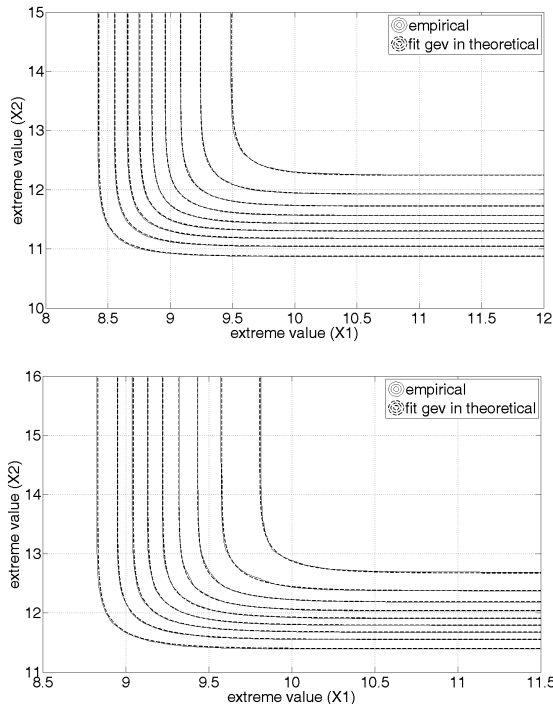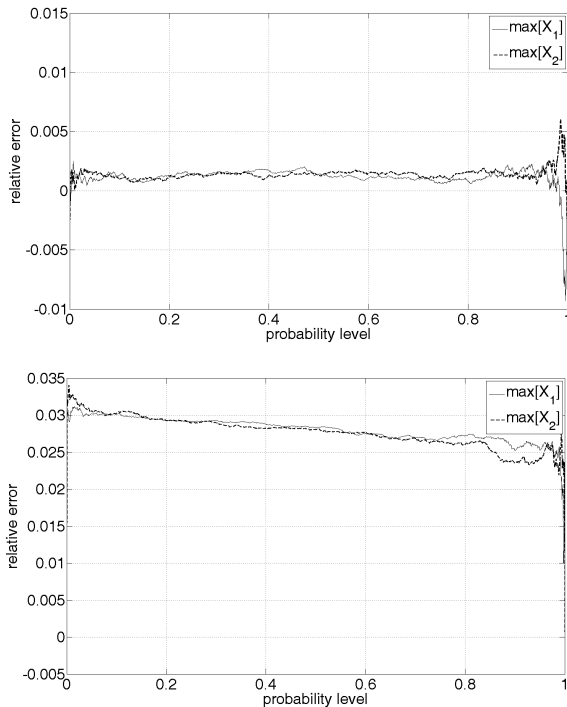$$m = (1 - \rho_m)^{-\frac{1}{2}}. \tag{17}$$

Figure 9: Comparison between empirical and analytical correlation coefficients, narrow-band (left) and wide-band (right) PSDs.

Figure 10: Bivariate extreme value distributions, narrow-band (top) and wide-band (bottom) cases.

Figure 11: Relative error between the empirical margins and their GEV fits, narrow-band (top) and wide-band (bottom) cases.

In this work, ρ_m is obtained from empirical data using Matlab (*corrcoef* function). Then Eq. (16) can be used with the GEV fit of the margins. The resulting distribution fits very well the empirical bivariate EVD, as shown by Fig. 10. This arises since fitting the

extreme value margins with GEV distributions yields little errors (see Fig. 11).

Application Considering an Hydrology Spectrum

Input data

Here we consider mechanical stresses involved in a combination of a traction and a torsion loading, respectively $S_{xx}(t)$ and $S_{xy}(t)$. Other components of the stress tensor are considered equal to zero. In the following, instead of simulating $S_{xx}(t)$ and $S_{xy}(t)$, we generate the associated stresses processes needed for a fatigue criterion (Lambert et al., 2010):

$$\begin{cases} X_1(t) = \frac{\sqrt{3}}{3} S_{xx}(t) \\ X_2(t) = S_{xy}(t) \end{cases}.$$

Statistical moments are supposed to be the following:

$$\begin{cases} E[S_{xx}(t)] = 0\text{MPa} \\ E[S_{xy}(t)] = 0\text{MPa} \\ V[S_{xx}(t)] = 2700\text{MPa}^2 \\ V[S_{xy}(t)] = 675\text{MPa}^2 \end{cases} \Rightarrow \begin{cases} E[X_1(t)] = 0\text{MPa} \\ E[X_2(t)] = 0\text{MPa} \\ V[X_1(t)] = 900\text{MPa}^2 \\ V[X_2(t)] = 675\text{MPa}^2 \end{cases}.$$

The previous considerations will now be applied to a realistic example which may be related to a tidal stream turbine. In this case, the mechanical stresses PSDs may be of the following type:

$$\begin{cases} S_{pq}(f) = 2\pi \frac{A_{pq}}{(2\pi f)^5} e^{-\left(\frac{3.11}{(2\pi f)^4 \times 8^2}\right)} & f > 0 \\ S_{pq}(f) = 0 & f = 0 \end{cases}. \quad (18)$$

Equation (18) represents a Bretschneider spectrum and is illustrated in Fig. 12. This kind of PSD is often encountered in the field of hydrology (Dalrymple, 1994). The values of the constants A_{pq} are set as follows: $A_{11} = 174.94$ S.I. units, $A_{22} = 131.21$ S.I. units, $A_{12} = 136.35$ S.I. units. It should be noted that in either case, A_{11} is such that $V[X_1(t)] = 900$ MPa^2, A_{22} is such that $V[X_2(t)] = 675$ MPa^2 and A_{12} is such that $\frac{\text{cov}[X_1(t), X_2(t)]}{\sqrt{V[X_1(t)]V[X_2(t)]}} = 0.9$. This situation is represented in Fig. 12.
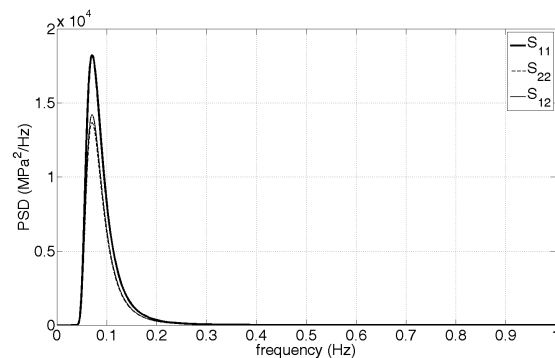
Figure 12: PSDs of the interested hydrology case.

Using the above-mentioned PSDs and properties, the signals samples look like those illustrated by Fig. 13. $n_f = 10^4$ frequencies are used. When comparing with Fig. 3, these signals correspond to wide-band processes. Moreover, it should be noticed that "short time" does not represent the same duration as in the academic cases. This is due to the difference of the related central frequencies: 10 Hz in the academic examples and 0.1 Hz in this application example.
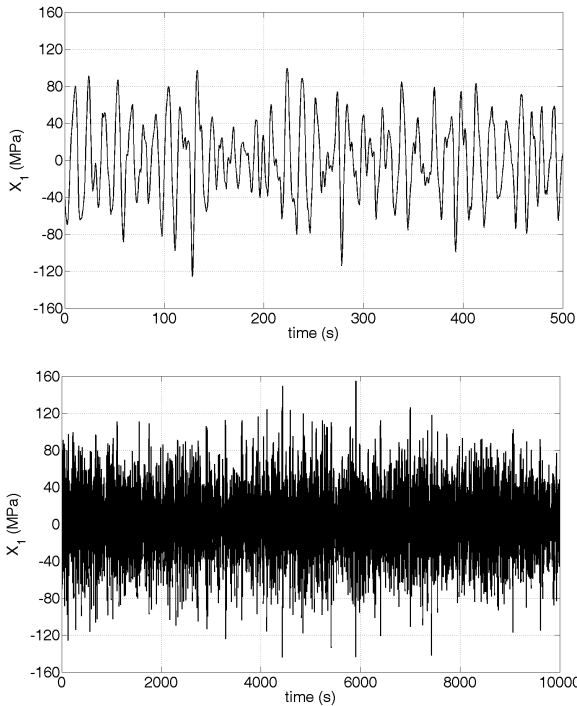
**Figure 13: Process (X_1) sample during a short (top) and a longer (bottom) time.**

Results analysis

In order to verify the adequateness of the computed samples, the empirical auto and cross-correlation functions are compared to the analytical ones in Fig. 14. $n_f = 10^4$ frequencies are used. Equations (12) to (14) are still valid in this situation. To compare the empirical and analytical correlation coefficients in a proper way, both the empirical and analytical correlation functions are normalized using the theoretical values of the variances. Due to the complexity of the involved PSDs, the analytical correlation functions cannot be evaluated analytically but they are computed by means of the IFFT algorithm. To quantify the precision of the simulation, let us compare in detail the theoretical and empirical values of the variances and cross-correlation coefficients. From Fig. 14, the empirical correlation coefficients agree well with the analytical ones. Moreover, Table 3 shows that the variances and cross-correlation coefficient of the computed processes are in total agreement with the theoretical values. This implies that the simulated processes actually correspond to their respective target PSDs and are correctly correlated.

As we did in the previous subsection Results analysis, marginal EVDs are fitted by GEV distributions. Figure 15 shows that the relative error is low. Then Eq. (16) can be used with the GEV approximations of the margins. The resulting distribution fits very well the empirical bivariate EVD, as shown by Fig. 16.

Table 3: Comparison between the theoretical and the empirical variances and cross-correlation coefficients.

	Empirical (target) value	Relative gap
$V[X_1]$	899.7903 (900)	-0.0233 %
$V[X_2]$	674.7338 (675)	-0.0394 %
ρ_{12}	0.8997 (0.9)	-0.0325 %
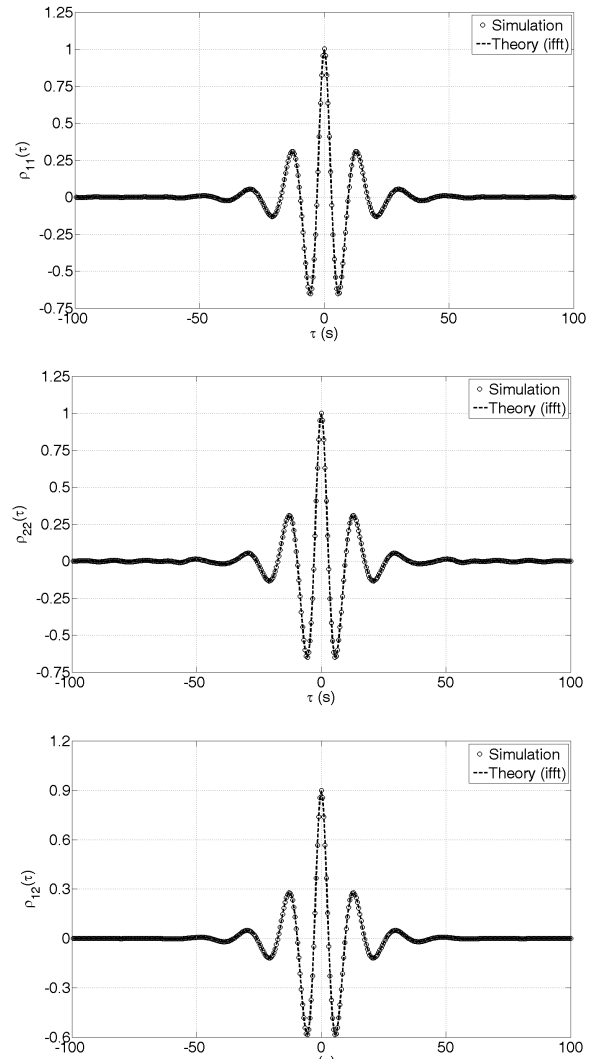
Figure 14: Comparison between empirical and analytical correlation coefficients.
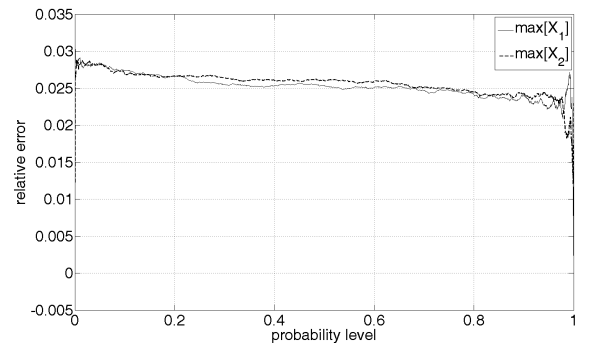
Figure 15: Relative error between the empirical margins and their GEV fits.

Figure 16: Bivariate extreme value distributions.

Conclusions and Perspectives

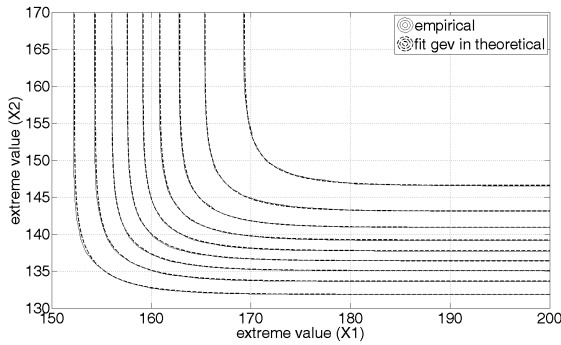We have presented an appropriate generation method for correlated Gaussian processes, dealing quite fast with a great amount of data thanks to the use of the IFFT algorithm. Moreover, it is able to give the processes extreme value distributions (univariate and multivariate) with a good quality, having a great confidence interval. Fitting the extreme value margins with GEV distributions provides excellent approximations and the model chosen for building the bivariate extreme value distribution is accurate. With this result, the structural reliability can then be evaluated by estimating the equivalent shear stress that is involved in multiaxial fatigue criteria; see Lambert et al. (2010) for an example of such an approach.

However, the bivariate extreme value distribution model involves the associated margins (or here their GEV fits) as well as the extreme value cross-correlation coefficient. Both these quantities are currently known from Monte-Carlo simulations, which are still time-consuming. Further work is then to be carried out to evaluate the cross-correlation coefficient of the extreme value distribution directly from the spectral description of the related processes. The same remark holds for the parameters of the GEV distributions fitting the margins. The influence parameters may be the spectral shapes and moments of the involved PSDs as well as the samples duration. The extension of the presented model for building bivariate EVD to more than two processes is also to be considered in order to deal with up to six-dimensional stress fields.

Appendix

Ergodicity for the mean

For more convenience, we introduce $\omega_i = 2\pi f_i$ in the following statements. At any given time t, the mathematical expectation of the process $X_p(t)$ is:

$$
\begin{aligned}
\mu_{X_p}(t) &= E\left[X_p(t)\right] \\
&= \sum_{i=1}^{n_f}\sum_{q=1}^{n} \sigma_{ipq} \left\{ E\left[A_{iq}\right]\cos\left(\omega_i t + \theta_{ipq}\right) \right. \\
&\quad \left. + E\left[B_{iq}\right]\sin\left(\omega_i t + \theta_{ipq}\right) \right\} = 0.
\end{aligned}
$$

For any given realization of the process $X_p(t)$, the temporal mean is:

$$
\begin{aligned}
\mu_{X_p}^{(r)} &= \frac{1}{T}\sum_{i=1}^{n_f}\sum_{q=1}^{n} \sigma_{ipq}\left\{ A_{iq}^{(r)}\int_0^T \cos\left(\omega_i t + \theta_{ipq}\right) \mathrm{d}t \right. \\
&\quad \left. + B_{iq}^{(r)}\int_0^T \sin\left(\omega_i t + \theta_{ipq}\right)\mathrm{d}t \right\} = 0.
\end{aligned}
$$

Both the mathematical expectation and the temporal mean are equal (to zero), hence the processes are ergodic for the mean value.

Non ergodicity for the correlation function

At any given time t, the correlation function between the processes $X_p(t)$ and $X_z(t)$ is:

$$
\begin{aligned}
R_{pz}(\tau) &= E\left[X_p(t)X_z(t+\tau)\right] \\
&= \sum_{i_1=1}^{n_f}\sum_{i_2=1}^{n_f}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n} \sigma_{i_1 p q_1}\sigma_{i_2 z q_2} \\
&\quad \times \left\{ E\left[A_{i_1 q_1}A_{i_2 q_2}\right]\cos\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\cos\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right) \right. \\
&\quad + E\left[A_{i_1 q_1}B_{i_2 q_2}\right]\cos\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\sin\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right) \\
&\quad + E\left[B_{i_1 q_1}A_{i_2 q_2}\right]\sin\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\cos\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right) \\
&\quad \left. + E\left[B_{i_1 q_1}B_{i_2 q_2}\right]\sin\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\sin\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right) \right\} \\
&= \sum_{i=1}^{n_f}\sum_{q=1}^{n} \sigma_{ipq}\sigma_{izq} \\
&\quad \times \left\{ E\left[A_{iq}^2\right]\cos\left(\omega_i t + \theta_{ipq}\right)\cos\left(\omega_i(t+\tau) + \theta_{izq}\right) \right. \\
&\quad \left. + E\left[B_{iq}^2\right]\sin\left(\omega_i t + \theta_{ipq}\right)\sin\left(\omega_i(t+\tau) + \theta_{izq}\right) \right\}.
\end{aligned}
$$

Since A_{iq} and B_{iq} are zero-mean random variables, their mean square values are equal to their (unit) variances $E\left[A_{iq}^2\right] = E\left[B_{iq}^2\right] = 1$. This yields:

$$
\begin{aligned}
R_{pz}(\tau) &= \sum_{i=1}^{n_f}\sum_{q=1}^{n} \sigma_{ipq}\sigma_{izq}\left\{ \cos\left(\omega_i t + \theta_{ipq}\right)\cos\left(\omega_i(t+\tau) + \theta_{izq}\right) \right. \\
&\quad \left. + \sin\left(\omega_i t + \theta_{ipq}\right)\sin\left(\omega_i(t+\tau) + \theta_{izq}\right) \right\} \\
&= \sum_{i=1}^{n_f}\sum_{q=1}^{n} \sigma_{ipq}\sigma_{izq}\cos\left(\omega_i t\tau + \theta_{izq} - \theta_{ipq}\right).
\end{aligned}
$$

For any given realization of the processes $X_p(t)$ and $X_z(t)$, the temporal correlation function is:

$$
\begin{aligned}
R_{pz}^{(r)}(\tau) &= \left\langle X_p^{(r)}(t)X_z^{(r)}(t+\tau)\right\rangle = \frac{1}{T}\int_0^T X_p^{(r)}(t)X_z^{(r)}(t+\tau)\,\mathrm{d}t \\
&= \frac{1}{T}\sum_{i_1=1}^{n_f}\sum_{i_2=1}^{n_f}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n} \sigma_{i_1 p q_1}\sigma_{i_2 z q_2} \\
&\quad \times \left\{ A_{i_1 q_1}^{(r)}A_{i_2 q_2}^{(r)}\int_0^T \cos\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\cos\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right)\mathrm{d}t \right. \\
&\quad + A_{i_1 q_1}^{(r)}B_{i_2 q_2}^{(r)}\int_0^T \cos\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\sin\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right)\mathrm{d}t \\
&\quad + B_{i_1 q_1}^{(r)}A_{i_2 q_2}^{(r)}\int_0^T \sin\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\cos\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right)\mathrm{d}t \\
&\quad \left. + B_{i_1 q_1}^{(r)}B_{i_2 q_2}^{(r)}\int_0^T \sin\left(\omega_{i_1}t + \theta_{i_1 p q_1}\right)\sin\left(\omega_{i_2}(t+\tau) + \theta_{i_2 z q_2}\right)\mathrm{d}t \right\} \\
&= \frac{1}{T}\sum_{i_1=1}^{n_f}\sum_{i_2=1}^{n_f}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n} \sigma_{i_1 p q_1}\sigma_{i_2 z q_2} \\
&\quad \times \left\{ A_{i_1 q_1}^{(r)}A_{i_2 q_2}^{(r)}I_1 + A_{i_1 q_1}^{(r)}B_{i_2 q_2}^{(r)}I_2 + B_{i_1 q_1}^{(r)}A_{i_2 q_2}^{(r)}I_3 + B_{i_1 q_1}^{(r)}B_{i_2 q_2}^{(r)}I_4 \right\}.
\end{aligned}
$$

After some calculations, we find:

$$I_1 = \frac{1}{2}\int_0^T \cos\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t$$

$$I_2 = -\frac{1}{2}\int_0^T \sin\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t$$

$$I_3 = \frac{1}{2}\int_0^T \sin\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t$$

$$I_4 = \frac{1}{2}\int_0^T \cos\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t.$$

Replacing I_1 to I_4 by their respective expressions gives:

$$R_{pz}^{(r)}(\tau) = \frac{1}{T}\sum_{i_1=1}^{n_f}\sum_{i_2=1}^{n_f}\sum_{q_1=1}^{n}\sum_{q_2=1}^{n}\sigma_{i_1pq_1}\sigma_{i_2zq_2}$$

$$\times\left\{ \frac{A_{i_1q_1}^{(r)}A_{i_2q_2}^{(r)} + B_{i_1q_1}^{(r)}B_{i_2q_2}^{(r)}}{2}\int_0^T \cos\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t \right.$$

$$\left. + \frac{B_{i_1q_1}^{(r)}A_{i_2q_2}^{(r)} - A_{i_1q_1}^{(r)}B_{i_2q_2}^{(r)}}{2}\int_0^T \sin\left((\omega_{i_1}-\omega_{i_2})t - \omega_{i_2}\tau + \theta_{i_1pq_1} - \theta_{i_2zq_2}\right)\mathrm{d}t \right\}$$

$$= \frac{1}{T}\sum_{i=1}^{n_f}\sum_{q=1}^{n}\sigma_{ipq}\sigma_{izq}\frac{A_{iq}^{(r)^2} + B_{iq}^{(r)^2}}{2}\int_0^T \cos\left(-\omega_i\tau + \theta_{ipq} - \theta_{izq}\right)\mathrm{d}t$$

$$= \sum_{i=1}^{n_f}\sum_{q=1}^{n}\sigma_{ipq}\sigma_{izq}\frac{A_{iq}^{(r)^2} + B_{iq}^{(r)^2}}{2}\cos\left(\omega_i\tau + \theta_{izq} - \theta_{ipq}\right).$$

Both the mathematical expectation and the temporal mean are not equal, hence the processes are not ergodic for the correlation function.

References

Dalrymple, R.A., 1994, "The theory of short-period waves," in M.B. Abbott and W.A. Price, editors, Coastal, estuarial, and harbour engineers' reference book, pp. 37-44.

De Haan, L. and Ferreira, A., 2006, "Extreme Value Theory: An Introduction," Springer, Ch. 9.

Elshamy, M., 1992, "Bivariate extreme value distributions," NASA Contractor Report No. 4444.

Fisher, R.A. and Tippet L.H.C., 1928, "Limiting forms of the frequency distribution of the largest and smallest member of a sample," Proceedings of the Cambridge Philosophical Society, Vol. 24, pp. 180-190.

Gmür, T., 1997, "Dynamique des structures: analyse modale numérique," Presses polytechniques et universitaires romandes, Ch. 2-3.

Guéna, F., Daviau, J.F., Majastre, H., Bischoff, V., Ruer, J. and Tartivel, C., 2006, "Design and operational features of a tidal stream turbine," Proceedings of the European Seminar on Offshore Wind and Other Marine Renewable Energies in Mediterranean and European Seas, OWEMES, Civitavecchia, Italy, 22nd April 2006, 8p.

Jenkinson, A.F., 1955, "The frequency distribution of the annual maximum (or minimum) values of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, John Wiley & Sons, Ltd, Vol. 81, pp 158-171.

Khalij, L., Pagnacco, E. and Lambert, S., (2010) "A measure of the equivalent shear stress amplitude from a prismatic hull in the principal coordinate system," *International Journal of Fatigue*, Vol. 32, No. 12, pp. 1977-1984.

Lambert, S., Pagnacco, E. and Khalij, L., 2010, "A probabilistic model for the fatigue reliability of structures under random loadings with phase shift effects," *International Journal of Fatigue*, Vol. 32, No. 2, pp. 463-474.

von Mises, R., 1936. "La distribution de la plus grande de n valeurs," reprinted in Selected Papers Volumen II, American Mathematical Society, Providence, R.I., 1954, pp. 271-294.

Raynal-Villasenor, J.A. and Salas, J.D., 1987, "Multivariate extreme value distributions in hydrological analyses." Proceedings of the Rome Symposium on "Water for the future: hydrology in perspective", IAHS Publ. No. 164, 10p.

Sun, J.Q., 2006, "Stochastic dynamics and control," in J.-Q. Sun, editor, Stochastic Dynamics and Control, Vol. 4 of Monograph Series on Nonlinear Science and Complexity, pp.229-235.

Theodorsen, T., 1934, "General theory of aerodynamic instability and the mechanism of flutter," National Advisory Committee for Aeronautics Report No. 496.

Thresher, R.W., Mirandy, L.P., Carne, T.G., Lobitz, D.W. and James, G.H., 2009, "Chapter 11: Structural dynamic behavior of wind turbines," Spera, D. A., ed. New York: American Society of Mechanical Engineers (ASME) pp. 605-662; NREL Report No. CH-500-45391.

Tucker, M.J., Challenor, P.G. and Carter, D.J.T., 1984, "Numerical simulation of a random sea: a common error and its effect upon wave group statistics," *Applied Ocean Research*, Vol. 6, No. 2, pp. 118-122.