Two Methods for Simulating the Bivariate Process of Wave Height and Direction

P. Ailliot and M. Prevosto
IFREMER
Brest, France

ABSTRACT

This paper describes two procedures for generating time sequences of sea state parameters. The results of these simulations could, for example, provide us scenarios for erosion modelling, or help us to decide the feasibility of an operation at sea.

The first procedure is based on the normal score transformation and has already been used by several authors for generating times sequences of wave height and period. In the second one we privilege the simulation of the significant height process and we try to take the specificity of the directional parameter into account: we first simulate the significant wave height process and then we use a Markov chain model to simulate the sea state direction.

After a presentation of these two methods, we will compare their abilities to respect the main statistical properties of the initial time series, namely the marginal and the joint density functions, the extreme values statistics and the duration statistics.

KEYWORDS

Sea state parameters, stochastic simulation, non Gaussian processes, marked point process, Markov chains.

INTRODUCTION

Time sequences of sea state conditions can be used, for example, as input in erosion models or for estimating the feasibility of offshore operations. These data can be obtained by direct measurements (buoys, satellites,...) or by models (hindcast...). These data are often available on a too short period of time to obtain accurate statistics. Thus, simulating realistic sea state histories on a longer period is necessary.

Several authors have been working on this problem. For instance, Scheffner and Borgman (1992) have proposed a method for simulating the trivariate process (\(H_s, T_p, \Theta_m\)) where \(H_s\) represents the significant wave height, \(T_p\) the spectral peak period and \(\Theta_m\) the mean direction of the sea state: they first transform this process in a Gaussian stationary process and then they simulate this Gaussian process. Monbet and Prevosto (1999) have proposed a method for simulating \((H_s, T_p)\), which is built on the same principle, but which has an improved ability to respect the physical constraints existing between \(H_s\) and \(T_p\).

In this paper, we look at the bivariate process \((H_s, \Theta_m)\), and then we try to fit a model to this time series. The main difficulties we have encountered were due to the specificity of the directional parameter, which takes its values in \([0, 360]\). We will first describe two models: the first one is an adaptation of the one proposed by Monbet and Prevosto, and the second one is based on marked point processes and Markov chains.

Then, we will fit these two models to our data and compare them through their ability to respect the main statistical properties of the data.

The data used in this work consist in a 20-year long time series of hindcast data from the AES North Atlantic Reference Wave Climatology produced by Oceanweather. These data describe the sea state conditions during the years 1979-1998 at several points off the French Atlantic coast. The sea state parameters are provided each six hours.

THE FIRST METHOD

In this part, instead of the polar co-ordinates \((H_s, \Theta_m)\), we have considered the Cartesian co-ordinates:

\[
(X, Y) = (H_s \cos(\Theta_m), H_s \sin(\Theta_m))
\]

Description of the statistical model

As mentioned in the introduction, the process \((X, Y)\) is non stationary and non Gaussian: the model must be able to describe these two characteristics.

The non-stationarity of the process is mainly induced by seasons. We have used a classical model in time series analysis to deal with this seasonal variability (Walton et Borgman (1990), Athanassoulis and Stephanakos (1995)). We will suppose that \(X\) (respectively \(Y\) and \(H_s\)) admits the following decomposition:

\[
X(t) = \bar{X}_{st}(t) + \mu_X(t) + \sigma_X(t)X_{stat}(t)
\]

where \(\bar{X}_{st}, \mu_X\) and \(\sigma_X\) are deterministic functions and \(X_{stat}\) is a stationary process. The function \(\bar{X}_{st}\) is linear and represents the long term trend, whereas the functions \(\mu_X\) and \(\sigma_X\) are periodic, with period one year and represent the seasonal mean value and the seasonal standard deviation, respectively.
The non-Gaussian nature of the residual time series \( \{X_{stat}, Y_{stat}\} \) can be successfully described by a translation model. This model has already been applied for sea state sequences simulation (Scheffner and Borgman (1992)) or for wind pressure simulation (Gioffré et al. (2000)). Let \( Z(t) = (Z_1(t), Z_2(t))^T \) be an ergodic Gaussian process with marginal distributions equal to the standard Gaussian distribution and a function \( \Phi : \mathbb{R}^2 \to \mathbb{R}^2 \). The translation vector \( U(t) = (U_1(t), U_2(t)) \) defined by \( Z \) and \( \Phi \) is given by the memoryless transformation \( U(t) = \Phi(Z(t)) \). In order to fit a translation model to \( \{X_{stat}, Y_{stat}\} \), Scheffner and Borgman proposed to take the normal score transformation \( \Phi \) given by:

\[
\Phi : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} F_{X_{stat}}^{-1}(G(x)) \\ F_{Y_{stat}}^{-1}(G(y)) \end{pmatrix}
\]

(2)

where \( G \) is the distribution function of the standard Gaussian distribution and \( F_{X_{stat}} \) and \( F_{Y_{stat}} \) are the distribution functions of the stationary process \( X_{stat} \) and \( Y_{stat} \), respectively. In order to obtain a better approximation of the joint instantaneous distribution of \( \{X_{stat}, Y_{stat}\} \), Monbet and Prevosto proposed to take:

\[
\Phi : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} F_{X_{stat}}^{-1}(G(x)) \\ F_{Y_{stat}}^{-1}(G(y)) \end{pmatrix}
\]

(3)

where \( F_{Y_{stat}}^{-1}(x) \) represents the distribution function of the conditional distribution of \( Y_{stat} \) given \( X_{stat} = x \).

**Simulation procedure**

According to this model, we can implement a simple procedure to simulate sequences of sea state parameters. This procedure can be divided in the following four steps:

1. **Stationarity procedure**: This first step consists in estimating the functions \( X_{ir}, Y_{ir}, \mu_X, \mu_Y, \sigma_X \) and \( \sigma_Y \). \( X_{ir} \) and \( Y_{ir} \) are estimated by fitting the least squares regression line to the series of the annual mean. In our data base, no significant trend has been found.

2. For the evaluation of the periodic functions, we have applied the methods proposed by Walton and Borgman, which is based on convolution with a Gaussian kernel. Athanasoulis and Stephanakos proposed a parametric model which consists in fitting a low order trigonometric polynomial. This method has the advantage to be parametric. After this first step, we can consider the residual time series defined by the relation (1), and suppose that it is a realisation of a stationary process.

3. **Normalization procedure**: Now, in order to obtain a Gaussian process, we have to estimate the function \( \Phi \). We preferred to take the function defined by (3) because the dependence between \( X_{stat} \) and \( Y_{stat} \) is strong. In order to estimate the distribution functions, we took the classical empirical estimators. Then, with (3) we can calculate a residual bivariate time series which is supposed to be a realisation of a Gaussian and ergodic process.

4. **Simulation of the Gaussian process**: In a third step we will simulate other realisations of this bivariate Gaussian process. First, with the realisation we obtain by transforming the data, we evaluate the spectrum of the Gaussian process; we first calculate the periodogram and then we smooth it with a kernel. Then we have adopted the simulation procedure proposed by Scheffner and Borgman, which permits to simulate realisations of a Gaussian process when the spectrum density is known. Other methods are described in the literature: see Dietrich and Newsman (1997) for a review of this different methods. If we want a parametric model, we can fit an ARMA model as proposed by Stephanakos and Athanasoulis (2000).

**Inverse normalisation and Stationary procedure**: This step consists in applying the inverse transforms of the first two steps to the simulated Gaussian sequence in order to obtain a time series with the same statistical properties than the initial one.

**THE SECOND METHOD**

**Transformation of the process \( H_t \)**

The first step of this method consists in fitting a linear spline curve to the time series \( H_t \). In this domain, several algorithms have been developed by mathematicians. The method chosen will depend on the nature of the time series. In our case, the data obtained with a hindcast model are smooth. This enabled us to use a simple algorithm to detect slope change times: we just take the time when \( H_t \) hits its local extremum.

![fig 1: example of linear curve fitting to \( H_t \)](image)

We can now define a discrete time bivariate process \( (H, D) \) (figure 2). \( D_{t+1} \) will represent the date of the i-th local minimum of \( H_t \) whereas \( D_2 \) will be the first date after \( D_{t+1} \) where \( H_t \) hits a local maximum (the date are expressed in hour with origin the 01/01/1979 at midnight). Of course, \( D_t \) takes its value in \( \mathbb{R}^* \) and is increasing. Then, we can define \( H_t = H_{D_t}(H) \). We have \( H_{2t-1} \leq H_{2t} \) and \( H_{2t-1} \leq H_{2t} \). Thus, \( D_t \) will represent the date of the events (local extremum), whereas \( H_t \) is called a mark and represents the intensity of this event.

From this new process, we can define a discrete time process \( T \) and a continuous time process \( N \) as follow:

\[
\forall t \in \mathbb{R}^*, \quad T_i = D_{t+1} - D_t \quad \forall t > 0 \quad N(t) = \max\{i \in \mathbb{R}^*: D_i < t\}
\]

\( T \) represents the duration between two successive events and \( N(t) \) is a point process which counts the number of local extremum before the date \( t \).

Let \( H_{lin} \) be the linear process we have fitted to the process \( H_t \). The following relation holds for \( T\)-\( D_t \):

\[
H_{lin}(t) = H(N(t)) + \frac{H(N(t)) - H(N(t)+1)}{T(N(t))}(t - D(N(t)))
\]

(4)
We will suppose that the linear curve, $H_{lin}$, we have fitted to our data is a good enough approximation of the initial time series, and thus that simulating the process $H_{lin}$ is sufficient for our applications.

**The statistical model**

We will now present a model for the bivariate process $(H_{lin}, O_m)$. We will first fit a model to $H_{lin}$ or equivalently to $(H, T)$, since the relation (4) holds.

**Model for the marked point process:** The process $(H, T)$ associated to $H_{lin}$ defines a so-called marked point process. This model has still been used for modelling insurance cost (Karyagina et al.) and rainfall (Cowpertwait et al.): they assume that $N$ is a renewal process (it means that $T_1, T_2, \ldots, T_m$) are independent and identically distributed. In our case, there is a significant correlation between two successive values of $T$, and thus we can not make this hypothesis.

Thus, we have made the following Markovian hypothesis:

$$L(H_{k+1}, T_{k+1}/H_k, T_1, \ldots, H_k, T_k) = L(H_{k+1}, T_{k+1}/H_k)$$

(5)

It means that the past of the process $(H, T)$ has no influence on the future of this process if the value of the intensity process $H$ is specified in the present. This condition implies in particular that $(H, T)$ is discrete-time, continuous state Markov chains. However, this Markov chain is not homogeneous. Indeed, the conditional law $L(H_{k+1}, T_{k+1}/H_k)$ depends on the parity of $k$ and of the season.

**Model for $(H_{lin}, O_m)$:** We will first deal with the seasonality of the directional parameter. We will suppose that $O_m$ can be decomposed in the following way:

$$O_m(t) = \mu(t) + O_{m}^{stat}(t)$$

(6)

where $\mu(t)$ is a deterministic function which represents the seasonal mean direction (see Mardia (1972)) and $O_{m}^{stat}$ is a stationary process.

Let $i$ be an integer and $k$ be the only integer which verifies $D_{2k-1} \leq i\Delta T < D_{2k+1}$ ($\Delta T = 6$ hours). We will suppose that:

$$L(O_m(i + 1)\Delta T/O_m(i\Delta T), \ldots, O_m(i\Delta T), H_1(\Delta T), \ldots, H_3(i\Delta T), \ldots) = L(O_m(i + 1)\Delta T/O_m(i\Delta T), H(k))$$

(7)

It means that on the stochastic time interval $[D_{2k-1}, D_{2k+1})$, $O_m$ is a Markov chain whose transitions probabilities depend on $H_{2k}$.

**Simulation procedure**

In order to simulate the bivariate process $(H_{lin}, O_m)$ we just have to estimate the transition probabilities of the Markov chains. As mentioned above, these Markov chains are not homogeneous, because of the seasonality of the process $(H_{lin}, O_m)$. Thus, we have used the following simulation procedure:

**Stationarity procedure:** We transform $(H_{lin}, O_m)$ in a stationary process $(H_{lin}^{stat}, O_m^{stat})$ by using the model (1) for $H_{lin}$ and (6) for $O_m$

**Linearisation of $H_{lin}^{stat}$:** We used the algorithm described above in order to fit a linear curve to $H_{lin}^{stat}$. We get a linear process $H_{lin}^{stat}$ and a marked point process $(H_{lin}^{stat}, T_{stat})$.

**Simulation of $(H_{lin}^{stat}, O_m^{stat})$:** We used the Markovian hypothesis described by (5) and (7) in order to simulate this bivariate process. We evaluate the probability transitions with their usual empirical estimators.

**Inverse stationarity procedure:** This step just consists in applying the inverse transform described by (1) and (6) to the simulated data.

**COMPARISON BETWEEN THE TWO METHODS**

The first method has been widely used by statisticians and has proved to be efficient. Furthermore, it has the advantage to work in a lot of domain and it can be used to simulate more than two parameters. Nevertheless, it assumed that the process obtained after the normal score transformation is Gaussian: this is not necessarily true. Moreover, this model has no easy physical interpretation. That is why we developed the second model.

In order to compare these two methods, we have simulated 20 years of data with these two procedures, and then we have compared the statistical properties of these simulated data with the ones of the initial time series. Because of the seasonality of the data, we have compared them season by season.

At first, we have compared the marginal laws (figure 3A and 5). Then, in order to check the temporal structure of the process, we have looked at two criteria. At first, we have compared the autocovariance functions of the initial data and of the simulated data (figure 6). Then, we have looked at the mean duration of the storms coming from a given direction (figure 7). Let $s$ be a significant wave height, which will represent the severity of the storm and $J$ be an angular interval which represents the direction of the storm. We have then calculated the average length of time that the significant wave height exceeds this threshold $s$ whereas the mean direction stays in $J$. This calculus has been made for different values of $s$ and $J$.

According to these figures, we can say that the marginal laws and the autocovariance functions of the simulated data and of the initial time series are in a good agreement for the months corresponding to the winter (December, January and February). The comparison for the other seasons gives comparable results. The both methods give results of the same quality.

The first method gives better results if we compare the bivariate distribution of $(H_{lin}, O_m)$. However, we get bad results with this method when we look at the duration statistics represented in figure 7. The second method has an improved ability to respect the statistical properties.
of H₄ and the storm duration in a given direction, but poorly restores the bivariate distribution of (H₄, Oₜₙ).

**Fig 3:** Comparison of marginal distribution for H₄ in winter

**Fig 4:** Zoom on the high values of the distribution of H₄ in winter

**Fig 5:** Comparison of marginal distribution for Oₜₙ in winter

**Fig 6:** Comparison of the autocovariance functions for H₄ in winter

**Fig 7:** Mean duration (in hour) of a storm for different severity. J=[0,360] for the first figure and J=[60,140] for the second one.

**CONCLUSION**

We have presented two methods for generating sea state sequences of H₄ and Oₜₙ. This two methods are of comparable complexity and are easy to implement on a computer. The first method respects some classical statistical properties of the process, namely the distribution of (H₄, Oₜₙ) and the autocovariance function. But when we look at the duration statistics, which is an important criterion for the applications, we get insufficient results. On the contrary, the second model is appropriate to respect the storm duration, and respects correctly the other criterions.

**REFERENCES**