EXTREME VALUE ESTIMATION OF MOORING LINES TOP TENSION
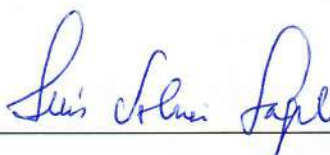
Marina Leivas Simão

Rio de Janeiro

Janeiro de 2019

EXTREME VALUE ESTIMATION OF MOORING LINES TOP TENSION

Marina Leivas Simão

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA CIVIL.

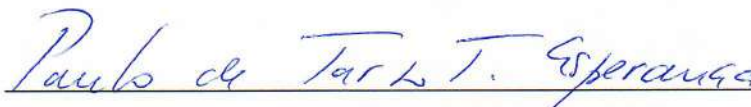Examinada por:

_____

Prof. Luis Volnei Sudati Sagrilo, D.Sc.

_____

Dr. Paulo Mauricio Videiro, Ph.D.

_____

Dr. Mauro Costa de Oliveira, Ph.D.

_____

Prof. Paulo de Tarso Themistocles Esperança, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

JANEIRO DE 2019

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed."

- Albert Einstein

ACKNOWLEDGMENTS

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ESTIMATIVA DE VALORES EXTREMOS DA TRAÇÃO DE TOPO EM LINHAS DE ANCORAGEM

Marina Leivas Simão

Janeiro/2019

Orientador: Luis Volnei Sudati Sagrilo

Programa: Engenharia Civil

Sabe-se que o sistema de ancoragem de plataformas flutuantes responde de forma não-linear aos carregamentos ambientais. Embora o modelo de excitação das ondas seja um processo Gaussiano, a tração nas linhas geralmente não o é devido aos movimentos de segunda ordem de baixa frequência do flutuador e às não-linearidades inerentes ao próprio sistema. Neste trabalho, a tração de curto prazo na linha é tratada como um processo ergódigo não-Gaussiano. A tração extrema é estimada com base na amostra de picos de uma única série temporal simulada de tração. Um número de modelos probabilísticos conhecidos são ajustados aos picos da série e a estatística de ordem clássica é aplicada para se determinar a tração extrema mais provável na linha, correspondente a um período de curto-prazo especificado (3-h), para identificar aquele com melhor performance. Os efeitos dos principais parâmetros da análise dinâmica, como a duração da simulação e o nível de discretização do espectro de ondas, também são investigados através da simulação de várias séries de tração. Além disso, o efeito da correlação entre picos consecutivos de tração na estimativa de valores extremos é investigado através da cadeia de Markov de primeira ordem (usando um modelo baseado na transformação de Nataf da distribuição de probabilidades conjunta de dois picos consecutivos) e através do método ACER. Mostra-se que esta consideração leva a estimativas extremas invariavelmente menores do que as obtidas pela estatística de ordem clássica. Mostra-se também que estas são mais próximas das estimativas obtidas diretamente de uma amostra de máximos episódicos obtida a partir de várias simulações numéricas distintas. Os exemplos numéricos abrangem dois casos de estudo de linhas de ancoragem pertencentes a unidades FPSO a serem instaladas no Brasil.

EXTREME VALUE ESTIMATION OF MOORING LINES TOP TENSION

Marina Leivas Simão

January/2019

Advisor: Luis Volnei Sudati Sagrilo

Department: Civil Engineering

It is known that the mooring system of floating platforms responds non-linearly to environmental loads. Even though the wave-frequency excitation can be assumed as a Gaussian process, the line tension generally is not due to second-order slow-drift floater motions and nonlinearities of the system itself. This work assumes the short-term line tension as a non-Gaussian ergodic process. The extreme tension is estimated based on the peaks sample of a single simulated tension time-history. A number of known probability distributions are fitted to the peaks of the time series and classic order statistics theory is applied to determine the most probable extreme line tension corresponding to a specified short-time period (3-h) in order to identify the one with best performance. The effects of major parameters in the dynamic analysis, such as simulation length and discretization level of the wave spectrum, are also investigated using several simulated tension time-histories. Furthermore, the effect of correlation between consecutive tension peaks in the extreme values estimation is investigated through the one-step Markov chain condition (using a Nataf transformation-based model for two consecutive peaks joint probability distribution) and through the ACER method. It is shown that this consideration leads to extreme value estimates that are invariably smaller than those obtained by standard order statistics. These estimates are also shown to be closer to the extreme estimates directly obtained from a sample of epochal maxima taken from several distinct numerical simulations. Numerical examples cover two study cases for mooring lines belonging to FPSO units to be installed offshore Brazil.

INDEX

ILLUSTRATION INDEX

TABLE INDEX

1 INTRODUCTION

In the design of floating production units and their mooring and production lines, it is of general understanding that the response is nonlinearly related to the applied environmental loads. Even though the short-term wave-frequency excitation can be assumed as a normal (or Gaussian) process, the mooring line tensions generally are not mainly due to the second-order effects associated to the floater slowly varying motions. It is clear that if an irregular wave elevation is modeled as a superposition of components at different frequencies then motions proportional to the square of that elevation will contain terms at both the sums and differences of the elementary wave frequencies [1]. The presence of these low difference frequency components in the excitation results in a resonant response of the floater-lines system in their horizontal modes, which is characterized for very high natural periods (order of 200~400s). In general, the frequency-domain analysis is not applicable since it relies on the linearization of the mooring system structural behavior. A nonlinear fully coupled dynamic analysis comprising both the floater and its risers and mooring lines in the time domain should be therefore a more appropriate approach.

Because of the nonlinearities in the behavior of the floater-lines system, one very challenging aspect of the mooring lines design is their extreme loads estimation, more specifically the estimation of characteristic short-term extremes values based on the sampled time-series. Taking the maximum observed line tension in each of independent short-term time-domain simulations (realizations) results in a set of distinct extreme values. Consequently, the ideal practice for estimating extreme tension values should be to perform a sufficiently large number N of independent simulations and fit an extreme value probability distribution to the sample of N maxima extracted from each simulation [2][3]. The extreme response usually adopted in the design is the most probable extreme value (MPV), i.e., the point where the fitted probability density function reaches its maximum value. As the number N of simulations increases, a more accurate distribution is obtained and hence a more accurate MPV response is estimated. However, this procedure is not feasible for the great majority of marine structures as it can be very time-consuming and cumbersome for everyday design applications.

On the other hand, by assuming the short-term line tension as a stationary and ergodic process, one can estimate extreme tensions from just a single response simulation.

This hypothesis is generally implicit in numerical simulations and methodologies used for marine structures design. Under this assumption, the short-term extreme response can be estimated using the peaks (maxima) probability distribution of a single tension time-history [2][3]. As there is not a theoretical solution for this distribution in the case of a non-Gaussian process, the procedure is to fit a given probability model to the peaks sample and use standard order statistics to obtain the short-term extreme response. In this work, searching for the model with best performance, three choices for the peaks probability model are investigated: (i) three-parameter Weibull model, (ii) two-parameter Weibull model fitted to the high probability level data (Weibull-tail model) and (iii) four-parameter Shifted Generalized Lognormal Distribution (SGLD) model, recently developed by Low [4]. A more recently developed extremes estimation method, the ACER (Average Conditional Exceedance Rate) method [5], is also investigated. The ACER method is not based on a procedure of probability model fitting to the peaks of a sample but rather on the estimation of rates of level crossing exceedances.

Moreover, this work also examines how some major parameters of a typical time-domain numerical simulation can influence the estimation of the MPV from a single tension time-history, such as the simulation length and the level of wave spectrum discretization. All results are compared to the 'true' value, i.e., the one obtained from fitting an extreme probability distribution model to an extreme short-term values sample taken from several (and distinct) time-domain simulations.

Another important aspect of the study of extreme mooring line tension response is that its low-frequency component is narrow-banded. An essential characteristic of a narrow-band process is the higher dependency (correlation) between the successive response peaks. This effect on the short-term extreme MPV is considered by means of the one-step Markov chain condition using a Nataf transformation-based model for the joint probability distribution of two consecutive peaks. This correlation is also investigated through the ACER method, considering the behavior of two and three consecutive peaks.

The numerical examples presented in this study comprise a chain-polyester-chain mooring line connected to a spread-moored FPSO in deep water and a chain-wire rope-chain mooring line connected to a turret-moored FPSO in shallow water.

Following Introduction, Chapter 2 presents basic probability and statistics concepts for better understanding of the theoretical basis and terminology used throughout the study. Chapter 3 presents concepts of the statistical analysis of random processes.

The most important theoretical basis of this study is presented in Chapter 4, where extremes statistics is explained in details for Gaussian and non-Gaussian random processes. In Chapter 5, a summary of the main characteristics and assumptions of a typical mooring line extreme load analysis is presented.

The methods for extremes estimation presented in Chapter 4 are employed in two case studies in Chapter 6. In the latter, it is presented the details and results of the performed analyses. Main conclusions of the present study and recommendations for future works are given in Chapter 7.

The References section lists all previous studies and literature that was taken as foundation for the present work.

In the Annex, further information and data regarding the analyses performed in this study can be found.

2 BASIC PROBABILITY AND STATISTICS CONCEPTS

2.1 Single Random Variables

In probability and statistics, a random or stochastic variable can be understood as a real function defined on a sample space. Given a random phenomenon Y, the random variable X describes its stochastic behavior. A realization x of the random variable X is any outcome of the random phenomenon Y. For every outcome x of the random variable there is a probability $P(X < x)$ associated with it. In this work, random variables are denoted by capital letters and the corresponding small letters denote their realizations.

A random variable is characterized by its probability density function $f_X(x)$, or simply its PDF, which satisfies the following condition

$$f_X(x)dx = P(x \leq X \leq x + dx) \tag{2.1}$$

where $P(A)$ indicates the probability that the event A occurs. The probability that the variable X assumes values between known limits a and b is given by

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx \tag{2.2}$$

This probability corresponds to the area under the curve $f_X(x)$ contained in the interval $[a, b]$. It is possible to demonstrate that every PDF satisfies the following properties:

i. The probability of occurrence of any event is always non-negative;

$$f_X(x) \geq 0 \tag{2.3}$$

ii. The sum of all possible results probabilities is 1 (or 100%).

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{2.4}$$

The integral of the PDF is known as the Cumulative Density Function, or CDF, and is defined as

$$F_x(a) = \int_{-\infty}^a f_X(x)dx = P(X \leq a) \tag{2.5}$$

For a continuous random variable X with PDF $f_X(x)$, the **mean** (or **expected value)** of the variable is

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx \tag{2.6}$$

and is also known as the first moment of the probability distribution $f_X(x)$.

The **mode** or the **most probable value** (**MPV**) of a continuous random variable is the point in which the probability density function reaches its maximum value, as shown in Figure 1.
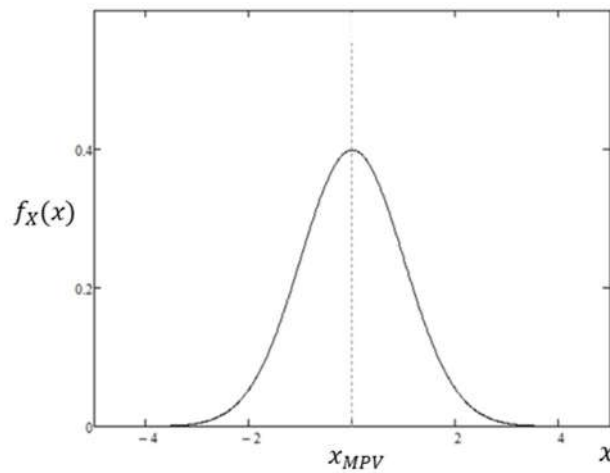
Figure 1 Mode or Most Probable Value of a generic random process

Besides the mean, the next most important quantity of a random variable is its measure of dispersion or variability, i.e., the quantity that gives a measure of how close (or far) the values of the variate are. For a continuous random variable X with PDF $f_X(x)$ and mean μ_X, the **variance** of X is

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \tag{2.7}$$

and is also known as the second moment of the probability distribution $f_X(x)$.

A more convenient measure of dispersion is known as the **standard deviation**, given by

$$\sigma_X = \sqrt{Var(X)} \tag{2.8}$$

5

High values of $Var(X)$ (and σ_X) indicate large dispersion of the variable outcomes relative to its mean, whereas low values of $Var(X)$ (and σ_X) indicate that the variable assumes values closer to its mean.

The **coefficient of variation**, identified as CoV_X or δ_X, is a dimensionless measure of X variability, defined as the ratio of the standard deviation to the mean (for $\mu_X \neq 0$):

$$\delta_X = \frac{\sigma_X}{\mu_X} \tag{2.9}$$

The **coefficient of asymmetry** (or **skewness**), identified as γ_X, is defined as

$$\gamma_X = \frac{E(X - \mu_X)^3}{\sigma_X^3} = \frac{\int_{-\infty}^{\infty}(x - \mu_X)^3 f_X(x)dx}{\sigma_X^3} \tag{2.10}$$

and it is also known as the third moment of the probability distribution $f_X(x)$. The coefficient of asymmetry, as the name indicates, is a parameter associated to the distribution asymmetry about its mean. A null skewness value indicates that $f_X(x)$ is symmetric. A positive skew indicates that the longest tail is on the right side of the probability distribution function and a negative skew indicates that the longest tail is on the left. This is illustrated in Figure 2.
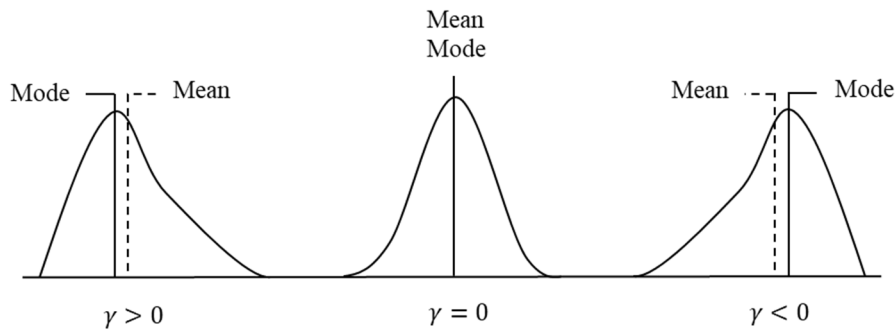
Figure 2 Skewness graphical interpretation

The **coefficient of kurtosis**, or simply **kurtosis**, identified as κ_X, is defined as

$$\kappa_X = \frac{E(X - \mu_X)^4}{\sigma_X^4} = \frac{\int_{-\infty}^{\infty}(x - \mu_X)^4 f_X(x)dx}{\sigma_X^4} \tag{2.11}$$

and it is also known as the fourth moment of the probability distribution $f_X(x)$. Figure 3 illustrates how kurtosis affects the probability distribution function curve. The kurtosis is

a measure of a distribution slenderness. The kurtosis of any univariate normal distribution is 3. It is common to compare the kurtosis of a distribution to this value.
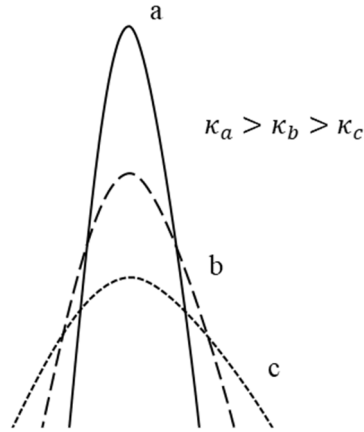
Figure 3 Kurtosis graphical interpretation

2.2 Statistical equivalence

X and Y are random variables with statistical equivalence at $X = x$ and $Y = y$ if their CDFs are identically evaluated at these points. This can be mathematically represented by

$$F_X(x) = F_Y(y) \tag{2.12}$$

$$x = F_X^{-1}\big(F_Y(y)\big) \tag{2.13}$$

$$y = F_Y^{-1}\big(F_X(x)\big) \tag{2.14}$$

where $F_X^{-1}(.)$ and $F_Y^{-1}(.)$ are the inverse of the cumulative probability functions of the variables X and Y, respectively. This equivalence is illustrated in Figure 4.
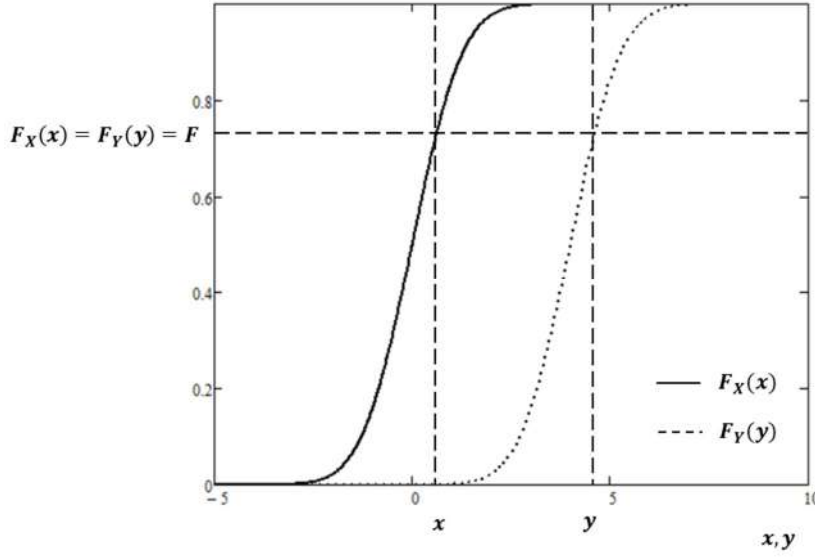
Figure 4 Statistical equivalence of X and Y at $X = x$ and $Y = y$

2.3 Jointly Distributed Random Variables

The concept of a single random variable and its probability distribution can be extended to two or more random variables. In order to quantify events that are associated to two or more physical processes, the events in the sample space may be mapped into two (or more) dimensions of the real space, requiring two or more random variables.

If X and Y represent physical phenomena, there are probabilities associated with any outcome, such as $\{X \leq x, Y \geq y\}$. The probabilities for all possible outcomes may be described with the joint cumulative probability function of the random variables X and Y, defined as

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \qquad (2.15)$$

This function defines the probability of the joint occurrences of the events identified by $X \leq x$ and $Y \leq y$. It is possible to demonstrate that it satisfies the following properties [6]:

i. $\quad F_{X,Y}(-\infty, -\infty) = 0; \qquad F_{X,Y}(\infty, \infty) = 1.0 \qquad\qquad (2.16)$

ii. $\quad F_{X,Y}(-\infty, y) = 0; \qquad F_{X,Y}(\infty, y) = F_Y(y)$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.17)$

$\quad F_{X,Y}(x, -\infty) = 0; \qquad F_{X,Y}(x, \infty) = F_x(x)$

iii. $F_{X,Y}(x,y)$ is nonnegative and a nondecreasing function of x and y. (2.18)

The joint probability density function, or PDF, is defined as

$$f_{X,Y}(x,y)dxdy = P(x \leq X \leq x + dx, y \leq Y \leq y + dy) \qquad (2.19)$$

then

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dvdu \qquad (2.20)$$

The statistical behavior of the X variable may depend on the values of Y, or vice versa. The conditional probability function is defined as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad (2.21)$$

Hence

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) \qquad (2.22)$$

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) \qquad (2.23)$$

However, if X and Y are statically independent, i.e., $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$, then

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad (2.24)$$

The covariance of X and Y is the joint second moment about the means μ_X and μ_Y i.e.,

$$Cov(X,Y) = E\big((X - \mu_X)(Y - \mu_Y)\big) = E(XY) - \mu_X\mu_Y \qquad (2.25)$$

If X and Y are statistically independent, then $Cov(X,Y) = 0$.

The physical interpretation of the covariance can be inferred from Equation 2.25 [6]. If the $Cov(X,Y)$ is large and positive, the values of X and Y tend to be both large or both small relative to their respective means. On the other hand, if the $Cov(X,Y)$ is large and negative, the values of X tend to be large when the values of Y are small, and vice

versa, relative to their respective means. If the $Cov(X, Y)$ is small or zero, there is no or little linear relationship between the values of X and Y.

Therefore, the $Cov(X, Y)$ is a measure of the degree of linear relationship between the variables X and Y. However, a more often used indicator for linear dependency is the Pearson correlation coefficient, or simply correlation coefficient. It can be interpreted as a normalized value of $Cov(X, Y)$:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.26}$$

The range of $\rho_{X,Y}$ is between -1 and +1. When $\rho_{X,Y} = \mp 1$, or for intermediate values of $\rho_{X,Y}$, X and Y are linearly related, or present some linear relationship, as shown in Figure 5(b), whereas, when $\rho_{X,Y} = 0$, values of X and Y may appear as in Figure 5(a). In the latter case, the variables are said to be independent, whereas in the first case they are said to be correlated. It is important to notice that the correlation coefficient is not capable of describing eventual nonlinear relationship between variables.

(a) x and y values
are uncorrelated

(b) x and y values
exhibit correlation
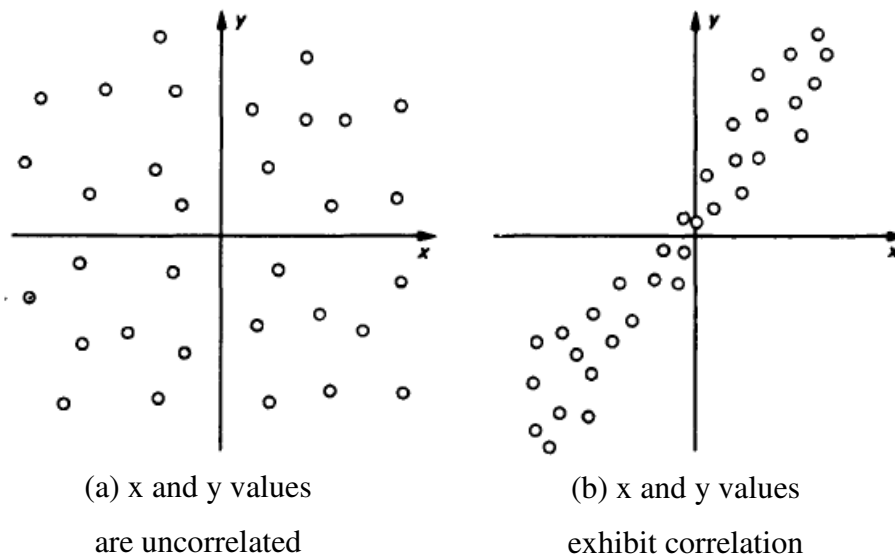
Figure 5 Illustrating correlation between two random variables x and y [7]

If the correlation coefficient between two random variables X and Y is known, the joint probability density function $f_{X,Y}(x, y)$ can be approximately determined through the Nataf transformation model [8]. According to the Nataf transformation model, the joint probability density function of correlated random variables X and Y is given by

$$f_{X,Y}(x,y) = \frac{f_X(x)f_Y(y)}{\phi(z)\phi(w)}\phi_2(z,w,\rho) \tag{2.27}$$

where $z = \Phi^{-1}(F_X(x))$, $w = \Phi^{-1}(F_Y(y))$, $\Phi^{-1}(.)$ is the inverse of cumulative probability distribution of a standard normal variable, $f_X(x)$ and $f_Y(y)$ are the marginal probability density functions of X and Y, $\phi(.)$ is the probability density function of a standard normal variable and $\phi_2(.)$ is the joint probability density function of two standard correlated normal variables with a correlation coefficient ρ given by

$$\phi_2(z,w,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(z^2 + w^2 - 2\rho zw)\right] \tag{2.28}$$

The joint cumulative distribution $F_{X,Y}(x,y)$ can be obtained by

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy \tag{2.29}$$

The above double integral can be solved in a fast way by proper use of numerical transformations [9] as

$$\begin{aligned} F_{X,Y}(x,y) = 1 - F_X(x) - F_Y(y) \\ + \int_{-\infty}^{\rho}\phi_2\left(-\Phi^{-1}(F_X(x)), -\Phi^{-1}(F_Y(y)), \xi\right)d\xi \end{aligned} \tag{2.30}$$

with $\Phi(.)$ standing for the cumulative probability distribution of a standard normal variable. A more detailed explanation on the Nataf transformation is given in Annex A.

2.4 Probability Distribution Models

One of the greatest challenges of stochastic and probabilistic analysis is correctly interpreting the analyzed data. Primarily, the aim of any study that involves variability, or uncertainty, is to adequately predict the probability of occurrence of each possible outcome of the variate in question. In general, that means fitting an adequate probability distribution model to the set of data available, i.e., finding a real function that is able to predict outcomes probabilities that are sufficiently close to the real ones.

This fitting process can be done through different methods such as the visual comparison of the theoretical and empirical PDFs and CDFs method, the method of moments or the linear regression method. The present work will focus on the method of moments and on the linear regression method.

Consider a sample of the random variable of interest X. In the **method of moments**, it is assumed that the moments (mean, variance, skewness and kurtosis) of the sample are the same as the probability distribution model chosen to represent it. In other words, the probability model parameters are straightforwardly determined by equaling its moments to the sample moments.

In the **linear regression method**, the parameters of the probability distribution model are estimated through adjusting its linearized CDF to the linearized empirical CDF of the sample. This adjustment is usually optimized by numerical techniques such as the least squares method. This method will become clearer in the following chapters of the present work.

In theory, any function that meets the requirements presented in Equations (2.3) and (2.4) can be understood as a probability distribution model, however there are many consolidated models (or functions) previously defined in literature. In this chapter, a few of the most used probability distributions models in offshore engineering will be presented.

2.4.1 Uniform Distribution

A variable X is uniformly distributed if its PDF is constant for any outcome of the variate. If the variable X is uniformly distributed in the interval $[a, b]$, its PDF and CDF will be given, respectively, by

$$f_X(x) = \frac{1}{b-a} \qquad a \leq x \leq b \tag{2.31}$$

$$F_X(x) = \frac{x-a}{b-a} \qquad a \leq x \leq b \tag{2.32}$$

There are many applications in which the uniform distribution is useful, such as running simulations experiments. Many programming languages come with

implementations (numerical routines) that are able to generate pseudo-random numbers that are distributed according to the standard uniform distribution, whose interval is [0,1]. If x is a value sampled from the standard uniform distribution, then the value $a + (b - a)x$ follows the uniform distribution parameterized by the interval of interest $[a, b]$. The uniform distribution is also useful for sampling from an arbitrary distribution, using the concept of statistical equivalence showed in section 2.2.

2.4.2 Normal or Gaussian Probability Distribution

A variable X is normally distributed if its PDF is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_X}{\sigma_X}\right)^2\right] \tag{2.33}$$

According to the above equation, this distribution depends only on the variable mean μ_X and standard deviation σ_X. This distribution is commonly designated by $N(\mu_X, \sigma_X^2)$.

As one can mathematically check, there is not a closed formula for $F_X(x)$ since it relies on an integral that can only be numerically estimated. $F_X(x)$ is usually denoted by $\Phi(x)$ and can be calculated with the use of numerical methods or tables available in literature. The gaussian probability density function curve defined by Equation 2.33 has the important particular of being at all times symmetric about its mean, as Figure 6 shows. This characteristic results in a null asymmetry (skewness) coefficient. Another particular of the gaussian PDF is that its kurtosis coefficient is invariably equal to 3.
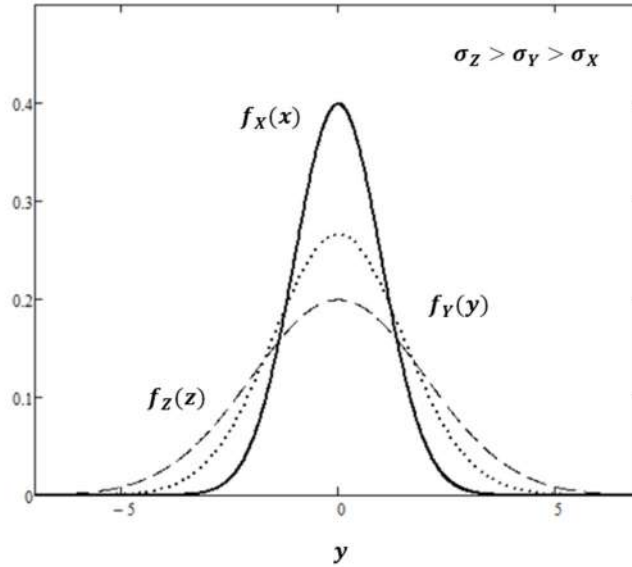
Figure 6 Family of normal or gaussian probability density functions

The simplest case of a normal distribution is known as the **standard normal distribution**. This is a special case when $\mu_X = 0$ and $\sigma_X = 1$. The standard normal probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \tag{2.34}$$

Every normal distribution is a version of the standard normal distribution whose domain has been stretched by a factor of σ_X and then translated by μ_X. This means that if Y is a standard normal variable, then $X = \sigma_X Y + \mu_X$ will have a normal distribution with parameters μ_X and σ_X. Then $Y = (X - \mu_X)/\sigma_X$ will have a standard normal distribution. This variate is called the standardized form of X.

2.4.3 Lognormal Probability Distribution

A random variable X is said to follow a lognormal distribution when its natural logarithm $ln(X)$ is normally distributed. Its PDF is hence given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}x\xi} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \lambda}{\xi}\right)^2\right] \tag{2.35}$$

14

In this case, the distribution also depends on two parameters, λ and ξ. This distribution is usually referred as $LN(\lambda, \xi)$. These parameters are related to the mean and standard deviation of X by

$$\mu_X = exp\left(\lambda + \frac{1}{2}\xi^2\right) \tag{2.36}$$

$$\sigma_X = \mu_X\sqrt{exp(\xi^2) - 1} \tag{2.37}$$

The cumulative probability distribution of this variable is related to the standard normal distribution by the following equation:

$$F_X(x) = \Phi\left(\frac{\ln(x) - \lambda}{\xi}\right) \tag{2.38}$$

2.4.4   3-Parameter Weibull Probability Distribution

The Weibull probability distribution is one of the most widely used distribution model in engineering applications due to its versatility and flexibility offered by its parameters [2]. A variable X follows a Weibull probability distribution when its PDF and CDF are respectively given by

$$f_X(x) = \frac{(x - \varepsilon)^{\lambda-1}}{\alpha^\lambda}\lambda exp\left[-\left(\frac{x - \varepsilon}{\alpha}\right)^\lambda\right] \tag{2.39}$$

$$F_X(x) = 1 - exp\left[-\left(\frac{x - \varepsilon}{\alpha}\right)^\lambda\right] \tag{2.40}$$

for $\alpha > 0$ and $\lambda > 0$. The parameters α, λ and ε are known as the scale, shape and location parameters, respectively. If the location parameter is set to zero then the above equations correspond to a 2-parameter Weibull distribution. The relationships between these parameters and the first four moments of the probability distribution $\mu_X, \sigma_X{}^2, \gamma_X$ and κ_X are given by

$$\mu_x = \varepsilon + \alpha\Gamma\left(1 + \frac{1}{\lambda}\right) \tag{2.41}$$

$$\sigma_x = \alpha \sqrt{\Gamma\left(1 + \frac{2}{\lambda}\right) - \Gamma^2\left(1 + \frac{1}{\lambda}\right)} \tag{2.42}$$

$$\gamma_x = \frac{\Gamma\left(1 + \frac{3}{\lambda}\right) - 3\Gamma\left(1 + \frac{1}{\lambda}\right)\Gamma\left(1 + \frac{2}{\lambda}\right) + 2\Gamma^3\left(1 + \frac{1}{\lambda}\right)}{\left[\Gamma\left(1 + \frac{2}{\lambda}\right) - \Gamma^2\left(1 + \frac{1}{\lambda}\right)\right]^{\frac{3}{2}}} \tag{2.43}$$

$$\kappa_x = \frac{\Gamma\left(1 + \frac{4}{\lambda}\right) - 4\Gamma\left(1 + \frac{1}{\lambda}\right)\Gamma\left(1 + \frac{3}{\lambda}\right) + 6\Gamma^2\left(1 + \frac{1}{\lambda}\right)\Gamma\left(1 + \frac{2}{\lambda}\right) - 3\Gamma^4\left(1 + \frac{1}{\lambda}\right)}{\left[\Gamma\left(1 + \frac{2}{\lambda}\right) - \Gamma^2\left(1 + \frac{1}{\lambda}\right)\right]^2} \tag{2.44}$$

where $\Gamma(.)$ is the Gamma function. This model is identified in the present work as the W3P approach.

A particular case of the Weibull distribution is the Rayleigh distribution, when $\varepsilon = 0$, $\lambda = 2$ and $\alpha = \sqrt{2}\alpha_R$. The PDF and CDF of the Rayleigh distribution are then given by

$$f_X(x) = \frac{1}{\alpha_R{}^2}\exp\left[-\frac{1}{2}\left(\frac{x}{\alpha_R}\right)^2\right] \tag{2.45}$$

$$F_X(x) = 1 - \exp\left[-\frac{1}{2}\left(\frac{x}{\alpha_R}\right)^2\right] \tag{2.46}$$

The Rayleigh distribution depends on the single parameter $\alpha_R > 0$.

2.4.4.1 Weibull Probability Distribution Fitted to the Tail of the Data (Weibull-Tail)

In an extreme analysis, the most important part of the data is that related to the upper tail of the probability distribution (see section 2.1). It is natural then to focus on the fitting of the probability distribution only to the upper part of the sample. This can be done using a linear regression analysis using empirical estimates for the cumulative probability distribution [10].

Particularly in this work, a 2-parameter Weibull is fitted to the tail of the data using various sub-sets located on the upper-tail of the sample. For the scope of this study, the sub-sets considered are those associated to the $60\%, 65\%, 70\%, 75\%, 80\%, 85\%$ and 90% percentiles of the empirical cumulative distribution. For each one of these levels, the data above them are used to get the parameters α and λ of the distribution through a linear regression, as explained in the following equations.

Taking the natural log of both sides of the Equation 2.40 two times yields

$$\ln\big(-\ln\big(1 - F_X(x)\big)\big) = \lambda \ln x - \lambda \ln \alpha \qquad (2.47)$$

This can be expressed as the linear equation

$$y = \lambda x\prime + a \qquad (2.48)$$

where $y = \ln\big(-\ln\big(1 - F_X(x)\big)\big)$, $x\prime = \ln x$ and $a = -\lambda \ln \alpha$. Given a sample of x and its empirical CDF, one can directly compute y and $x\prime$. The 2-P Weibull parameters can be then straightforwardly estimated via linear regression.

The final parameters of the Weibull-tail distribution are taken as the corresponding mean values of all sub-sets parameters. An example of the final fitting for the Weibull Tail procedure is given in Figure 7. This procedure is identified in this work as the WT approach. The Weibull-tail fitting procedure calculations are better explained in Annex B. More details about the WT approach can be found in Sagrilo et al [10].
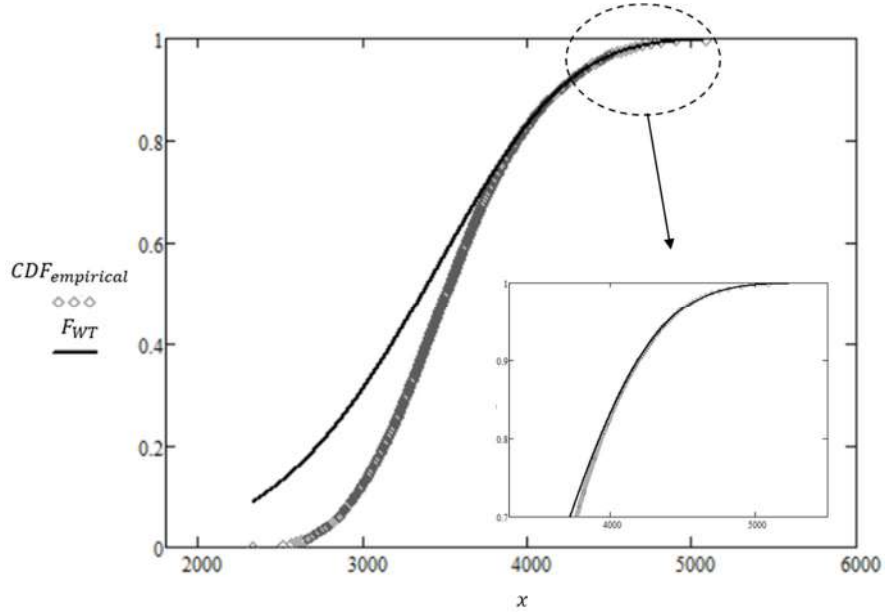
Figure 7 Weibull-tail cumulative distribution in comparison to the empirical CDF of the random variable

2.4.5 Shifted Generalized Lognormal Distribution

The Shifted Generalized Lognormal Distribution, or simply SGLD, was recently proposed by Low [4]. It is a four-parameter distribution whose PDF and CDF are respectively given by

$$f_X(x) = \frac{\alpha}{x-b} exp\left[-\frac{1}{r\sigma^r}\left|\ln\left(\frac{x-b}{\theta}\right)\right|^r\right] \tag{2.49}$$

$$F_X(x) = \frac{1}{2} + \frac{1}{2} \text{sgn}\left[\frac{x-b}{\theta} - 1\right] g\left(\frac{1}{r}, \frac{\left|\ln\left(\frac{x-b}{\theta}\right)\right|^r}{r}\right) \tag{2.50}$$

for $x > b$, $r > 0$ and $\sigma > 0$. The parameters b and θ are known as the location and scale parameters and r and σ are the two shape parameters of the distribution. sgn(.) is the signal function. The parameter α is defined by

$$\alpha = \frac{1}{2r^{1/r}\sigma\Gamma\left(1+\frac{1}{r}\right)} \tag{2.51}$$

and $g(.)$ is the incomplete gamma function, given by

18

$$g(v, x) = \frac{\int_0^x t^{v-1}e^{-t}dt}{\Gamma(v)} \tag{2.52}$$

In order to estimate the statistical moments of this distribution, a new auxiliary variable Y is defined from X as the following equation [4]

$$Y = \frac{X - b}{\theta} \tag{2.53}$$

It can be demonstrated the following relationships between the first four moments of the variables X and Y:

$$\mu_Y = \frac{\mu_X - b}{\theta} \tag{2.54}$$

$$\sigma_Y = \frac{\sigma_X}{\theta} \tag{2.55}$$

$$\gamma_Y = \gamma_X \tag{2.56}$$

$$\kappa_Y = \kappa_X \tag{2.57}$$

Additionally, Low [4] shows that the first four moments of Y can be obtained by

$$\mu_Y = \mu_Y(r, \sigma) = E[Y^1] \tag{2.58}$$

$$\sigma_Y = \sigma_Y(r, \sigma) = \sqrt{E[Y^2] - (E[Y^1])^2} \tag{2.59}$$

$$\gamma_Y = \gamma_Y(r, \sigma) = \frac{E[Y^3] - 3\mu_Y \cdot \sigma_Y^2 - \mu_Y^3}{\sigma_Y^3} \tag{2.60}$$

$$\kappa_Y = \kappa_Y(r, \sigma) = \frac{E[Y^4] - 4\mu_Y E[Y^3] + 6\mu_Y^2 \sigma_Y^2 + 3\mu_Y^4}{\sigma_Y^4} \tag{2.61}$$

where

$$E[Y^k] = \frac{1}{\Gamma\left(\frac{1}{r}\right)} \sum_{n=0}^{\infty} \left[\frac{(k\sigma)^{2n}}{(2n)!} r^{2n/r} \Gamma\left(\frac{2n+1}{r}\right) \right] \qquad k = 1,2,3,4 \tag{2.62}$$

The statistical parameters of the variable X can be obtained through the method of moments. Since the SGLD parameters r and σ are directly related to the distribution skewness and kurtosis, the following two-variable nonlinear equation system can be defined:

$$\gamma_X = \gamma_Y(r, \sigma) \qquad (2.63)$$

$$\kappa_X = \kappa_Y(r, \sigma) \qquad (2.64)$$

This system can be solved with the help of numerical methods. In the present work, it was chosen the Newton-Raphson root-finding algorithm. Once r and σ are determined, the remaining θ and b parameters can be directly calculated by

$$\theta = \frac{\sigma_X}{\sigma_Y(r, \sigma)} \qquad (2.65)$$

$$b = \mu_X - \theta\mu_Y(r, \sigma) \qquad (2.66)$$

Low [4] highlights that the relations presented in Equations 2.49 and 2.50 are only valid for positive γ_X skewnesses; however, that is not a limiting factor of the method. In the case of data sets with negative skewnesses ($\gamma_X < 0$), the SGLD parameters are first determined considering the skewness absolute value. Hence a positive skewed SGLD distribution is determined. SGLD's PDF is then rotated about its mean, as Figure 8 shows. Mathematically this means substituting $y = 2\mu_Y - y$ and determining $f_Y(2\mu_Y - y)$ instead of $f_Y(y)$.
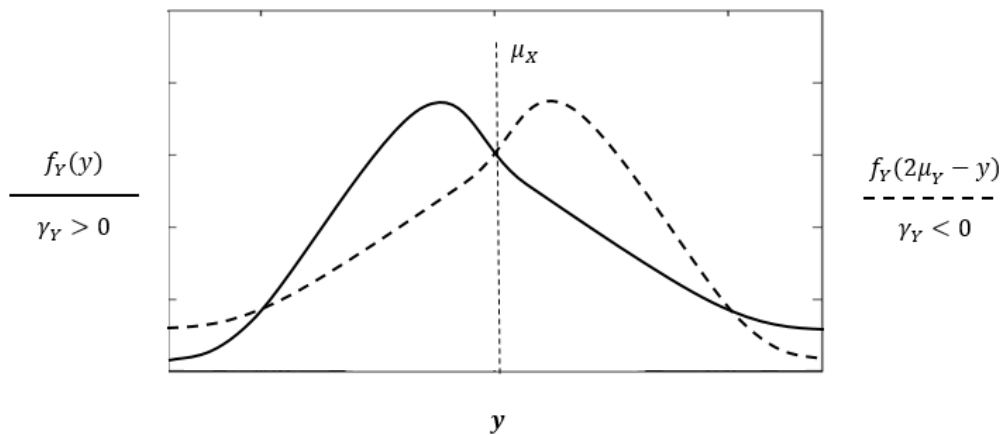
Figure 8 Positive skewed SGLD and its equivalent negative skewed after rotation

The inverse of the SGLD cumulative density function, for a probability $p \neq 1/2$, is given by

$$x = F_x^{-1}(p) =$$

$$b + \theta\exp\left(sgn\left(p - \frac{1}{2}\right)\sigma\left(rg^{-1}\left(\frac{1}{r},\frac{2p-1}{sgn\left(p - \frac{1}{2}\right)}\right)\right)^{1/r}\right) \tag{2.67}$$

where p is accumulated probability, g^{-1} is the Inverse of Incomplete Gamma Function. $g^{-1}(.)$ can be determined from Equation 2.52 as $z = g^{-1}(s,w)$ corresponds to $w = g(s,z)$. For $p \neq 1/2$ the above equation becomes $x = b + \theta$.

As mentioned before, the main contribution for r and σ parameters is to control the distribution kurtosis and skewness, respectively. In practice, SGLD is capable of modeling a vast range of skewness and kurtosis combinations, which gives it a great flexibility of use. SGLD is also equivalent to a few of known distributions as its particular cases, as shown in Table 1.

Table 1 Particular cases of SGLD

Symmetrical PDFs ($\gamma = 0$)	Asymmetrical PDFs ($\gamma \neq 0$)
Normal Distribution ($r = 2$)	Lognormal Distribution ($r = 2, b = 0$)
Laplace Distribution ($r = 1$)	Log-Laplace Distribution ($r = 1, b = 0$)
Uniform Distribution ($r \to \infty$)	Log-Uniform Distribution ($r \to \infty, b = 0$)

2.5 Central Limit Theorem

The central limit theorem states that the limiting distribution of the sum of independent identically distributed random variables of arbitrary distributions is the normal distribution (see, for instance, [6]). Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables with means $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$. Let S_n be the sum of this sequence:

$$S_n = \sum_{i=1}^{n} X_i \tag{2.68}$$

Noting that the X_i are independent, the mean and variance are given by

$$\mu_{Sn} = \sum_{i=1}^{n} \mu_i \tag{2.69}$$

$$\sigma_{Sn}^{2} = \sum_{i=1}^{n} \sigma_i^{2} \tag{2.70}$$

Then the limit as n goes to infinity, the standardized variable of S_n, $Z_n = (S_n - \mu_{Sn})/\sigma_{Sn}$, has the standard normal distribution:

$$\lim_{n \to \infty} f_{Z_n}(z) = \phi(z) \tag{2.71}$$

The conditions under which this occurs are [6]:

1. The individual terms in the sum contribute a negligible amount to the variation in the sum;

2. It is very unlikely that any single term will contribute a disproportionately large amount to the sum.

This means that no single variable (or very few variables) dominates either the total uncertainty or the total sum. All variables in the sum contribute comparably to the sum.

The central limit theorem is illustrated in Figure 9, for the case of the sum of uniformly distributed random variables.
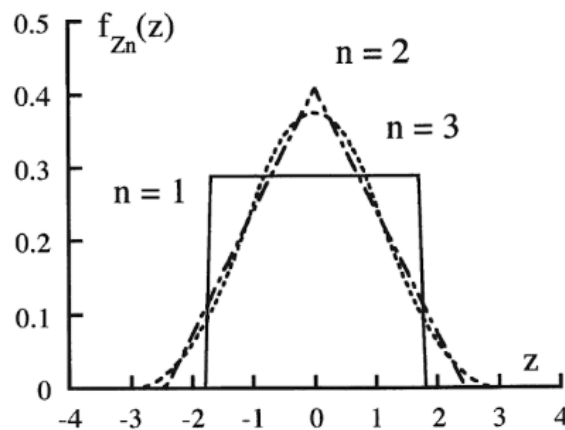
Figure 9 Convergence of the distribution of the sum of uniform random variables to the normal distribution (densities of the standardized sums) [6]

2.6 Distribution of the Sample Mean

Suppose that X is a random variable with mean μ_X, variance $\sigma_X{}^2$ and a known distribution $f_X(x)$. If we carefully conduct an experiment in which we sample this random variable n times, then the ith sample X_i will be independent and identically distributed. The sample mean or sample average $\bar{X}$ is a random variable given by the sum

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.72}$$

If the observed variable X is normally distributed, then $\bar{X}$ will be normally distributed. If X is not normally distributed, then according to the central limit theorem $\bar{X}$ will be approximately normally distributed if the sample size n is large enough [6]. The mean and variance of $\bar{X}$ are:

$$\mu_{\bar{X}} = \frac{1}{n}n\mu_X = \mu_X \tag{2.73}$$

$$\sigma_{\bar{X}}{}^2 = \frac{1}{n^2}n\sigma_X{}^2 = \frac{\sigma_X{}^2}{n} \tag{2.74}$$

The standardized variable Z

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} \tag{2.75}$$

has a standard normal distribution. An important remark can be made to the fact the variability $\sigma_{\bar{X}}$ of the variable average value estimator $\bar{X}$ is inversely proportional to $\sqrt{n}$, i.e., as the sample size n increases, the standard deviation of $\bar{X}$ decreases by a factor of $\sqrt{n}$. Thus, naturally, the larger the sample of X is, the more accurate its estimators will be.

3 STATISTICAL ANALYSIS OF RANDOM PROCESSES

3.1 Random Processes in Time Domain

3.1.1 Random Processes Definitions

A random process can be simply defined as a sequence of random variables $X_1, X_2, \cdots, X_n$. In the case of a random process in the time domain, this sequence is defined in time thus it can be tracked by the index t. Then $X(t)$ will refer to the random process itself and each random variable $x_1, x_2, \cdots, x_n$ will be the outcome of $X(t)$ for $t = 1, t = 2, \cdots, t = n$. Just as the random process $X(t)$ can be discrete or continuous, the index t may also be discrete or continuous. Because $X(t)$ is here defined in the time domain, an outcome sequence $x_1, x_2, \cdots, x_n$ can be named as a time-series. A set of $X(t)$ finite outcomes $x_1, x_2, \cdots, x_n$, or any sub-set of interest, is named a realization of the random process $X(t)$. In Figure 10 different realizations of a generic random process $X(t)$ with a given length T are shown.
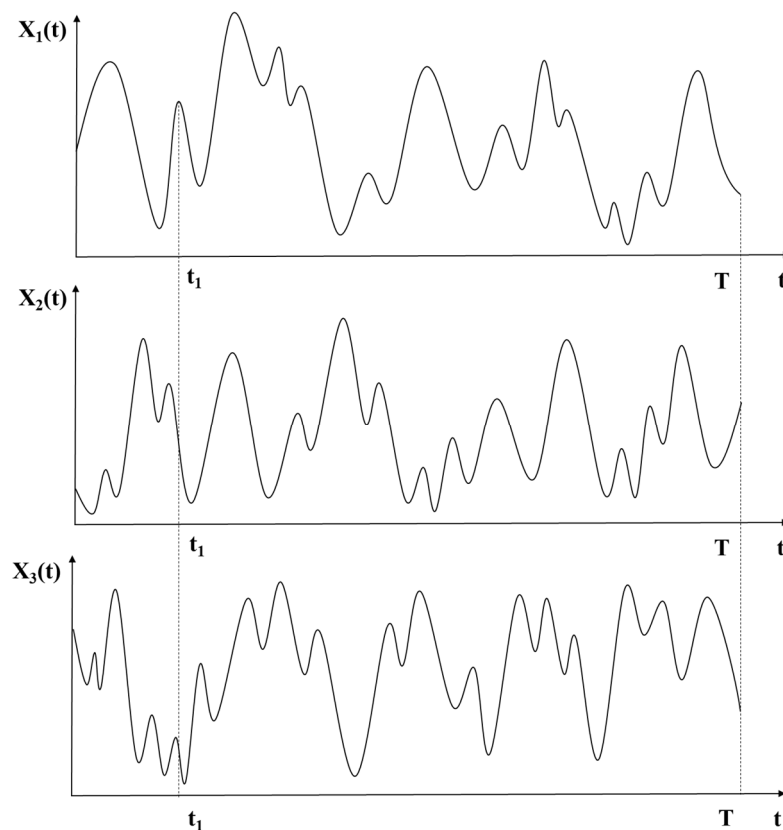
Figure 10 Generic random process $X(t)$ realizations in time domain

There are infinite examples of random processes in offshore engineering. Some of them are the sea water surface elevation at a given location, the wind velocity at a given height above the sea surface and the motions of a floating unit submitted to wave, wind and current environmental loads.

The statistics of the random process of interest $X(t)$ can be calculated across different realizations at a specific given time $t = t_1$. Taken that $X(t)$ is here defined by the discrete index t, its mean and variance at a specific instant can be estimated by

$$\mu_1 = E[X(t_1)] = \lim_{n \to \infty} \sum_{i=1}^{n} \frac{x_i(t_1)}{n} \tag{3.1}$$

$$E[X(t_1)^2] = \lim_{n \to \infty} \sum_{i=1}^{n} \frac{x_i(t_1)^2}{n} \tag{3.2}$$

$$Var(X_1) = {\sigma_1}^2 = E[X(t_1)^2] - {\mu_1}^2 \tag{3.3}$$

Higher order moments of $X(t)$ can be similarly calculated. When not specified otherwise, this work will always refer to a random process with zero mean.

The joint distribution of $X(t)$ at all times t is the ultimate product of the study of the random process. Any number of times may be considered for a joint distribution, for example, $X(t_1)$, $X(t_2)$, $X(t_3)$,..., $X(t_n)$, but the most practical case is to consider $X(t)$ at two times, t_1 and t_2. As seen in Section 2.3, the correlation between two variables is one important parameter for characterizing their joint distribution. Likewise, a very important joint measure in a random process is the correlation of the process with itself at two different times, $X(t_1)$ and (t_2) [6]. Denoted $R_X(t_1, t_2)$, this measure of correlation is called the **autocorrelation function** and is given by

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \sum_{t_1}^{n} x(t_1)x(t_2) \tag{3.4}$$

Related to the autocorrelation function, the autocovariance function $COV(t_1, t_2)$ is defined as

$$COV(t_1, t_2) = E[(X(t_1) - \mu_1)(X(t_2) - \mu_2)] = R_X(t_1, t_2) - \mu_1\mu_2 \tag{3.5}$$

From the autocovariance function, the autocorrelation coefficient function $\rho_{XX}(t_1, t_2)$ can be defined as

$$\rho_{XX}(t_1, t_2) = \frac{COV(t_1, t_2)}{\sigma_1 \sigma_2} \tag{3.6}$$

If the statistical parameters of the random process $X(t)$ are constant over time, i.e., if above presented statistical parameters are the same, independently of the time instant $t_1, t_2, \cdots, t_n$, then the random process is called a **stationary process**.

In other words, the random process will be stationary when the expected value and other higher order statistical parameters for $t = t_1$ is the same as any other time instant $t_2, t_3, \ldots, t_n$. In this case, the autocorrelation function will only depend on the time lag τ, or the difference between two given time instants, for example $t_2 - t_1 = \tau$. Then $R_X(t_1, t_2)$ will simply be $R_X(\tau)$. Similarly, other statistical parameters can be expressed by

$$E[X(t_1)] = E[X(t_2)] = \cdots = E[X(t_{\mathrm{n}})] = \mu_X \tag{3.7}$$

$$\sigma[X(t_1)] = \sigma[X(t_2)] = \cdots = \sigma[X(t_{\mathrm{n}})] = \sigma_X \tag{3.8}$$

$$\gamma[X(t_1)] = \gamma[X(t_2)] = \cdots = \gamma[X(t_{\mathrm{n}})] = \gamma_X \tag{3.9}$$

$$\kappa[X(t_1)] = \kappa[X(t_2)] = \cdots = \kappa[X(t_{\mathrm{n}})] = \kappa_X \tag{3.10}$$

In practice, it is usual to consider environmental random processes as stationary processes for a short-term period.

A particular case of a stationary process is the process called ergodic. An **ergodic random process** is the process in which the statistical parameters calculated over time for a single realization are the same as the statistical parameters calculated across different realizations at a specific time instant. This process can hence be characterized from a single sample realization of the random process (see, for example, [7]).

Let $X(t, k)$ be the kth sample realization of the stationary random process $X(t)$. The temporal average of $X(t, k)$ is defined as

$$\langle X(t,k) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T X(t,k) dt \tag{3.11}$$

The symbol $\langle \cdot \rangle$ indicates a temporal average. A process is ergodic in the mean if

$$\langle X(t,k) \rangle = E[X(t)] = \mu_X \tag{3.12}$$

Similarly, a process is ergodic in the variance if

$$\langle [X(t,k) - \mu_X]^2 \rangle = E[[X(t) - \mu_X]^2] = \sigma_X{}^2 \tag{3.13}$$

And a process is ergodic in the autocorrelation function if

$$\langle X(t,k)X(t+\tau,k) \rangle = E[X(t)X(t+\tau)] = R_X(\tau) \tag{3.14}$$

For a process to be ergodic, it must be stationary. However, the converse is not true; a stationary process is not necessarily ergodic.

In offshore engineering, in general, it is practically impossible to obtain the sufficient amount of data in order to study its statistical properties through different realizations. However, short-term data can be considered as approximately stationary and ergodic. One example of this is the usual statistical treatment of ocean waves: for a short period of 3-h, the sea surface elevation can be modelled as a stationary and ergodic random process. For this short-term period of $T_{st} = 3$-h, this condition is usually referred as a sea state. Then, in this dissertation it is assumed that the random processes of interest are characterized as stationary and ergodic processes.

3.1.2 Characteristics of Autocorrelation Function

With Figure 10 in mind, the autocorrelation function $R_X(\tau)$ allows to statistically predict the random variable behavior in the time instant $t + \tau$ if its value is known in the time instant t. This function is defined as the expected value of the product $X(t)X(t + \tau)$, which is equivalent to writing

$$R_X(\tau) = E[X(t)X(t+\tau)] = \lim_{T \to \infty} \frac{1}{T} \int_0^T X(t)X(t+\tau) dt \tag{3.15}$$

If the process mean is null, $\mu_X = 0$, when $\tau = 0$ the above equation becomes equivalent to Equation 3.3. Then the following equality is valid

$$R_X(0) = \sigma_X{}^2 \tag{3.16}$$

If the process mean is not null, $\mu_X \neq 0$, its mean can be subtracted from the original process for the above calculations and later added back.

As τ increases, it becomes harder to predict the process behavior and therefore the autocorrelation tends to become lower. Figure 11 shows a typical autocorrelation function.
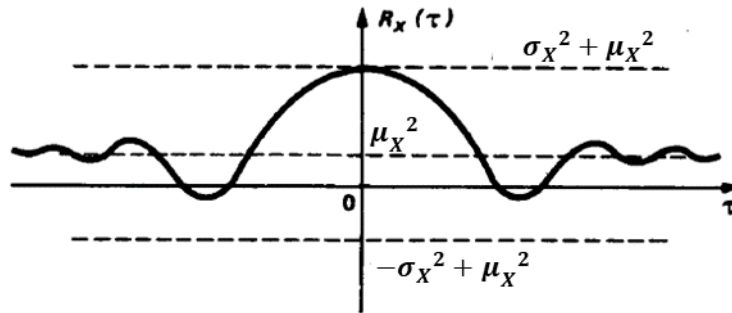
Figure 11 Typical autocorrelation function $R_X(\tau)$ [7]

3.2 Random Processes in Frequency Domain

3.2.1 Spectral Density Function

The spectral density function or simply the spectrum of a random process $X(t)$ is defined as the Fourier transform of the autocorrelation function of the process:

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau) e^{-i\omega\tau} d\tau \tag{3.17}$$

where ω denotes angular frequency, in rad/s.

The complementary function is given by

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega) e^{i\omega\tau} d\omega \tag{3.18}$$

Equations 3.17 and 3.18 are often called the Wiener-Khinchine relations. In order for the transform pair to exist, the autocorrelation function must be absolutely integrable and the mean of the random process must be equal to zero. The Fourier transform and the spectral density function of a random process is illustrated in Figure 12.

From Equation 3.18 it is possible to deduce that, when $\tau = 0$ and $\mu_X = 0$, the process variance is equal to the area under the spectral density function curve, as shows Equation 3.19.

$$\sigma_X{}^2 = \int_0^\infty S_X(\omega)d\omega \qquad (3.19)$$
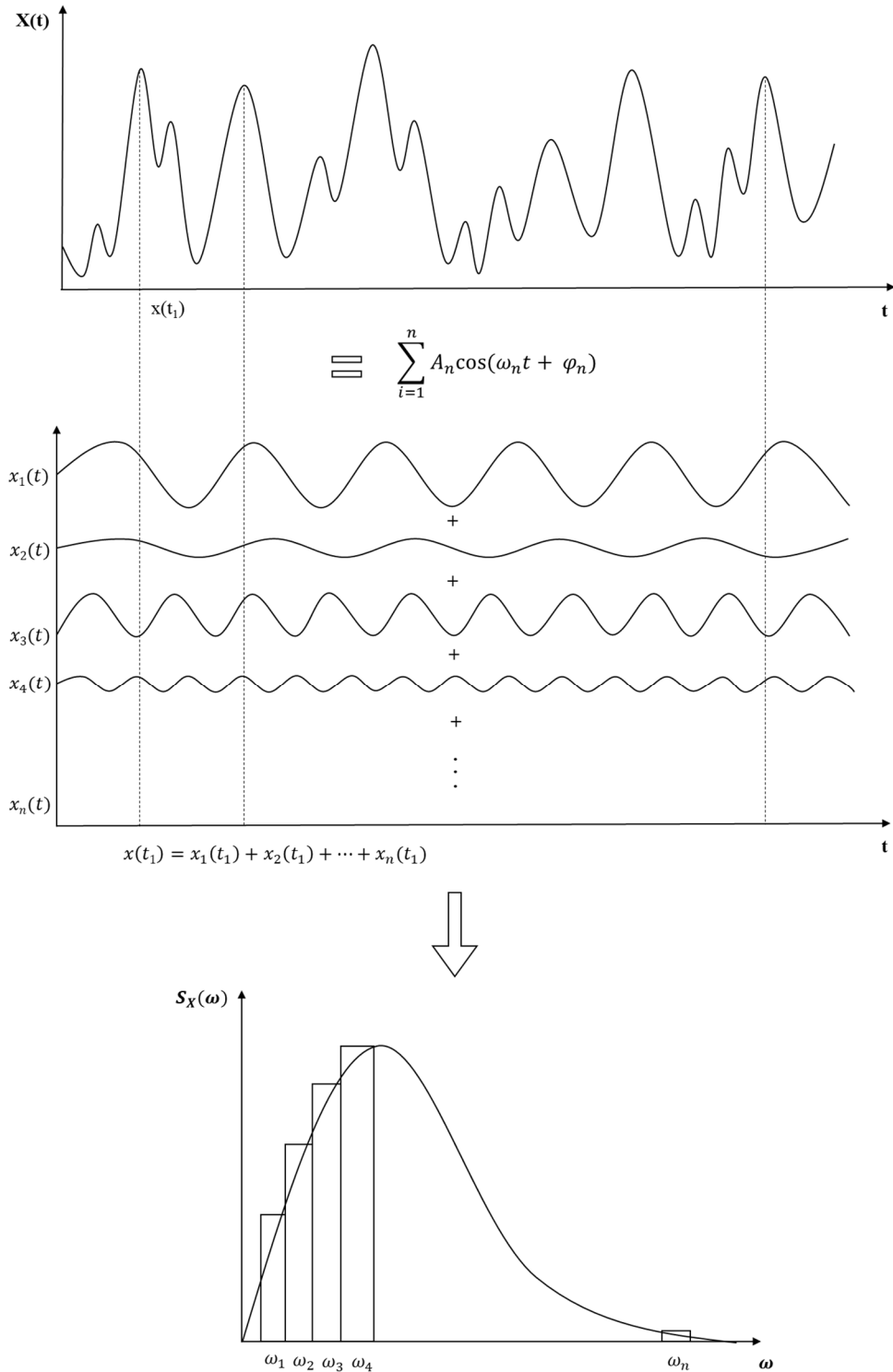
Figure 12 Illustration of the Fourier transform and the spectral density function of a random process

3.2.2 Level Crossings

Given an arbitrary time interval T in a random process realization (time-series) $X(t)$, one can compute the number of times there is a crossing of an arbitrary level $x = a$ and call this number $N_a(t, t + T)$. This number of level a crossing is naturally also a random variable given that each realization of the same random process will present a different counting for N_a. This is valid for any level a of interest, including the zero mark. Figure 13 illustrates levels crossings of interest in a generic random process realization.
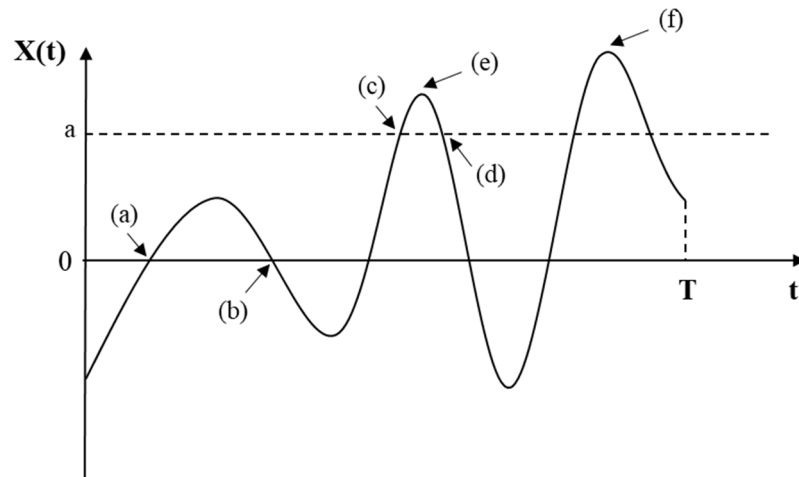
Figure 13 Generic random process $X(t)$ realization of length T: (a,b) level $x = 0$ up-crossing (positive velocity) and down-crossing (negative velocity), respectively; (c,d) level $x = a$ up-crossing and down-crossing, respectively; (e,f) positive peaks above a level. Adapted from [6]

From Figure 13 one can observe that there are two obvious correlations between level crossings. First, the crossings usually occurs in pairs, one up crossing followed by one down crossing. Second, there is an apparent correlation between level crossings when the signal resembles a sine wave, or when the process is narrow banded, concept to be introduced in the next chapter. In other words, if $X(t)$ up-crosses the level $x = a$ in one cycle, it is highly likely that it will up-cross it again in the next cycle: level up crossings tend to occur in clusters or clumps. The latter observation will be of great relevance in the present study.

Since up and down-crossings may be correlated, and since the scope of this study is related to extreme values, it is natural then to focus only on the up crossings, with positive slope. The expected level up-crossing rate ν_{a^+} in a time-series can be defined as

$$\nu_{a^+} = \frac{N_{a^+}}{T} \qquad (3.20)$$

where N_{a^+} is the number of up-crossings of the level $x = a$ in the time interval of interest T.

For any generic stationary random process, the expected level up-crossing rate of the level $x = a$ can also be estimated by [6]

$$\nu_{a^+} = \int_0^\infty \upsilon f_{X\dot{X}}(a,\upsilon)d\upsilon \qquad (3.21)$$

where $f_{X\dot{X}}$ is the joint probability density function of $X(t)$ and its first derivative $\dot{X}(t)$. It is clear that Equation (3.20) is more straightforward since $f_{X\dot{X}}$ is difficult to be determined. However for some very specific scenarios, $f_{X\dot{X}}(x,\dot{x})$ has a closed-form solution, as it will be presented later in this chapter.

When the level of crossing of interest is the null level, i.e., a = 0, the level up-crossing rate ν_{a^+} is simply referred as the zero up-crossing rate ν_{0^+}.

With the concepts presented above, the peaks in a given time-series can now be defined. Given a process $X(t)$ realization, for each mean up-crossing followed by a mean down-crossing there is a maximum value, identified as a maximum peak. Likewise, for each mean down-crossing followed by a mean up-crossing there is a minimum value, identified as a minimum peak. This is illustrated in Figure 14.
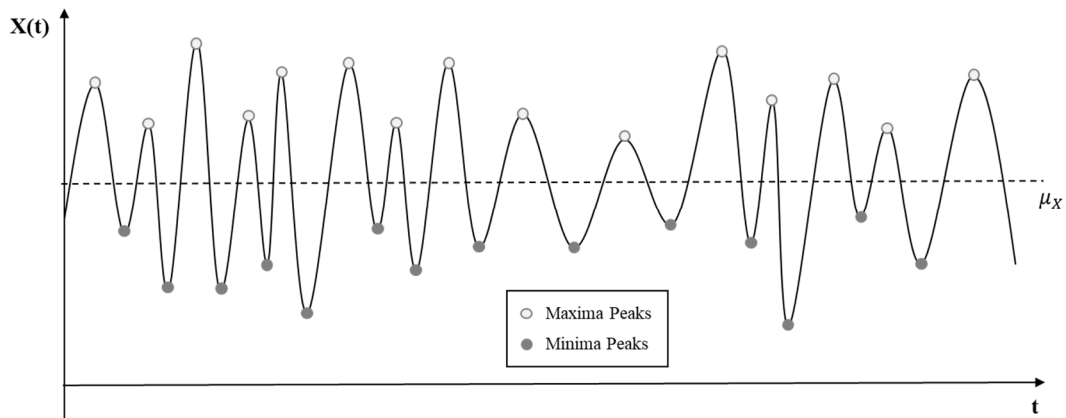
Figure 14 Maxima and minima peaks definition

The present work will focus on the maxima peaks of random processes.

3.2.3 Bandwidth Factor

The bandwidth factor is a measure that allows to identify how the energy contained in the spectrum $S_X(\omega)$ is distributed among the frequencies ω. This factor is represented by ε and is calculated by:

$$\varepsilon = \sqrt{1 - \frac{m_2^2}{m_0 m_4}} \tag{3.22}$$

where the spectral moment of order n, m_n, is given by

$$m_n = \int_0^\infty \omega^n S_X(\omega) d\omega \tag{3.23}$$

Depending on the ε value, a random process can be classified as a narrow band process or as a broad (or wide) band process.

In a **narrow band process**, most of the spectral density is concentrated in a narrow range of frequencies, hence ε assumes values closer to zero ($\varepsilon \to 0$). In this situation, each positive up-crossing in the signal, or time-series, corresponds to only one maxima (or peak), as Figure 15 shows.
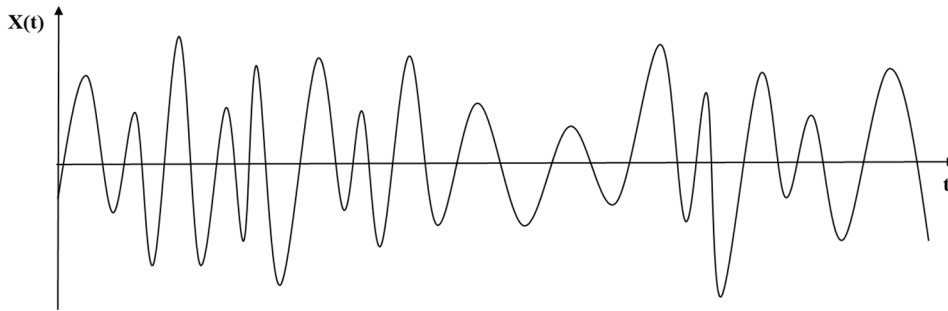
Figure 15 Generic narrow banded random process $X(t)$

On the other hand, in a **broad band process** the spectral density is distributed among a broad range of frequencies and ε assumes values greater than zero ($\varepsilon > 0$). A broad banded process is characterized by presenting more than one local maxima between one up-crossing and one down-crossing at the process mean level, as shows Figure 16.
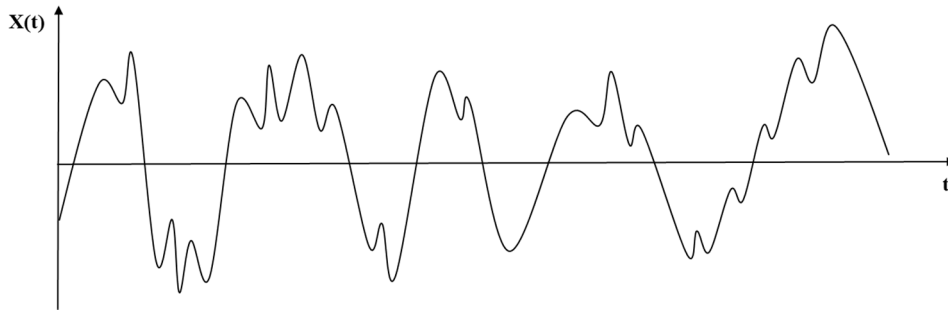
Figure 16 Generic broad banded random process $X(t)$

Then, in a broad banded process, one can define two categories of maxima, or peaks, in a time-series: the local peaks and the global peaks, as shown in Figure 17 for a zero reference level. Local peaks are all maxima points in the random process. Global peaks is the set of the largest peaks between two mean level up-crossings of the random process.
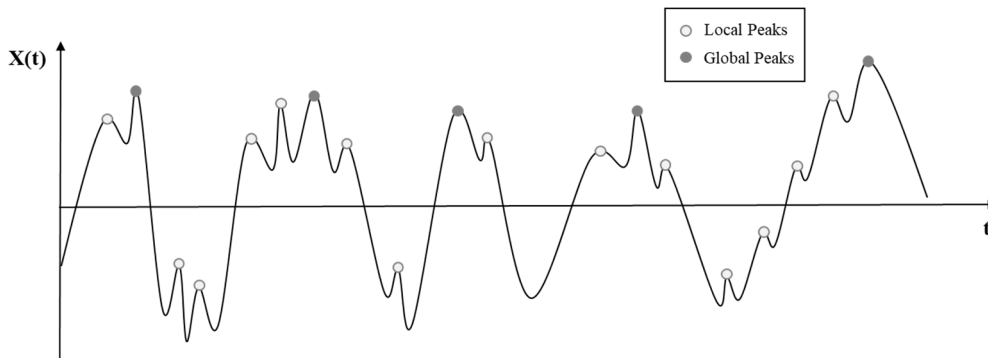
Figure 17 Definition of local and global peaks in a generic broad banded random process $X(t)$

Typical spectral density functions for narrow banded and broad banded processes are presented in Figure 18.
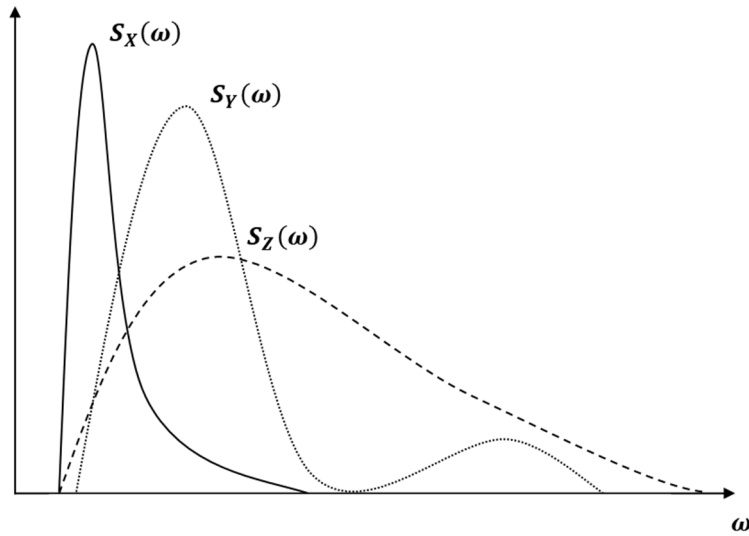
Figure 18 Typical spectral density function of a narrow banded process $X(t)$ and two broad banded process $Y(t)$ and $Z(t)$

3.3 Probability Distributions of Random Processes

3.3.1 Generic Random Processes

As mentioned before, in a stationary random process, statistical parameters are constant, i.e., they are independent of time. It is natural then to associate the process to a probability distribution, with the same statistical characteristics.

Using Equation (2.2), the probability of a process $X(t)$ to be contained between two given values x_a and x_b is given by

$$P(x_a \leq X(t) \leq x_b) = \int_{x_a}^{x_b} f_X(X(t))dx \tag{3.24}$$

For a continuous process $X(t)$, the above integral can be understood as the ratio between the sum of all time intervals Δt_i in which the process assumes values in the interval $[x_a, x_b]$ in the total observed time T, as illustrated in Figure 19 and presented in Equation (3.25).
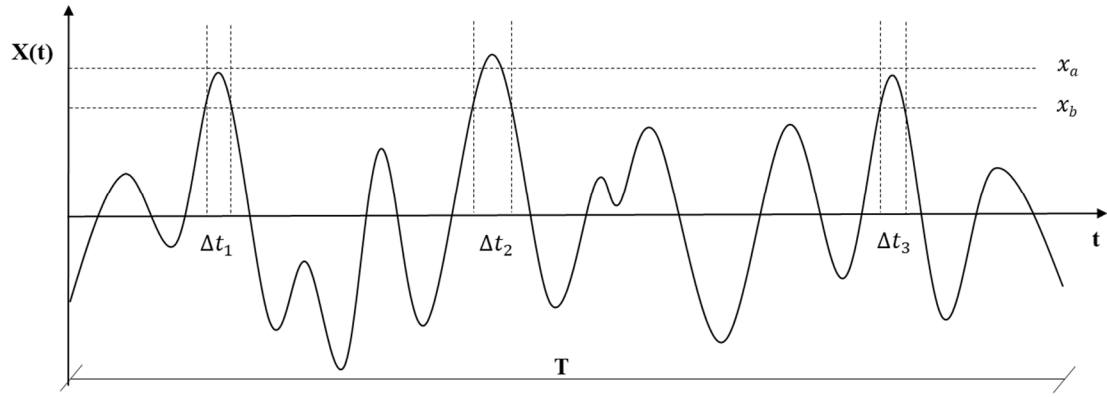
Figure 19 Time intervals Δt_i in which the continuous process assumes values in the level interval $[x_a, x_b]$

$$P(x_a \leq X(t) \leq x_b) = \int_{x_a}^{x_b} f_X(X(t))dx = \frac{1}{T}\sum_{i=1}^{N} \Delta t_i \qquad (3.25)$$

In practice, $X(t)$ is usually a discrete process having N points spaced by a constant Δt with a total length of $T = N\Delta t$, as illustrated in Figure 20.
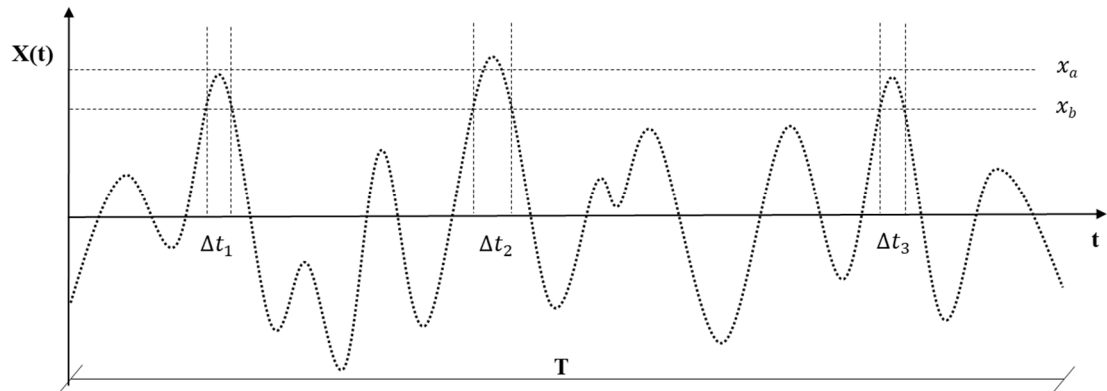
Figure 20 Time intervals Δt_i in which the discrete process assumes values in the level interval $[x_a, x_b]$

Since $X(t)$ has a finite number of defined points, the probability expressed in Equation (3.25) will simply be given by the ratio between the number of points that fall in the interval of interest $\Delta x = [x_a, x_b]$ and the total number of points N in the period of observation T, as

$$P(x_a \leq X(t) \leq x_b) = f_X(x_i) \cdot \Delta x = \frac{1}{T}\sum_{i=1}^{N} \Delta t_i = \frac{n_i}{N} \qquad (3.26)$$

35

with $x_i = (x_a + x_b)/2$. The probability distribution of the process will hence be given by

$$f_X(x_i) = \frac{n_i}{N\Delta x} \tag{3.27}$$

Apart from the probability distribution of the process as a whole, another important distribution associated with the process is the distribution of its peaks (since the peaks of a random process can be considered as a random process itself). As shown in [6], the general formula for determining the distribution of peaks of a random process is:

$$f_{X_p}(x_p) = \frac{\int_{-\infty}^{0} -w f_{X\dot{X}\ddot{X}}(x_p, 0, w) dw}{\nu_p} \tag{3.28}$$

where x_p represent a peak of the random process $X(t)$, $f_{X\dot{X}\ddot{X}}(x, \dot{x}, \ddot{x})$ is the joint probability density function of $X(t)$, its first derivative $\dot{X}(t)$ and its second derivative $\ddot{X}(t)$ and ν_p is the peak frequency, given by

$$\nu_p = \int_{-\infty}^{0} -w f_{\dot{X}\ddot{X}}(0, w) dw \tag{3.29}$$

where $f_{\dot{X}\ddot{X}}(x, \dot{x})$ is the joint probability density function of the first derivative of the process $\dot{X}(t)$ and its second derivative $\ddot{X}(t)$. Equation (3.29) is the result of the same level-crossing analysis made in Section 3.2.2, but this time applied on the 'velocity' process $\dot{X}(t)$. A zero down-crossing in $\dot{X}(t)$, i.e., a change in velocity from positive to negative, corresponds to the occurrence of a peak in $X(t)$.

It is clear that $f_{X\dot{X}\ddot{X}}(x, \dot{x}, \ddot{x})$ theoretical estimation seems to be a non-easy work. However, Equation (3.29) has a closed-form solution for very specific cases, as it will be presented in the next chapter.

3.3.2 Gaussian Processes

A process can be named as a Gaussian process when its probability density function is the normal (or Gaussian) distribution. In this case, the process probability distribution is defined as

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2\right] = \frac{1}{\sqrt{m_0}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x}{\sqrt{m_0}}\right)^2\right] \qquad (3.30)$$

where m_0 is the variance of the random process, or the area under the spectral density function of the process, as shown in Equation (3.23).

In a Gaussian process, Equation (3.29) has a closed-form and the distribution of peaks of the process is defined as [6]

$$\begin{aligned} f_{X_p}(x_p) &= \frac{\varepsilon}{\sqrt{m_0}\sqrt{2\pi}} exp\left[-\frac{1}{2}\frac{x_p{}^2}{m_0\varepsilon^2}\right] \\ &\quad + \frac{p}{m_0}\sqrt{1-\varepsilon^2}\, exp\left[-\frac{1}{2}\frac{x_p{}^2}{m_0}\right]\Phi\left[\frac{x_p}{\sqrt{m_0}\varepsilon}\sqrt{1-\varepsilon^2}\right] \end{aligned} \qquad (3.31)$$

where ε is the bandwidth factor defined in Equation (3.22), m_0 is the spectral moment of order zero defined in Equation (3.23) and $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. Equation (3.31) is known as the Rice distribution [11].

In the limiting case of a narrow band process ($\varepsilon \to 0$), the first term of Equation (3.31) vanishes and the expression becomes the distribution known as the Rayleigh distribution (see section 2.4.4).

The peak frequency ν_p for a gaussian process is given by

$$\nu_p = \frac{1}{2\pi}\sqrt{\frac{m_4}{m_2}} \qquad (3.32)$$

It is important to notice that, in the case of Gaussian processes, the spectral density function is sufficient to statistically describe the whole process.

Processes that are not Gaussian are simply called non-Gaussian. For these processes, as mentioned before, there is no theoretical solution for the distribution of peaks. In Chapter 4 a few methodologies for representing the distribution of peaks of non-Gaussian processes are presented.

3.4 Statistical Parameters Estimation Based on a Single Realization

As presented earlier in this chapter, if a random process is ergodic then its statistical parameters can be estimated based on just a single realization. However, one must be cautious: the accuracy of these estimates depends on the finite length of the time series. Naturally, the larger the sample of points is, the more accurate statistical parameters will be.

In a discrete random process realization, the number of points N is a finite amount given by $N = T/\Delta t$, where T is the total time of the realization and Δt is the constant sampling space between points. Hence, given a random process sample $X = \{x_1, x_2, x_3, \dots, x_N\}$, the estimators for the mean, variance, skewness and kurtosis of $X(t)$ are:

$$m_X = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3.33}$$

$$s_x{}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m_X)^2 \tag{3.34}$$

$$g_x = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - m_X)^3}{s_x{}^3} \tag{3.35}$$

$$\kappa_x = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - m_X)^4}{s_x{}^4} \tag{3.36}$$

With the support of above equations, the notion of improved stability of statistical parameters estimators when T, and consequently N, is increased becomes clearer. Figure 21 illustrates this for the case of the mean estimator. From this figure, it is evident that the mean estimator only begins to converge to a stable value when the duration T is sufficiently large. In other words, the time-series has to be sufficiently long in order to correctly estimate the statistical parameters of the random process being sampled.
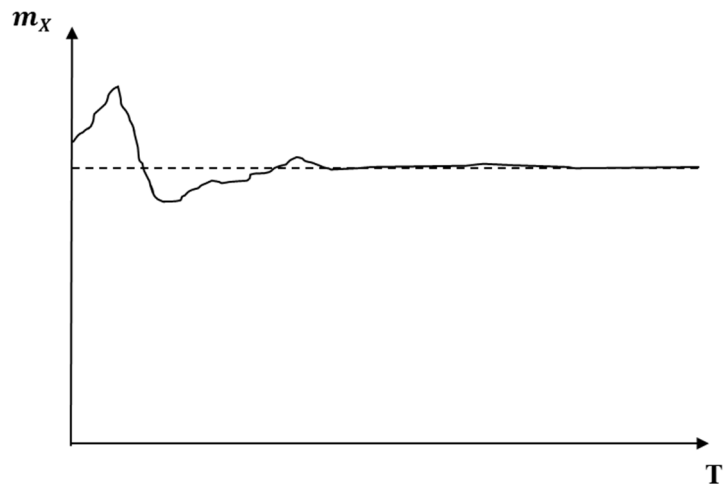
Figure 21 Mean estimator for the random process $X(t)$ based on discrete realizations of duration T

4 STATISTICS OF EXTREMES

It is of particular interest of everyday engineering design to estimate the maximum response possible that can exist in a given structure due to the set of load conditions. This applies to any structure being designed, since bridges to airplanes. Naturally, this is of extreme importance, the design has to ensure that the loading is being resisted, according to given standards, for the whole design life of the structure. At the same time, engineering design can also be interested in estimating the minimum possible resistance of the structure, given that manufacturing and construction introduce uncertainties that have to be accounted for.

In the case of offshore structures engineering, considering environmental conditions as random variables is a recurrent assumption. Therefore, structures responses can also be understood as random. Since maxima, or the time-series peaks, are different for each realization of a random process, extreme peaks themselves can be treated as random processes. In this chapter, Extreme Value Statistics, i.e., statistics aimed at estimating maximum or minimum values of a random variable, will be presented. As the scope of this study is in the maximum values estimation, a greater emphasis will be given for the latter.

4.1 Classical Extreme Value Theory

As any random process, in order to one estimate its statistical parameters and probability distribution, one has to have a sample of the process being studied. In the case of the extremes random process this is no simple task. Imagine that the random process of interest is the maximum wind velocity V_w for a 10 years time span at a given location, i.e., the 10 years return-period wind velocity. In order to estimate the extreme wind velocity probability distribution at this location, one needs a sample of N extreme values of 10 years, i.e., a sample of extremes $v_e = \{v_1, v_2, v_3, ..., v_N\}$ where v_1 is the maximum observed wind velocity in the first 10 years of observation, v_2 is the maximum observed wind velocity in the next 10 years of observation and so on. Considering a minimum of $N = 30$ extreme values in the sample, for minimum accuracy of the estimated statistical parameters, a 300 years observation period would be necessary. It is needless to say that this is, for the vast majority of cases, impossible to achieve, especially when the return-

periods of interest are even larger, such as 100 or 1000 years. However, it is still the ideal procedure for determining a variable's characteristic extreme value, especially because of its validity independent of the process stationarity or original distribution. In the context of environmental random processes, this method also enables taking account of possible climate changes in the long-term.

On the other hands, depending on the investigated random process, some simplifications can be made. In order to overcome the issue described above, a theory named Classical Extreme Value Theory (or Order Statistics) was developed [12]. In this theory, in order to estimate the extreme values probability distribution of a random process, it is only necessary to previously know the initial process probability distribution and establish a few assumptions, as it will be explained in this chapter.

Let $X = \{X_1, X_2, X_3, \dots, X_N\}$ be a sample of size N of a generic random variable $X(t)$, where X_i represents the statistical properties of the ith realization of $X(t)$. Given that each X_i represents a realization of the same process $X(t)$, all probability distributions are the same as $X(t)$ probability distribution:

$$F_{X_1}(x) = F_{X_2}(x) = F_{X_3}(x) = \cdots = F_{X_N}(x) = F_X(x) \tag{4.1}$$

Let now $Y_N = max\{X_1, X_2, X_3, \dots, X_N\}$ be the maximum extreme value of X in a sample of size N. Assuming that a specific value of $X = y$ belongs to the population of extreme values of X, the following condition can be set:

$$P(Y_N \leq y) = P(X_1 \leq y \cap X_2 \leq y \cap X_3 \leq y \cap \dots \cap X_N \leq y) \tag{4.2}$$

Assuming that the observed values of $X(t)$ are statistically independent, above equation can also be written as:

$$P(Y_N \leq y) = P(X_1 \leq y) \cdot P(X_2 \leq y) \cdot P(X_3 \leq y) \cdots P(X_N \leq y) \tag{4.3}$$

Considering that $P(X \leq x) = F_X(x)$ and that realizations are identically distributed, the maximum extreme value of $X(t)$ cumulative density function is given by

$$F_{Y_N}(y) = [F_X(y)]^N \tag{4.4}$$

and therefore, its probability density function is given by

$$f_{Y_N}(y) = \frac{dF_{Y_N}(y)}{dy} = N \cdot [F_X(y)]^{N-1} \cdot f_X(y) \tag{4.5}$$

where $F_X(.)$ and $f_X(.)$ are the CDF and PDF of the original variable $X(t)$. The probability distribution of the original variable is frequently identified as the marginal distribution.

Thus it has been demonstrated that, if the probability distribution of X is known, one can determine its maximum extreme values probability distribution under the hypothesis of statistical independence between its realizations. Equation (4.5) is illustrated in Figure 22.
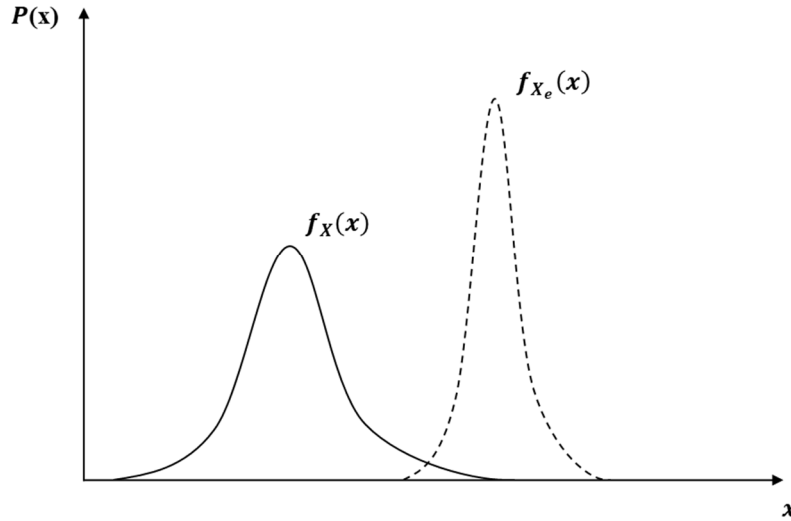
Figure 22 Original probability density function of the variable X, $f_X(x)$, and the probability density function of maximum extreme values of X, $f_{X_e}(x)$

In the particular case of an ergodic process, a single realization is sufficient to determine the statistics of extremes [13][6][3][2]. If the sample analyzed corresponds to the process peaks, i.e., $X_p = \{x_p^1, x_p^2, x_p^3, \dots, x_p^{Np}\}$, then $F_{X_e}(x)$ denotes the extreme probability distribution of the peaks. Following the same methodology presented above, one can conclude that

$$F_{X_e}(x) = \left[F_{X_p}(x)\right]^{N_p} \tag{4.6}$$

42

where $F_{X_p}(x)$ is the CDF that represents the process peaks and N_p is the expected number of peaks in the observation period T.

The characteristic extreme value of a random process is usually taken as the most probable value in the extreme probability distribution of the process, i.e., the value where the latter distribution reaches its maximum. This value will be referenced from now on as the MPV (Most Probable Value) and is illustrated in Figure 23.
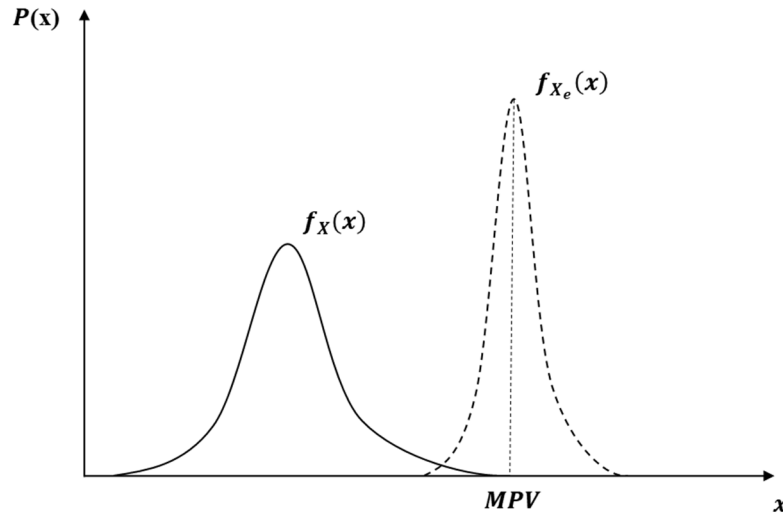
Figure 23 Most Probable Value in the extreme distribution of X, $f_{X_e}(x)$

4.2 Correlated Random Variables

The usual procedure adopted in an extreme analysis is to assume that individual maxima in a random process are statistically independent. In a narrow-band time-series, these individual maxima tend to occur in groups, or 'clumps', thus they are characterized by a strong correlation between peaks of the same 'clump'. It is known that correlation between neighboring peaks in a process will influence the prediction of the largest maximum of the random process for a given time interval [14]. In general, and especially for narrow banded random processes, accounting for these groupings of maxima leads to extreme values estimates that are lower than those predicted by the standard order statistics [14].

If variables are assumed correlated, equations presented in the previous chapter are not valid. Let $X_p = \{x_p^1, x_p^2, x_p^3, \dots, x_p^{Np}\}$ be a sample of size N_p of the peaks of a generic random variable $X(t)$ realization where N_p is the expected number of peaks in

the observation period T. Assuming that only two subsequent peaks in the sample are correlated, i.e., assuming that peaks present a significant correlation two by two, Equation (4.6) now becomes

$$F_{X_e}(x) = \frac{\left(F_{X_p,X_p}(x,x)\right)^{N_p-1}}{\left(F_{X_p}(x)\right)^{N_p-2}} \tag{4.7}$$

where $F_{X_p,X_p}(x,x)$ is the joint cumulative probability distribution for two consecutive peaks of $X(t)$ and $F_{X_p}(x)$ is the peaks marginal CDF. Equation (4.7) is a consequence of applying the one-step Markov chain condition [14] on Equation (4.2). It is clear that the challenge here is to determine $F_{X_p,X_p}(x,x)$. This can be achieved through the use of the Nataf transformation [8], as shown in Section 2.3, which requires the peaks probability distribution $F_{X_p}(x)$ and the empirical correlation coefficient between two consecutive peaks. The latter requisite can be straightforwardly obtained from the peaks sample as

$$\rho = \left[\frac{\sum_{i=1}^{N_p-1} x_i^p x_{i+1}^p}{N_p - 1} - \mu_{Xp}^2\right]\frac{1}{\sigma_{Xp}^2} \tag{4.8}$$

where N_p is the total number of peaks in the sampled time-series, μ_{Xp} is the peaks sample average value and σ_{Xp} is the peaks sample standard deviation. Equation (4.8) is the discrete form of Equation (2.26).

According to the Nataf transformation model previously shown, the joint probability density function of two consecutive peaks is given by

$$f_{X_p,X_p}(x,x) = \frac{f_{X_p}(x)^2}{\phi(z)^2}\phi_2(z,z,\rho) \tag{4.9}$$

where $z = \Phi^{-1}\left(F_{X_p}(x)\right)$, $\Phi^{-1}(.)$ is the inverse of cumulative probability distribution of a standard normal variable, $f_{X_p}(x)$ is the peaks probability density function, $\phi(.)$ is the probability density function of a standard normal variable and $\phi_2(.)$ is the joint probability function of two standard correlated normal variables with a correlation coefficient ρ. The joint cumulative distribution $F_{X_p,X_p}(x,x)$ can be obtained by

$$F_{X_p, X_p}(x, x) = \int_{-\infty}^{x} \int_{-\infty}^{x} f_{X_p, X_p}(y, w) dy dw \qquad (4.10)$$

As mentioned before, the above double integral can be solved in a fast way by proper use of numerical transformations [9] as

$$F_{X_p, X_p}(x, x) = 1 - 2\Phi(z) + \int_{-\infty}^{\rho} \phi_2(-z, -z, y) dy \qquad (4.11)$$

with z as defined before and $\Phi(.)$ standing for the cumulative probability distribution of a standard normal variable.

The methodology here presented considers only correlation between two consecutive peaks. However, the same procedure, using a higher level of the Markov chain and employing the Nataf transformation model, can be applied for considering a larger number of consecutive peaks correlated. Since the mathematics involved for considering more than two consecutive peaks correlated are significantly more complex, this procedure will not be presented nor investigated here. Correlation between three consecutive peaks will, however, be investigated through a different methodology presented later.

4.3 Asymptotic Distributions

In Equations (4.6) and (4.7), the parameter N represents the number of observations of $X(t)$ or the number of peaks N_p in a single realization of a time length of interest. It is well established in literature that the theoretical extremes distributions defined in the previous chapter converge to asymptotic functions for large values of N, i.e., when $N \to \infty$. The asymptotic functions to which the extremes distributions converge are known as [6]:

- Type I distribution (or Gumbel distribution);

- Type II distribution (or Fréchet distribution);

- Type III distribution (or Weibull distribution)

This study will focus only on the Type I and Type II distributions, since Type III distribution is connected to truncated samples of data

45

4.3.1 Type I or Gumbel Distribution

Variables that present an exponential decay converge to the asymptotic Type I or Gumbel distribution. A random variable X presents exponential decay if

$$\frac{f_X(x)}{1 - F_X(x)} = -\frac{f_X{}'(x)}{f_X(x)} \tag{4.12}$$

Among the most known distributions, there are a few that meet the above requisite: the Normal and Weibull distributions and the Gumbel distribution itself when applied as a marginal probability distribution. These distributions will asymptotically converge, as N increases, to the Gumbel distribution of extremes given by

$$f_X(x) = \alpha \cdot exp\big[-\alpha(x - u) - exp[-\alpha(x - u)]\big] \tag{4.13}$$

$$F_X(x) = exp\big[-exp[-\alpha(x - u)]\big] \tag{4.14}$$

where u is the location parameter and α is the scale parameter. The mean and standard deviation are related to these parameters by

$$\mu_X = u + \frac{0.57722}{\alpha} \tag{4.15}$$

$$\sigma_X = \frac{\pi}{\alpha\sqrt{6}} \tag{4.16}$$

The location parameter u is also the mode of the distribution, or the most probable extreme value of X. This parameter is related to N by [12]

$$F_X(u) = 1 - \frac{1}{N} \tag{4.17}$$

4.3.2 Type II or Fréchet Distribution

Variables that present a polynomial decay converge to the asymptotic Type II or Fréchet distribution. A random variable X presents polinomial decay if

$$\lim_{x \to \infty} x^k[1 - F_X(x)] = a = cte \tag{4.18}$$

46

Among the most known distributions, the distribution that meets the above requisite is the Lognormal distribution. The Fréchet distribution is given by

$$f_X(x) = \frac{k}{v} \cdot \left(\frac{v}{x}\right)^{k+1} exp\left[-\left(\frac{v}{x}\right)^k\right]$$

(4.19)

$$F_X(x) = exp\left[-\left(\frac{v}{x}\right)^k\right]$$

(4.20)

where v is the scale parameter and k is the shape parameter. The mean and standard deviation are related to these parameters by

$$\mu_X = v\Gamma\left[1 - \frac{1}{k}\right]$$

(4.21)

$$\sigma_X = v\left[\Gamma\left[1 - \frac{2}{k}\right] - \Gamma^2\left[1 - \frac{1}{k}\right]\right]^{1/2}$$

(4.22)

where $\Gamma(.)$ is the Gamma function.

4.4 Extremes Estimation for Gaussian Processes

In a Gaussian process, the peaks distribution is given by the Rice distribution, as shown in Section 3.3.2. If it is assumed that the peaks are non-correlated, i.e., independent, Equation (4.6). will return the distribution for the extreme value of the process peaks. It is possible to demonstrate that, for $N \to \infty$, the latter probability distribution will converge to the Gumbel distribution [9], independently of the bandwidth of the process spectrum, with the following parameters

$$u = \sqrt{m_0}\sqrt{2\ln(v_0 T)}$$

(4.23)

$$\alpha = \frac{\sqrt{2\ln(v_0 T)}}{\sqrt{m_0}}$$

(4.24)

where v_0 is the zero crossing frequency of the process, m_0 is the spectral moment of order zero (or the variance of the process) and T is the interval of time in seconds considered in the extreme value analysis.

According to the above equations, the extreme distribution of the peaks of a Gaussian process is independent of the bandwidth factor of the process. Additionally, it depends only on its spectral density function and the time interval considered for the extremes analysis.

As previously mentioned, in the case of a non-Gaussian process, there is not an analytical solution for the extreme value distribution. The usual procedure in this scenario, if one can assume that the process is ergodic, is to determine the extreme value probability distribution from the peaks distribution of a single realization. The latter is usually obtained through a fitting process to the sample of the peaks of the realization considered. As this is the case for the studied phenomena, i.e., the top tension in mooring lines, the topic will be explored in the following sections.

4.5 Methods for Extremes Estimation of Non-Gaussian Processes

When a random process is not normally distributed, that is when it is non-Gaussian, its most probable extreme value cannot be straightforwardly obtained, as it can be for Gaussian processes. In this case, there are a few methods available in the literature in order to obtain the corresponding extreme MPV.

The ideal method, presented in the beginning of Chapter 4, should be to perform a sampling of extremes values of the random process of interest and then fit an extreme distribution to the histogram of maxima obtained from these simulations. Since the accuracy of this method depends only on the number of simulations performed, not on the process stationarity, original process distribution or the possibility of correlation between maxima, it is applicable to any random process of interest, including non-Gaussian processes. Although it is ideal, this method can be very time-consuming and cumbersome for everyday design applications.

To overcome this issue, there are a few other methods available in the literature that do not involve a great number of simulations sampling in order to obtain a process extreme MPV. These methods can be classified as (i) methods based on the process

transformation, (ii) methods based on process crossing frequencies and (iii) methods based on the process peaks distribution.

The methods based on the process transformation use the concept of statistical equivalence (Section 2.2) in order to transform a non-Gaussian process in a normally distributed process. The extreme values estimates are calculated for the equivalent Gaussian process and later transformed back to the original process. These methods will not be explored in the present study.

The methods based on crossing frequencies assume that, for the high levels, the crossings can be modeled as a Poisson process [15]. Recent researches also point out a method based on the average exceedance rates of high levels [16][5], named *ACER* (Average Conditional Exceedance Rate). This method can automatically account for correlation between peaks in a time series and will be better explored in the following chapters.

In the methods based on the peaks distribution of the process, a probability distribution is fitted to the peaks sample of a single realization and the extremes distribution is then estimated through classical extreme statistics. There are a few possible procedures for the fitting of a probability distribution to the peaks sample. This study will focus on the method of moments. For the peaks probability modelling, this study will cover the three-parameter Weibull distribution, the Weibull-tail distribution and the SGLD distribution. Correlation between subsequent peaks will also be explored through the theoretical assumptions described in Section 4.2.

4.5.1 Extremes Sampling

As mentioned before, this method of extreme value estimation is applicable to any random process of interest, be it stationary or non-stationary, Gaussian or non-Gaussian. It is also independent of any possible correlation that can exist between successive peaks in a time-series of the process. Therefore, it is considered to be the ideal method for estimating a process extreme MPV [2].

In this method, a sufficiently large number N of sample realizations of the random process of interest is generated (or there already exists a sufficiently large number of measured observation of the random process). From each of the realizations, the largest

value, or the maximum, is extracted, resulting in an extreme values sample of size N. A known asymptotic distribution (Type I, Type II or Type III distribution) is then fitted to this extreme values sample. The extreme MPV can hence be directly determined. Figure 24 illustrates the procedure for estimating a random process extreme MPV through the sampling of maxima method. Thinking about structural analysis of marine structures, as they are the focus of the present work, this technique is very time-consuming and hence only indicated for benchmarking other more efficient procedures.
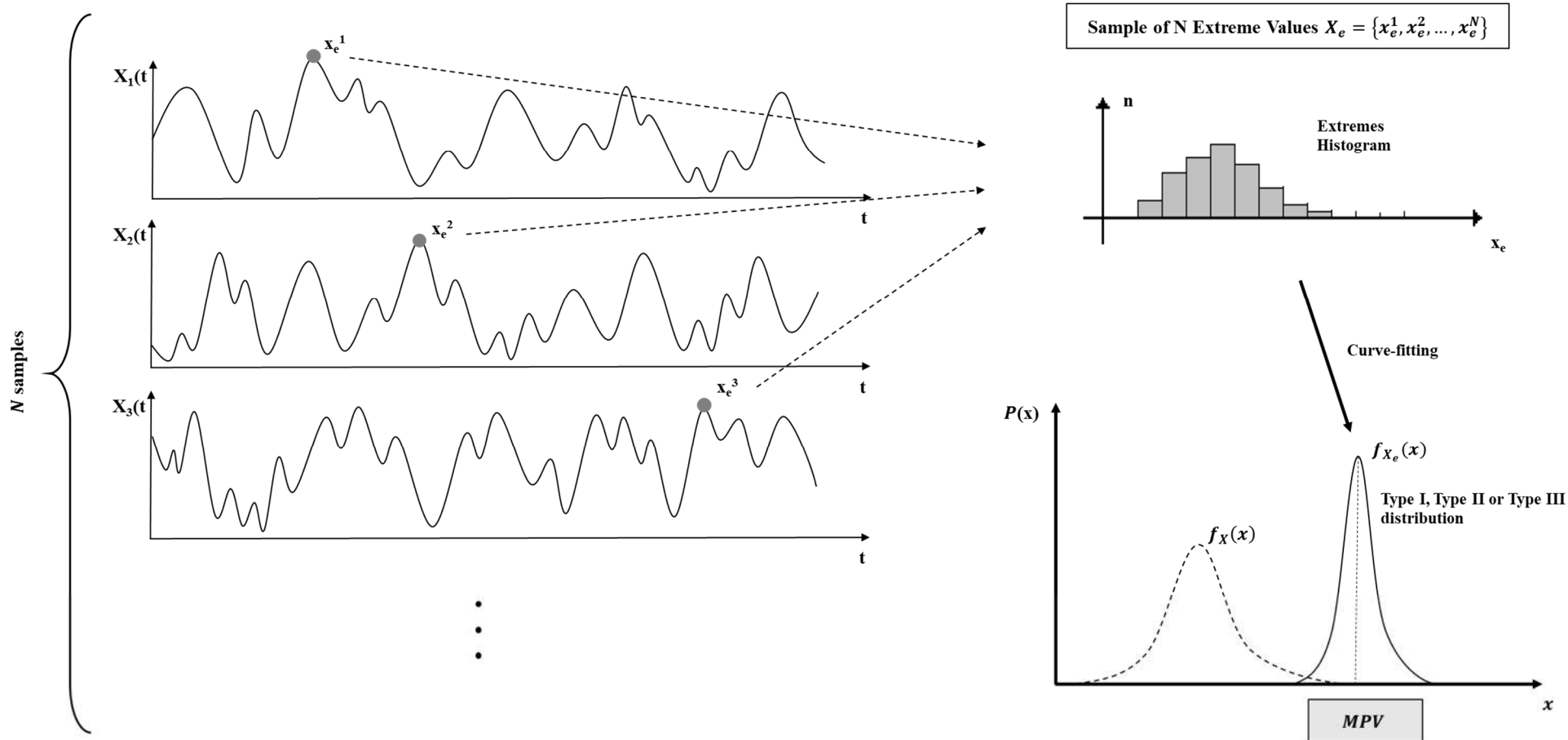
Figure 24 Extreme sampling method procedure for estimating a random process characteristic extreme value MPV

4.5.2 Peaks Probability Distribution Models

The methods based on the process peaks probability distribution model consist of fitting a probability distribution $F_{X_p}(x)$ to the observed peaks in a realization of the random process $X(t)$ (Figure 25) and later determine the extremes distribution through extreme value statistics for the interval of time T, i.e.,

$$F_{X_e}(x) = \left[F_{X_p}(x)\right]^{\nu_p T} \tag{4.25}$$

where ν_p is the peaks frequency in the observed realization.
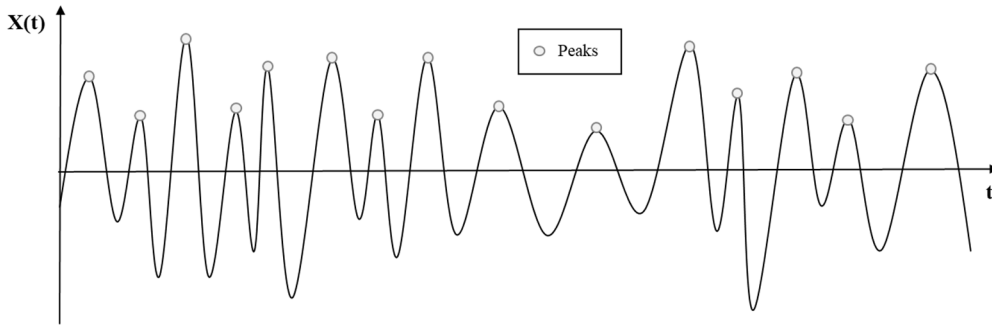
Figure 25 Peaks in a realization of $X(t)$

As previously mentioned, the value of interest in a usual extreme analysis is the most probable extreme value, or simply the MPV. This value can be obtained by solving the following equation:

$$F_{X_p}(x_{MPV}) = 1 - \frac{1}{\nu_p T} \tag{4.26}$$

The above equations are valid when the peaks are assumed to be statistically independent. As shown earlier in this dissertation, when considering correlation between two consecutive peaks, the extremes distribution is given by

$$F_{X_e}(x) = \frac{\left(F_{X_p,X_p}(x,x)\right)^{\nu_p T - 1}}{\left(F_{X_p}(x)\right)^{\nu_p T - 2}} \tag{4.27}$$

where the parameters were previously defined. The MPV, in this case, is obtained by its definition, i.e., the point where $f_{X_e}(x)$ reaches its maximum value. The MPV will be the point where the first derivative of $f_{X_e}(x)$, or the second derivative of $F_{X_e}(x)$, is null.

In the present study, three probability models were investigated for the peaks probability $F_{X_p}(x)$ modelling: (i) the three-parameter Weibull distribution, (ii) the Weibull-tail distribution and (iii) the SGLD distribution. In each of them, the correlation between consecutive peaks was also considered. The fitting process of the probability distribution to the peaks sample was done by the method of moments.

There is not a general rule for the definition of the peaks sample since a generic random process can present local and global peaks, as shown in Figure 26. A global peak in a process time-series is the largest one between two mean up-crossings. Based on practical experience and previous works [17][10], in the present study the peaks sample is represented by all global peaks belonging to the sampled time-series.
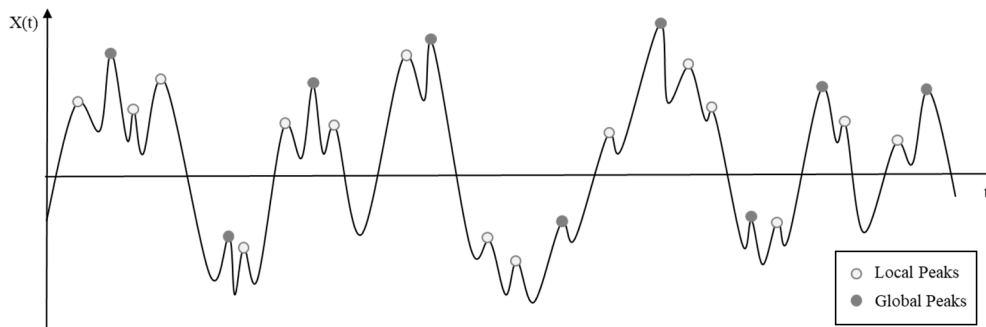
Figure 26 Definition of global and local peaks in a realization of $X(t)$

A summary for the peaks distribution model method for estimating extremes in a non-Gaussian and ergodic process using a single realization is illustrated in Figure 27. Since this study focus only on the maxima of the studied phenomena, top tension loads in mooring lines, the sample of peaks is only referred to the peaks above the mean level.
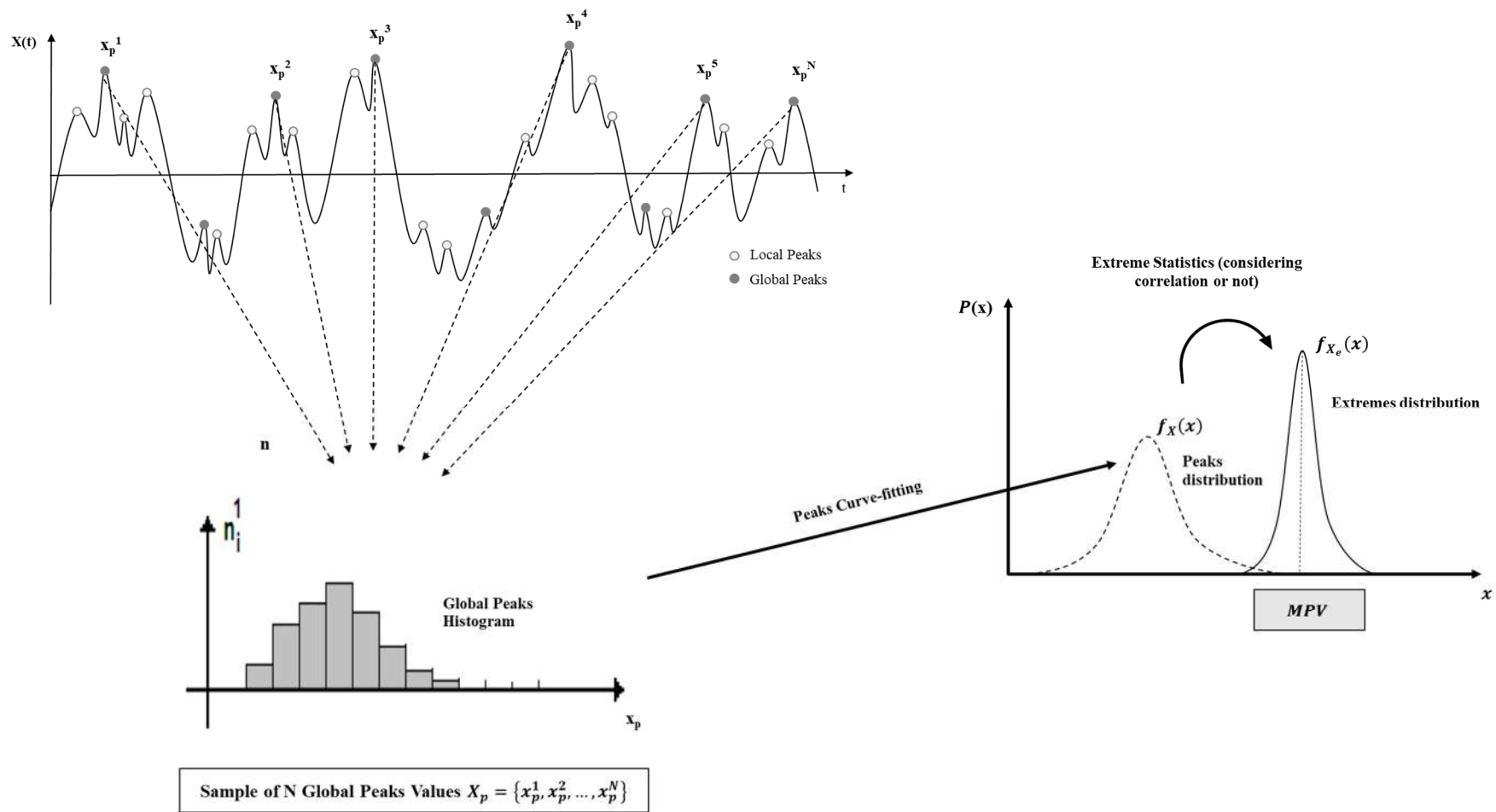
Figure 27 Peaks distribution model method for estimating extreme values in a non-gaussian and ergodic process using a single realization: the peaks sample includes only global peaks above the mean level

4.5.3 Average Conditional Exceedance Rate Method (ACER)

Another methodology that enables the consideration of dependency between the peaks in a time-series when estimating extreme values was developed by researchers from the Norwegian University of Science and Technology NTNU [16][5] and is known as the Average Conditional Exceedance Rate, or ACER. This methodology uses a chain of statistical dependency between successive peaks in order to create an extrapolated function of the extremes distribution tail. With this function, one can estimate the most probable extreme value taking into account the dependency between the peaks. The consideration of independent peaks is a particular case of the methodology.

The method consists in identifying the peaks sample in a realization of size N of the random process $X(t)$ and calculating the ACER function for a specific observed peak X_p^j and a exceedance level η, represented by $\varepsilon_{kj}(\eta)$:

$$\varepsilon_{kj}(\eta) = Prob\big(X_p^j > \eta \big| X_p^{j-1} \leq \eta, \dots, X_p^{j-k+1} \leq \eta\big), \qquad 1 \leq k \leq j \leq N \qquad (4.28)$$

$\varepsilon_{kj}(\eta)$ denotes the probability of the jth peak X_p^j exceeding the value η given that all $j-1$ previous peaks did not exceeded. The positive integer index k denotes the number of conditional exceedances to be considered in the peaks sample. Thus the indexes k represent each of the kth ACER function. For example, the first three ACER functions are given by

$$\varepsilon_{1j}(\eta) = Prob\big(X_p^j > \eta\big), k = 1 \qquad (4.29)$$

$$\varepsilon_{2j}(\eta) = Prob\big(X_p^j > \eta \big| X_p^{j-1} \leq \eta\big), k = 2 \qquad (4.30)$$

$$\varepsilon_{3j}(\eta) = Prob\big(X_p^j > \eta \big| X_p^{j-1} \leq \eta \cap X_p^{j-2} \leq \eta\big), k = 3 \qquad (4.31)$$

In practice, the ACER functions are determined empirically by counting the number of conditional exceedance occurrences in the peaks sample and dividing it by the total number of peaks. It is important to remark that the peaks sample used by the ACER method comprises both local and global peaks.

Figure 28 presents the behavior of the first three ACER functions ($k = 1, 2, 3$) in a logarithmic vertical scale. The functions decrease as the index k increases or as the level of exceedance η increases: naturally, the probability of existing a peak that satisfies the condition for high values of k or η is lower.
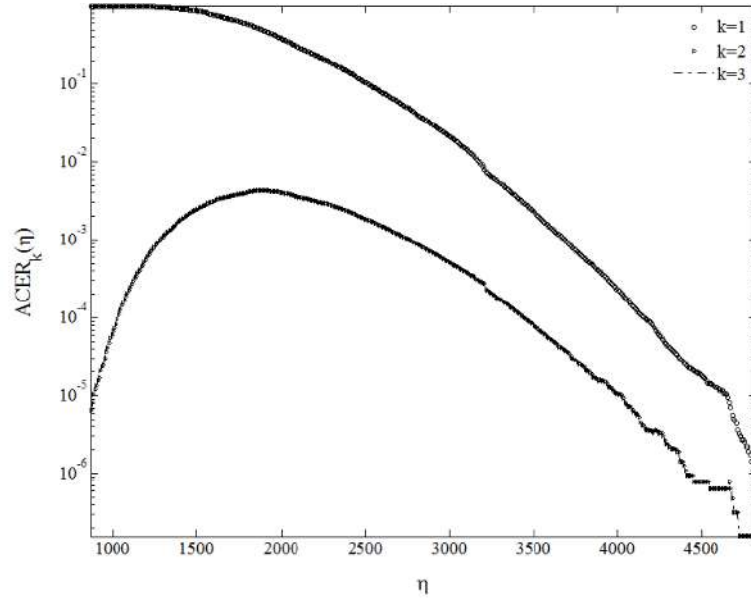
Figure 28 Plot of exceedance levels η versus ACER functions $\varepsilon_k(\eta)$ for $k = 1, 2, 3$ on the logarithmic scale [16]

The next step of the ACER method is to obtain continuous functions $\varepsilon_{kj}(\eta)$ that allow to extrapolate its behavior for larger values of the level η. The behavior of the ACER function tail is exponential thus the following model can be used to represent it [5]

$$\varepsilon_k(\eta) = q_k exp(-a_k(\eta - b_k)^{c_k}) \tag{4.32}$$

where η is the level of exceedances and the constants a_k, b_k, c_k and q_k are those that better fit the behavior of the empirical ACER function tail. These parameters can be determined with the help of optimization methods, such as the Levenberg-Marquardt algorithm [5].

Finally, the extreme cumulative density function of the peaks is given by:

$$F_k(\eta) = exp\left[-\left(N_p\big(\varepsilon_k(\eta)\big)\right)\right] \tag{4.33}$$

where N_p is the total number of peaks for the period of time considered in the extreme value estimation.

In practice, the initial level η to be considered is the one that stabilizes the function $\varepsilon_{kj}(\eta)$. In this study, the ACER method was investigated through the use of the software developed by Karpa [18]. In addition, the first three dependencies ($k = 1, 2, 3$) between peaks were considered, i.e., the case where consecutive peaks are independent of each other ($k = 1$), the case where two consecutive peaks are correlated ($k = 2$) and the case where three consecutive peaks are considered correlated ($k = 3$).

5 MOORING LINES EXTREME TOP TENSION ANALYSIS

5.1 Environmental Loads: Ocean Waves, Wind and Current

One of the main random loads that acts on an offshore moored floating unit is associated to the ocean wave elevations. The standard approach to the statistical modeling of ocean waves is to assume that the ocean surface constitutes a stochastic process that can be assumed stationary in time [15]. In engineering practice, stationarity is assumed only for limited periods of time, e.g. three hours, which is then referred to as the short-term description of the process, or simply the sea state. Thus, the sea elevation can be simplified as a sequence of these stationary processes, or sea states.

Of particular importance of this simplification is the possibility to represent the random process by its spectral density. In practice, each sea state is represented by a significant wave height H_s and a characteristic period T_p or T_z, both associated to an appropriate spectral density function $S_\eta(\omega)$ for the location of interest. The significant wave height H_s of a sea state is defined as the mean wave height of the highest third of the waves heights H_i and the characteristic period T_z is defined by the average of the mean level crossing periods T_i, as illustrated in Figure 29.
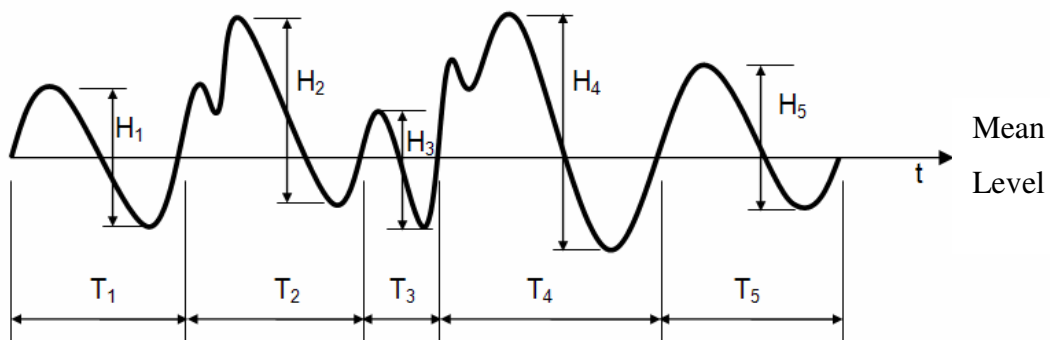
Figure 29 Random wave observation [20]

Having known the spectral density $S_\eta(\omega)$ of the location of interest, as previously commented in section 3.2.1, one can, through the Fourier Transform (see, for instance, [6]), represent this random process in the time-domain. Thus, the wave elevation can be approximately represented by a superposition of N harmonic components, as follows

$$\eta(t) = \sum_{k=1}^{N} a_k \cos(\omega_k t + \phi_k) \tag{5.1}$$

where ϕ_k are random phases uniformly distributed in the interval $[0, 2\pi]$ and the magnitudes of a_k are related to the spectral density function of the waves $S_\eta(\omega)$ by

$$a_k = \sqrt{2 S_\eta(\omega_k) \Delta\omega} \tag{5.2}$$

where ω_k is the representative frequency of the kth harmonic component and $\Delta\omega$ is the frequency interval of the spectrum $S_\eta(\omega)$ discretization.

Due to the central limit theorem (Section 2.5), it can be expected that $\eta(t)$ is approximately Gaussian when the number of harmonic components used to represent it is large enough, i.e., when $N \to \infty$.

It is important to notice that one must be careful when using above expressions. If the representative frequency of the kth harmonic component ω_k is taken, for example, as the mean frequency in the frequency interval of the spectrum $S_\eta(\omega)$ discretization $\Delta\omega$, it is expected that the simulated wave elevation repeats itself within a period $2\pi/\Delta\omega$, i.e., becomes a periodic function. To avoid such inconvenient, in the present study ω_k is taken randomly in the interval of the spectrum discretization.

From Equations (5.1) and (5.2), one can naturally assume that the number of harmonic components used to represent the spectral density function $S_\eta(\omega)$ will directly influence in how accurately the generated signal in time $\eta(t)$ represents the energy contained in the original wave spectrum. There are several references in the literature, as for instance, [15] and [19], on the recommended wave spectrum discretization level one must employ for a given type of structure and analysis. This study tries to investigate the

influence of using lower and higher levels of discretization of the wave spectrum in the extreme analysis of top tension loads in mooring lines.

There a few well established spectral density functions $S_\eta(\omega)$ in the literature, such as the JONSWAP (Joint North Sea Wave Project) spectrum, specially developed for the North Sea in joint studies with offshore industries, or the Pierson-Moskowitz spectrum, initially developed from measurements in the North Atlantic [21][22]. The spectrum that best represents the Brazilian offshore locations here investigated is the JONSWAP spectrum. Further explanations about other types of wave spectra can be found in [15].

The parameters that characterize the adjusted JONSWAP spectrum are classified as a function of H_s and T_p. The peak angular frequency ω_p (or the period T_p) corresponds to the frequency at the maximum value of $S_\eta(\omega)$. JONSWAP spectrum can be defined as

$$S_{JS}(\omega) = \alpha \frac{g^2}{2\pi^4\omega^5} exp\left[-1.25\left(\frac{\omega}{\omega_p}\right)^{-4}\right]\gamma^{exp\left[\frac{(\omega-\omega_p)^2}{2\sigma^2\omega_p^2}\right]} \tag{5.3}$$

where α and γ are the spectrum shape parameters. They are related to H_s and T_p, by expressions previously defined for a location of interest. For instance, for the Campos Basin, located offshore Brazil, parameters α and γ are related to H_s and T_p by the following equations:

$$\gamma = exp\left(1.0394 - 0.01966\frac{T_p}{\sqrt{H_s}}\right) \tag{5.4}$$

$$\alpha = 5.0609\frac{H_s^2}{T_p^4}(1 - 0.287\ln\gamma) \tag{5.5}$$

The third shape parameter σ is fixed and only dependent on ω_p:

$$\sigma = \begin{cases}\sigma_a = 0.07, & \omega \leq \omega_p \\ \sigma_b = 0.09, & \omega > \omega_p\end{cases} \tag{5.6}$$

The usual shape of the JONSWAP spectrum is shown in Figure 30.
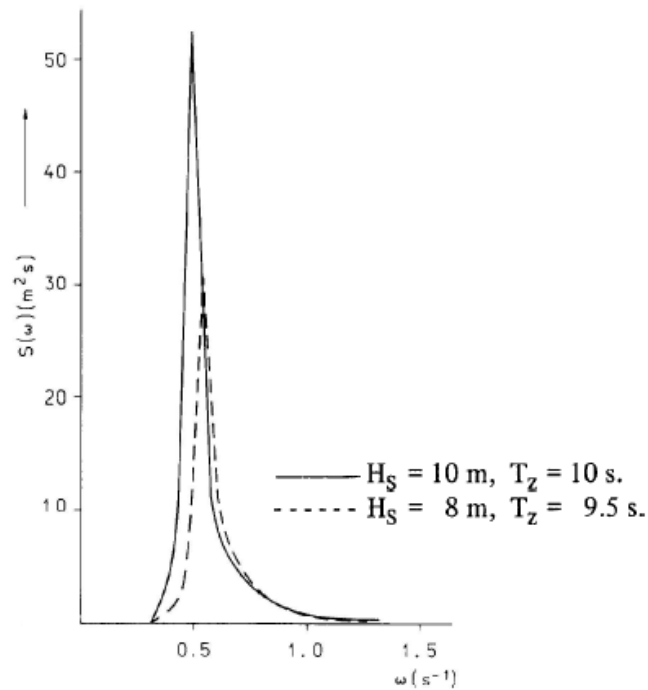
Figure 30 JONSWAP spectra [15]

Alongside with the ocean waves, wind velocities can also be interpreted as random processes. Therefore, the wind velocity may be statistically treated in a similar manner of that presented for the ocean waves. However, in the present study, wind was considered as a deterministic variable associated to the sea state in question. More information on wind statistical treatment can be found, for example, in [15].

The current was also considered as a deterministic variable, for each of the loading directions investigated. The current considered was associated with a triangle profile defined by the current velocity at the surface level v_0, at a single direction plane, as shows Figure 31. The current velocity at the bottom level was considered zero.

$$v_c = v_0$$
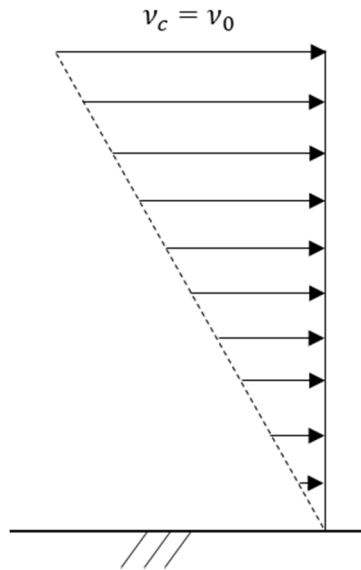
Figure 31 Current velocity triangle profile

5.2 Forces Induced by Ocean Waves, Wind and Current

The interaction between the ocean waves and the floating unit to which the mooring lines are connected to, here considered as a rigid body, induces hydrodynamic forces that will generate dynamic motions in the floating unit, and therefore in the mooring lines system, in the six degrees of freedom. Due to the complexity of this interaction, the hydrodynamic forces acting on the mooring lines are nonlinearly related to the sea surface elevation.

Even though the short-term wave-frequency excitation can be assumed to be a Gaussian process, the mooring line tensions generally are not mainly due to the second-order effects associated to the floater motions. These motions are nonlinearly related to the sea surface elevation, more specifically, they are associated with the square of the sea surface elevation, as it will be shown in this chapter.

In practice, one can represent the acting forces in the floater unit as a bi-dimensional Fourier series [23]. Thus, the hydrodynamic force generated by the ocean waves in a given degree of freedom of the floating unit can be represented by

$$F(t) = F_0 + \sum_{k=1}^{N} C(\omega_k) A_k \cos\left(\omega_k t + \phi_k + \varphi(\omega_k)\right)$$

$$+ \sum_{k=1}^{N} \sum_{j=1}^{N} Q^-(\omega_k - \omega_j) A_k A_j \cos\left((\omega_k - \omega_j)t + (\phi_k - \phi_j) + \psi^-(\omega_k - \omega_j)\right) \qquad (5.7)$$

$$+ \sum_{k=1}^{N} \sum_{j=1}^{N} Q^+(\omega_k + \omega_j) A_k A_j \cos\left((\omega_k + \omega_j)t + (\phi_k + \phi_j) + \psi^-(\omega_k + \omega_j)\right)$$

where

F_0 Mean drift force $F_0 = \sum_{k=1}^{N} C(0) A_k{}^2$;

$C(\omega_k)$ First order (linear) coefficient of hydrodynamic force;

$Q^-(\omega)$ and $Q^+(\omega)$ Second order quadratic coefficients of hydrodynamic force correspondent to the wave frequencies differences and the wave frequencies sum, respectively;

$\varphi(\omega_k)$ Phase angle of the components of the hydrodynamic force linear portion;

$\psi^-(\omega)$ and $\psi^+(\omega)$ Phase angles of the components of the hydrodynamic force nonlinear portion.

All of the above parameters basically depend on the floater geometry and must be estimated by specific computational routines, such as the software WAMIT [24].

It is clear that if an irregular wave elevation is modeled as a superposition of components at different frequencies then forces proportional to the square of that elevation will contain terms at both the sums and differences of the elementary wave frequencies [1], as it can be observed in Equation (5.7).

Apart from the ocean waves, the interaction between the wind and current and the floater-line system will also induce hydrodynamic forces that will generate response of the system.

The wind forces act on the exposed portion of the floating unit. The wind velocity can be considered as a deterministic variable, i.e., having a constant value independent of

time, or, along with the ocean waves, as random variables characterized by a specific spectral density function. As mentioned above, this study only contemplates the consideration of constant velocity wind.

A constant wind velocity v_w will generate only a static load, proportional to the wind exposed area of the floating unit, given by

$$f_w = \frac{\rho_a}{2} C_w A_{wind} {v_w}^2 \tag{5.8}$$

where ρ_a is the air density, A_{wind} is the total exposed area of the floating unit subjected to the wind action, C_w is the area shape coefficient (dependent on how perpendicular the wind area is to the wind acting direction) and v_w is the mean wind velocity.

On the other hand, the current induces not only forces on the submerged portion of the floating unit but it also induces hydrodynamic forces on the hanging portion of risers and mooring lines connected to the floating unit. Estimation of forces induced by fluids flow on slender bodies, such as the forces induced by the current on the risers and lines system, can be done through the Morison's formulation [25]. The Morison's formulation assumes that, for a sufficiently slender body, the forces induced by a flowing fluid can be estimated as an approximation in which important parameters of the fluid flow, such as pressure, velocity and acceleration, are approximated as the correspondent value in the cross section axis of the slender body. Morison's formulation is given, as a force per unit length, by

$$f_c = \frac{1}{2} \rho_w D C_d |\dot{u} - \dot{x}| (\dot{u} - \dot{x}) + \rho_w \frac{\pi D^2}{4} C_m \ddot{u} - \rho_w \frac{\pi D^2}{4} C_a \ddot{x} \tag{5.9}$$

where ρ_w is the fluid density (in this case, the seawater density), D is the characteristic dimension of the cross section of the body (usually taken as the external diameter of a cylindrical body), $\dot{u}, \dot{x}, \ddot{u}$ and $\ddot{x}$ are the fluid and slender body velocities and accelerations, respectively, and C_d, C_m and C_a are the dimensionless coefficients of drag, inertia and additional mass. The first portion of Morison's formulation is usually referred as the drag portion, associated with the viscous effects of the fluid and body interaction. The other portion of the formulation is usually referred as the inertia portion, proportional to the fluid and body accelerations.

Morison's formulation is usually a sufficiently accurate approach for the analysis of risers and mooring lines. These type of structures do not interfere in the waves profile and are hence considered as sufficiently slender bodies.

Since the current velocities are considered constant in the present work, they will only generate static loads in the hull, proportional to the submerged area of the floating unit. Likewise, in the risers and mooring lines system, since there is no variation of the current velocities, there are no inertia forces caused by the current flow. On the other hand, there will be inertia forces induced by the portion of the random ocean waves acting directly on the mooring lines and risers system.

5.3 Time-Domain Dynamic Response Analysis

The dynamic model for determining the motions $x(t)$ of a floating unit can be expressed by the dynamic equilibrium equation:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F(t) + F_0 \tag{5.10}$$

where m, c and k denote the mass (structural + inertial mass) of the floater-lines system, the damping (hydrodynamic damping + viscous damping due to the mooring lines) of the system and the stiffness (hydrostatic + mooring lines) of the system, respectively. $F(t)$ denotes the sum of all time-defined loads that act on the floating unit and F_0 denotes the sum of all constant loads that act on the floating unit. In the present study, only the loads generated by the ocean waves are treated as dependent of time.

Usually, the second order hydrodynamic force components are small when compared to the first order ones. However, the presence of the low difference frequency components in the excitation results in a resonant response of the floater-lines system in their horizontal degrees of freedom (surge, sway and yaw), which are characterized for very high natural periods (order of 200~400s). This resonance is characterized by motions of large amplitudes, known as slow-drift motions, or simply low frequency motions enabled by lower damping in the low frequencies range. This way, the floater unit motions can be represented by

$$x(t) = x_0 + x_{LF}(t) + x_{WF}(t) \tag{5.11}$$

where

x_0 static motion or offset due to the mean drift force;

$x_{LF}(t)$ dynamic motion component proportional to the low (difference) frequency components;

$x_{WF}(t)$ dynamic motion component in the wave frequencies range.

Likewise, a typical mooring line tension response will present three main contributions, given by

$$T(t) = T_0 + T_{LF}(t) + T_{WF}(t) \tag{5.12}$$

where

T_0 steady offset component due to the static offset of the floater + line pre-tension;

$T_{LF}(t)$ tension dynamic component proportional to the low (difference) frequency components;

$T_{WF}(t)$ tension dynamic component associated to the first order motions of the floater, in the wave frequencies range.

Due to the nonlinear contributions in the floater-lines system behavior, a frequency-domain analysis is not applicable since it relies on the linearization of the mooring system structural behavior. A nonlinear fully coupled dynamic analysis comprising both the floater and its risers and mooring lines in the time domain should be therefore a more appropriate approach.

As mentioned above, the floater-line system response cannot be treated as a Gaussian process, since it is nonlinearly related to the sea surface elevation. Analytical solutions for the response process probability distribution, or for the probability distribution of response peaks, are not available. In this case, the procedure for estimating

extreme tensions in the mooring lines has to follow one of the methods for extreme estimation for non-Gaussian processes described in Section 4.5.

On the other hand, although the floater-line system response cannot be considered as a Gaussian process, it can be considered ergodic and consequently stationary. Since the sea surface elevation model described in Equation (5.1) is stationary and ergodic, the response of the system can be considered stationary and ergodic as well due to the system stability. Dynamic equilibrium equation (Equation (5.6)) ensures that there is no instability point in the response, i.e., there is no instant of time where the response becomes unpredictable. Thus the extremes MPV's can be estimated by the single realization approach presented in Section 4.5.

In the present work, ocean waves, wind velocity and current velocity were considered to be acting in the same direction. Since the loads from both wind and current were treated as constant ones, they will only contribute to the static offset of the floater-line system and therefore to the steady offset component of the line tension.

5.4 Extremes Analysis: Long-Term and Short-Term Design Approaches

Since a wave spectrum can be completely defined by the pair of parameters H_s and T_z, a long-term model of the ocean waves can be based on the occurrence frequency of these parameters for the location were the floating unit is installed [15]. From a scatter diagram that provides the probability of occurrence of a specific pair of H_s and T_z, one can predict the joint distribution function of these environmental parameters. The long-term sea states representation will then be considered as a series of short-term sea states. The complete long-term environmental process will be given by the joint distribution function of the ocean waves significant heights, periods and directions, the wind velocities and directions and the current velocities and directions.

At first, the most accurate design approach to determine extreme load effects for design checks is based on the long-term statistics of response, but it may clearly not be the most economic method from the computational point of view since it also involves the response calculations for sea states that contribute little or nothing to the structural response [15]. Nowadays, the short-term design approach, using nonlinear time domain

simulations, is a procedure accepted by various modern design standards for the extreme response analysis of risers and mooring lines connected to floating production units [26]. The most recommended procedure to select the short-term environmental conditions for the wave design approach is based on the environmental contour method [3].

In the environmental contour method, an arbitrary number of environmental conditions is selected from the environmental contour relative to a return-period of interest, usually 100 years or 1000 years. For each environmental condition a short-term extreme response analysis is performed. The extreme characteristic response value for design verification is taken as the largest one among all short-term extreme MPVs associated with the environmental conditions on the contour. According to [27], in this method, the variability of this short-term extreme response value in the long-term needs to be artificially accounted for. This is usually achieved by multiplying the obtained extreme response with a predetermined factor. In this work, the factor was considered equal to 1.

In this study, the short-term design approach denotes a short-term period T_{st} of 3 hours. For research purposes, the short-term environmental conditions are defined by the waves in the 100-yr environmental contour directionally aligned with the associated wind and current conditions.

6 CASE STUDIES

Previous chapters evidenced that when studying a non-Gaussian random process such as the top tension response in mooring lines connected to floating units, one cannot obtain analytical results for the description of the statistics of such process. An appropriate statistical treatment is essential and, for an extreme response analysis, a specific methodology is required.

In Chapter 4 it was presented that if the studied random process can be treated as an ergodic process, its most probable extreme value can be estimated from a single realization of the process. However, this extreme MPV will depend on the probability model used for representing the peaks of the process, on the time-series length and on the discretization level employed in the wave spectrum. It will also depend on whether correlation between the peaks of the time-series is important or not.

It was previously shown that the top tension load in mooring lines connected to floating platforms cannot be treated as a Gaussian process, even though some environmental load-generator processes are. This results from the nonlinear behavior of the floater-lines system itself, the second-order random wave loading applied to the system, etc. However, floater-lines system response can be treated as a stationary and ergodic stochastic process [2], naturally including the top tension response of the mooring lines.

This study hence investigates the influence of the probability distribution model choice for the line tension peaks, the influence of the numerical simulation (time-series) length, the influence of the wave spectrum discretization level and the effect of considering correlation between consecutive peaks on the top tension extreme response analysis. The study investigates these effects for two distinct mooring lines connected to FPSOs located in different water depths offshore Brazil.

The first case study is a chain-polyester-chain mooring line connected to a spread-moored FPSO in 1980m water depth (Case Study A). The second is a chain-wire rope-chain mooring line connected to a turret-moored FPSO in 180m water depth (Case Study B). The investigated top tension time-series are generated from the weakly coupled models of the floaters and its systems of risers and mooring lines using the DYNASIM

computer code [28]. In this weakly coupled model, the lines are represented by their corresponding geometric catenary equations and hydrodynamic loads on the lines are modeled by a simplified approach [28].

For the present studies, wave, current and wind environmental loads are considered to be directionally aligned. Wave parameters (and corresponding ones for wind and current) are taken from the short-term condition on 100-yr environmental contour of the location of interest, producing the largest extreme response on the most loaded line. For each case study, the investigated mooring line will be the one that presented the largest peak for all environmental conditions belonging to the environmental contour for the corresponding unit location. The environmental condition that generated this largest response will be the condition used for the statistical investigations.

In order to investigate the statistical influence of the finite length of the sampled time-series on the extreme response simulation lengths T_s of 3, 6, 9, 12 and 15 hours are considered. These simulation lengths are obtained after neglecting the first 5000s of simulated time-series, in order to properly disregard any dynamic transient effect on the response, e.g. a simulation length of 3-h (10800s) implies that the time-series was generated with a 15800s duration in DYNASIM (10800s + 5000s). Different levels of wave spectrum discretization N_w of 300, 1000 and 2000 harmonic waves are also considered. For each pair of simulation length and spectrum discretization level $\{T_s, N_w\}$ 200 distinct top tension realizations are generated by changing the seed number for random wave phase generation. In total, 200 realizations x 3 discretization levels for the wave spectrum x 5 time-series lengths = 3000 distinct simulations are performed for each case study.

For each sampled realization, the 3-h extreme MPV is estimated for each peaks probability model investigated, including or not the effect of correlation between peaks. The extreme MPV is also estimated through the ACER method, for $k = 1, 2, 3$. From the set of 200 MPV results for each specific case, the average and coefficient of variation of the MPV sample mean estimator are obtained. The latter value is compared with the 'true' extreme MPV, estimated through a sampling of extremes, in order to establish the biases

associated to each set of 'simulation length - spectrum discretization - peaks probability model' condition. This will be better explained in the next sections.

Assuming that the process is ergodic, just a single realization is enough to statistically evaluate it. However, since numerical simulations of random processes are associated to initial seed numbers, usually provided by the user, and of finite time length, it is natural to deduce that each simulated time-series has a different outcome and therefore distinct statistical characteristics. When the values of interest of the process are its extremes values, the latter observation is even more important since extremes methodologies relies on a sub-group of the realization outcomes: its peaks. The bias of a method, or the measure of how accurate the method is if compared to benchmark values, therefore relies not only on the method itself but also on these simulations details. Thus, in order to assess the effect of the parameters on the performance of each method, a large enough number of 200 distinct realizations are simulated. An accurate model is expected to converge on average (the mean value of the estimated MPVs) to the 'true' one and to have a low standard deviation of its MPV estimates.

The statistical analyses performed considered only the environmental loading portion of the mooring line top tension response, i.e., the mooring line pre-tension is previously subtracted from the top tension time-series. Estimated MPV's are then associated only with the top tension loads generated by the effects of the random ocean waves and deterministic wind and current loads.

6.1 Short-Term Extreme Top Tension MPV Benchmark Values

In order to establish a reliable benchmark to which the results can be compared to, i.e., the one considered 'ideal' in section 4.5.1, a sampling of extremes is generated in order to estimate the 'true' extreme top tension MPV. The general procedure is illustrated in Figure 24.

For the extremes sampling, 1000 independent 3-h long ($T_s = 3$h) time-domain simulations are generated with the largest level of the discretization for the wave spectrum employed in the study, i.e., $N_w = 2000$. The maximum value encountered in each simulated most loaded line top tension time-series is extracted, resulting in a sample of

1000 extreme short-term values $X_e = \{x_e^1, x_e^2, x_e^3, \ldots, x_e^{1000}\}$. Then, a Gumbel distribution is fitted to this extreme sample through the method of moments. The Gumbel distribution (or extreme Type I distribution) choice is supported by the fact that in extreme statistics theory the vast majority of marginal distributions asymptotically converge to this extreme distribution [12], as mentioned before.

Using the method of moments for the model fitting, the parameters u_G and α_G of the Gumbel distribution are given by

$$\alpha_G = \frac{\pi}{\sigma_{X_e}\sqrt{6}} \tag{6.1}$$

$$u_G = \mu_{X_e} - \frac{0.57722\sqrt{6}}{\pi}\sigma_{X_e} \tag{6.2}$$

where σ_{X_e} and μ_{X_e} are the standard deviation and the mean of the extremes sample, respectively. Gumbel's parameter u_G is also the most probable extreme value. The benchmark or reference value for the extreme top tension MPV of each case study will be referred from this point as MPV_{REF}.

6.2 Associated Biases

For each group of N realizations of each set of 'simulation length T_s - spectrum discretization N_w - peaks probability model' conditions, the estimator of the average value of the extreme MPVs is calculated, as well as the estimator of the standard deviation, as described in the following equations for a given peaks probability model:

$$\mu_{\{T_s,\, N_w\}} = \frac{\sum_{i=1}^{N} MPV_i}{N-1} \tag{6.3}$$

$$\sigma_{\{T_s,\, N_w\}} = \sqrt{\frac{\sum_{i=1}^{N}(MPV_i - \mu_{\{T_s,\, N_w\}})^2}{N-1}} \tag{6.4}$$

In the present work, as previously mentioned, $N = 200$. In addition, the coefficient of variation of the MPVs sample of size N is taken:

$$CoV_{\{T_s, N_w\}} = \frac{\sigma_{\{T_s, N_w\}}}{\mu_{\{T_s, N_w\}}} \tag{6.5}$$

In order to measure how close to the benchmark value the estimator $\mu_{\{T_s, N_w\}}$ is, for each peaks probability model, the bias tool is used. In statistics, the bias (or bias function) of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated. An estimator or zero bias is called unbiased. Otherwise, the estimator is said to be biased. The 'true' value of the extreme MPV for each case study is taken as the benchmark value, explained in the previous section. The bias for each peaks probability model in this study is defined as the ratio between the expected value of the extreme MPV for each set of (T_s, N_w) and the benchmark value, as Equation (6.6) shows.

$$Bias_{\{T_s, N_w\}} = \frac{\mu_{\{T_s, N_w\}}}{MPV_{REF}} \tag{6.6}$$

Figure 32 illustrates the estimators investigated and the calculation of the bias for each set of conditions.

	Ts = 10800s (3-h) / Nw = 300																	
Sample	Signal				Peaks						Extreme 3-h MPV								
											Independent Peaks				Correlated Peaks				
	μ_x	σ_x	Υ_x	κ_x	μ_p	σ_p	Υ_p	κ_p	ν_m	ρ	W3P	WT	SGLD	ACER (k = 1)	W3P - corr	WT - corr	SGLD - corr	ACER (k = 2)	ACER (k = 3)
1	386	427	0.56	3.08	813	407	0.75	2.47	0.03	0.67	1965	1987	1788	2690	1882	1904	1805	2091	2090
2	387	487	0.44	2.43	1079	376	0.00	2.47	0.04	0.73	1890	1976	1858	2551	1859	1935	1777	1945	1944
3	399	516	0.68	3.48	1046	503	0.68	2.83	0.04	0.82	2377	2523	2316	2701	2215	2324	2239	2102	2101
4	368	371	0.45	2.82	760	336	0.54	2.28	0.03	0.76	1651	1670	1538	2539	1630	1648	1555	1893	1893
-																			
-																			
-																			
200	367	384	0.44	2.79	738	276	0.94	3.65	0.04	0.68	1823	1866	1851	2682	1713	1766	1682	1872	1871

$\mu_{W3P}, \sigma_{W3P}, CoV_{W3P}$ $\mu_{W3P-corr}, \sigma_{W3P-corr}, CoV_{W3P}$

$$Bias\ (W3P) = \frac{\mu_{W3P}}{Benchmark\ MPV} \qquad Bias\ (W3P - corr) = \frac{\mu_{W3P-corr}}{Benchmark\ MPV}$$
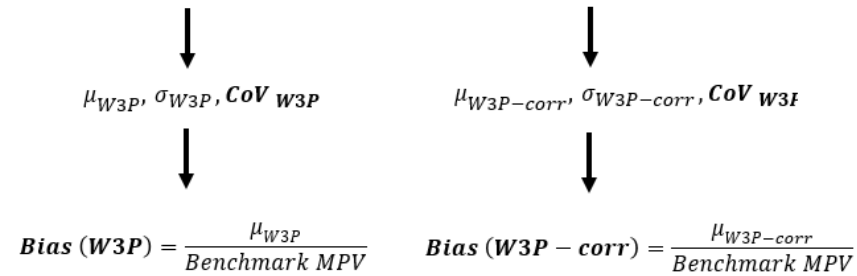
Figure 32 Procedure for characterizing a chosen pair of $\{T_s, N_w\}$ for each peaks probability model: example for the 3-parameter Weibull probability model for $T_s = 3h$ and $N_w = 300$

6.3 Study Case A – FPSO in Deep Water

In this case study, a chain-polyester-chain mooring line connected to a spread-moored FPSO in 1980m water depth under a 100-yr environmental condition offshore Brazil is investigated. The line configuration and description is given in Figure 33. The selected environmental condition and mooring line orientations are presented in Figure 34. Study case A corresponds to a short-term condition in the environmental 100-yr contour with H_s = 6.62m and T_p = 10.80s (modelled with JONSWAP wave spectrum), and associated surface current velocity = 0.34m/s and wind velocity = 22.10m/s acting as indicated in Figure 34.
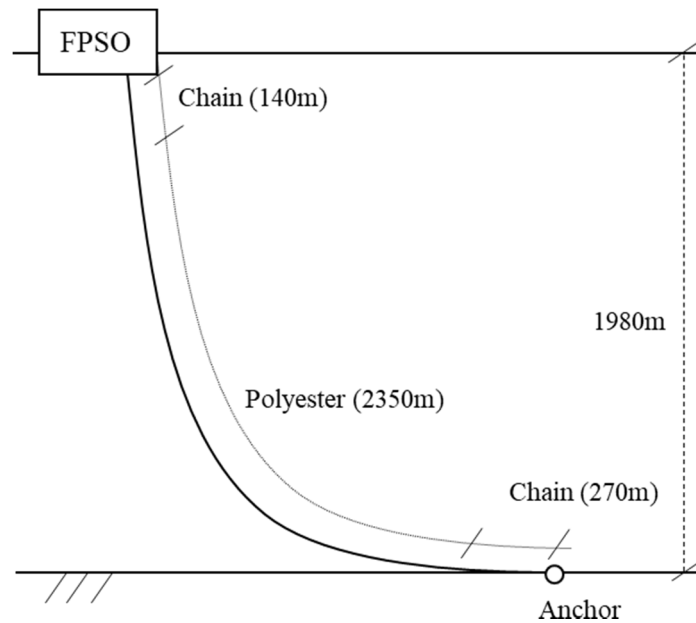
Figure 33 Mooring line configuration for case study A
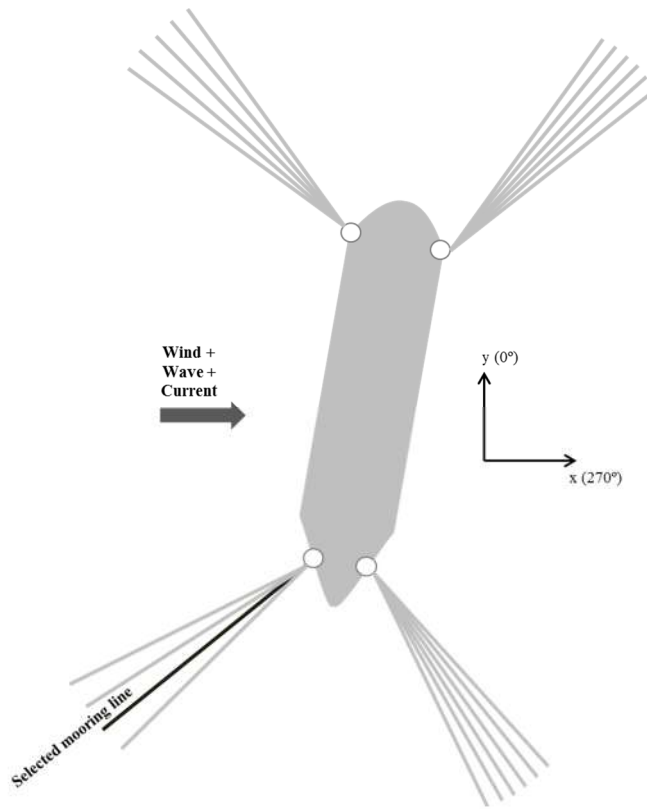
Figure 34 Environmental condition and selected mooring line orientations for case study A

The typical spectral density of the selected mooring line top tension response is shown in Figure 35. In this figure, the spectrum is obtained by Fast Fourier Transform (FFT) applied to a generated random time-series. For better visualization, the result of the FFT is smoothed by the Hanning window procedure [29].
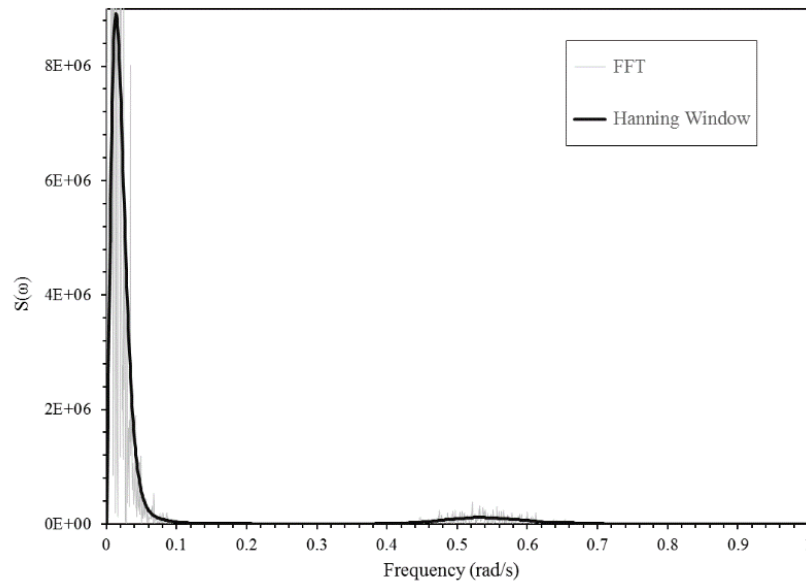
Figure 35 Typical spectrum of the mooring line top tension response for model A

From Figure 35 it is clear that the response main component is associated to the low-frequency region, due to the second-order effects mentioned before. There is also a significant portion of the response energy in the wave-frequency region, characterizing a broadband response spectrum.

6.3.1 Extreme Top Tension MPV Benchmark Value

Figure 36 presents the Gumbel distribution fitting for the 1000 maxima sample obtained from the same number of independent 3-h realizations. The good adjustment observed in Figure 36 indicates that the Gumbel distribution is an adequate choice for estimating the benchmark MPV.
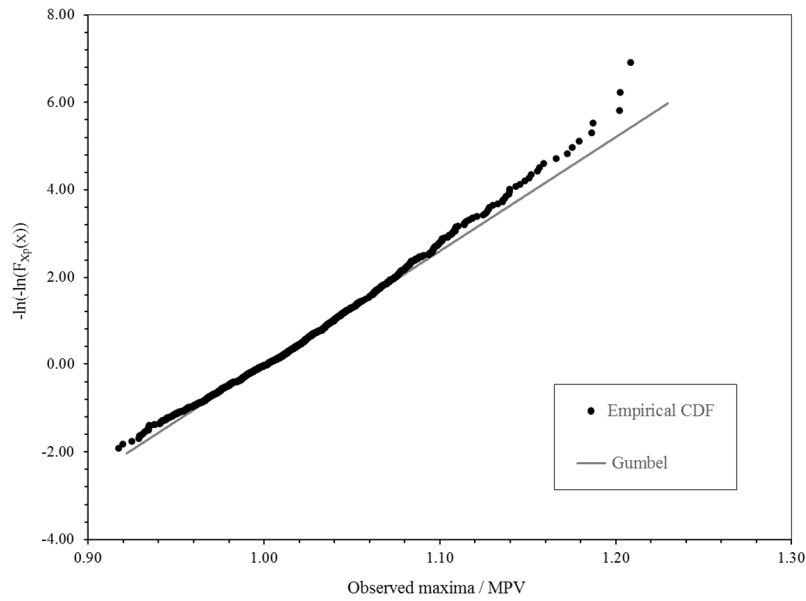
Figure 36 Benchmark value estimation through fitting the Gumbel distribution to the empirical CDF of the extremes sample

For case study A, the reference value encountered for the top tension most probable extreme value is $MPV_{REF} = 3644kN$. Again, it is worth mentioning that for this estimation the mooring line pre-tension was subtracted from each extreme value observed in the simulated time-series ($T_0 = 1562kN$).

For comparison purposes, the Fréchet distribution was also investigated for modelling the maxima sample. Figure 37 presents the Fréchet distribution fitting for the 1000 maxima sample. Although the fitting of Fréchet seems as accurate as the Gumbel one, the Gumbel distribution was chosen since the vast majority of probability distributions asymptotically converge to this model, as was mentioned before.
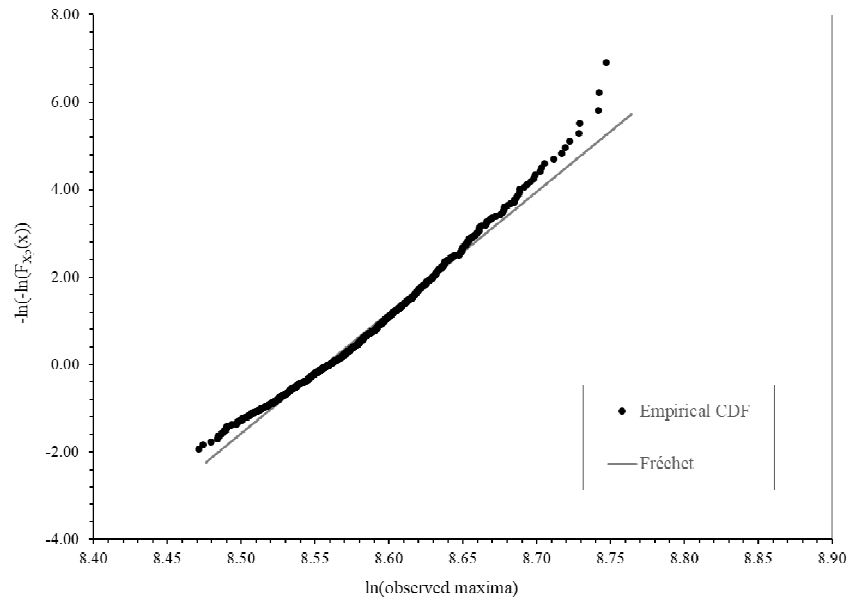
Figure 37 Benchmark value estimation through fitting the Fréchet distribution to the empirical CDF of the extremes sample

6.3.2 Statistical Analysis of the Top Tension Response

For each of the 200 simulations for each pair of $\{T_s, N_w\}$, the main statistical parameters (mean, standard deviation, skewness and kurtosis) of the time-series are calculated, as illustrated in Figure 32. These results are shown in more details in Annex C. In order to summarize these results, Table 2 shows the mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series for the pair $\{T_s = 54000s \ (15h), N_w = 2000\}$.

Table 2 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Mean μ_x	Standard Deviation σ_x	Skewness γ_x	Kurtosis κ_x
Mean	1959.71	487.11	0.37	3.15
Standard Deviation	14.49	12.15	0.07	0.15
CoV	0.01	0.02	0.19	0.05

From Table 2 it is clear that the process distribution is non-Gaussian, as it was expected. This is evident because the expected value of the process skewness is greater than zero and the expected value of the process kurtosis is greater than 3, which are the values for a normal distribution. This is more clear when the empirical PDF of a random time-series from the 200 realizations set is plotted. Figure 38 presents the comparison between the empirical PDF of four time-series generated with $T_s = 54000s$ and $N_w = 2000$ and the normal distribution fitted to the signals through the method of moments.
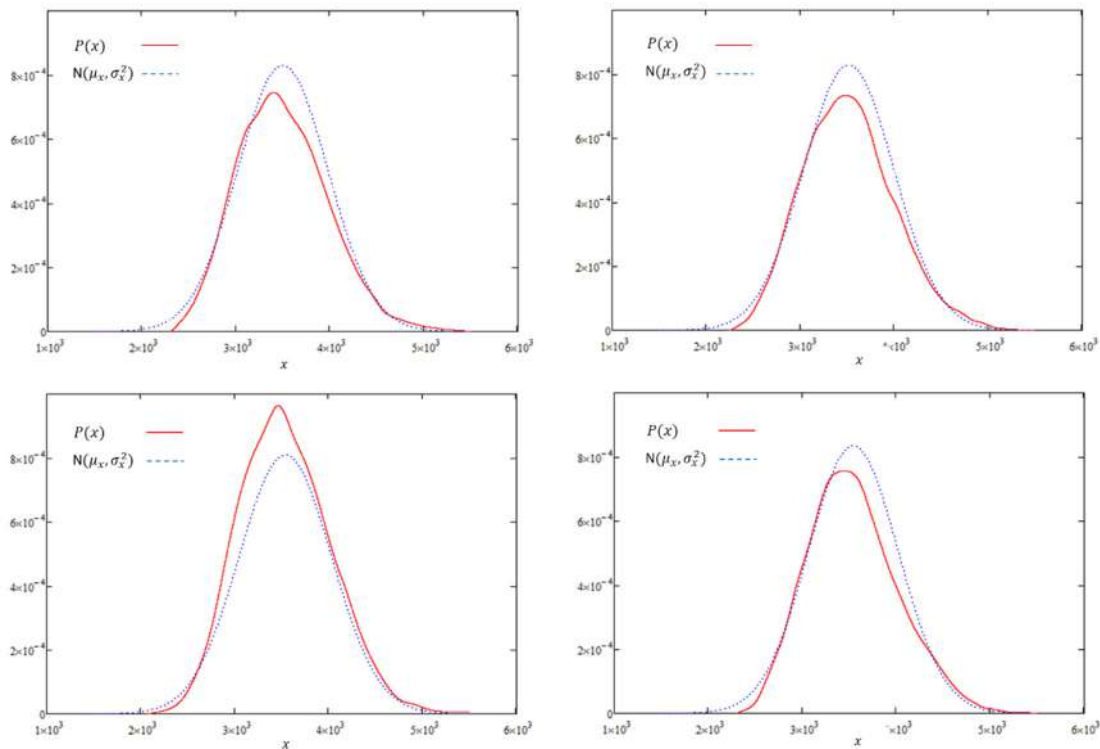
Figure 38 Comparison between the sampled process empirical PDFs $P(x)$ (solid line) and the adjusted normal distributions $N(\mu_x, \sigma_x^2)$ (dashed lines) for case study A

Figures 39, 40 and 41 present three random realizations from the set of 200 samples for the pair $\{T_s = 10800s\ (3h), N_w = 2000\}$. The line pre-tension is subtracted from these time-series.
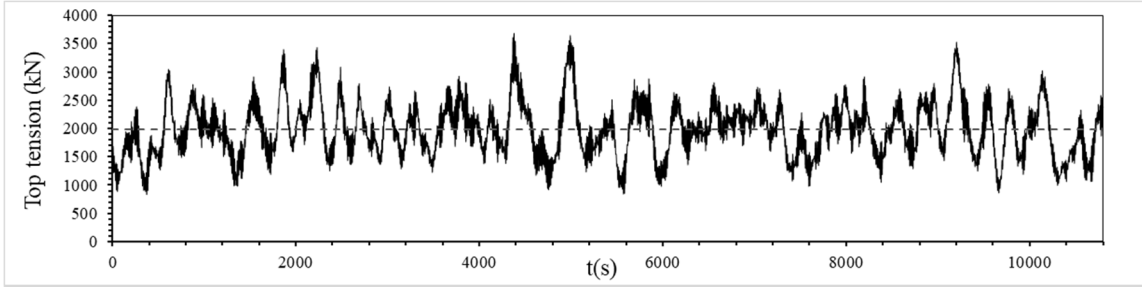
Figure 39 Model A: Top tension time-series #1 ($T_s = 10800s$ and $N_w = 2000$)
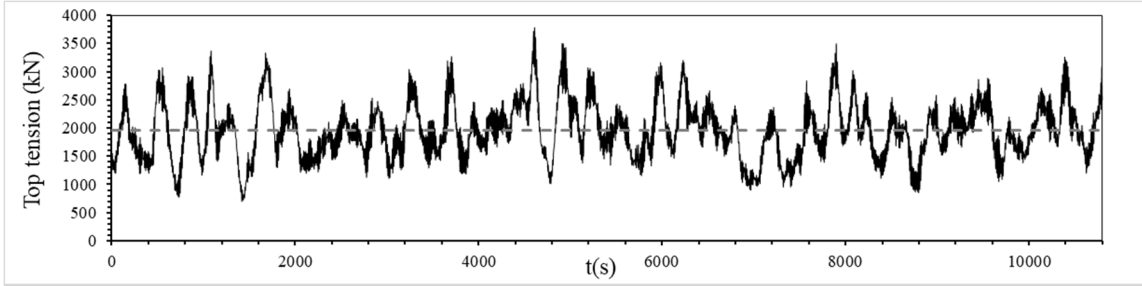
Figure 40 Model A: Top tension time-series #50 ($T_s = 10800s$ and $N_w = 2000$)
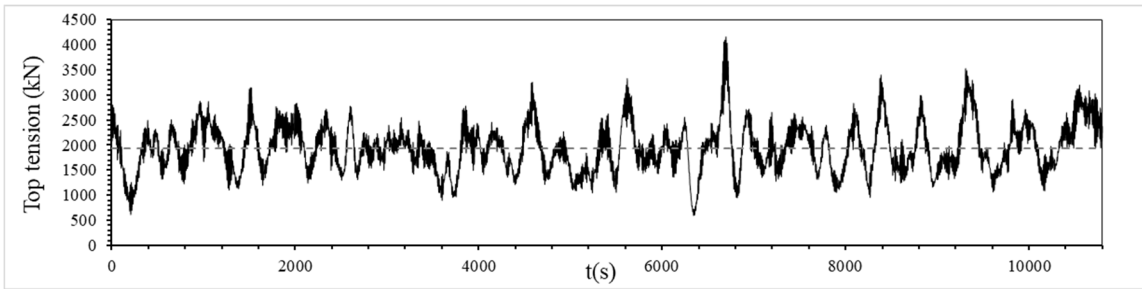
Figure 41 Model A: Top tension time-series #150 ($T_s = 10800s$ and $N_w = 2000$)

6.3.3 Statistical Analysis of the Top Tension Response Peaks

In this study, three probability models are investigated for the peaks of the line top tension time-series (W3P, WT and SGLD), following the procedure explained in Section 4.5.2.

As in the previous section, for each of the 200 simulations and for each pair of $\{T_s, N_w\}$, the main statistical parameters (mean, standard deviation, skewness and kurtosis) are calculated but this time for the global peaks sample of the time-series. These results are shown in more details in Annex C. In order to summarize these results, Table 3 shows the mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series global peaks for the pair

$\{T_s = 54000s\ (15h), N_w = 2000\}$. It is also shown the mean, standard deviation and coefficient of variation of the 200 obtained values for the peaks frequency and the empirical correlation between two consecutive peaks.

Table 3 Statistical parameters of the global peaks of the global 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Mean μ_p	Standard Deviation σ_p	Skewness γ_p	Kurtosis κ_p	Peaks Frequency ν_p	Correlation ρ
Mean	2269.04	297.48	2.04	8.35	0.03	0.25
Standard Deviation	16.51	9.55	0.13	0.90	0.00	0.03
CoV	0.01	0.03	0.06	0.11	0.04	0.13

For a random sample of the set $\{T_s = 54000s\ (15h), N_w = 2000\}$, the comparison between the empirical CDF of the global peaks sample and the fitted probability distributions are given in the Figures 42, 43 and 44.
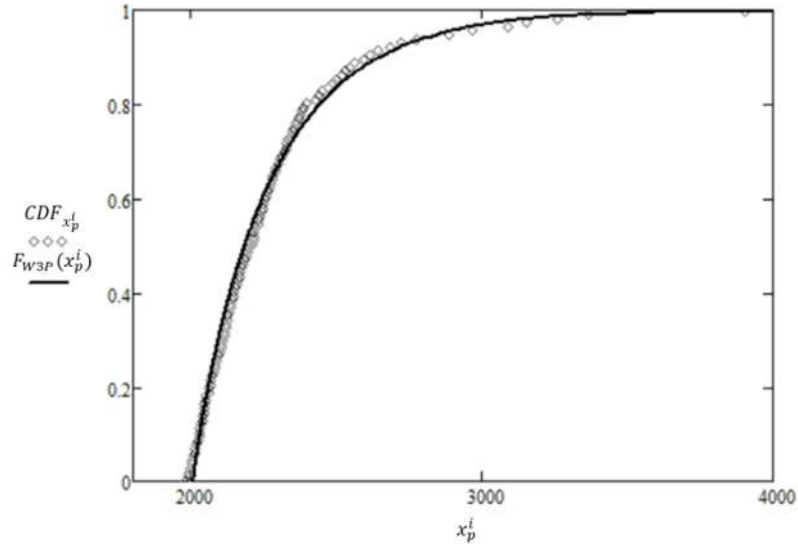
Figure 42 Comparison between empirical CDF of the peaks and fitted 3-parameter Weibull probability model for a random time-series of model A
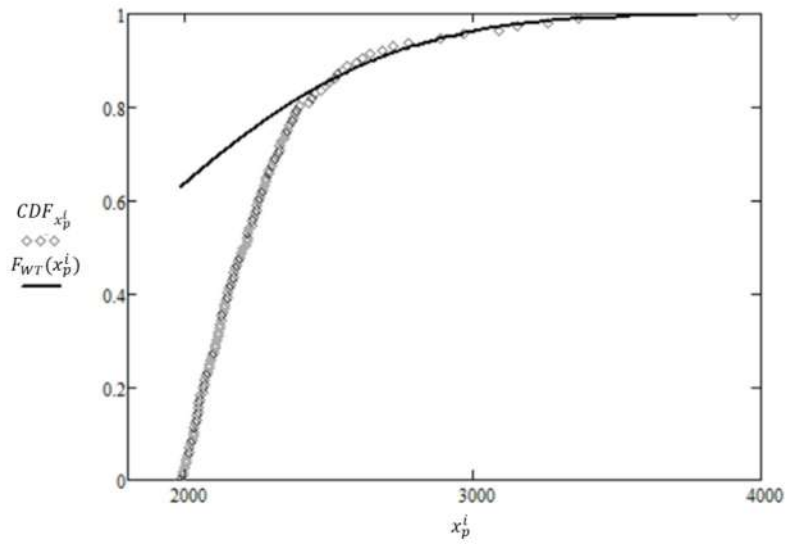
Figure 43 Comparison between empirical CDF of the peaks and fitted Weibull-tail probability model for a random time-series of model A
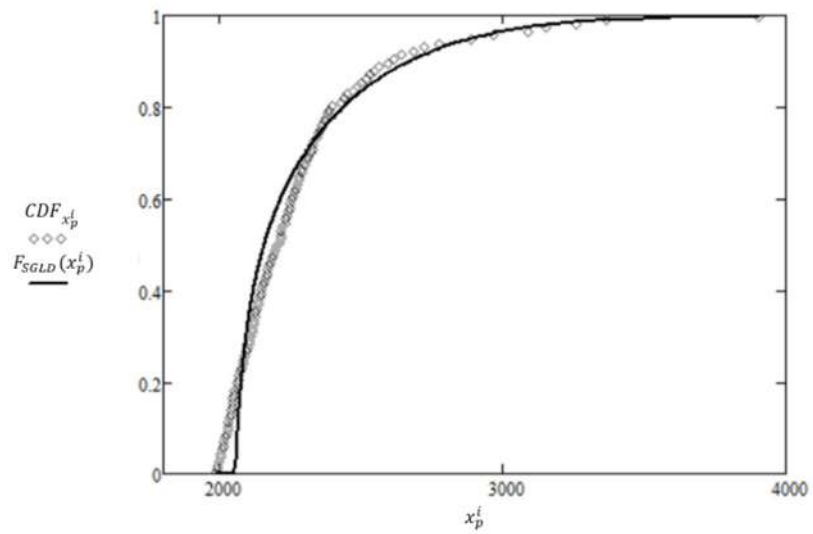
Figure 44 Comparison between empirical CDF of the peaks and fitted SGLD probability model for a random time-series of model A

The probability models W3P, WT and SGLD are also analyzed when two consecutive peaks are considered correlated (see Section 4.2). Besides these, the ACER method is also investigated, for $k = 1, 2, 3$. Results obtained from the ACER method will be presented in the next section.

6.3.4 General Results

Each one of the methods described in Section 4.5 were investigated in order to determine the extreme most probable value of the line top tension response. As these methods are based on a single realization of the process, the bias and coefficient of variation for each method and pair of $\{T_s, N_w\}$ were estimated in order to seek the method with the best performance.

Figure 45 to Figure 47 show, as function of the simulation length, the biases estimates for the top tension 3-h MPV, i.e., the ratio between the average MPV from the 200 realizations sample and the benchmark value for each peaks probability model, pair of $\{T_s, N_w\}$ and the consideration or not of the correlation between consecutive peaks. Figure 48 to Figure 50 show the corresponding CoVs of the MPVs.
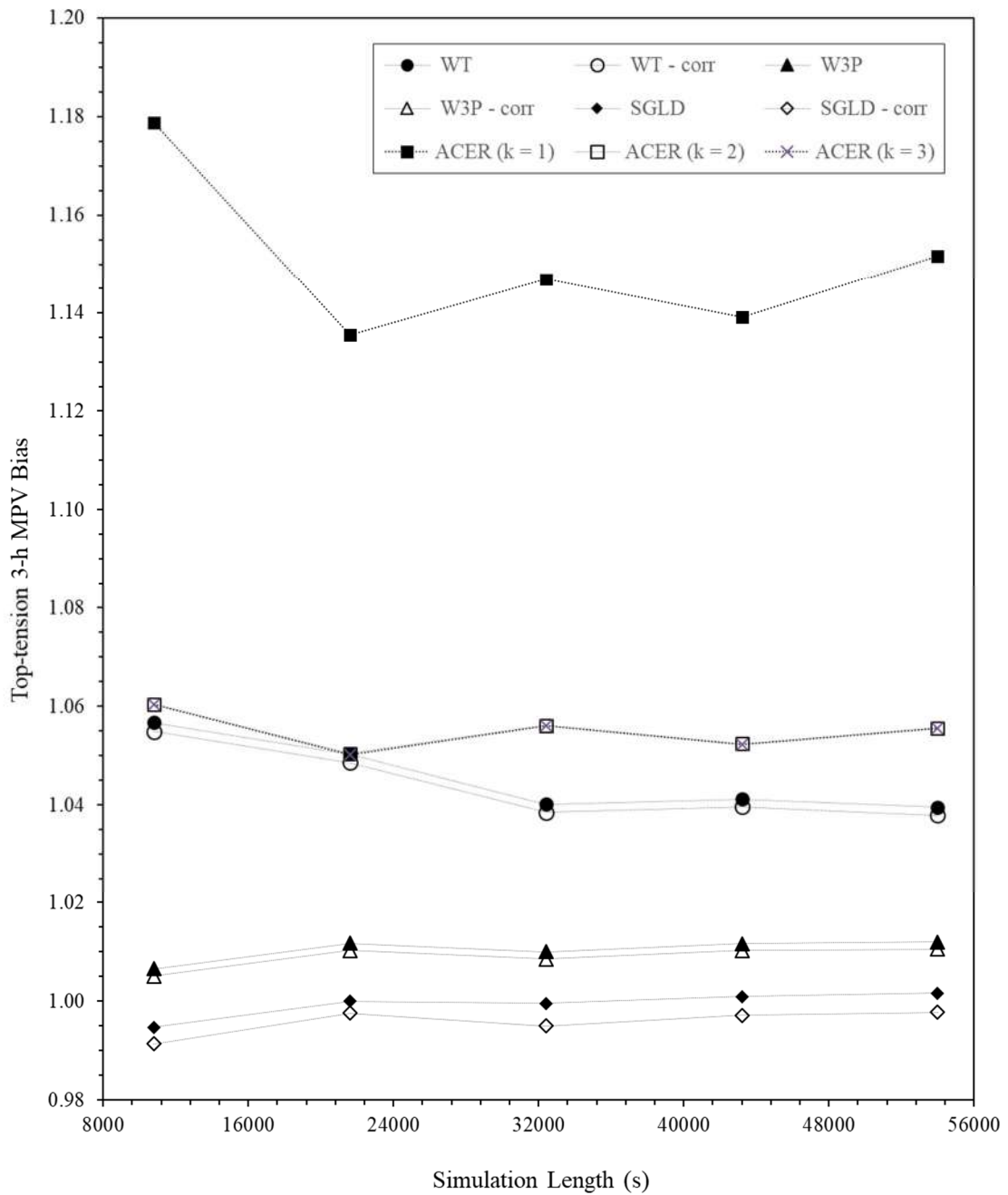
Figure 45 Bias estimates for the top tension 3-h MPV – Case study A ($N_w = 300$).
('corr' indicates the consideration of peaks correlation)
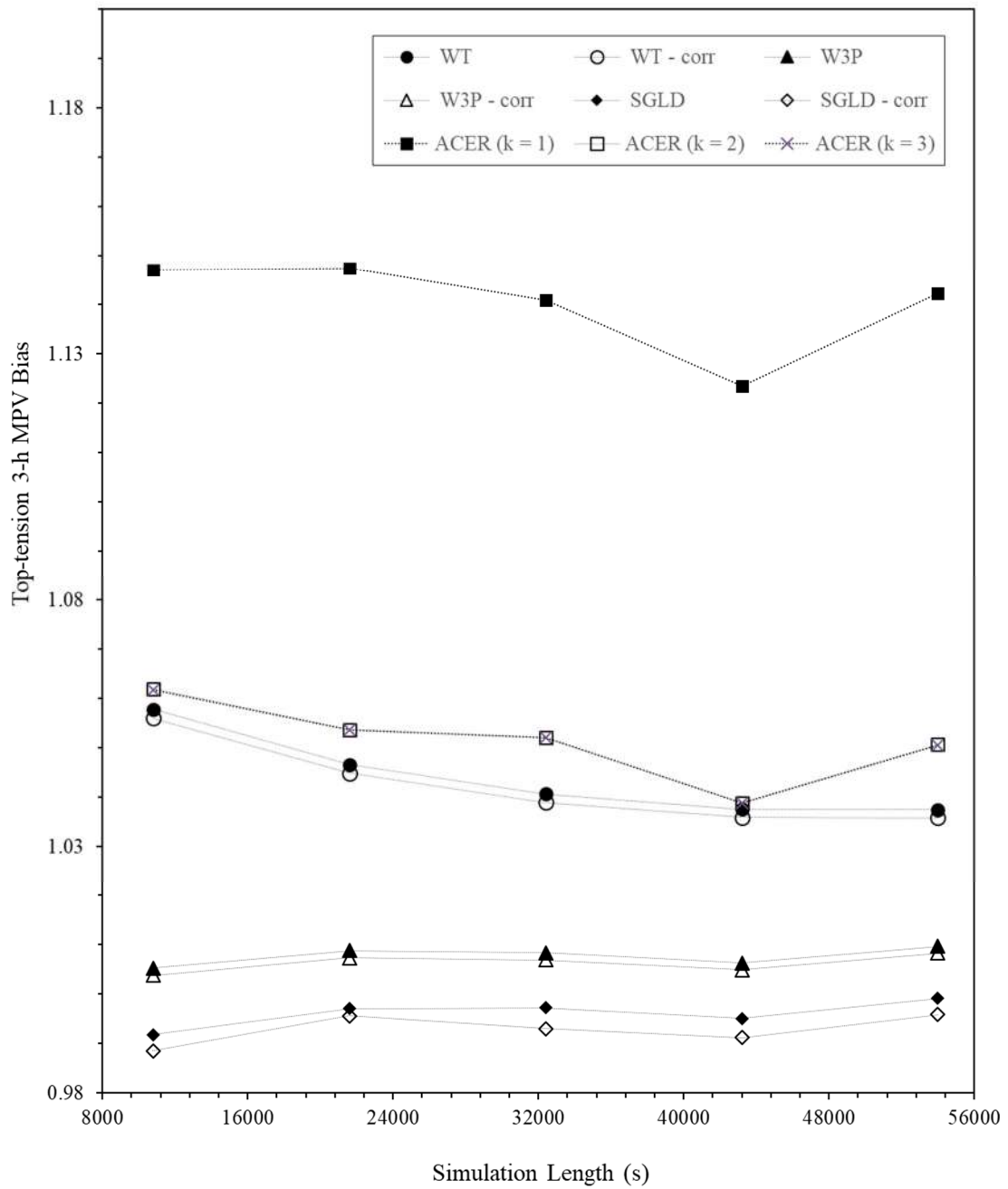
Figure 46 Bias estimates for the top tension 3-h MPV – Case study A ($N_w = 1000$).
('corr' indicates the consideration of peaks correlation)
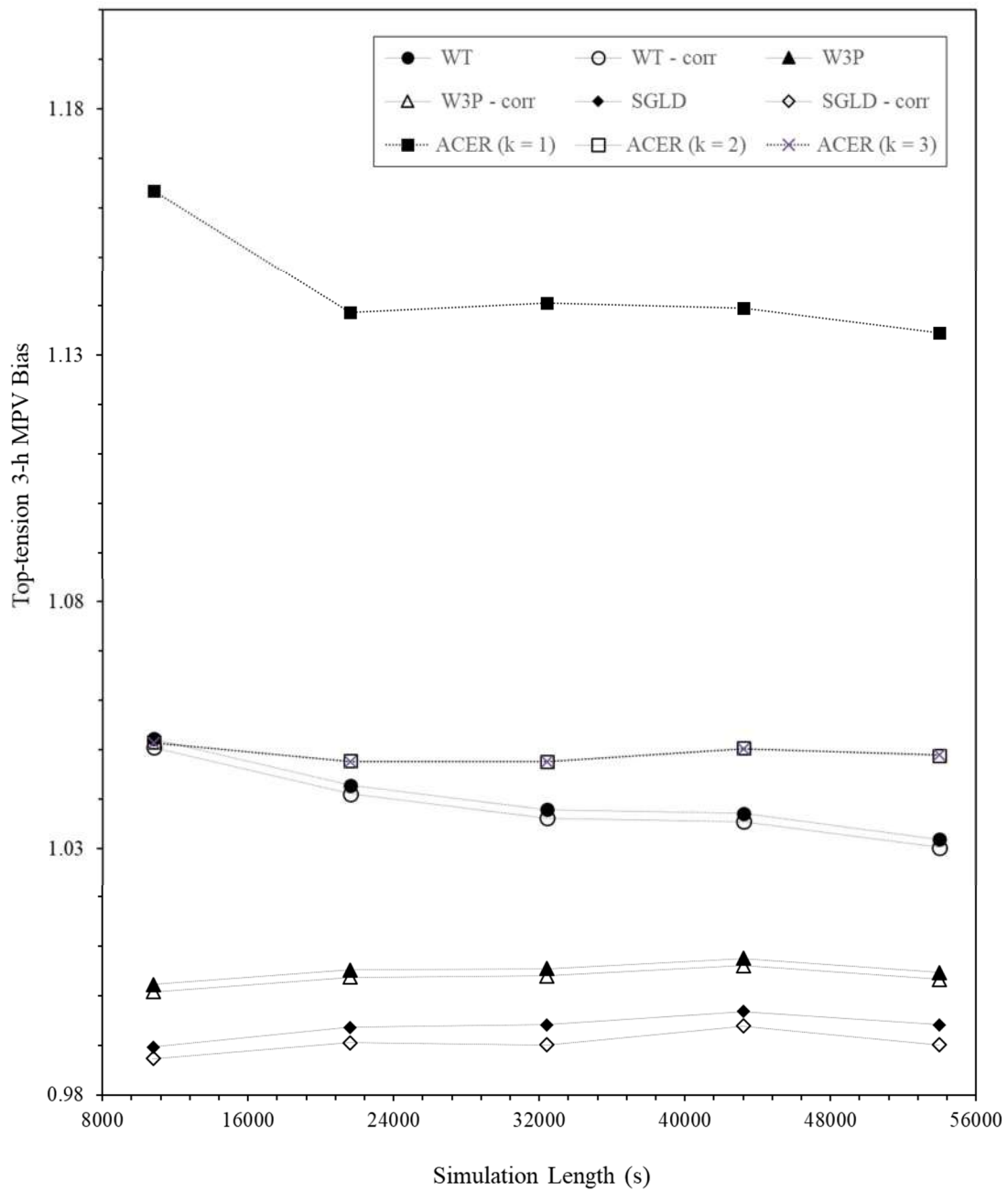
Figure 47 Bias estimates for the top tension 3-h MPV – Case study A ($N_w = 2000$). ('corr' indicates the consideration of peaks correlation)
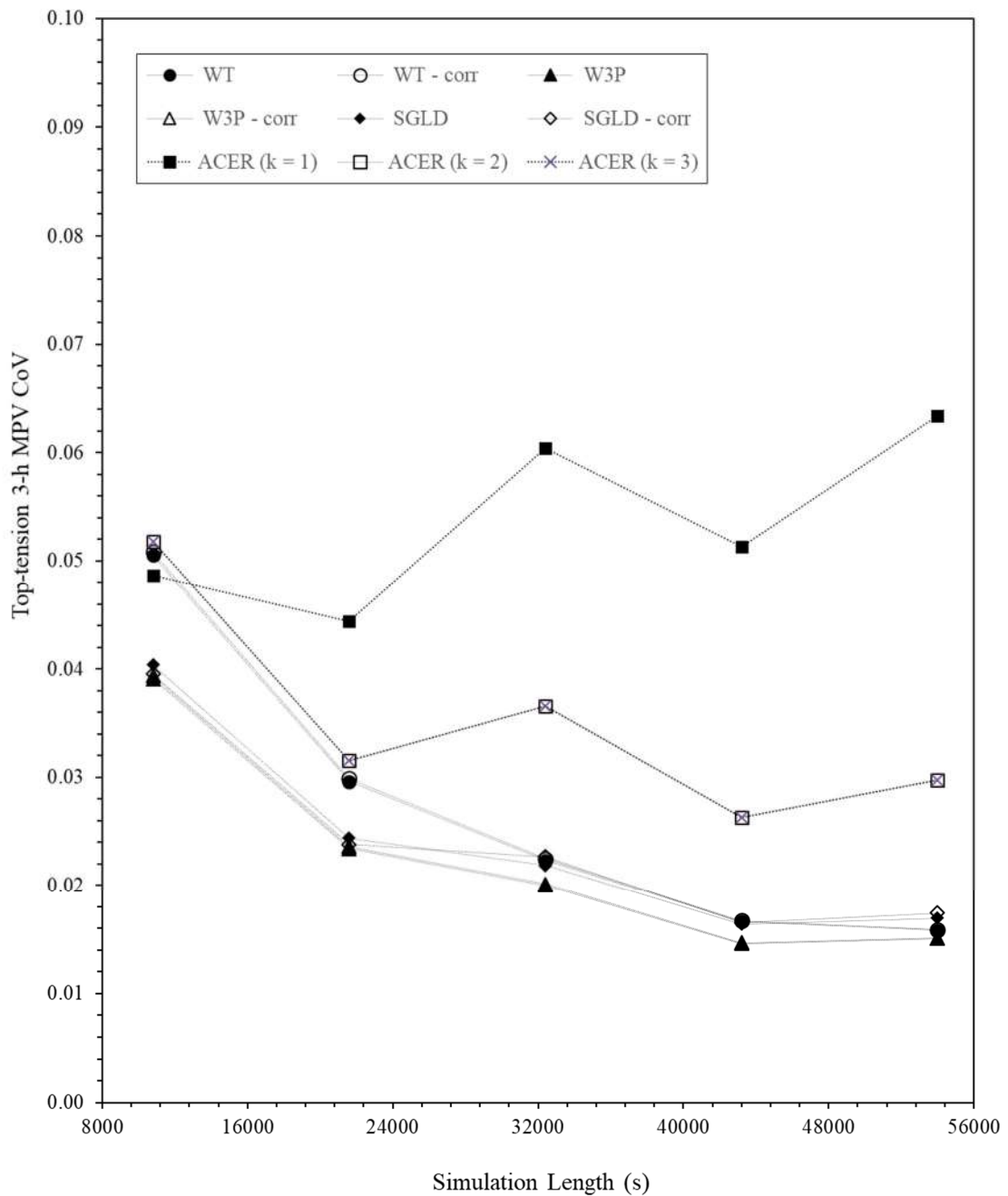
Figure 48 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 300$).
('corr' indicates the consideration of peaks correlation)
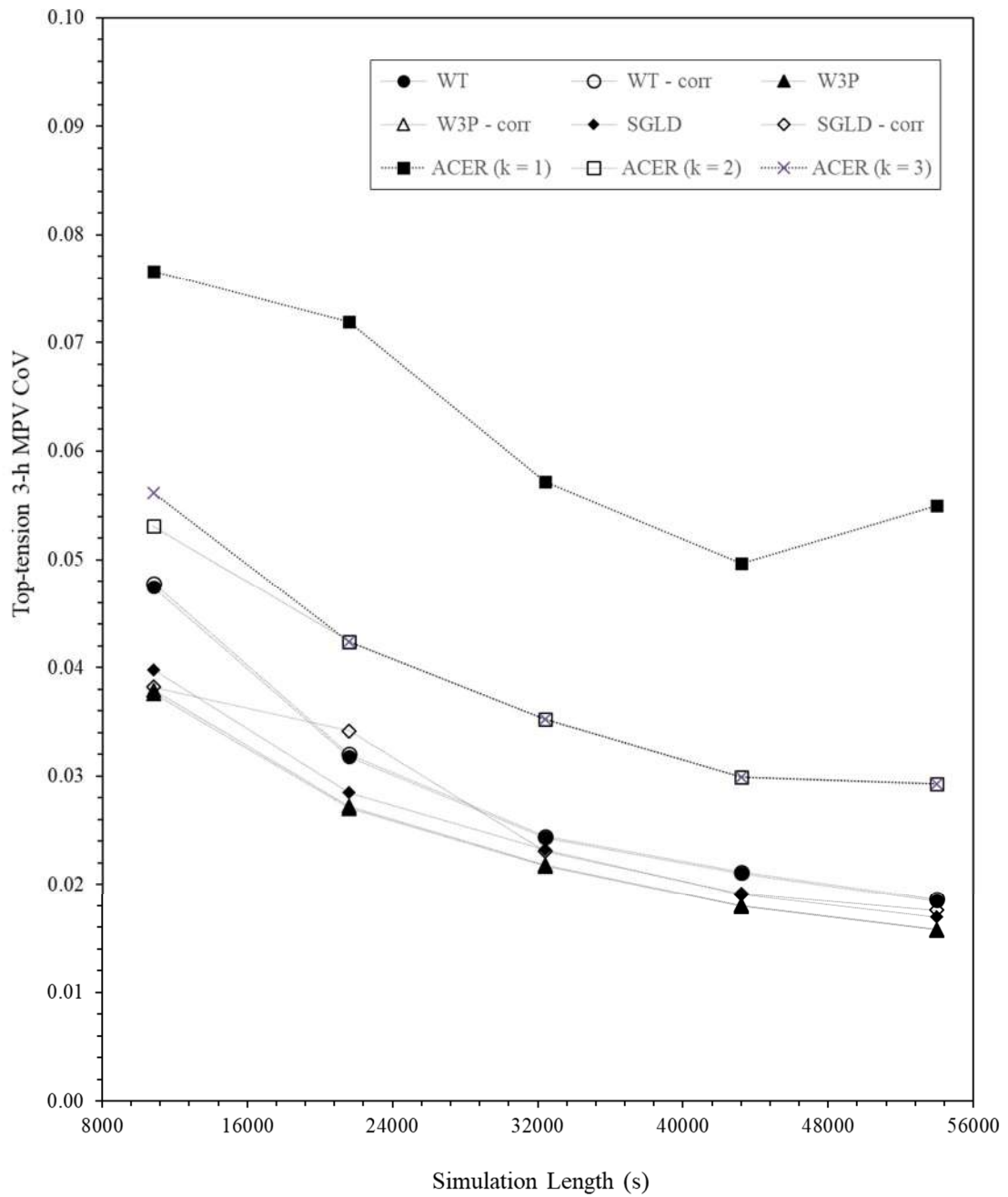
Figure 49 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 1000$).
('corr' indicates the consideration of peaks correlation)
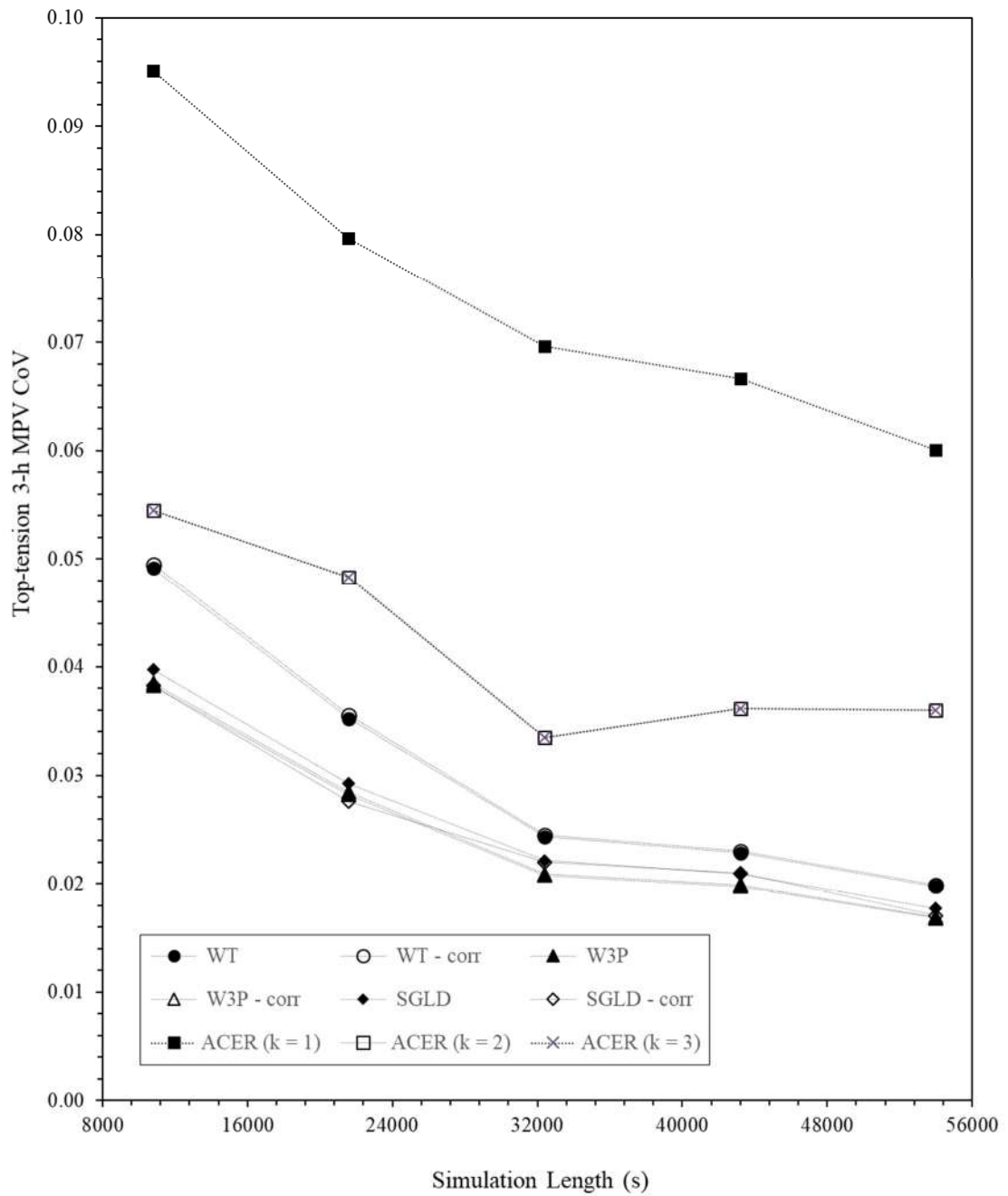
Figure 50 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 2000$).
('corr' indicates the consideration of peaks correlation)

From Figure 45 to Figure 50 it is possible to notice that:

i. The level of discretization of the wave spectrum seems to be of low importance for the extreme response estimation since it has little impact on the biases and coefficient of variation of the extreme MPVs. However, higher levels of discretization seems to result in more stabilized values for the ACER method;

ii. The expected value of the extreme MPV is not highly dependent on the simulation length, for all methods considered;

iii. However, shorter numerical simulations are associated to higher CoVs of the short-term extreme MPV. As the simulation length increases, the number of peaks extracted from the series increases as well, leading to less variability in the statistical estimates. The CoVs are very similar among the probability distribution methods investigated and they seem to be smaller than those obtained through the ACER method. 6-h or longer simulations lead to CoVs lower than 4% for the probability distribution model methods;

iv. MPV estimates are often overestimated. In other words, the short-term extreme MPVs calculated through fitting a probability model to the peaks sample extracted from a single time-series seem to be often greater than the extreme response reference value. This overestimation is higher for the WT model, reaching a 5% bias;

v. Short-term extreme MPV estimates taking into consideration two subsequent peaks correlation tend to be practically the same as MPV estimates that do not consider it. This indicates that for this case the correlation between subsequent global peaks does not take an important role in the extremes statistics. In accordance with this, the expected value of the correlation coefficient between two consecutive peaks is $\rho = 0.3$, as shown in Table 3;

vi. The SGLD-based procedure and the W3P approach seem to predict practically unbiased estimators;

vii. Comparatively to others, the ACER method seems to predict highly overestimated extreme MPVs when $k = 1$ (or when the peaks are considered independent of each other). The overestimation drastically decreases when correlation

between two local peaks is considered, $k = 2$. However, for $k = 3$ the extremes MPVs are no longer altered. This shows that the correlation between three consecutive peaks is low and do not influence the extremes estimation. It is important to notice that, for the ACER method, the consideration of correlation between two-consecutive peaks ($k = 2$) has a larger role on the extremes estimation as a result of the fact that the peaks sample in this case comprises both local and global peaks. For this case, the empirical correlation coefficient between two consecutive local peaks is in the $\rho = 0.8$ level.

6.4 Study Case B – FPSO in Shallow Water

In this case study, a chain-wire rope-chain mooring line connected to a turret-moored FPSO in 180m water depth offshore Brazil under a 100-yr environmental condition is investigated. The environmental loading condition and mooring line orientations are presented in Figure 51. The mooring line response corresponds to a short-term condition in the environmental 100-yr contour with $H_s = 6.53$m and $T_p = 11.63$s (JONSWAP Spectrum), with associated surface current velocity = 0.80m/s and wind velocity = 18.96m/s acting as indicated in Figure 52.
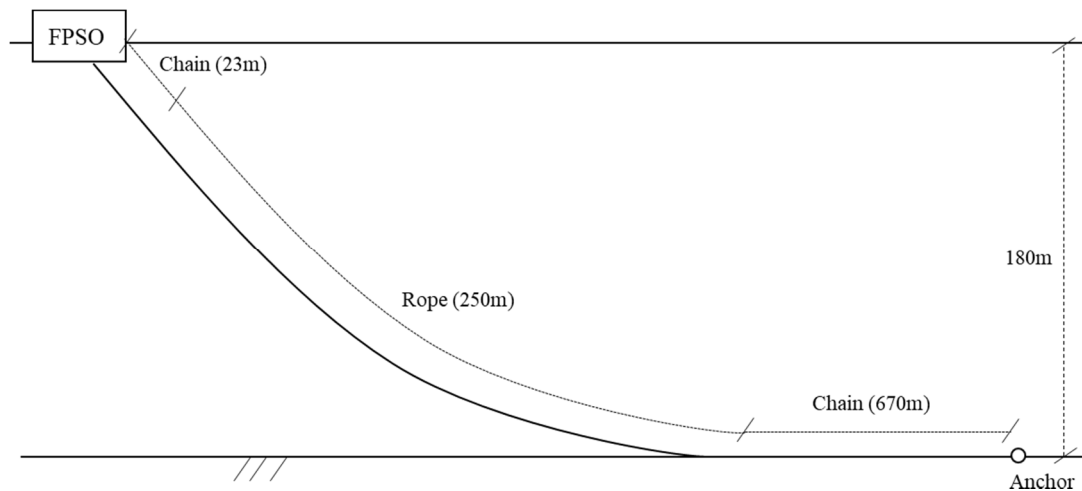
Figure 51 Mooring line catenary configuration for case study B
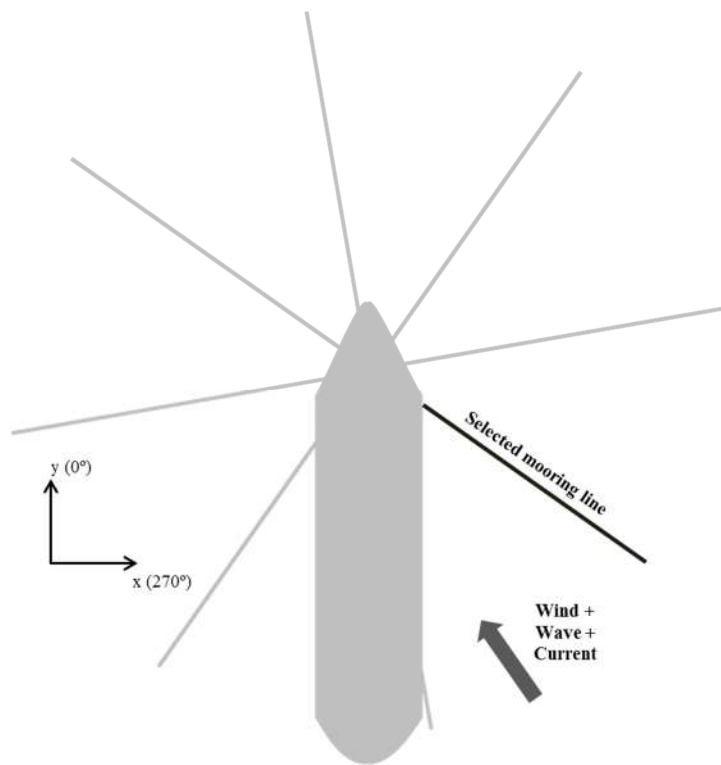
Figure 52 Environmental condition and selected mooring line orientations for case study B

The typical spectral density of the selected mooring line top tension response is shown in Figure 53. This figure shows that the role of first-order effects in this scenario seems to be negligible since almost the total response energy is in the low-frequency region, associated with the second-order effects. This characterizes a narrow band spectrum.
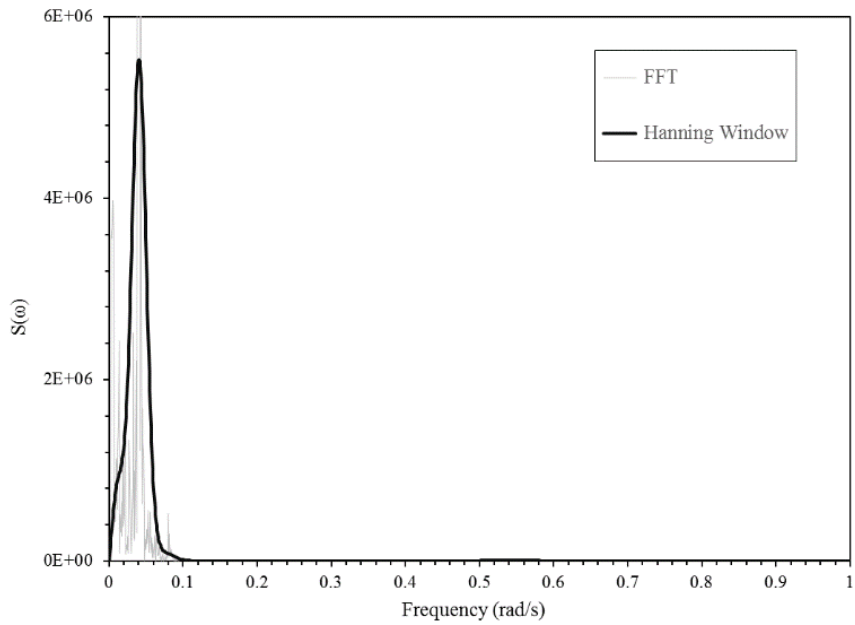
Figure 53 Typical spectrum of the mooring line top tension response for model B

6.4.1 Extreme Top Tension MPV Benchmark Value

Figure 54 presents the Gumbel distribution fitting for the 1000 maxima sample obtained from the same number of independent 3-h realizations. The good fitting observed in Figure 54 indicates that the Gumbel distribution is an adequate choice for estimating the benchmark MPV.
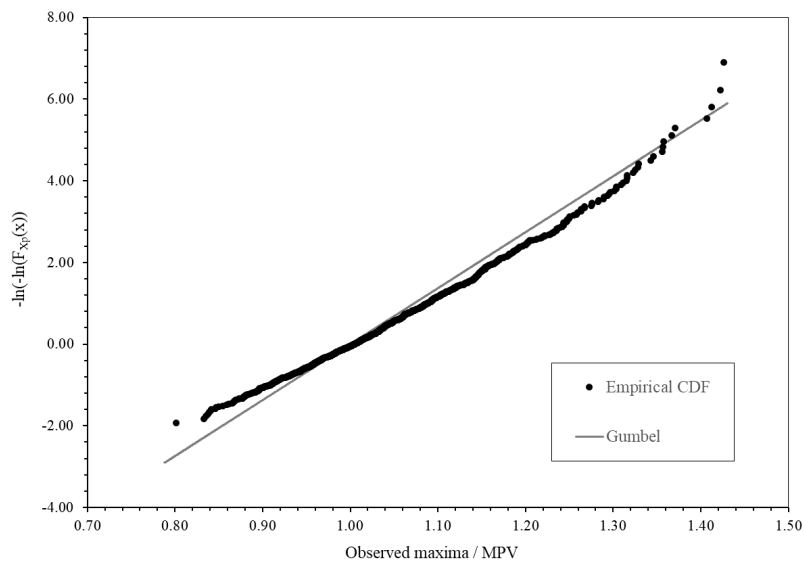
Figure 54 Benchmark value estimation through fitting the Gumbel distribution to the empirical CDF of the extremes sample

For case study B, the reference value encountered for the top tension most probable extreme value is $MPV_{REF} = 1862kN$. The line pre-tension ($T_0 = 1563kN$) was subtracted of the top tension time-series.

As in the other case study previously described, the Fréchet distribution was also investigated for the modelling of the maxima sample for sake of comparison. Figure 55 presents the Fréchet distribution fitting for the 1000 maxima sample. Differently from case study A, here it is clear that the Gumbel distribution fitting to the maxima sample is more accurate than the Fréchet one. Hence the reference value for the MPV is the one estimated through Gumbel model.
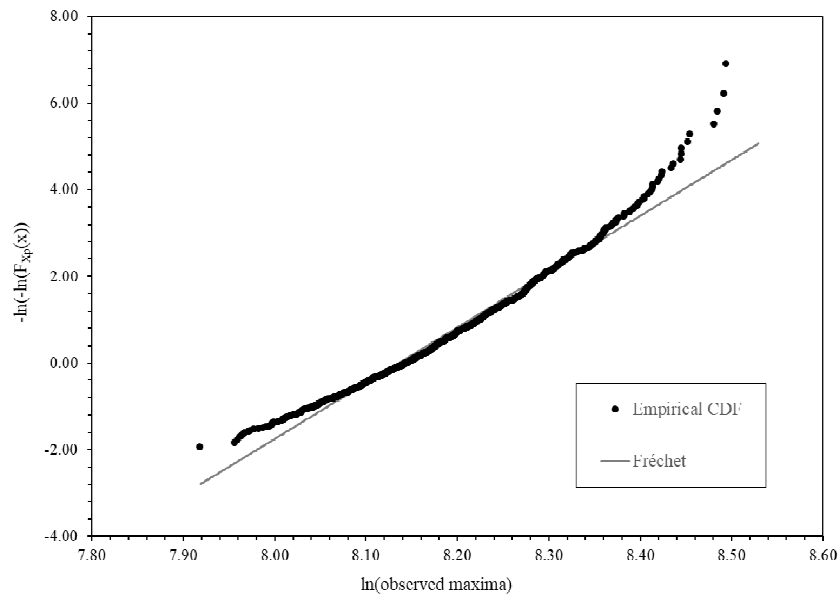
Figure 55 Benchmark value estimation through fitting the Fréchet distribution to the empirical CDF of the extremes sample

6.4.2 Statistical Analysis of the Top Tension Response

In a similar manner as showed in Section 6.3.2, for each of the 200 simulations for each pair of $\{T_s, N_w\}$, the main statistical parameters (mean, standard deviation, skewness and kurtosis) of the time-series were calculated. These results are also shown in more details in Annex C. In order to summarize these results, Table 4 shows the mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series for the pair $\{T_s = 54000s\ (15h), N_w = 2000\}$.

Table 4 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Mean μ_x	Standard Deviation σ_x	Skewness γ_x	Kurtosis κ_x
Mean	377.03	445.59	0.64	3.52
Standard Deviation	6.27	29.45	0.09	0.48
CoV	0.02	0.07	0.14	0.13

From Table 4 it is clear that once more the process distribution is non-Gaussian, as it was expected. This is more evident in this case study than in the previous because the skewness and kurtosis values are even greater than the values for the normal distribution. Figure 56 presents the comparison between the empirical PDFs of four time-series generated with $T_s = 54000s$ and $N_w = 2000$ and the normal distribution fitted to the signals through the method of moments.
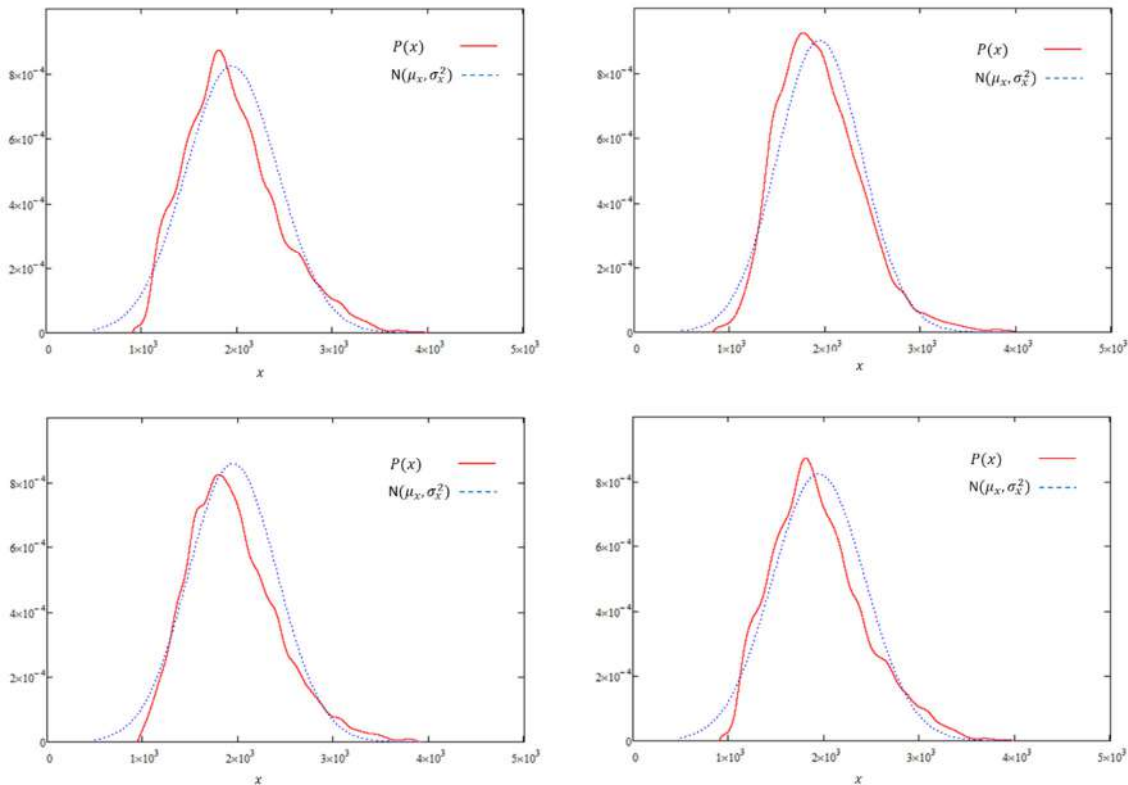
Figure 56 Comparison between the sampled process empirical PDFs $P(x)$ (solid line) and the adjusted normal distributions $N(\mu_x, \sigma_x^2)$ (dashed line) for case study B

Figures 57, 58 and 59 present three random realizations from the set of 200 samples for the pair $\{T_s = 10800s\ (3h), N_w = 2000\}$. The line pre-tension is subtracted from the time-series shown in the figures.
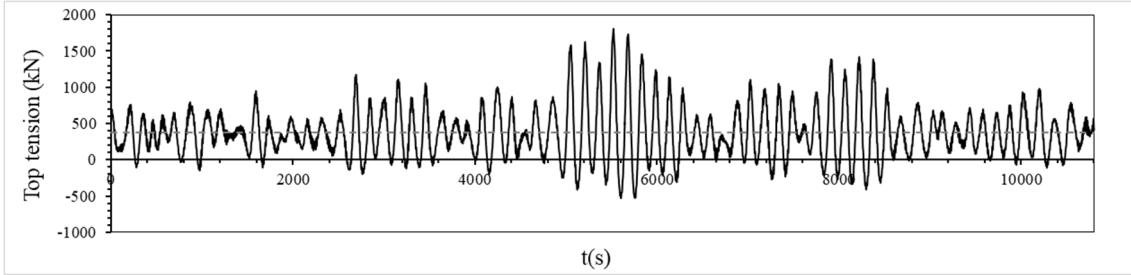
Figure 57 Model B: Top tension time-series #1 ($T_s = 10800s$ and $N_w = 2000$)
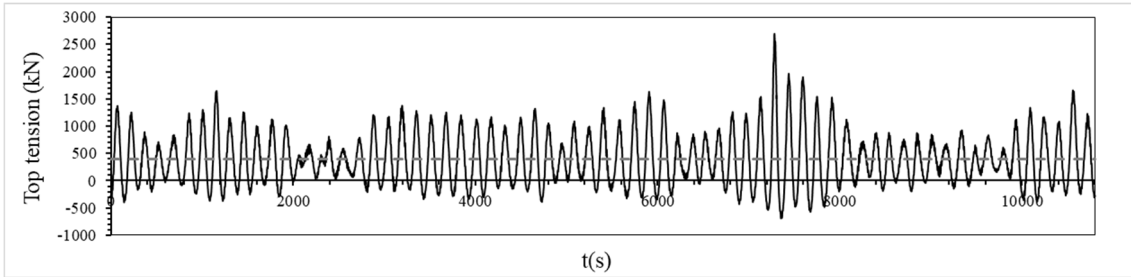
Figure 58 Model B: Top tension time-series #50 ($T_s = 10800s$ and $N_w = 2000$)
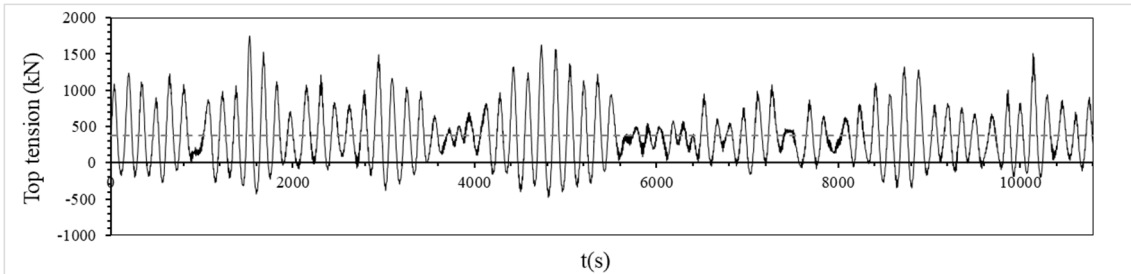
Figure 59 Model B: Top tension time-series #150 ($T_s = 10800s$ and $N_w = 2000$)

6.4.3   Statistical Analysis of the Top Tension Response Peaks

As in the previous section, for each of the 200 simulations for each pair of $\{T_s, N_w\}$, the main statistical parameters (mean, standard deviation, skewness and kurtosis) are calculated but this time for the global peaks sample of the top tension time-series. These results are shown in more details in Annex C. In order to summarize these results, Table 5 shows the mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series peaks for the pair

$\{T_s = 54000s\ (15h), N_w = 2000\}$. It is also shown the mean, standard deviation and coefficient of variation of the 200 obtained values for the global peaks frequency and the empirical correlation between two consecutive peaks.

Table 5 Statistical parameters of the global peaks of the 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Mean μ_p	Standard Deviation σ_p	Skewness γ_p	Kurtosis κ_p	Peaks Frequency ν_p	Correlation ρ
Mean	863.20	428.69	0.88	3.43	0.01	0.73
Standard Deviation	56.16	36.04	0.24	0.92	0.00	0.05
CoV	0.07	0.08	0.27	0.27	0.05	0.06

The higher average correlation between global peaks found in the table above, in comparison to the value encountered in case study A, evidences the narrow characteristic of the process band. As mentioned before, narrow band processes are associated to higher correlations between the process maxima.

The comparison between the empirical CDFs of the peaks sample and the fitted probability distributions investigated in this work are shown in the Figures 60 , 61 and 62.
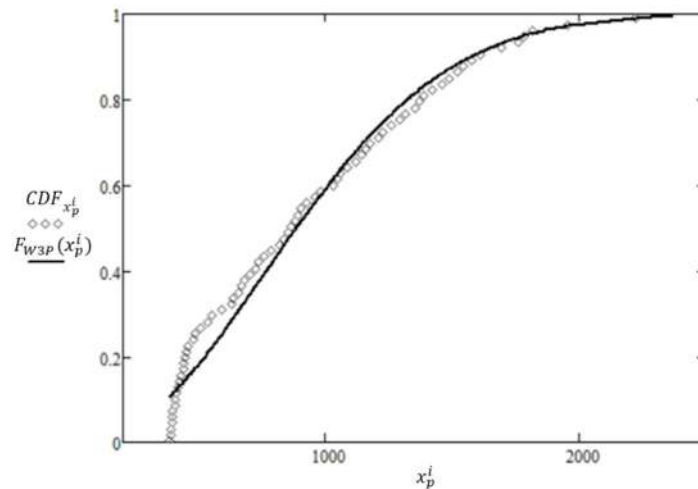
Figure 60 Comparison between empirical CDF of the peaks and fitted 3-parameter Weibull probability model for a random time-series of model B
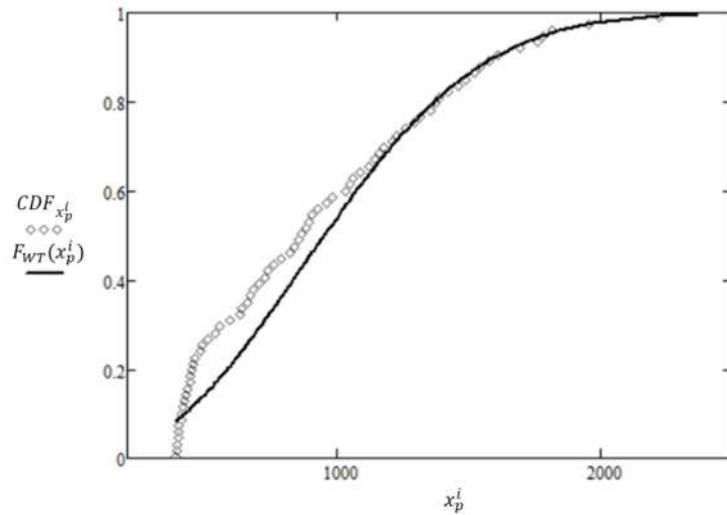
Figure 61 Comparison between empirical CDF of the peaks and fitted Weibull-tail probability model for a random time-series of model B
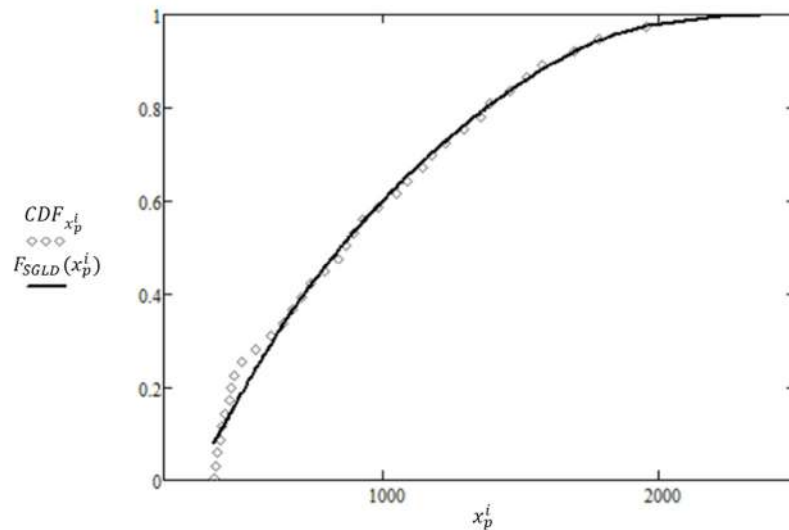
Figure 62 Comparison between empirical CDF of the peaks and fitted SGLD probability model for a random time-series of model B

6.4.4 General Results

Figure 63 to Figure 65 show, as function of the simulation length, the biases estimates for the top tension 3-h MPV, i.e., the ratio between the average MPV from the 200 realizations sample and the benchmark value for each peaks probability model, pair of $\{T_s, N_w\}$ and the consideration or not of the correlation between consecutive peaks. Figure 66 to Figure 68 show the corresponding CoVs of the MPVs.
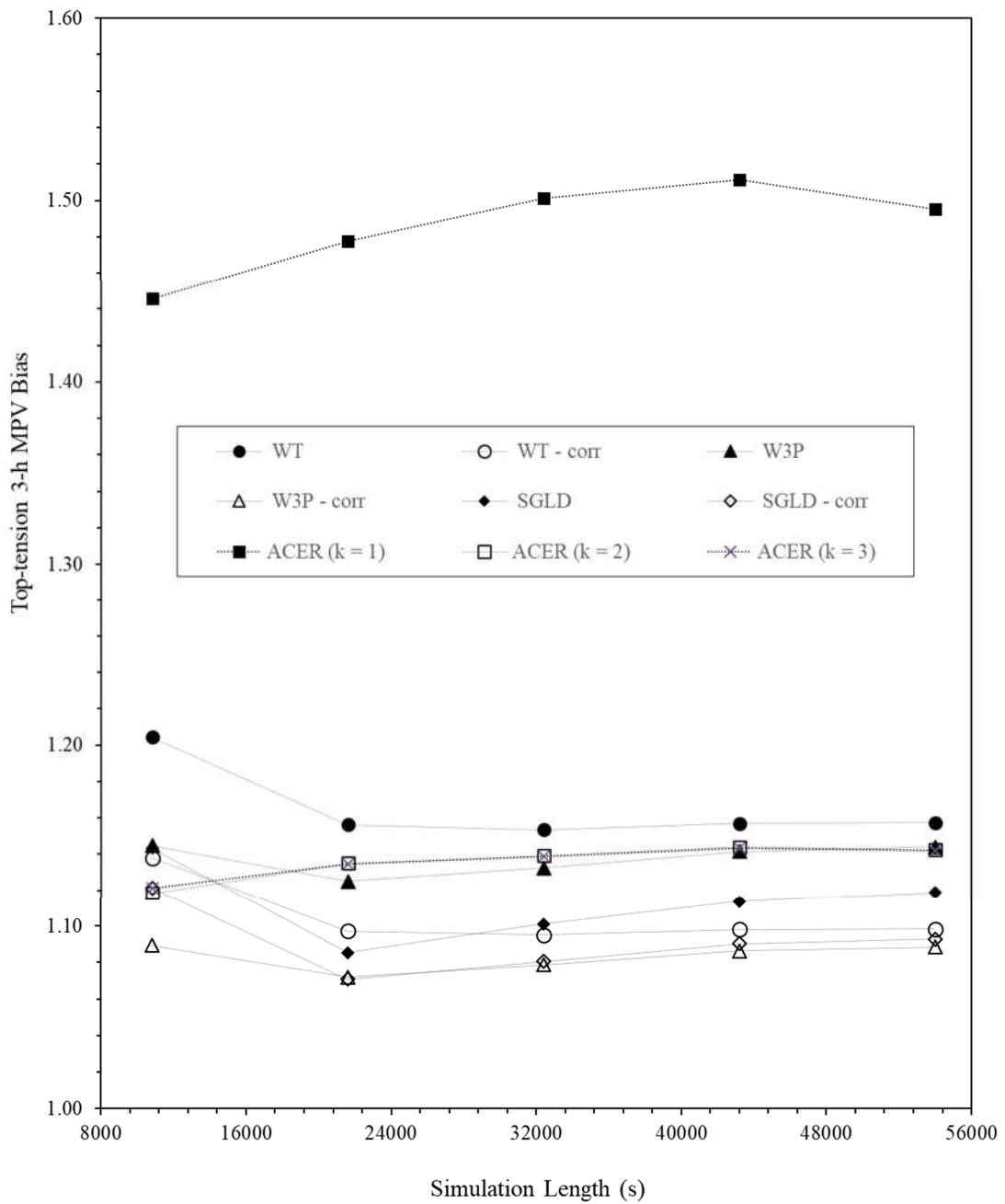
Figure 63 Bias estimates for the top tension 3-h MPV – Case study B ($N_w = 300$).
('corr' indicates the consideration of peaks correlation)
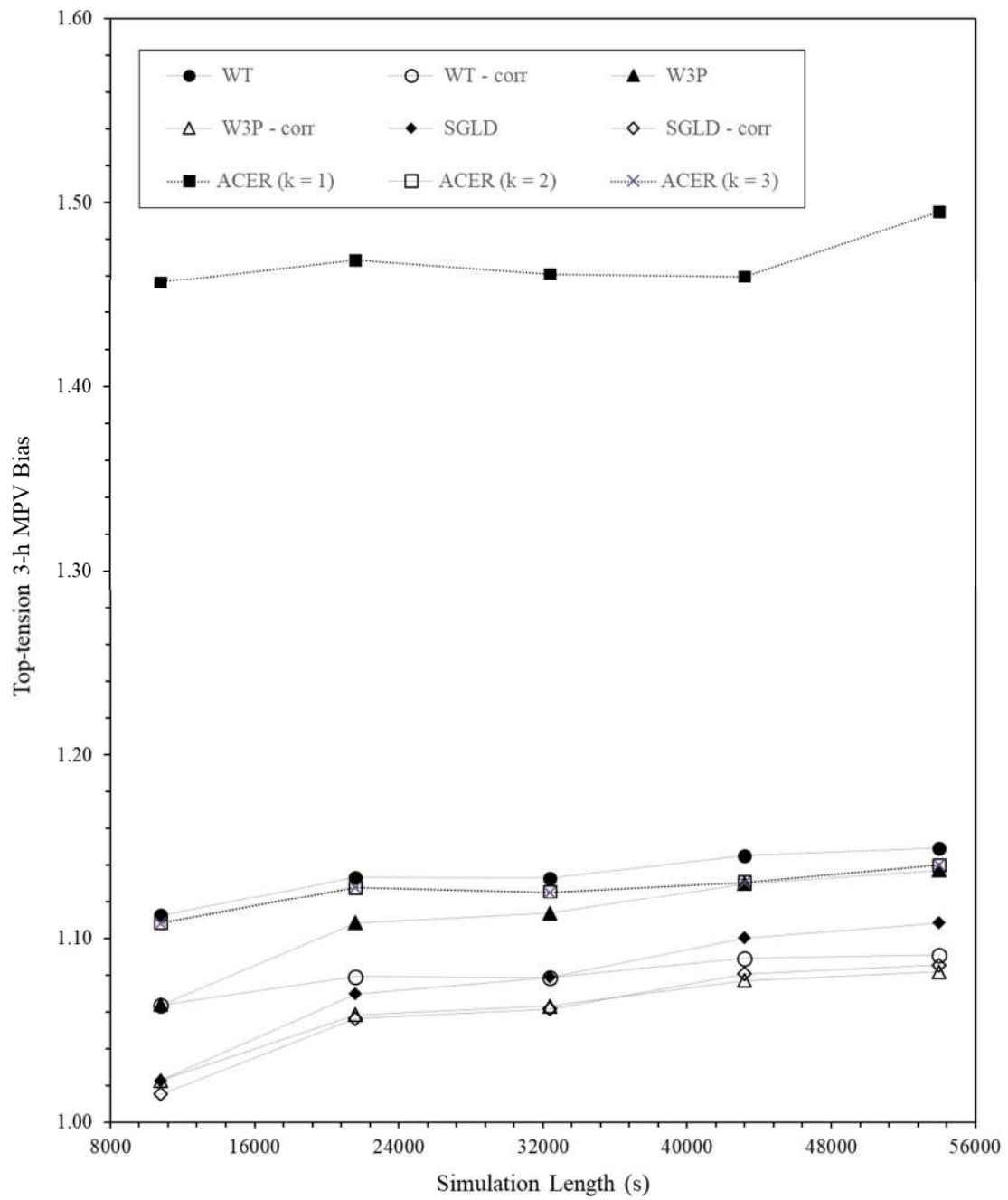
Figure 64 Bias estimates for the top tension 3-h MPV – Case study B ($N_w = 1000$). ('corr' indicates the consideration of peaks correlation)
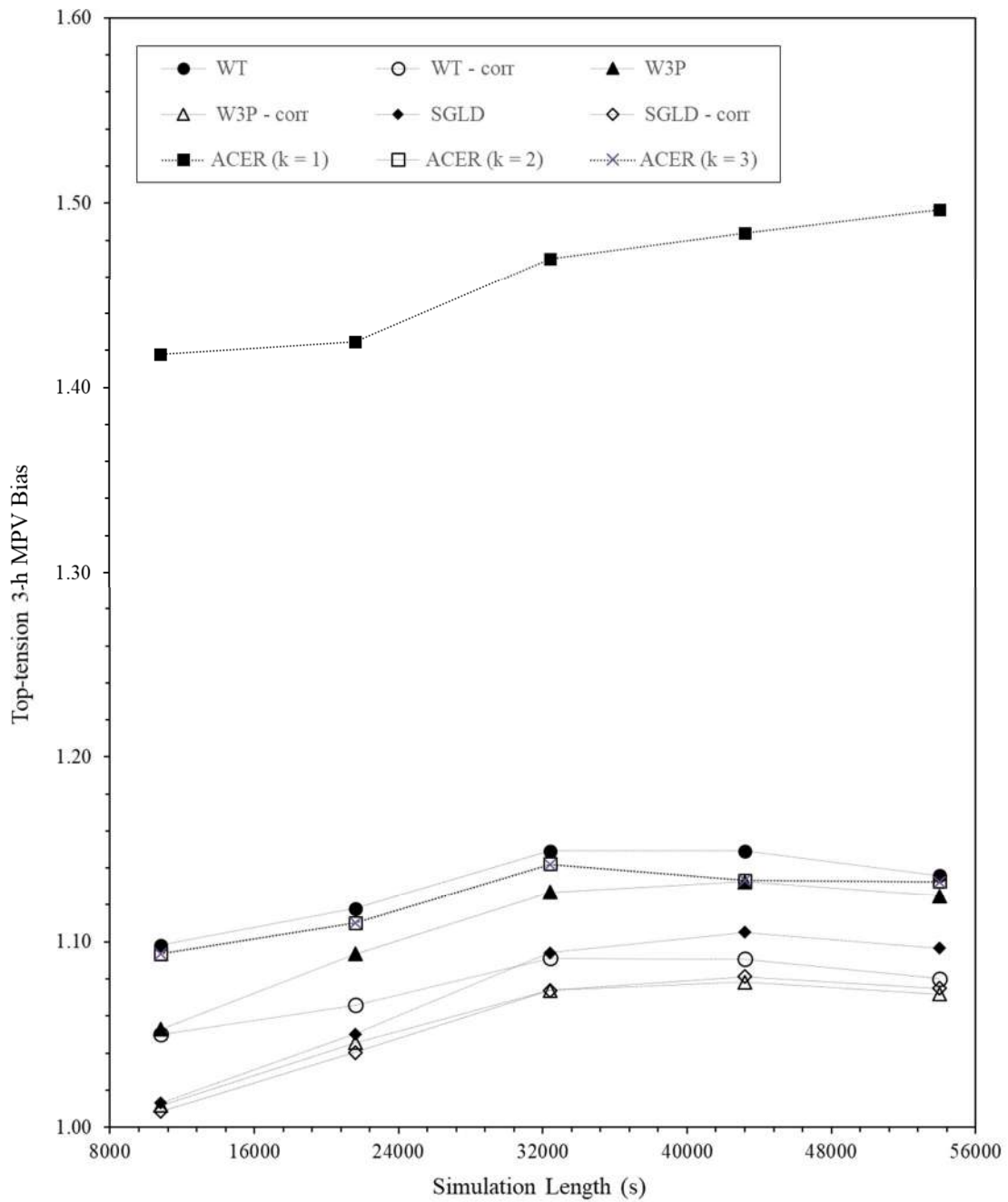
Figure 65 Bias estimates for the top tension 3-h MPV – Case study B ($N_w = 2000$). ('corr' indicates the consideration of peaks correlation)
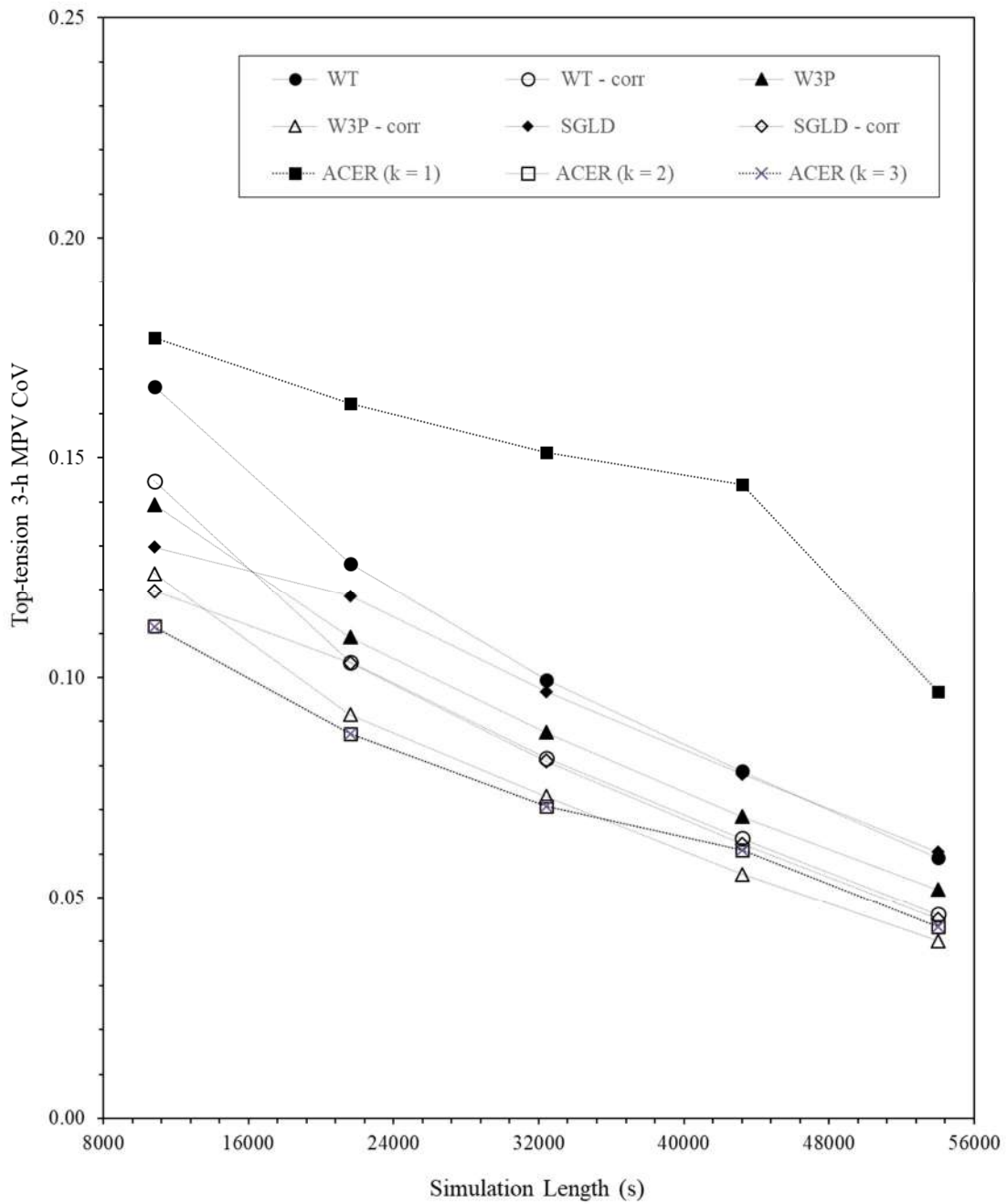
Figure 66 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 300$).
('corr' indicates the consideration of peaks correlation)
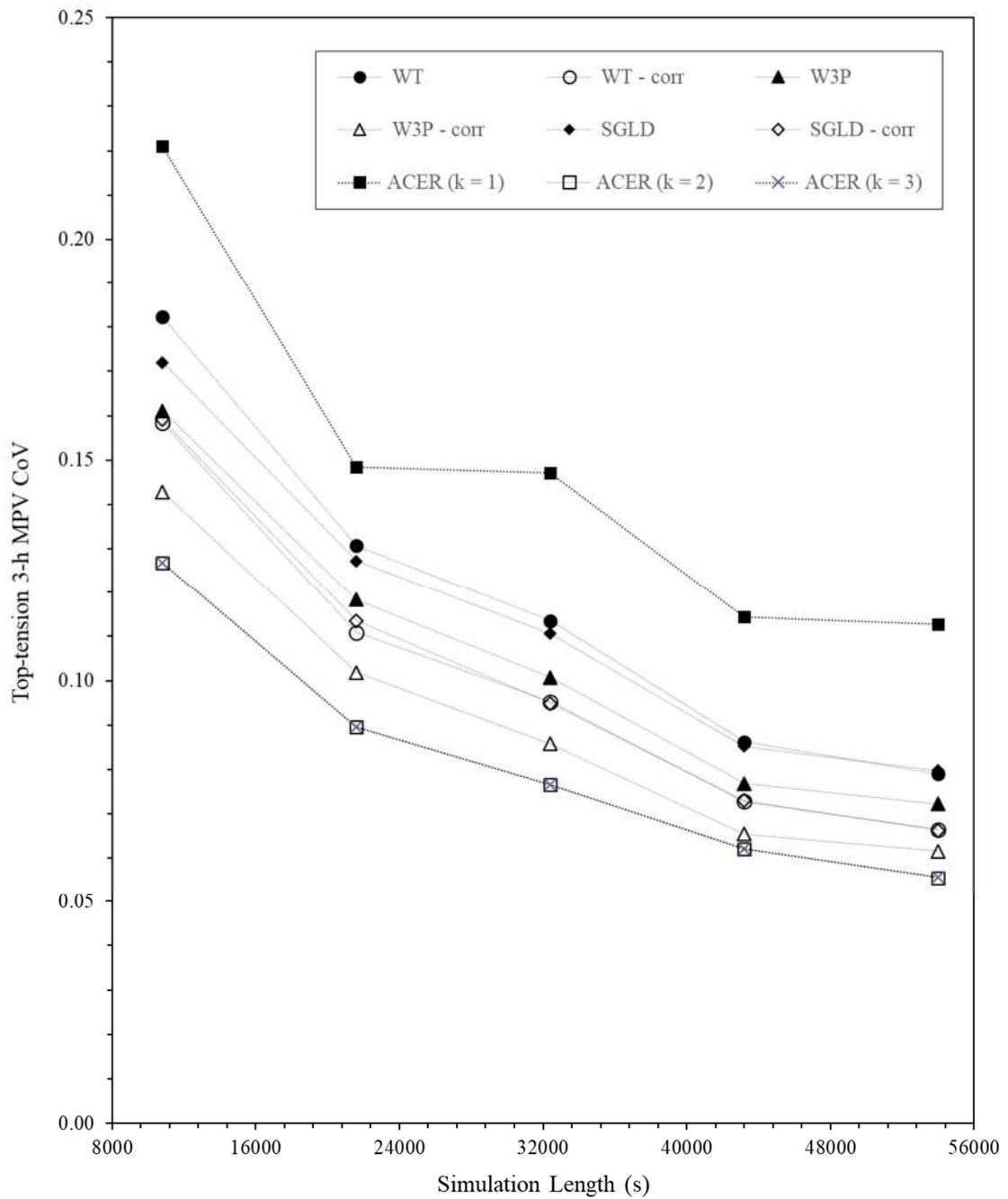
Figure 67 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 1000$).
('corr' indicates the consideration of peaks correlation)
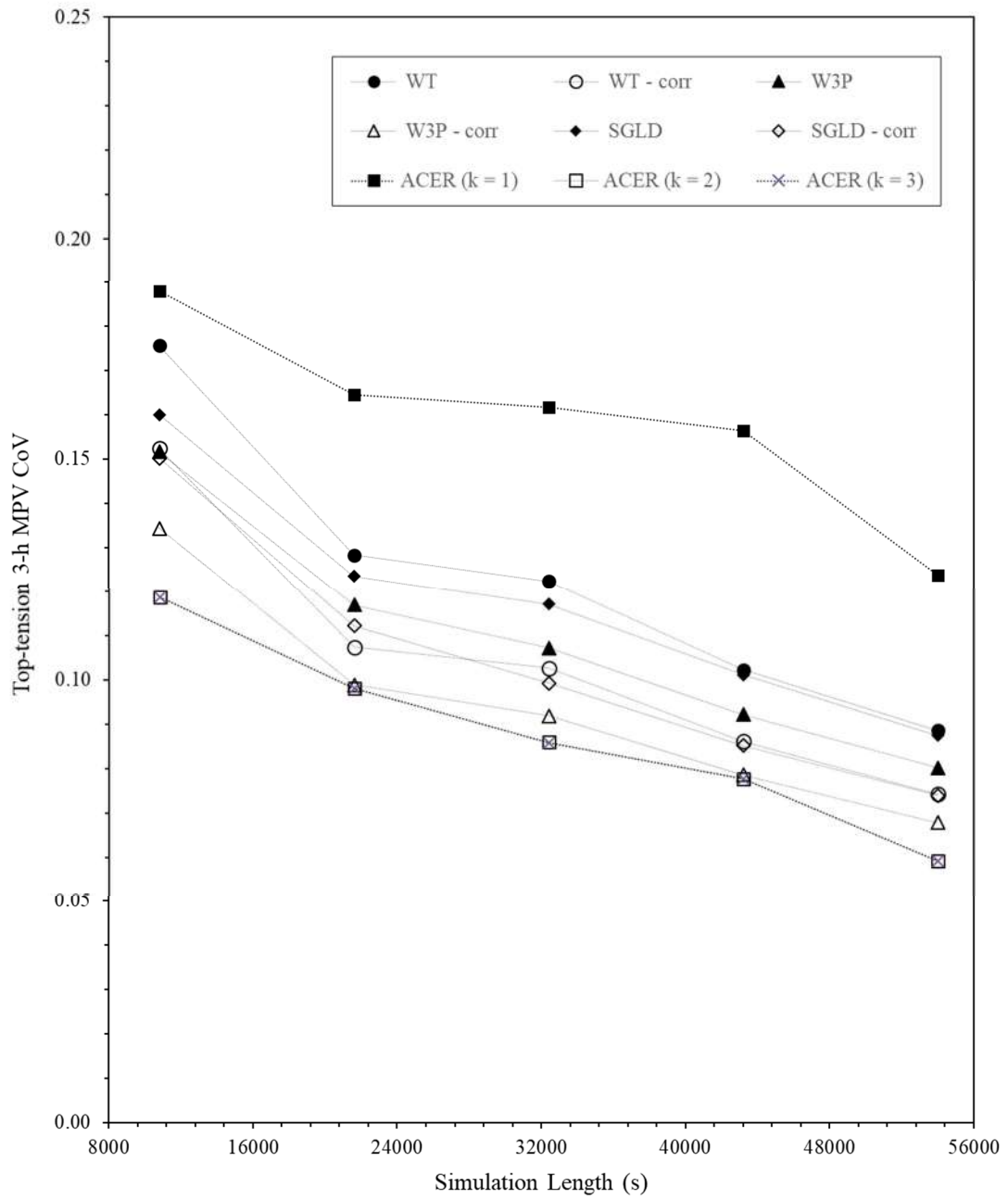
Figure 68 CoVs estimates for the top tension 3-h MPV - Case study A ($N_w = 2000$).
('corr' indicates the consideration of peaks correlation)

Likewise case study A, the following observations can be drawn from Figure 63 to Figure 68:

i. The level of spectrum discretization is again not crucial for the extremes estimation;

ii. On average, the short-term 3-h MPV estimates converge to a stable value for simulations longer than 3-h;

iii. For a given simulation length, the CoVs for the MPVs are higher than those obtained in the previous case study; only 15-h long simulations lead to CoVs near the 5% level;

iv. MPV estimates are here invariably overestimated, reaching a level of 22% bias for Weibull-Tail model;

v. Short-term extreme MPV estimates taking into consideration two subsequent peaks correlation tend to be circa 3% ~ 4% less than the MPV estimates that do not consider it, hence closer to the benchmark value. This indicates that the correlation takes a significant role in the extremes statistics. In model B, the expected value of the correlation coefficient between two consecutive peaks is $\rho = 0.7$, as shown in Table 5. The correlation effect helps in lowering the overestimation level; however, even for the probability models that show the best performances; i.e., W3P and SGLD models, the overestimation remains significant, at a 5 ~ 6% level. This larger variability can be due to the dynamic behavior of the floater-lines system: low-frequency second-order effects have a greater impact on model B top tension series mainly due to its 10 times smaller water depth.

vii. In model B, the ACER method also seems to predict highly overestimated extreme MPVs when $k = 1$. The overestimation drastically decreases when correlation between two peaks is considered, $k = 2$, evidencing the important role correlation takes in the prediction of the extreme values in model B. However, for $k = 3$ the extremes MPVs are equal to values obtained for $k = 2$. This shows that the correlation between three consecutive peaks is low in model B as well and hence do not affect the extremes estimation.

7 CONCLUSION AND FUTURE WORK RECOMMENDATION

This study investigated the influence of several aspects in the extremes top tension estimation in mooring lines, such as the probability distribution model choice for the tension global peaks, the numerical simulation length, the level of wave spectrum discretization, as well as the effect of considering correlation between two and three consecutive peaks. Three choices for the peaks probability model were investigated: (i) three-parameter Weibull model; (ii) two-parameter Weibull model fitted to the high probability level data (Weibull-tail model) and (iii) four-parameter Shifted Generalized Lognormal Distribution (SGLD) model. The ACER method was also investigated for $k = 1, 2, 3$.

For each simulation length and wave spectrum discretization level investigated, 200 realizations were generated in order to obtain an average MPV for each probability model chosen, including or not the peaks correlation consideration. The bias for each 'simulation length - wave spectrum discretization level - probability model' condition was determined through the comparison between the average MPV (from the 200 realizations) and the benchmark MPV estimate. This 'true' MPV is determined by fitting the Gumbel distribution to a set of 1000 epochal maxima taken from 1000 3-h realizations with the largest level of wave spectrum discretization (N_w = 2000). The parameters of the Gumbel distribution are determined through the method of moments.

By analyzing two distinct mooring lines, connected to two FPSOs, located in different water depths offshore Brazil, and under different extreme environmental conditions, some remarks could be made:

- The level of discretization of the wave spectrum seems to have low effect in the extremes response estimation. Bias estimates did not greatly differ for N_w = 300, N_w = 1000 or N_w = 2000;

- As expected, higher simulation lengths are associated to lower CoVs of the extreme MPV estimate. For the situation where second-order effects are more

expressive, an acceptable variability of 5% is only achieved when 15-h long simulations are generated;

- The correlation between successive peaks can influence extremes statistics: when it is significant, MPV estimates are lowered and can better approach benchmark values;

- The MPV values are often overestimated, i.e. for the great majority of obtained results, MPV estimates are conservative, especially in the situation when low-frequency second-order effects are more expressive. When this is not the case, MPV estimates are practically the same as the benchmark values. The three-parameter Weibull and the SGLD models presented the best performances when 'corrected' by the consideration of peaks correlation, predicting low biased MPV estimators or even practically unbiased estimators. Among the three probability models investigated, the Weibull-tail is the one that presented the largest level of over prediction in all cases investigated;

- The ACER method, for $k = 1$, returned the most overestimated extremes MPVs for both case studies investigated. MPVs calculated for $k = 2$ were closer to the reference values, showing lower biases. The consideration of three consecutive peaks being correlated, i.e., $k = 3$, resulted in values practically equal to those obtained for the consideration of two consecutive peaks correlated;

- Finally, from all simulated tension time-series it was possible to reassure the non-Gaussian characteristic of the top tension response in mooring lines connected to floating units. This fact prevents the possibility of determining extreme values of tension through the use of closed formulas since the process probability distribution is not known. Methods such as the ones presented in this dissertation are therefore imperative in determining values for an extremes design.

Based exclusively in the two case studies presented in this work, some general practical recommendations for the extreme statistics analysis of top tension of mooring lines connected to FPSOs can be made:

- Minimum use of wave spectrum discretization level of $N_w = 1000$ or higher;

- Weibull-3P or SGLD models choice for modelling global peaks distribution: (with the reservation that Weibull-3P is easier to be numerically implemented);

- Minimum simulation length (after cutting-off transient effects) of 15-h;

- Consideration of correlation between two consecutive global peaks in the extremes statistics calculations.

For future research is this area, the following topics are suggested:

- Extend the global peaks modelling by Weibull-3P and SGLD models to perform long-term response statistics of mooring lines;

- Taking in account the results of the present work, by choosing dynamic analysis parameters and an extreme value estimation method that will result in more accurate estimates of extreme top tension values, in order to develop a complete reliability analysis and define more efficient design safety factors.

REFERENCES

[1]   TAYLOR, R. E. and KERNOT, M.P., 1999, "On Second Order Wave Loading and Response in Irregular Seas", *Advances in Coastal and Ocean Engineering*, v. 5, pp. 155-212.

[2]   SAGRILO, L.V.S., GAO, Z., NAESS, A., LIMA, E.C.P., 2011, "A Straightforward Approach for Using Single Time Domain Simulation to Assess Characteristic Extreme Response", *Ocean Engineering*, v. 38, n. 13, pp. 1464-1471.

[3]   NAESS, A. and GAIDAI, O., 2009, "Estimation of Extreme Values from Sampled Time Series", *Structural Safety,* v. 31, pp. 325-334.

[4]   LOW, Y.M., 2013, "A New Distribution Model for Fitting Four Moments and its Application to Reliability Analysis", *Structural Safety,* v. 42, pp. 12-25.

[5]   NAESS, A., GAIDAI, O. AND KARPA, O., 2013, "Estimation of Extreme Values by the Average Conditional Exceedance Rate Method", *Journal of Probability and Statistics*, v. 2013, Article ID 797014, 15 p.

[6]   WIRSHING, P.H., PAEZ, T.L., ORTIZ, K., 1995, *Random Vibrations Theory and Practice*, 2 ed, Dover Publications, Inc., New York, USA.

[7]   NEWLAND, L.S., 1993, *An Introduction to Random Vibrations*, Spectral & Wavelet Analysis, 3rd ed, Longman Scientific and Technical, Harlow, England.

[8]   LIU, P. and KIUREGHIAN, A.D., 1986, "Multivariate Distribution Models with Prescribed Marginals and Covariances", *Probabilistic Engineering Mechanics*, v.1, n. 2, pp. 105-112.

[9]   MADSEN, H.O., KRENK, S. and LIND, N.C., 1986, *Methods of Structural Safety*, Dover Publications, Inc., Mineola, New York, USA.

[10] SAGRILO, L.V.S., SIQUEIRA, M.Q., ELLWANGER, G.B., Lima, E.C.P., Ferreira, M.D.A.S and MOURELLE, M.M., 2002, "A Coupled Approach for Dynamic Analysis of CALM Systems", *Applied Ocean Research*, v. 24, pp. 47-58.

[11] RICE, S.O., 1945, "Mathematical Analysis of Random Noise", *Bell System Technical Journal* v. 24, pp. 46-156, American Telephone and Telegraph Company (AT&T), New York, USA.

[12] ANG, A.H.S. and TANG, W.H., 1984, *Probability Concepts in Engineering Planning and Design,* 2 ed, John Wiley and Sons, New York, USA

[13] SAGRILO, L.V.S., NAESS, A. and GAO, Z., 2012, "On the Extreme Value Analysis of the Response of a Turret-Moored FPSO", *Journal of Offshore Mechanics and Artic Engineering,* v. 134 n. 2.

[14] NAESS, A., 1984, "The Effect of the Markov Chain Condition on the Prediction of Extreme Values", *Journal of Sound and Vibration*, v.94, n.1, pp. 87-103.

[15] NAESS, A. & MOAN, T., 2013, *Stochastic Dynamics of Marine Structures*, Cambridge University Press, New York, USA.

[16] KARPA, O., 2012, Development of Bivariate Extreme Value Distributions for Applications in Marine Technology, Ph.D. Thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

[17] FARNES, K.A., 1990, *Long-term Statistics of Response in Non-linear Marine Structures*, Ph.D Thesis, Department of Marine Structures, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

[18] KARPA, O., 2015, ACER User Guide, Centre for Ships and Ocean Structures (CeSOS), Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

[19] FALTINSEN, O.M., 1990, *Sea Loads on Ships and Offshore Structures*, Cambridge University Press, New York, USA.

[20] ELLWANGER, G.B., 2010, Tecnologias de explotação de petróleo, Apostila cursos de Mestrado e Doutorado PEC, COPPE/UFRJ.

[21] HASSELMANN, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., et al., 1973, *Measurements of Wind-Wave Growth and Swell Decay during the Joint North Sea Wave Project (JONSWAP),* Deutsches Hydrographisches Institut, Hamburg, Germany.

[22] MOSKOWITZ, L. & PIERSON, W.J., 1963, *A Proposed Spectral Form for Fully Developed Wind Seas Based on the Similarity Theory of S. A. Kitaigorodskii*, New York University, New York, USA.

[23] PINKSTER, J.A., 1980, *Low Frequency Second Order Wave Exciting Forces on Floating Structures*, Ph.D. Thesis, TU Delft, Delft, Netherlands.

[24] WAMIT, 2016, User Manual v. 7.2, WAMIT, Inc, Massachussets, USA.

[25] MORISON, J.R., 1953, "The Force Distribution Exerted by Surface Waves on Piles", Berkeley, California, USA.

[26] DNGL AS, 2015, DNVGL-OS-E301, *Position Mooring*, Offshore Standard, Hovik, Norway.

[27] NORSOK, 2007, N-003, *Actions and Action Effects*, Offshore Standard, Norway.

[28] DYNASIM System Manual – Version 1.4, 2001, *Dynamic Analysis of Anchoring Systems*, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil.

[29] BLACKMAN, R.B. and TUKEY, J.W., 1958, *The Measurement of Power Spectra*, Dover Publications, New York, USA.

ANNEX A – NATAF TRANSFORMATION PROCEDURE

If the marginal distributions and correlation coefficient between any random variables is known, the joint probability density function of these variables can be approximately determined through the Nataf transformation model [8]. The Nataf transformation is performed in two steps. In the first step, variables of any arbitrary distributions are transformed to correlated standard normal variables by means of statistical equivalence concepts (see Section 2.2). In the second step, the latter are transformed into uncorrelated standard normal variables. The model can be applied to any desired number of original variables. However, this annex will focus only on the two-dimensional case.

Transformation of arbitrary variables into correlated standard normal variables

Let X_i, $i = 1, 2, \ldots, n$ be a set of n correlated random variables whose marginal cumulative distribution functions $F_{X_i}(x_i)$ and correlation matrix $\boldsymbol{\rho}$ are known. Each term $\rho_{i,j}$ in the matrix $\boldsymbol{\rho}$ denotes the correlation coefficient between the random variables X_i and X_j. The first step of the Nataf transformation generates correlated standard normal variables U_i^0, with correlation matrix $\boldsymbol{\rho}^0$, obtained as follows:

$$
\begin{aligned}
\Phi(u_1^0) &= F_{X_1}(x_1) \\
\Phi(u_2^0) &= F_{X_2}(x_2) \\
&\vdots \\
\Phi(u_n^0) &= F_{X_n}(x_n)
\end{aligned}
\tag{A.1}
$$

where $\Phi(.)$ denotes the cumulative standard normal distribution.

For the two-dimensional case, the joint density function of the variables U_i^0 and U_j^0 is given by the joint normal probability density function:

$$f_{U_i^0,U_j^0}(u_i^0, u_j^0) = \phi_2(u_i^0, u_j^0, \rho_{i,j}^0) \tag{A.2}$$

where $\phi_2(u_i^0, u_j^0, \rho_{i,j}^0)$ is the bivariate standard normal probability density function. The expression of $\phi_2(.)$ is:

$$\phi_2(u_i^0, u_j^0, \rho_{i,j}^0) = \frac{1}{2\pi\sqrt{1-\left(\rho_{i,j}^0\right)^2}} exp\left[-\frac{(u_i^0)^2 - 2\rho_{i,j}^0 u_i^0 u_j^0 + \left(u_j^0\right)^2}{2\left(1-\left(\rho_{i,j}^0\right)^2\right)}\right] \tag{A.3}$$

Naturally, since the original variables X_i and X_j are correlated with the correlation coefficient $\rho_{i,j}$, the transformed standard normal variables U_i^0 and U_j^0 will also be correlated, but with the correlation coefficient $\rho_{i,j}^0$. The relation between the two coefficients is later presented.

The Joint Probability Distribution of Random Variables through the Nataf Transformation

Through classical formulation of variables transformation, for the two-dimensional case, the joint distribution of the original variables $f_{X_iX_j}(x_i, x_j)$ and the joint distribution of the correlated standard normal variables $f_{U_i^0,U_j^0}(u_i^0, u_j^0)$ is given by:

$$f_{X_i,X_j}(x_i, x_j) = f_{U_i^0,U_j^0}(\Phi^{-1}(F_{X_i}(x_i)), \Phi^{-1}(F_{X_j}(x_j)), \rho_{i,j}^0)J(x_i, x_j) \tag{A.4}$$

where $J(x_i, x_j)$ is the Jacobian matrix of the transformation which is defined by its general term as $\partial u_j/\partial x_i$. For the two-dimensional case, the Jacobian matrix is given by

$$J(x_i, x_j) = \left\Vert\begin{matrix}\dfrac{\partial u_i}{\partial x_i} & \dfrac{\partial u_i}{\partial x_j} \\ \dfrac{\partial u_j}{\partial x_i} & \dfrac{\partial u_j}{\partial x_j}\end{matrix}\right\Vert \tag{A.5}$$

where $\Vert\ \Vert$ denotes the determinant. According to Eq. (A.1), the crossed derivative terms are null, then

$$J(x_i, x_j) = \left\| \begin{matrix} \dfrac{\partial u_i}{\partial x_i} & 0 \\ 0 & \dfrac{\partial u_j}{\partial x_j} \end{matrix} \right\| \tag{A.6}$$

The Jacobian can also be written as

$$J(x_i, x_j) = \left\| \begin{matrix} \dfrac{\partial \Phi^{-1}(F_{X_i}(x_i))}{\partial x_i} & 0 \\ 0 & \dfrac{\partial \Phi^{-1}(F_{X_j}(x_j))}{\partial x_j} \end{matrix} \right\| \tag{A.7}$$

It can be shown that [12]:

$$\frac{\partial \Phi^{-1}(F_{X_i}(x_i))}{\partial x_i} = \frac{f_{X_i}(x_i)}{\phi(\Phi^{-1}(F_{X_i}(x_i)))} \tag{A.8}$$

where $\phi(.)$ corresponds to the density probability function of a standard normal variable:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}x^2\right) \tag{A.9}$$

Then, since probability density functions are nonnegative, one gets:

$$J(x_i, x_j) = \frac{f_{X_i}(x_i)f_{X_j}(x_j)}{\phi(\Phi^{-1}(F_{X_i}(x_i)))\phi(\Phi^{-1}(F_{X_j}(x_j)))} \tag{A.10}$$

From Eq. (A.4), finally it follows that:

$$f_{X_i,X_j}(x_i, x_j)$$
$$= \phi_2(\Phi^{-1}(F_{X_i}(x_i)), \Phi^{-1}(F_{X_j}(x_j)), \rho_{i,j}^0) \frac{f_{X_i}(x_i)f_{X_j}(x_j)}{\phi\left(\Phi^{-1}\left(F_{X_i}(x_i)\right)\right)\phi\left(\Phi^{-1}\left(F_{X_j}(x_j)\right)\right)} \tag{A.11}$$

The relation between the terms of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^0$ is derived as follows:

$$\rho_{i,j} = \frac{Cov(X_i, X_j)}{\sigma_{X_i}, \sigma_{X_j}}$$

$$= \frac{E(X_i, X_j) - E(X_i)E(X_j)}{\sigma_{X_i}, \sigma_{X_j}}$$

$$= \frac{1}{\sigma_{X_i}, \sigma_{X_j}} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j \, f_{X_i, X_j}(x_i, x_j) dx_i dx_j \right.$$
$$\left. - \int_{-\infty}^{\infty} x_i \, f_{X_i}(x_i) dx_i \int_{-\infty}^{\infty} x_j \, f_{X_j}(x_j) dx_j \right)$$

(A.12)

where $f_{X_i, X_j}(x_i, x_j)$ is given by Eq. (A.11). The above double integral needs to be solved in order to determine $\rho_{i,j}^0$ for a given $\rho_{i,j}$. It is important to notice that the above calculations need to be performed for every element in the $\boldsymbol{\rho}^0$ matrix. For engineering applications, $\rho_{i,j}^0$ is usually numerically approximated by polynomials. However, for many cases, $\rho_{i,j}^0 \approx \rho_{i,j}$ [8].

It is then demonstrated that for the determination of the joint probability density function of two arbitrary correlated random variables, one only needs the original random variables distributions and the correlation between them.

Rosenblatt transformation – Determination of uncorrelated standard normal variables

Many times in engineering applications, as in the case of structural reliability analysis, the use of standard normal variables that are statistically independent is necessary. In other words, it is frequently necessary to transform a set of correlated random variables to uncorrelated standard normal ones.

Initially considering the case of normal correlated variables, once all the elements of the $\boldsymbol{\rho}^0$ correlation matrix are determined, the variables U_i^0 can be manipulated in order to obtain the uncorrelated standard normal variables U_i. Let the following linear transformation be considered:

$$U^0 = AU + B \tag{A.13}$$

where U^0 and U are vectors containing the correlated and uncorrelated standard normal variables, respectively. A and B are constant matrices. Taking the expected value of both sides of Eq. (A.13) yields:

$$E(U^0) = E(AU + B) = AE(U) + B = 0 \tag{A.14}$$

Because both variables are standard normal variables, $B = 0$. Considering that $E(UU^T) = Cov(U, U) = I$ and $E\left(U^0 U^{0^T}\right) = Cov(U^0, U^0) = \rho^0$ (this relation can be proven by the fact that $\rho^0_{i,j} = E(u^0_i u^0_j)$), it can be shown that:

$$AA^T = \rho^0 \tag{A.15}$$

Therefore, A can be obtained by the Cholesky decomposition of ρ^0, which is a positive definite matrix. For a two-dimensional problem, A is given by:

$$A = \begin{bmatrix} 1 & 0 \\ \rho^0_{i,j} & \sqrt{1 - \left(\rho^0_{i,j}\right)^2} \end{bmatrix} \tag{A.16}$$

The variables U_i and U_j are hence related to U^0_i and U^0_j by the following equations:

$$U_i = U^0_i$$
$$U_j = \frac{U^0_j - \rho^0_{i,j} \cdot U^0_i}{\sqrt{1 - \left(\rho^0_{i,j}\right)^2}} \tag{A.17}$$

Hence, in the general case, a set of n independent standard normal variables U are related to a set of X correlated random variables by the following relationship:

$$\begin{bmatrix} U_i \\ \vdots \\ U_n \end{bmatrix} = \mathbf{\Gamma} \begin{bmatrix} \Phi^{-1}\left(F_{X_1}(x_1)\right) \\ \vdots \\ \Phi^{-1}\left(F_{X_n}(x_n)\right) \end{bmatrix} \tag{A.18}$$

with $\mathbf{\Gamma} = \boldsymbol{A}^{-1}$.

ANNEX B – WEIBULL-TAIL FITTING PROCEDURE

In a Weibull-tail fitting, a 2-parameter Weibull is fitted to the tail of the data of interest. The tail is defined by sub-sets of the data associated to percentiles of the empirical cumulative distribution. For each one of these levels, the data above them are used to estimate the parameters α and λ of the distribution through a linear regression fitting procedure, as explained in the following equations.

The CDF of the 2-parameter Weibull distribution is given by

$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^\lambda\right] \tag{B.1}$$

Taking the natural log of both sides of the above equation two times yields

$$\ln\left(-\ln\left(1 - F_X(x)\right)\right) = \lambda \ln x - \lambda \ln \alpha \tag{B.2}$$

The above equation can be expressed as the linear equation

$$y = \lambda x\prime + a \tag{B.3}$$

where $y = \ln\left(-\ln\left(1 - F_X(x)\right)\right)$, $x\prime = \ln x$ and $a = -\lambda \ln \alpha$. Given a sample of x and its empirical CDF, one can directly compute y and $x\prime$. Weibull parameters can be then straightforwardly estimated via linear regression. In the present work, the linear regression is performed through the Least Mean Squares (LMS) method.

In the linear LMS regression method, the objective consists of adjusting the parameters of a linear function to best fit a data set. The model function has the form $f(x, \beta)$, where m adjustable parameters are held in the vector β. The goal is finding estimates for the parameters that best fits the data. The fit of a model to a data is measured by the difference between the actual value of the dependent variable and the value predicted by the model, or simply its residual:

$$r_i = y_i - f(x, \beta) \tag{B.4}$$

In the LMS method the parameters estimates are those that minimize the sum, S, of squared residuals:

$$S = \sum_{i=1}^{n} r_i^2 \tag{B.5}$$

where n is the number of points in the data set.

In this case, the function $f(x, \beta)$ is the linear function corresponding to Equation A.3.

In this work, the sub-sets of data considered are associated to the $60\%, 65\%, 70\%, 75\%, 80\%, 85\%$ and 90% percentiles of the empirical cumulative distribution. Hence for each of these data sub-sets the linear LMS regression method is performed and the parameters for the Weibull distribution are estimated. The result is a group of N pairs of adjusted Weibull parameters $\{(\lambda_{60\%}, \alpha_{60\%}), (\lambda_{65\%}, \alpha_{65\%}), \ldots, (\lambda_{90\%}, \alpha_{90\%})\}$. The final parameters of the Weibull-tail distribution are taken as the corresponding mean values of all sub-sets parameters:

$$\lambda_{WT} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \tag{B.6}$$

$$\alpha_{WT} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \tag{B.7}$$

ANNEX C – STATISTICAL PARAMETERS OF SIMULATED TIME HISTORIES

Mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series, as well as the mean, standard deviation and coefficient of variation of the 200 obtained values for the main statistical parameters of the time-series peaks, peaks frequency and correlation between two consecutive peaks for each pair $\{T_s, N_w\}$.

Study Case A

Table 6 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1967.13	486.27	0.37	3.06	2276.28	296.75	2.05	8.54	0.03	0.25
σ	36.94	26.43	0.16	0.38	42.28	21.91	0.31	2.17	0.00	0.07
CoV	0.02	0.05	0.43	0.12	0.02	0.07	0.15	0.25	0.09	0.27

Table 7 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1971.74	490.42	0.37	3.02	2281.59	298.98	2.06	8.44	0.03	0.25
σ	23.17	16.06	0.11	0.24	27.06	13.17	0.20	1.35	0.00	0.05
CoV	0.01	0.03	0.29	0.08	0.01	0.04	0.10	0.16	0.06	0.20

Table 8 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1968.04	488.76	0.37	3.06	2277.02	297.40	2.07	8.62	0.03	0.25
σ	16.24	10.56	0.09	0.22	21.06	10.22	0.18	1.32	0.00	0.04
CoV	0.01	0.02	0.25	0.07	0.01	0.03	0.09	0.15	0.04	0.14

Table 9 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1969.28	489.29	0.38	3.06	2278.15	298.46	2.08	8.59	0.03	0.25
σ	9.27	7.04	0.08	0.19	12.59	7.83	0.14	1.07	0.00	0.03
CoV	0.00	0.01	0.22	0.06	0.01	0.03	0.07	0.12	0.04	0.11

Table 10 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1969.25	490.01	0.37	3.07	2278.02	298.24	2.08	8.66	0.03	0.25
σ	6.85	5.95	0.08	0.21	10.68	6.93	0.15	1.11	0.00	0.02
CoV	0.00	0.01	0.23	0.07	0.00	0.02	0.07	0.13	0.03	0.10

Table 11 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1956.04	487.50	0.38	3.03	2264.14	298.88	2.05	8.40	0.03	0.25
σ	42.13	27.58	0.15	0.35	48.10	20.56	0.29	2.10	0.00	0.07
CoV	0.02	0.06	0.40	0.12	0.02	0.07	0.14	0.25	0.09	0.28

Table 12 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1963.93	489.45	0.37	3.02	2274.47	300.09	2.04	8.31	0.03	0.25
σ	27.08	20.52	0.11	0.23	32.63	15.08	0.21	1.40	0.00	0.05
CoV	0.01	0.04	0.29	0.08	0.01	0.05	0.10	0.17	0.07	0.21

Table 13 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1964.31	488.62	0.37	3.04	2273.20	298.24	2.06	8.47	0.03	0.25
σ	21.78	14.18	0.10	0.23	25.51	11.52	0.17	1.23	0.00	0.04
CoV	0.01	0.03	0.26	0.07	0.01	0.04	0.08	0.15	0.05	0.15

Table 14 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1961.62	487.72	0.36	3.02	2269.95	297.56	2.06	8.41	0.03	0.25
σ	16.13	12.63	0.09	0.20	19.97	10.22	0.16	1.15	0.00	0.03
CoV	0.01	0.03	0.24	0.07	0.01	0.03	0.08	0.14	0.05	0.12

Table 15 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1964.14	489.89	0.37	3.05	2273.80	298.54	2.07	8.53	0.03	0.25
σ	14.14	9.97	0.07	0.18	16.26	8.12	0.13	0.99	0.00	0.03
CoV	0.01	0.02	0.19	0.06	0.01	0.03	0.06	0.12	0.04	0.12

Table 16 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1955.92	485.78	0.38	3.04	2265.57	298.17	2.02	8.27	0.03	0.25
σ	41.33	27.31	0.16	0.42	48.63	21.38	0.28	1.91	0.00	0.07
CoV	0.02	0.06	0.42	0.14	0.02	0.07	0.14	0.23	0.09	0.27

Table 17 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1963.61	485.96	0.37	3.04	2272.11	297.11	2.05	8.37	0.03	0.26
σ	29.88	20.24	0.10	0.26	35.15	14.88	0.22	1.51	0.00	0.05
CoV	0.02	0.04	0.28	0.08	0.02	0.05	0.11	0.18	0.07	0.21

Table 18 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1961.28	486.02	0.37	3.03	2269.50	297.38	2.05	8.39	0.03	0.26
σ	24.10	13.74	0.10	0.23	26.69	11.05	0.17	1.23	0.00	0.04
CoV	0.01	0.03	0.26	0.08	0.01	0.04	0.08	0.15	0.05	0.15

Table 19 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1961.39	488.59	0.37	3.04	2271.71	298.67	2.05	8.42	0.03	0.25
σ	19.80	13.86	0.08	0.22	23.90	10.53	0.15	1.11	0.00	0.04
CoV	0.01	0.03	0.22	0.07	0.01	0.04	0.08	0.13	0.04	0.14

Table 20 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	1959.71	487.11	0.37	3.15	2269.04	297.48	2.04	8.35	0.03	0.25
σ	14.49	12.15	0.07	0.15	16.51	9.55	0.13	0.90	0.00	0.03
CoV	0.01	0.02	0.19	0.05	0.01	0.03	0.06	0.11	0.04	0.13

Study Case B

Table 21 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	380.91	450.60	0.58	3.22	892.99	420.77	0.67	2.82	0.01	0.69
σ	13.43	60.95	0.14	0.63	122.27	68.26	0.35	0.84	0.00	0.10
CoV	0.04	0.14	0.25	0.19	0.14	0.16	0.53	0.30	0.11	0.15

Table 22 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 300$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	380.55	449.75	0.63	3.45	874.15	429.81	0.83	3.22	0.01	0.73
σ	8.62	42.45	0.12	0.60	88.49	50.06	0.33	0.94	0.00	0.07
CoV	0.02	0.09	0.20	0.17	0.10	0.12	0.40	0.29	0.09	0.09

Table 23 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 300$

| | Time-series | | | | Peaks | | | | | |
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	380.03	449.68	0.64	3.53	873.83	431.55	0.86	3.43	0.01	0.73
σ	6.17	29.55	0.12	0.65	62.11	37.28	0.32	1.12	0.00	0.05
CoV	0.02	0.07	0.18	0.18	0.07	0.09	0.37	0.33	0.07	0.07

Table 24 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 300$

| | Time-series | | | | Peaks | | | | | |
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	380.35	451.59	0.65	3.56	876.57	435.09	0.88	3.47	0.01	0.74
σ	3.47	16.72	0.10	0.58	38.27	26.00	0.27	1.04	0.00	0.04
CoV	0.01	0.04	0.16	0.16	0.04	0.06	0.31	0.30	0.05	0.06

Table 25 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 300$

| | Time-series | | | | Peaks | | | | | |
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	380.08	450.61	0.65	3.59	872.67	434.98	0.91	3.56	0.01	0.74
σ	1.30	6.85	0.09	0.54	26.24	16.71	0.25	1.05	0.00	0.03
CoV	0.00	0.02	0.14	0.15	0.03	0.04	0.27	0.30	0.04	0.04

Table 26 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	377.97	438.40	0.57	3.22	870.82	405.69	0.70	2.95	0.01	0.67
σ	15.50	70.96	0.16	0.74	147.24	75.23	0.40	1.06	0.00	0.11
CoV	0.04	0.16	0.28	0.23	0.17	0.19	0.57	0.36	0.13	0.17

Table 27 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	378.01	445.95	0.61	3.39	870.70	424.64	0.80	3.14	0.01	0.72
σ	10.27	49.52	0.12	0.56	94.17	56.46	0.29	0.91	0.00	0.07
CoV	0.03	0.11	0.19	0.17	0.11	0.13	0.37	0.29	0.08	0.10

Table 28 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	377.95	445.58	0.62	3.43	868.66	426.12	0.82	3.21	0.01	0.72
σ	8.10	39.61	0.10	0.50	75.08	45.96	0.27	0.82	0.00	0.06
CoV	0.02	0.09	0.17	0.15	0.09	0.11	0.33	0.26	0.06	0.09

Table 29 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 1000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	378.95	450.12	0.64	3.48	874.35	431.98	0.85	3.32	0.01	0.74
σ	6.24	29.80	0.08	0.42	62.31	34.77	0.23	0.76	0.00	0.04
CoV	0.02	0.07	0.13	0.12	0.07	0.08	0.27	0.23	0.06	0.06

Table 30 Statistical parameters of the 200 realizations for $T_s = 10800s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	375.29	432.43	0.57	3.22	855.73	401.08	0.71	2.93	0.01	0.68
σ	13.81	66.25	0.15	0.64	136.95	69.85	0.40	0.95	0.00	0.11
CoV	0.04	0.15	0.26	0.20	0.16	0.17	0.57	0.32	0.12	0.16

Table 31 Statistical parameters of the 200 realizations for $T_s = 21600s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	376.88	443.77	0.60	3.32	868.79	419.91	0.76	3.02	0.01	0.71
σ	9.72	47.46	0.12	0.56	91.61	55.00	0.30	0.88	0.00	0.07
CoV	0.03	0.11	0.20	0.17	0.11	0.13	0.40	0.29	0.09	0.10

Table 32 Statistical parameters of the 200 realizations for $T_s = 32400s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	377.84	448.62	0.63	3.43	874.87	428.69	0.81	3.20	0.01	0.73
σ	9.58	44.59	0.12	0.60	86.65	53.79	0.30	0.96	0.00	0.06
CoV	0.03	0.10	0.19	0.17	0.10	0.13	0.37	0.30	0.08	0.09

Table 33 Statistical parameters of the 200 realizations for $T_s = 43200s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	377.64	447.97	0.62	3.39	873.21	426.86	0.80	3.15	0.01	0.73
σ	8.75	41.61	0.10	0.49	80.88	49.01	0.26	0.79	0.00	0.06
CoV	0.02	0.09	0.17	0.14	0.09	0.11	0.32	0.25	0.07	0.08

Table 34 Statistical parameters of the 200 realizations for $T_s = 54000s$ and $N_w = 2000$

	Time-series				Peaks					
	μ_x	σ_x	γ_x	κ_x	μ_p	σ_p	γ_p	κ_p	ν_p	ρ
μ	377.03	445.59	0.64	3.52	863.20	428.69	0.88	3.43	0.01	0.73
σ	6.27	29.45	0.09	0.48	56.16	36.04	0.24	0.92	0.00	0.05
CoV	0.02	0.07	0.14	0.13	0.07	0.08	0.27	0.27	0.05	0.06