Point-in-time and Extreme-Value Probability Simulation Technique for Engineering Design

R. Mínguez, Y. Guanche, F. J. Méndez

Environmental Hydraulics Institute "IH Cantabria", Universidad de Cantabria, Cantabria Campus Internacional, Spain

Abstract

Engineering design of structural elements entails the satisfaction of different requirements during each of the phases that the structure undergoes: construction, service life and dismantling. Those requirements are settled in form of limit states, each of them with an associated probability of failure. Depending on the consequences of each failure, the acceptable probability varies and also the denomination of the limit state: ultimate, damage, serviceability, or operating stop. This distinction between limit states forces engineers to: i) characterize both the point-in-time and extreme probability distributions of the random variables involved (agents), which are characterized independently, and ii) use the appropriate distribution for each limit state depending on the failure consequences. This paper proposes a Monte Carlo simulation technique, which allows the generation of possible outcomes for agents holding the following conditions: i) both the point-in-time and the extreme value distributions are appropriately reproduced within the simulation procedure, and ii) it maintains the temporal dependence structure of the stochastic process. In addition, a graphical representation of both distributions on a compatible scale is given, this graph clarifies the link between point-in-time and extreme regimes and helps quantifying the degree of accuracy of the simulation results. In addition, new insights for the development of First-Order-Reliability methods (FORM) combining point-in-time and extreme distributions simultaneously are provided. The method is illustrated through several simulation examples from well-known distributions, whereas its skill over real data is shown using the significant wave height data record

Email address: roberto.minguez@unican.es (R. Mínguez)

from a buoy located on the Northern Spanish coast.

Keywords: Extreme-value distribution, Mixture distribution, Monte Carlo simulation, Level III method, Point-in-time distribution, Temporal autocorrelation

1. Introduction

Engineering structures undergo different phases during their lifetime: construction, service life and dismantling. During each of these phases, the structure must satisfy different requirements, which from the engineering design point of view, are defined as limit states. The objective of the design is to verify that the structure satisfies those project requirements in terms of acceptable failure rates and costs (see [1] and [2]).

Acceptable failure rates are established by codes and expert committees [3, 4, 5, 2] on the basis of the consequences of failure for each limit state, and trying to counter-balance safety and costs (direct, societal and environmental). Since the consequences of failure might be very different depending on the limit state considered, these limit conditions are classified in different categories: ultimate, damage, serviceability, or operating stop. The acceptable probabilities of failure for each category depends on the type of structure and environmental conditions, but in all cases it increases from the minimum acceptable probability of failure related to the ultimate limite state, up to the maximum acceptable probability associated with the operating stop limit state.

From the design perspective, these different probability thresholds encompass the consideration of different probability distributions for agents. Serviceability or operating stop limit conditions depend on regular, central or mean values of those agents, whereas damage and ultimate limit states require extreme conditions, i.e. to pay attention to singular values. The statistical theory for dealing with mean values (point-in-time distribution) is different from the theory for extreme values [6, 7, 8, 9]. Traditionally, both problems are treated independently, which makes difficult to understand the link between point-in-time and extreme distributions and their implications from the practical point of view.

There are several attempts in the literature to incorporate both the point-in-time (central) and extreme information in the same probability distribution model (mixture models), see for instance, [10, 11, 12, 13, 14, 15, 16].

The common characteristic of these works is that all are applied to specific distributions, and the parameter estimation is fuzzy, not providing a general framework to deal with the problem. This work is intended to fill this niche.

As previously mentioned, safety of structures is the fundamental criterion for design, and once limit states and required probabilities are defined, engineering design must ensure satisfaction of the safety constraints. There are several methods to check the satisfaction of the safety requirements which can be classified in two main groups: (a) the classical safety factor approach, and (b) the probability based approach. The latter deals with probabilities of failure, which are difficult to deal with because (a) it requires the definition of the joint probability of all variables involved, and (b) the evaluation of the failure probability is not an easy task. The problem becomes even more difficult if several failure modes are analyzed, because the failure region is the union of the different failure mode regions, and regions defined as unions are difficult to work with because of their irregular and non-differentiable boundaries [17]. A method widely used by engineers to overcome these difficulties is Monte Carlo simulation technique. Once the probability distributions are defined, long records of the random variables involved may be sampled [18, 19, 20] and used to check whether the safety constraints are satisfied in terms of probabilities of failure. The simplicity on its implementation has increased the development of different methods for structural reliability analysis [21, 22, 23], such as directional simulation techniques [24, 19], importance sampling [25, 26, 27], or techniques which allows reproducing on multidimensional settings, not only the marginal distributions but the temporal dependence of the stochastic processes involved as well [28, 29, 30].

The aim of this paper is threefold: i) to develop a Monte Carlo simulation method for reproducing both the point-in-time (mean values) and extreme value distributions of random variables, while keeping the temporal dependence structure of the stochastic process involved, and valid for any kind of probability distribution function, ii) to present a graphical interpretation of simulation results to merge both distributions on a compatible scale, and iii) to provide new insights for the use of the point-in-time and extreme regimes simultaneously within First-Order-Reliability methods (FORM). The theoretical and practical material developed in this paper is intended to support engineers on the design process and help understanding the relationship between both distributions, freeing engineers of deciding which conditions, average or extreme, must be used for each failure mode, because both conditions

are considered into the proposed distribution.

The rest of the paper is structured as follows. Section 2 introduces the concept of order statistics and extremes. In Section 3, the graphical interpretation to show the behavior of both distributions in a compatible scale is provided. The Monte Carlo simulation method reproducing both distributions is developed in Section 4. Section 5 introduces the autocorrelation into the simulation method and different examples of the simulation performance are shown. Section 6 discusses some implications of the method proposed in this paper with respect to FORM methods. Finally, some conclusions are duly drawn in Section 7.

2. Order Statistics and Extremes

Let consider the point-in-time probability density and distribution functions of a random variable X, i.e. $f_X(x)$ and $F_X(x)$. If we draw from this distribution a sample $x_1, x_2, \ldots, x_n$ of size n, and arrange it in increasing order $x_{1:n}, x_{2:n}, \ldots, x_n$, we could obtain the probability distribution of the rth element of this sequence, $X_{r:n}$, so-called the rth order statistic of a sample of size n. The first and last order statistics are the minimum $X_{1:n}$ and maximum $X_{n:n}$ respectively, and are called extremes [31, 8].

This maximum and minimum are very important for the design considering ultimate and damage limit states, and assuming that the point-in-time distribution of the variable of interest $F_X(x)$ (loads, significant wave height, strength, etc.) is known, the cumulative distribution functions of the maximum and minimum order statistics of a sample of size n are, respectively:

$$F_X^{\max}(x) = \left[F_X(x)^n\right], \tag{1}$$

and

$$F_X^{\min}(x) = 1 - \left[1 - F_X(x)\right]^n. \tag{2}$$

When n tends to infinity, distributions (1) and (2) are degenerate, only taking values equal to 0 or 1. For these cases linear transformations of x, consisting on location and scale changes, are looked for to avoid degeneracy. Note that when this is possible, $F_X(x)$ is considered to belong to the domain of attraction of the limit distribution.

[32] proved that there is only one parametric family for each of the limit distributions of maxima and minima, which correspond to the Generalized

Extreme Value Distributions for maxima (GEV) and minima (GEVm), respectively. For instance, the cumulative distribution function (CDF) for maxima is given by:

$$F_X^{\max}(x;\mu,\psi,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{x-\mu}{\psi}\right)\right]_+^{-\dfrac{1}{\xi}}\right\}; \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\psi}\right)\right]\right\}; \xi = 0, \end{cases} \tag{3}$$

where μ, ψ, and ξ are the location, scale and shape parameter, $[a]_+ = \max(0, a)$, and the support is $x \leq \mu - \psi/\xi$, if $\xi < 0$, or $x \geq \mu - \psi/\xi$, if $\xi > 0$. The GEV family includes three distributions corresponding to the different types of tail behavior: Gumbel ($\xi = 0$) with a light tail decaying exponentially; Fréchet distribution ($\xi > 0$) with a heavy tail decaying polinomially; and Weibull ($\xi < 0$) with a bounded tail.

Note that this result has two very important practical implications:

1. The complexity to characterize the point-in-time regime $F_X(x)$ of a given random variable X, which allows using multiple distributions as possible candidates, contrasts with respect to the apparent simplicity to characterize the probability distributions for maxima and/or minima, which only requires the estimation of the three parameters μ, ψ, and ξ from the corresponding limit distribution family.
2. Since different point-in-time distributions may have the same domain of attraction, the best way to characterize the tail (upper/lower) of the distribution is using data belonging to the corresponding tail (maxima/minima) and estimate the parameters of the corresponding limit distribution.

From the practical point of view, the use of the GEV distribution for maxima is not appropriate in many cases because it uses small samples for the fitting process. For those cases, it is preferable to use the Pareto-Poisson model, which is valid for independent and identically distributed processes, or the Peaks Over Threshold (POT) method, suitable for dependent and identically distributed processes. The method presented in this paper is valid for those distributions or any other distribution for maxima.

Traditionally, engineers treat both point-in-time and extreme value distributions independently depending on the kind of limit state under consideration. The method proposed in this paper facilitates the engineering task as follows:

1. By presenting a graphical interpretation which makes easier to check if the right-tail of the distribution is appropriately reproduced or fitted by the point-in-time distribution, and then decide if an additional analysis of those extremes is required.
2. For those cases where both analysis are relevant and required, we present the methodology to link both distributions and use them simultaneously. Thus avoiding the decision to choose one or the other depending on the limit state considered.

3. Relationship between point-in-time and extreme value distributions: Graphical representation

From the practical point of view, it would be very useful for engineers to establish the relationship between the point-in-time and the extreme value distributions for random agents, or even to have a graphical visualization of this relationship, which would allow them to quantify the skill of any Monte Carlo simulation technique to deal with both central and extreme conditions at the same time.

The aim of this section is to present a graphical representation to accomplish the aforementioned task. Let assume an stochastic process X_t with associated sampling or occurrence frequency $f = 1/T_x$ (T_x is the sampling period, for instance 1 hour, 2 hours, etc.) and whose point-in-time distribution is $F^{\mathrm{PT}}(x)$. If we simulate samples of size n from the stochastic process X_t and calculate their maximum values, this maximum is a random variable X_M with probability distribution function $F^{\mathrm{EV}}(x)$. Both distributions may be plotted on the same return period graph as follows:

1. Calculate the "equivalent return period" from the point-in-time distribution, i.e. $T^{\mathrm{PT}} = \dfrac{1}{1 - F^{\mathrm{PT}}(x)}$.
2. Plot T^{PT} versus x.
3. Calculate the "return period" from the extreme value distribution, i.e. $T^{\mathrm{EV}} = \dfrac{1}{1 - F^{\mathrm{EV}}(x)}$.

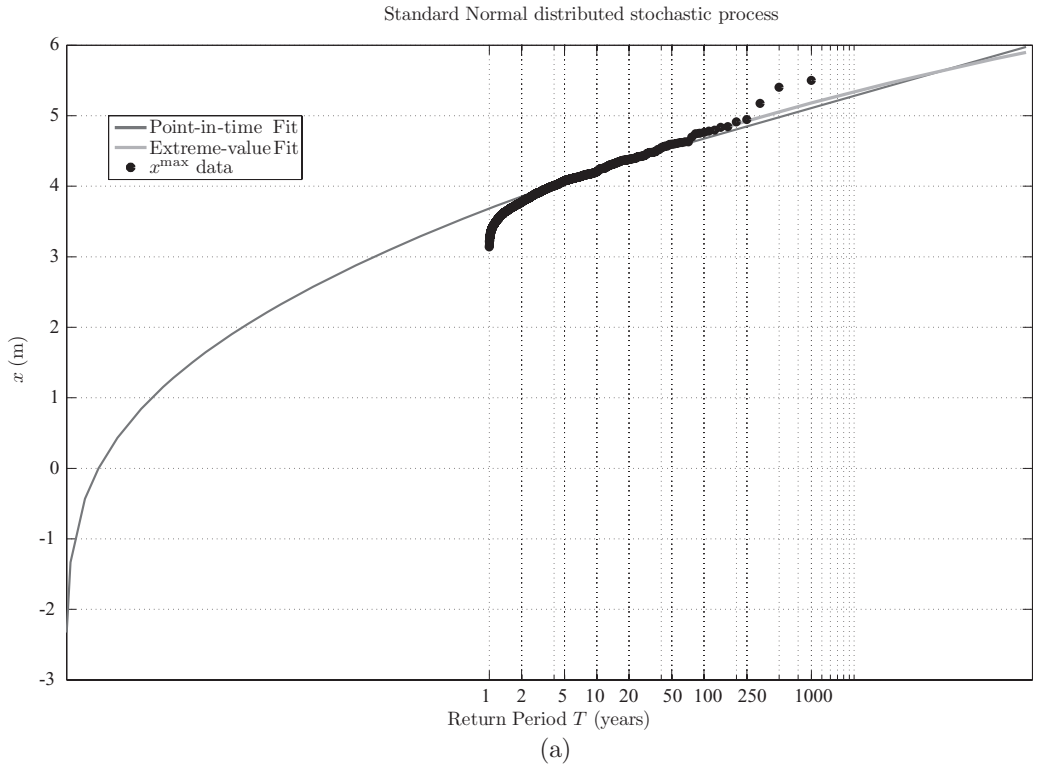
4. Plot the re-scaled return period using the sample size n, i.e. $T_r^{\mathrm{EV}} = n \cdot T^{\mathrm{EV}}$ versus x.

Figure 1 (a) shows the proposed graphical interpretation associated with an hourly stochastic process ($T_x = 1$ hour) without temporal correlation, and whose marginal (point-in-time) distribution corresponds to the standard normal ($X_t \sim N(0, 1^2)$). Dark gray line corresponds to (T^{PT}, x). We sample $n_y = 1000$ years of data and look for the annual maximum $x^{\max}$. Black dots correspond to $(T_r^{\mathrm{EV}},\ x^{\max})$. The re-scaled return period is calculated as $T_r^{EV} = \frac{n}{1-\hat{F}^{\mathrm{EV}}(x)}$, where $\hat{F}^{\mathrm{EV}}(x_i) = \frac{i}{n_y+1};\ \forall i = 1, \ldots, n_y$ is the empirical annual maxima probability distribution for the sample, and $n = 8766$. Finally, the light gray line represents $(T_r^{\mathrm{EV}}, x^{\max})$, where T_{EV_r} has been calculated using the GEV fitted distribution to annual maxima. Note that both the point-in-time and the extreme regimes converge on the tail of interest, however, there are slight differences between the point-in-time and the maxima fitted distribution due to the simulation and fitting process uncertainty. These differences tend to zero as the sample size tends to infinity. Note that the true abscissas axis units in Figure 1 are hours, however, we have re-scaled the ticks to years to facilitate the interpretation.

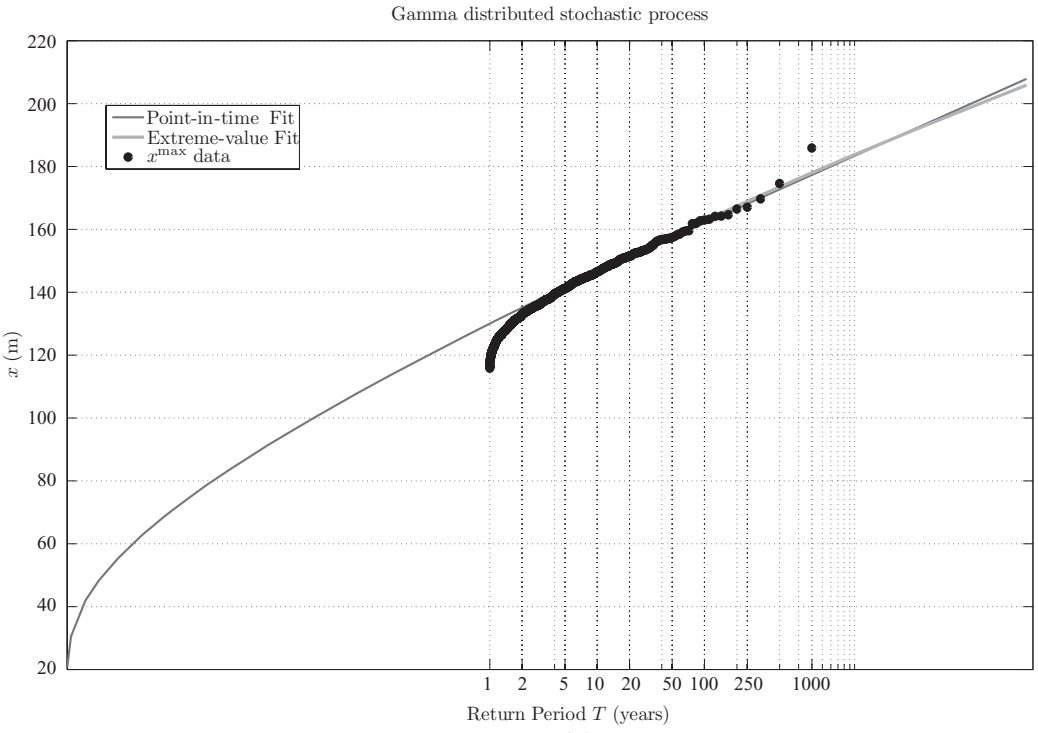
Analogous results are shown in Figure 1 (b) for a gamma distributed stochastic process with scale and shape parameters $\theta = 5$ and $\kappa = 10$, respectively. Note that as in the previous case, both the point-in-time and extreme-value probability distributions converge on the tail of interest.

These results are not surprising, since we are sampling from given point-in-time distributions, and thus the sampled data reproduce appropriately the tail of interest, especially if large samples are used. However, when dealing with real data sets, the point-in-time distribution does not usually fit appropriately the tail of interest. This is the case for the significant wave height instrumental record (gray line) associated with Bilbao buoy, shown in Figure 2. Their corresponding annual maxima (triangle dots) and peaks over the threshold $u = 4.2$ m (circle dots) are also shown. Note that the latter correspond to maximum values during independent storms. The independence assumption is considered assuming that the minimum distance in time between peaks must be 3 days. This data set consists of an hourly time series of significant wave height in meters from February 21, 1985 to July 13, 2009.

Significant wave height is a very important parameter for harbor design. Average conditions of significant wave height are relevant to analyze operat-



(a)



(b)

Figure 1: Graphical representation of the point-in-time and extreme regimes for: a) an standard normal and b) a gamma distributed stochastic processes.

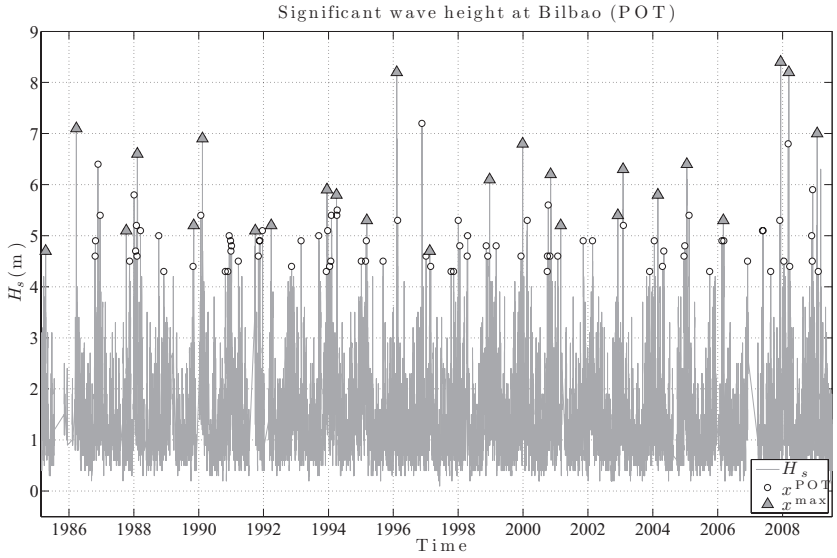


Figure 2: Hourly significant wave height record at Bilbao buoy from February 21, 1985 to July 13, 2009, annual maxima and peaks over the threshold $u = 4.2$.

ing conditions for ships, whereas extreme significant wave heights are used for the stability design of protection structures, such as, vertical breakwaters, dikes, etc.. Thus the importance of characterizing both distributions.

We fit both the significant wave height record and the corresponding peaks over the selected threshold to different parametric distributions: i) a Gaussian Mixture with 4 components for the point-in-time distribution, and ii) a POT model for the annual maxima (extreme-value) distribution, it is possible to plot i) the histograms, ii) the fitted densities, iii) the empirical cumulative and iv) fitted cumulative distributions (see Figure 3). Note that they all present very good fitting diagnostic plots. However, it is difficult to establish whether the fitted point-in-time distribution is capable of reproducing the tail of interest.

If data and fits from Bilbao buoy are plotted using the proposed graphical representation, results shown in Figure 4 are obtained. Note that this representation allows establishing the range of validity of the fitted point-in-time distribution, which starts distorting results above 4.8 meters of significant wave height approximately. The hourly probability of not exceeding this value within the year is 0.996. Above these quantile and probability thresh-

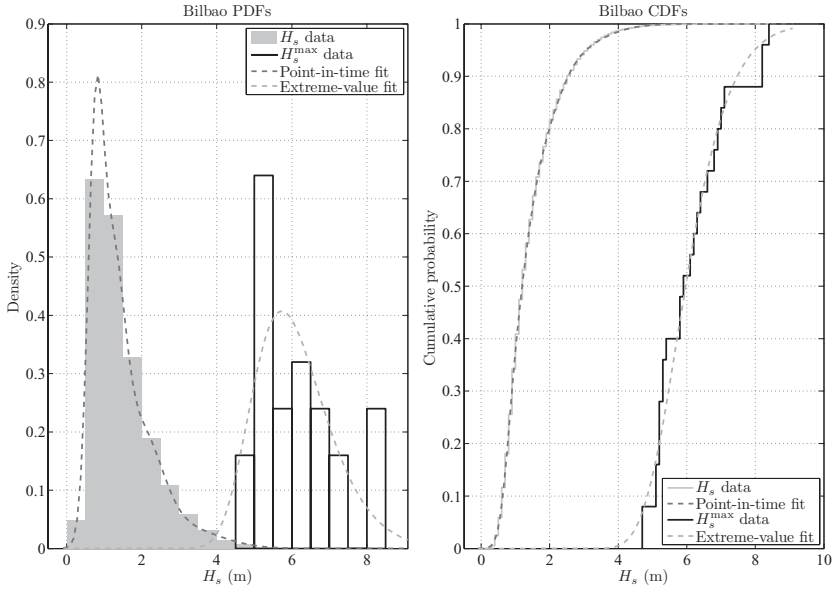


Figure 3: Graphical representation of the point-in-time and extreme (annual maxima) distributions for the significant wave height record at Bilbao buoy.

olds, the point-in-time distribution is no longer valid. It can be observed that the extreme value fit allows reproducing appropriately the tail of the distribution, especially for long equivalent return periods.

These results confirm the appropriateness of using the graphical representation to help understanding the relationship between both the point-in-time and extreme regimes, posing a new challenge for LEVEL III reliability methods based on Monte Carlo simulation techniques: is it possible to simulate, from given point-in-time and extreme-value fitted distributions, samples reproducing both regimes simultaneously? The answer to this question is given in the next section.

4. Point-in-time and extreme-value simultaneous Monte Carlo simulation technique

Consider the stochastic process X_t, whose point-in-time and extreme-value probability distributions are $F^{\mathrm{PT}}(x)$ and $F^{\mathrm{EV}}(x)$, respectively. In Figure 5 the PDFs and CDFs of both distributions in case of maxima are plotted. The first important issue in order to reproduce both distributions is to select

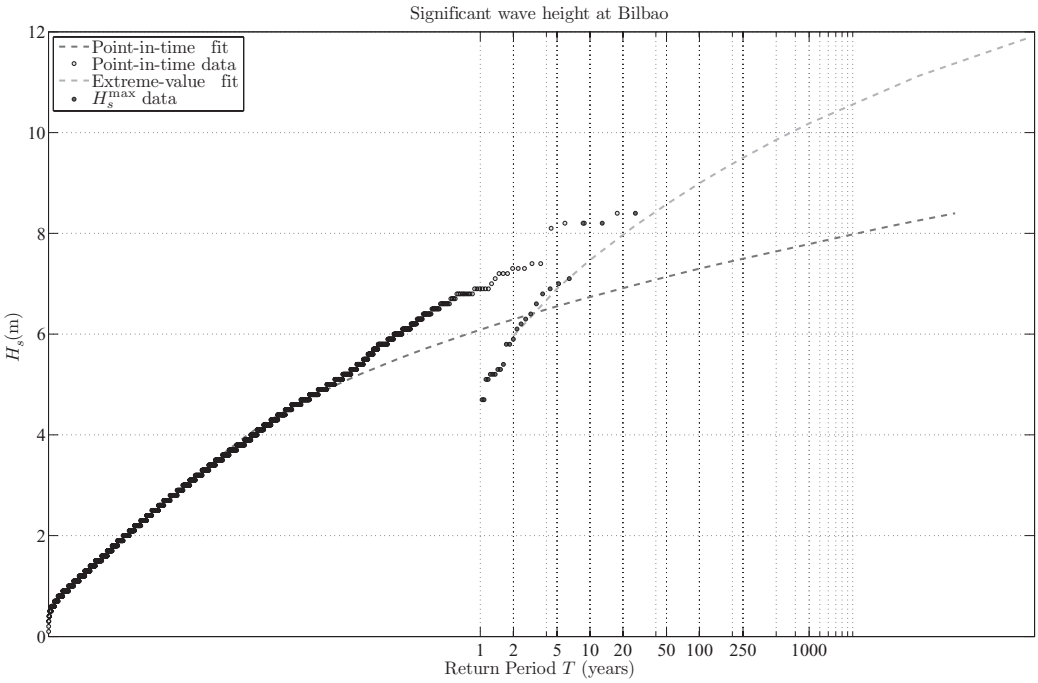


Figure 4: Graphical representation of the point-in-time and extreme distributions for the significant wave height record at Bilbao buoy using the proposed method.

the threshold x_{lim}, this limit corresponds to the maximum value which is governed by the point-in-time distribution. From the practical point of view, it is established based on the proposed graphical representation, as shown in panel left-below from Figure 5, being the x-value whose associated return periods T^{PT} and T_r^{EV} are closer. This condition can be mathematically defined as:

$$\underset{x}{\text{Minimize}} \quad \left(T^{\text{PT}} - T_r^{\text{EV}}\right)^2, \tag{4}$$

which in case of dealing with maxima becomes:

$$\underset{x}{\text{Minimize}} \quad \left(\frac{1}{1 - F^{\text{PT}}(x)} - \frac{n}{1 - F^{\text{EV}}(x)}\right)^2. \tag{5}$$

Note that in case both regimes intersect, as it is shown in left-bottom panel from Figure 5, the optimal solution from problem (4) corresponds to zero, i.e. x_{lim} is the solution of the implicit equation $T^{\text{PT}} = T_r^{\text{EV}}$. Nevertheless, we advocate this approach to overcome the difficulties of solving the implicit equation for those cases where there is no solution (no intersection of regimes). In case of multiple solutions, we take the minimum solution if we are dealing with maxima.

The probability of not exceeding the maximum value x_{lim} within the point-in-time distribution is equal to $p_{\text{lim}}^{\text{PT}} = F^{\text{PT}}(x_{\text{lim}})$, thus the simulation technique uses $F^{\text{PT}}(x)$ for probabilities lower than or equal to $p_{\text{lim}}^{\text{PT}}$ (which is equivalent to x lower than x_{lim}), and $F^{\text{EV}}(x)$ otherwise. However, for the extreme regime the probability must be re-scaled considering:

1. The extreme distribution is related to the maximum of n elements from the point-in-time distribution.
2. There is a probability $p_{\text{lim}}^{\text{EV}} = F^{\text{EV}}(x_{\text{lim}})$ of not exceeding the x_{lim}-value within the extreme distribution, which is usually different from zero. This is the case shown in Figure 5. Thus, those values are not sampled again because they are already considered within the point-in-time distribution.

Finally, when dealing with maxima, and for given uniformly distributed random number u^{PT} representing a probability, the corresponding simulated

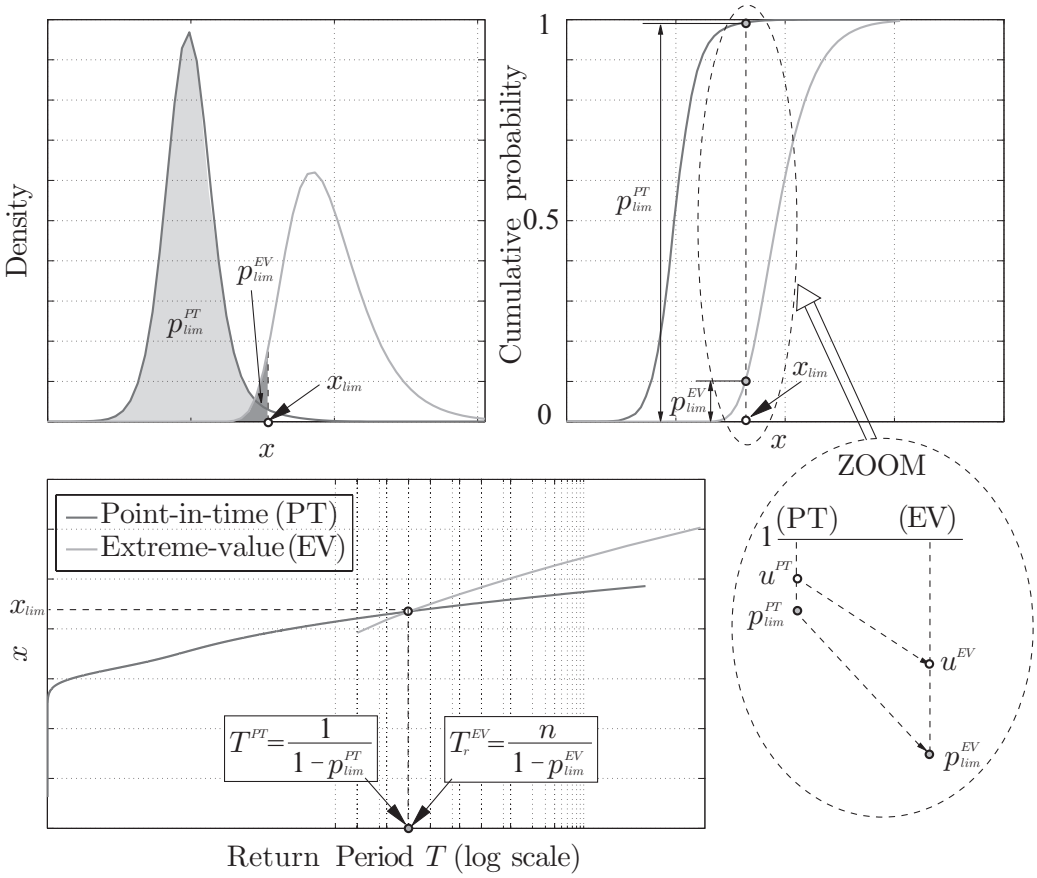


Figure 5: Graphical illustration of the simulation process.

value is obtained as follows:

$$x = \begin{cases} F^{\mathrm{PT}^{-1}}\left(u^{\mathrm{PT}}\right) & \text{if} \quad u^{\mathrm{PT}} \leq p_{\mathrm{lim}}^{\mathrm{PT}} \ (x \leq x_{\mathrm{lim}}) \\ F^{\mathrm{EV}^{-1}}\left(u^{\mathrm{EV}}\right) & \text{if} \quad u^{\mathrm{PT}} > p_{\mathrm{lim}}^{\mathrm{PT}} \ (x > x_{\mathrm{lim}}), \end{cases} \tag{6}$$

$$x = \begin{cases} F^{\mathrm{PT}^{-1}}\left(p^{\mathrm{PT}}\right) & \text{if} \quad p^{\mathrm{PT}} \leq p_{\mathrm{lim}}^{\mathrm{PT}} \ (x \leq x_{\mathrm{lim}}) \\ F^{\mathrm{EV}^{-1}}\left(p^{\mathrm{EV}}\right) & \text{if} \quad p^{\mathrm{PT}} > p_{\mathrm{lim}}^{\mathrm{PT}} \ (x > x_{\mathrm{lim}}), \end{cases} \tag{7}$$

where the re-scaled probability, is equal to:

$$u^{\mathrm{EV}} = p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{u^{\mathrm{PT}} - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}). \tag{8}$$

$$p^{\mathrm{EV}} = p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{p^{\mathrm{PT}} - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}). \tag{9}$$

The bottom-right panel of Figure 5 shows the graphical interpretation of the probability re-scaling, which constitutes a distorted zoom of the panel above. Note that expression (6) allows reproducing both the point-in-time and extreme-value distributions simultaneously.

To shown the functioning of the proposed simulation technique, 1000 years of hourly significant wave height data has been sampled using (6) and the fitted distributions at Bilbao buoy location. For this particular case $n = 8766$ corresponds to the mean number of hours per year used to evaluate the annual maxima.

For the significant wave record, the solution of equation (5) is $x_{\mathrm{lim}} = 6.404$, and the associated probabilities are $p_{\mathrm{lim}}^{\mathrm{PT}} = 0.99996$ and $p_{\mathrm{lim}}^{\mathrm{EV}} = 0.66$. These values correspond to return periods $T_{\mathrm{lim}} \approx 25817$ hours and $T_{\mathrm{lim}} = 2.94$ years, respectively, which are equivalent.

Results from the simulation process are shown in Figure 6. Note that the sample fits appropriately the point-in-time distribution up to the probability related to $T_{\mathrm{lim}} = 2.94$ years return period, and finally the data fits to the extreme distribution for larger return periods. In addition, results related to the annual maxima are also shown. Note also the good fitting shown with respect to the theoretical extreme value distribution above $T_{\mathrm{lim}} = 2.94$ years return period.

These results confirm the validity and good performance of the proposed procedure.

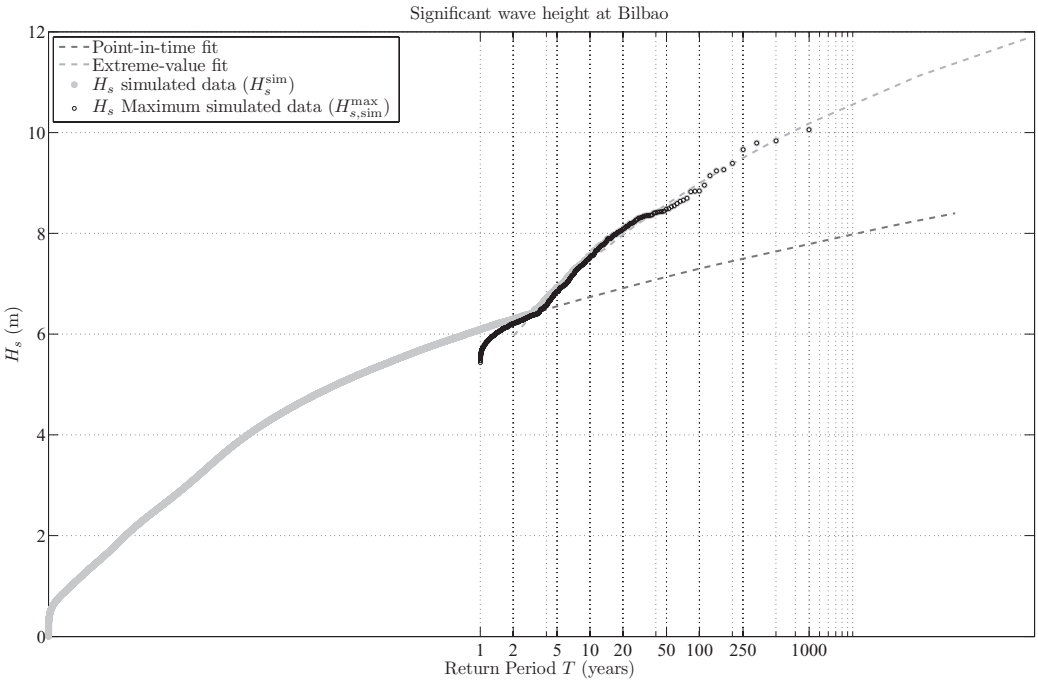


Figure 6: Graphical representation of the point-in-time and extreme simulation results for the significant wave height at Bilbao buoy location using the proposed method.

5. Autocorrelation

The Monte Carlo simulation method provided in the previous section focusses on the marginal distribution associated with an stochastic process. However, it has been recognized by different authors the importance of the temporal correlation of any stochastic process, or even the cross correlation between different stochastic processes (see, for instance, [33, 14, 29]).

An appropriate description of any stochastic process requires recognizing its time dependent nature. For this particular case, the proposed method given in (6) is combined with results from [28, 29]. Basically, the method encompasses the following sequential procedure:

1. Using the point-in-time marginal distribution function, transform the time series of historical values $\boldsymbol{x}_t$ into a normalized Gaussian time series using the following transformation [34]:

$$\Phi(\boldsymbol{z}_t) = F^{\mathrm{PT}}(\boldsymbol{x}_t). \tag{10}$$

   Transformation (10) allows preserving the marginal distribution of the random variables involved.
2. Fitting of a time series model (e.g., an ARMA process) to the transformed historical values obtained in step 1 above. The obtained model allows taking into account temporal correlations.

The time series theory based on *autoregressive moving average* (ARMA) models allows incorporating the temporal structure. An ARMA(p, q) process $\boldsymbol{Z}$ is mathematically expressed as

$$z_t = \sum_{j=1}^{p} \phi_j z_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \tag{11}$$

where $\phi_i;\ i = 1, \ldots, p$ are the autoregressive parameters, and $\theta_j;\ j = 1, \ldots, q$ are the moving average parameters. The term ε_t stands for an uncorrelated normal stochastic process with mean zero and variance σ_ε^2, and it is also uncorrelated with $z_{t-1}, z_{t-2}, \ldots, z_{t-p}$. This process is so-called *white noise*, *innovation term*, or *error term*.

Observe in (11) that z_t boils down to a linear combination of white noises, and as such, the marginal distribution associated with the stochastic process $\boldsymbol{Z}$ is necessarily normal, which is in accordance with the first assumption (10).

Note that a stationary process is assumed. In case of dealing with seasonal behaviors, which could wreck the stationarity, non-stationary probability distributions could be used instead [35, 36, 37, 38] to easily overcome this difficulty.

It is important to point out that only the data belonging to the point-in-time distribution is used to characterize the autocorrelation structure of the stochastic process, because the extreme data is by definition independent, and has no information about autocorrelations. Note also that using ARMA models, only second-moment properties are preserved by the autocorrelation, and this might not provide a complete description in the case of a non-Gaussian process. That is the reason why non-gaussian processes are transformed into gaussian processes using (10). References [28, 29] proved that this approach reproduces autocorrelations in the original domain with a high degree of accuracy.

5.1. The algorithm

Once the parameters of the ARMA model are estimated from the transformed time series, it is very simple to incorporate the autocorrelation structure to the final series. The overall method consists of the following sequential procedure:

- *Step 1:* Estimate the parameters of the probability distributions that best fits both the point-in-time and the extreme-value regimes. This is done using the available historical data.

- *Step 2:* Apply transformation (10) to the historical time series using the point-in-time marginal cumulative distribution function. This way, a *transformed* series is obtained with an associated standard normal marginal distribution.

- *Step 3:* Adjust a univariate ARMA model to the corresponding *transformed* series (obtained in Step 2 above). The fitting process to be performed in this step is well known (see, e.g., [39]) and yields uncorrelated normal residuals (historical errors) with zero mean and constant variance σ_ε^2.

- *Step 4:* Simulate independent normal errors with zero mean and variance σ_ε^2, i.e. $\boldsymbol{\varepsilon}_t^{\text{sim}}$.

- *Step 5:* Introduce the simulated error series $\boldsymbol{\varepsilon}_t^{\text{sim}}$ into the ARMA model fitted in Step 3, obtaining $\boldsymbol{z}_t^{\text{sim}}$.
- *Step 6:* Calculate the time series $\boldsymbol{u}_t^{\text{sim}} = \Phi(\boldsymbol{z}_t^{\text{sim}})$, which is uniformly distributed.
- *Step 7:* In this step, the inverse transformation (6) is applied to this serie in order to enforce the actual marginal distribution that has been estimated in Step 1.

Note that the method proposed in this paper has the following advantages with respect to existing Monte Carlo simulation methods:

1. It reproduces the autocorrelation function as in [21] or [22].
2. It preserves the statistical properties of the stochastic process in terms of the marginal distribution, reproducing appropriately not only the central part of the distribution (point-in-time) but also the right-tail (extremes).

In addition, as proposed in [29], the method could be easily extended to simulate different stochastic processes at the same time. This would allow replicating the main cross-correlations coefficients characterizing those stochastic processes, and not just the contemporaneous. However, since it is not clear how the cross correlation in the point-in-time and extreme distributions behaves, this is a subject for further research.

5.2. Illustrative example

To show the functioning of the proposed algorithm, the following ARMA process $(1,1)$ with parameters $\phi_1 = -0.8$, $\theta_1 = 0.3$, and variance $\sigma_\varepsilon^2 = 1$ is considered. According to [40], the variance of the process is $\sigma_Z^2 = \dfrac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_\varepsilon^2 \approx 4.3611$ ($\sigma_Z \approx 2.0883$). One hundred years of hourly data $\boldsymbol{x}^{\text{sim}}$ ($n = 100 \times 24 \times 365.25 = 876600$) is sampled from this stochastic process. This sample is considered as our initial data set. The idea is to use this sample and the algorithm presented in the previous subsection to generate one thousand years ($n_y = 1000$) of hourly data $\boldsymbol{y}^{\text{sim}}$ considering the autocorrelation, the fitted point-in-time and extreme-value distributions, and compare results with the reference values from the original stochastic process and the initial sample data $\boldsymbol{x}^{\text{sim}}$.

The results of the application of the algorithm are the following:

- *Step 1:* The point-in-time distribution of $\boldsymbol{x}^{\text{sim}}$ is normal, and its estimated parameters and 95% confidence bands are:

$$\begin{array}{rcl} \hat{\mu} & = & 2.7840 \times 10^{-4}\ (-0.0041, 0.0047) \\ \hat{\sigma} & = & 2.0910\ (2.0879, 2.0941), \end{array} \tag{12}$$

  which both contain the true values 0 and 2.0883, respectively. For the extreme value distribution, the annual maxima from the sample $\boldsymbol{x}^{\text{sim}}$, i.e. $\boldsymbol{x}^{\max}_{\text{sim}}$, follows a GEV distribution with estimated parameters and 95% confidence bounds:

$$\begin{array}{rcl} \hat{\mu}_e & = & 7.7050\ (7.7007, 7.7093) \\ \hat{\psi} & = & 0.6020\ (0.5989, 0.6051) \\ \hat{\xi} & = & -0.0178\ (-0.0227, -0.0129). \end{array} \tag{13}$$

- *Step 2:* Apply transformation (10) to the historical time series (x^{sim}) using the normally distributed point-in-time marginal distribution.

- *Step 3:* Adjust an univariate ARMA model to the corresponding *transformed* series $\boldsymbol{z}^{\text{sim}}$, obtaining the following parameter estimates: $\hat{\phi} = -0.8011$ and $\hat{\theta} = 0.2984$. The residuals standard deviation is $\hat{\sigma}_\varepsilon = 0.4781$.

Figure 7 shows the proposed graphical interpretation applied to the sample data $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x}^{\max}_{\text{sim}}$. Dark gray line corresponds to (T^{PT}, x) for the point-in-time fitted distribution. Light gray line corresponds to (T^{EV}_r, x) associated with the GEV fitted distribution for annual maxima. Note that both fitted distributions differ at the right tail of the distribution (see the corresponding zoom in the figure, where data has been remove to ease visualization), which is usually the case when fitting real data. Light gray circle dots correspond to the sample data values, and black dots are related to sample annual maxima.

- *Step 4:* In order to obtain $n_y = 1000$ years of hourly data, $n = 1000 \times 365.25 \times 24$ independent normal errors are sampled with standard deviation $\hat{\sigma}_\varepsilon = 0.4781$, i.e. $\boldsymbol{\varepsilon}^{\text{sim}}_y$.

- *Step 5:* Introduce the sampled error time series into the ARMA model fitted in Step 3, obtaining $\boldsymbol{z}^{\text{sim}}_y$.

- *Step 6:* Calculate the uniformly distributed time series of probabilities $\boldsymbol{u}^{\text{sim}}$.

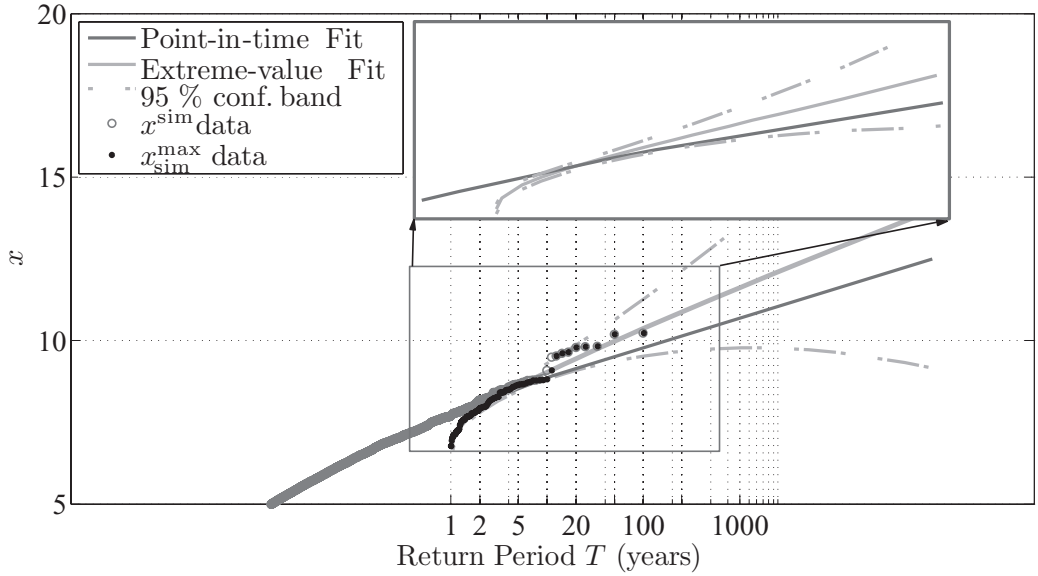


Figure 7: Graphical representation of the point-in-time and extreme regimes for the illustrative autocorrelated normal stochastic process.

- *Step 7:* In this step, the inverse transformation (6) is applied to get $\boldsymbol{y}^{\text{sim}}$. Note that for this particular example, the solution of equation (5) is $x_{\text{lim}} = 8.3396$, and the associated probabilities are $p^{\text{PT}}_{\text{lim}} = 0.9999667$ and $p^{\text{EV}}_{\text{lim}} = 0.7082156$. These values correspond to return periods $T_{\text{lim}} = 30042.7$ hours and $T_{\text{lim}} = 3.427$ years, respectively, which are equivalent.

The graphical illustration of the 1000 years simulated sample is given in Figure 8. Note that it shows the same results as Figure 7 but replacing the sample data $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x}^{\max}_{\text{sim}}$ used to fit the distributions, by the 1000 years simulated samples $\boldsymbol{y}^{\text{sim}}$ and $\boldsymbol{y}^{\max}_{\text{sim}}$ using the proposed procedure. Note the accuracy of the method to reproduce both the point-in-time distribution up to $x_{\text{lim}} = 8.3396$, where the simulated sample starts following the extreme-value distribution. In addition, several tests have been performed to check simulation results :

1. For the point-in-time distribution, a two-sample Kolmogorov-Smirnov test with 0.05 significance level is performed to compare the distributions of the initial sample data ($\boldsymbol{x}^{\text{sim}}$) and the simulation results ($\boldsymbol{y}^{\text{sim}}$).

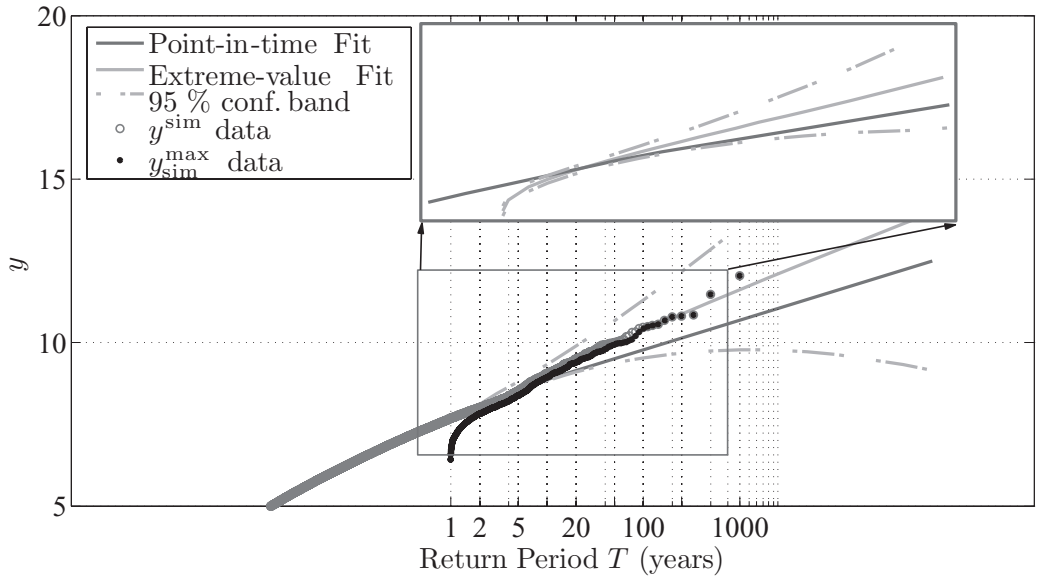


Figure 8: Graphical representation of the point-in-time and extreme regimes for the illustrative autocorrelated normal stochastic process.

Note that the p-value obtained is 0.5966, which is higher than the significance level, i.e. the null hypothesis that both samples come from the same continuous distribution is accepted.

2. Analogously, the two-sample Kolmogorov-Smirnov test is applied to compare the samples related to annual maxima, i.e. $\boldsymbol{x}^{\max}_{\text{sim}}$ versus $\boldsymbol{y}^{\max}_{\text{sim}}$. Note that the p-value obtained is 0.1183, so that the null hypothesis that both samples come from the same extreme-value continuous distribution is accepted.
3. Finally, an ARMA model is fitted to the simulated sample, obtaining the following parameter estimates: $\hat{\phi} = -0.8002$ and $\hat{\theta} = 0.2998$. The corresponding residuals standard deviation is $\hat{\sigma}_{\varepsilon} = 1.00015$, which almost coincide with the one from the initial ARMA process.

These results confirm the appropriate performance of the proposed procedure to reproduce i) the point-in-time and extreme-value distributions and ii) the temporal dependence structure of any stochastic process.

6. New insights into structural reliability methods

Besides providing a new Monte Carlo simulation method for dealing with point-in-time and extreme-value distributions, new insights about how to incorporate this methodology within alternative reliability analysis methods, such as *First Order Reliability Methods* (FORM), are also given. Note that we assume that the reader is familiar with LEVEL III methods ([41, 42, 43, 44, 45]) for evaluating the reliability index associated with any mode of failure:

$$\beta = \underset{\boldsymbol{z}}{\text{Minimum}} \quad \sqrt{\sum_{\forall i} z_i^2} \tag{14}$$

subject to

$$g(\boldsymbol{x}, \boldsymbol{\eta}) = 0, \tag{15}$$

$$\mathbf{T}(\boldsymbol{x}, \boldsymbol{\eta}) = \boldsymbol{z}, \tag{16}$$

where $g(\boldsymbol{x}, \boldsymbol{\eta}) = 0$ is the failure condition, and $\mathbf{T}(\boldsymbol{x}, \boldsymbol{\eta})$ is the transformation ([46]) giving the values of the standard and independent normal variables $\boldsymbol{z}$ as a function of the values of the random $\boldsymbol{x}$ and design $\boldsymbol{\eta}$ variables. The probability of failure p_f is related to the reliability index by the approximate relation $p_f = \Phi(-\beta)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

The key issue when dealing with structural risk problems where the point-in-time and extreme-value distributions may coexist, is to decide which one is more relevant for the corresponding limit state. The Monte Carlo method proposed in this paper deals with the simulation process giving more importance to the point-in-time probability distribution, and it uses the re-scaled extreme-value regime to improve accuracy in the right tail of the distribution. According to (6)-(8) and considering $u = \Phi(z^{\mathrm{PT}})$, the Rosenblatt transformation (16) becomes:

$$\begin{aligned} \Phi(z^{\mathrm{PT}}) &= F^{\mathrm{PT}}(x) \quad \text{if} \quad x \leq x_{\mathrm{lim}} \text{ or } z^{\mathrm{PT}} \leq z_{\mathrm{lim}} \\ p_{\mathrm{lim}}^{\mathrm{EV}} + \frac{\Phi(z^{\mathrm{PT}}) - p_{\mathrm{lim}}^{\mathrm{PT}}}{1 - p_{\mathrm{lim}}^{\mathrm{PT}}}(1 - p_{\mathrm{lim}}^{\mathrm{EV}}) &= F^{\mathrm{EV}}(x) \quad \text{if} \quad x > x_{\mathrm{lim}} \text{ or } z^{\mathrm{PT}} > z_{\mathrm{lim}}, \end{aligned} \tag{17}$$

where $z_{\mathrm{lim}} = \Phi^{-1}\left(p_{\mathrm{lim}}^{\mathrm{PT}}\right)$. It is important to point out that transformation (17) takes into account the point-in-time distribution, but improving accuracy on

the upper tail by using the re-scaled extreme-value distribution. Probabilities of failure obtained from this approach are related to the point-in-time frequency sampling, i.e. hours.

Alternatively, if only the extreme-value distribution is considered, transformation (16) becomes:

$$\Phi(z^{\mathrm{EV}}) = F^{\mathrm{EV}}(x)\,. \tag{18}$$

In this case, probabilities of failure are associated with the extreme-value frequency sampling, i.e. years.

From the practical point of view, we advocate the use of transformation (17) and consider probabilities related to the point-in-time frequency sampling, because it allows the consideration of any kind of limit state equation. However, it is important to define the maximum probabilities of failure for each failure mode in terms of the point-in-time frequency sampling. For instance, if an inner harbor must be designed so that ships might not maneuver during no more than 1000 hours per year, then the acceptable probability of failure must be equal to $p_f = 1000/(365.25 \times 24)$. Besides, if the offshore breakwater of the same harbor must be designed to fail on average once every 25 years, the acceptable probability of failure must be equal to $p_f = 1/(25 \times 365.25 \times 24)$. Considering those probability values, transformation (17) may be used for both operating and ultimate limit states without any further consideration.

7. Conclusions

The method proposed in this paper provides new insights on the relationship between the point-in-time and extreme-value distributions associated with any stochastic process, and a possible way to deal with both distributions at the same time. The advances with respect to the state-of-the-art can be summarized as follows:

1. A new graphical representation to help understanding the relationship between both distributions is proposed.
2. A new Monte Carlo simulation technique holding the following requirements is provided:
    - It is able to reproduce both the point-in-time and extreme-value regimes.

- It maintains the temporal dependence structure of the stochastic process through ARMA models.

3. In addition, some hints about extending the method into FORM techniques are given. In this case, the method frees the engineer to decide about what regime should be used within the design process.

All the methods have been tested using different synthetically generated samples and an example based on real data. Results confirm the good behavior of the proposed methods, and their suitability to i) support engineers on the design process and ii) help understanding the relationship between both point-in-time and extreme-value regime.

Further research must be done on cross-correlations between different stochastic processes, however, this paper constitute a clear advance on the knowledge of point-in-time and extreme-value distributions.

Note that although all the material developed in this paper is related to the upper tail of the point-in-time distribution (maxima), alternative formulations can be straightforwardly obtained for dealing with minima.

Acknowledgements

This work was partly funded by projects "AMVAR" (CTM2010-15009) from Spanish Ministry MICINN, and by project MARUCA (E17/08) from the Spanish Ministry MF. Y. Guanche is indebted to the Spanish Ministry of Science and Innovation, FPI Program (BES-2009-027228). R. Mínguez is also indebted to the Spanish Ministry MICINN for the funding provided within the "Ramon y Cajal" program. The authors thank Puertos del Estado (Spanish State Port) for providing the buoy data information.

References

[1] M. A. Losada, Recent development in the design of mound breakwaters, in: J. Herbich (Ed.), Chapter 21 in Handbook of Coastal and Ocean Engineering, Vol. I, Gulf Publishing, 1990.

[2] ROM 0.0, Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias., Tech. rep., Puertos del Estado, Madrid, España, pp 245 (Noviembre 2001).

[3] M. J. Baker, Evaluation of partial safety factors for level i codes. Example of application of methods to reinforced concrete beams, in: Bulletin d'Information No. 112, Comite Européen due Béton, Vol. 112, Paris, 1976, pp. 190–211.

[4] N. C. Lind, Application to design of level i codes, in: Bulletin d'Information No. 112, Comite Européen due Béton, Vol. 112, Paris, 1976, pp. 73–89.

[5] M. R. Horne, P. P. H., Commentary on the level 2 procedure, rationalization of safety and serviceability factors in structural codes, Report 63, Construction Industry Research and Information Association, London (1977).

[6] E. Castillo, Extreme Value Theory in Engineering, Academic Press, New York, 1988.

[7] S. Coles, An introduction to statistical modeling of extreme values, Springer Series in Statistics, 2001.

[8] E. Castillo, A. S. Hadi, N. Balakrishnan, J. M. Sarabia, Extreme Value and Related Models in Engineering and Science Applications, John Wiley & Sons, New York, 2005.

[9] E. Castillo, C. Castillo, R. Mínguez, Use of extreme value theory in engineering design, in: G. S. C. Martorel, S., J. Barnett (Eds.), Proceedings of the European Safety and Reliability Conference 2008 (ESREL 2008), Safety, Reliability and Risk Analysis: Theory, Methods and Applications, Vol. 3, Taylor & Francis Group, Valencia, 2008, pp. 2473–2488.

[10] A. Frigessi, O. Haug, H. Rue, A dynamic mixture model for unsepervised tail estimation without threshold selection, Extremes 5 (2002) 219–235.

[11] B. Vaz de Melo Mendes, H. Freitas Lopes, Data driven estimates for mixtures, Computational Statistics and Data Analysis 47 (2004) 583–598.

[12] B. Behrens, H. F. Lopes, D. Gamerman, Bayesian analysis of extreme events with threshold estimation, Statistical Modelling 4 (2004) 227–244.

[13] A. Tancredi, C. Anderson, A. Ohagan, Accounting for threshold uncertainty in extreme value estimation, Extremes 9 (2006) 87–106.

[14] Y. Cai, B. Gouldby, P. Hawkes, P. Dunning, Statistical simulation of flood variables: incorporating short-term sequencing, J. Folld Risk management 1 (2008) 3–12.

[15] E. M. Furrer, R. W. Katz, Improving the simulation of extreme precipitation events by stochastic weather generators, Waters Resources Research 44 (12) (2001) 1–13.

[16] S. Solari, Metodologías de simulación de agentes naturales y desarrollo de sistemas. Modelo de verificación y gestión de terminales portuarias, Ph.D. thesis, University of Granada, Córdoba y Málaga, Spain (2011).

[17] R. E. Melchers, Structural reliability analysis and prediction, 2nd Edition, John Wiley & Sons, New York, 1999.

[18] B. V. Rubinstein, Simulation and the Monte Carlo Method, John Wiley & Sons, New York, 1981.

[19] P. Bjerager, Probability integration by directional simulation, J. Engineering Mechanics, ASCE 114 (8) (1988) 1285–1302.

[20] B. V. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo Method, 2nd Edition, John Wiley & Sons, New York, 2007.

[21] A. Harbitz, An efficient sampling method for probability of failure calculation, Structural Safety 3 (2) (1986) 109–115.

[22] M. Shinozuka, Stochastic methods in structural dynamics, Stochastic fields and their digital simulations, Martinus Nijhoff, The Hague (1987).

[23] E. Castillo, C. Solares, P. Gómez, Estimating extreme probabilities using tail simulated data, International Journal of Approximate Reasoning: official publication of the North American Fuzzy Information Processing Society 17 (2-3) (1996) 163–190.

[24] O. Ditlevsen, R. Olesen, G. Mohr, Solution of a class of load combination problems by directional simulation, Structural Safety 4 (1987) 95–109.

[25] O. Ditlevsen, P. Bjerager, Plastic reliability analysis by directional simulation, J. Engineering Mechanics Div., ASCE 115 (6) (1989) 1347–1362.

[26] R. E. Melchers, Improved importance sampling for structural reliability calculation, in: Proceedings 5th International Conference on Structural Safety and Reliability, ASCE, New York, 1989, pp. 1185–1192.

[27] R. E. Melchers, Simulation in time-invariant and time-variant reliability problems, in: Proceedings 4th IFIP Conference on Reliability and Optimization of Structural Systems, Springer, Berlin, 1991, pp. 39–82.

[28] C. Guedes Soares, C. Cunha, Bivariate autoregressive models for the time series of significant wave height and mean period, Coastal Engineering 40 (2000) 297–311.

[29] J. M. Morales, R. Mínguez, A. J. Conejo, A methodology to generate statistically dependent wind speed scenarios, Applied Energy 87 (2010) 843–855, 10.1016/j.apenergy.2009.09.022.

[30] S. Solari, M. A. Losada, Non-stationary wave height climate modeling and simulation, Journal of Geophysical Research 116 (2011) –. doi:10.1029/2011JC00711.

[31] J. Galambos, The Asymptotic Theory of Extreme Order Statistics, Robert E. Krieger Publishing Company, Malabar, Florida, 1987.

[32] R. A. Fisher, L. H. C. Tippett, Limiting forms of the frequency distributions of the largest or smallest member of a sample, Journal of Coastal Research 24 (1928) 180–190.

[33] C.-H. Chang, Y.-K. Tung, J. C. Yang, Monte carlo simulation for correlated variables with marginal distributions, Journal of Hydraulic Engineering 120 (3) (1994) 313–331.

[34] A. Nataf, Détermination des distribution de probabilités dont les marges sont données, Comptes Rendus de l'Academie des Sciences 225 (1962) 42–43.

[35] F. J. Méndez, M. Menéndez, A. Luceño, I. J. Losada, Estimation of the long term variability of extreme significant wave height using a time-dependent POT model, J. of Geophys. Res. 111, doi:10.1029/2005JC003344.

[36] F. J. Méndez, M. Menéndez, A. Luceño, I. J. Losada, Analyzing monthly extreme sea levels with a time-dependent gev model, J. Atmos. Ocean. Technol. 24 (2007) 894–911.

[37] M. Menéndez, F. J. Méndez, C. Izaguirre, I. J. Losada, The influence of seasonality on estimating return values of significant wave height, Coastal Engineering 56 (3) (2009) 211–219.

[38] R. Mínguez, F. J. Méndez, C. Izaguirre, M. Menéndez, I. J. Losada, Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables, Environmental Modelling & Software 25 (2010) 1592–1607. doi:DOI: 10.1016/j.envsoft.2010.05.008.

[39] D. Peña, G. C. Tiao, R. S. Tsay, A Course in Time Series Analysis, Probability and Statistics, Wiley, New York, NY, 2001.

[40] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time Series Analysis: Forecasting and Control, Prentice-Hall International, New Jersey, NJ, 1994.

[41] A. M. Freudenthal, Safety and the probability of structural failure, Transactions, ASCE 121 (1956) 1337–1397.

[42] A. M. Hasofer, N. C. Lind, Exact and invariant second moment code format, J. Engrg. Mech. 100 (EM1) (1974) 111–121.

[43] R. Rackwitz, B. Fiessler, Structural reliability under combined load sequences, Comput. Struct. 9 (1978) 489–494.

[44] M. Hohenbichler, R. Rackwitz, Non-normal dependent vectors in structural safety, J. Engineering Mechanics Div., ASCE 107 (6) (1981) 1227–1238.

[45] O. Ditlevsen, Principle of normal tail approximation, J. Engineering Mechanics Div., ASCE 107 (6) (1981) 1191–1208.

[46] M. Rosenblatt, Remarks on a multivariate transformation, Ann. Math. Stat. 23 (3) (1952) 470–472.