**Applied Ocean Research**
June 2001, Volume 23, Issue 3, Pages 139-145
http://dx.doi.org/10.1016/S0141-1187(01)00011-6




# Bivariate simulation of non stationary and non Gaussian observed processes Application to sea state parameters

Valérie Monbet[a, *] and Marc Prevosto[b]

[a] SABRES, UBS, rue Yves Mainguy, 56000 Vannes, France
[b] IFREMER, TMSI/IDM/COM, BP70, 29280 Plouzané, France

*: Corresponding author : V. Monbet, email address : valerie.monbet@univ-ubs.fr

**Abstract:**

A method for artificially generating operational sea state histories has been developed. This is a distribution free method to simulate bivariate non stationary and non Gaussian random processes. This method is applied to the simulation of the bivariate process ($H_s$, $T_p$) of sea state parameters. The time series respects the physical constraints existing between the significant wave height and the peak period. Furthermore, we show that the persistence properties of the simulations match to those of the observations.

**Keywords:** Non parametric simulation; Wave data; Non Gaussian processes; Bivariate simulation

## 1. Introduction

Many offshore operations are sensitive to the frequent changes in sea conditions that occur during operations. Offshore storage and tanker loading efficiency will depend on the frequency, intensity and duration of storms. The total time necessary to complete underwater pipelines installations or other construction operations will also depend on the sea state conditions over the operating period of interest.

Let us consider a practical application as an example. Suppose that a company has to install an underwater pipeline. It is known that this operation needs *d* days of effective work and *h* hours to prepare the material at each arrival on the site. Furthermore, the operation has to stand by if the significant wave height $H_s$ is over the level *h* or if the peak period $T_p$ belongs to the interval $[\tau_{inf}, \tau_{sup}]$ (e.g. the wave period is too close to the resonant heave period). Now, the company responsible for the installation of the pipeline wants to evaluate the risk of not completing the mission by the date of the beginning of the operation. There is an analytical model and no analytical response for such a problem. Then, a solution consists of calculating statistical predictions given a data set of observations and evaluating the time duration of $[H_s<h]$ and $[T_p \notin [\tau_{inf},\tau_{sup}]]$ events.

In nature, the time sequence of sea conditions is random and not usually repeatable. Often, the required long-term information can be estimated from observations, but, in this case we can only repeat already observed scenarios and the durations are not always sufficient to produce statistics with enough accuracy. An alternative to observations is the application of hindcasting techniques. However, it requires expensive and time-consuming studies to produce theoretical models, and the histories may be biased due to the inadequacies of the models.

So, we need alternative techniques to describe and predict the pattern of sea conditions for use in operational simulation studies. We propose in this paper a method of bivariate simulation for sea-state parameters such as significant wave height and peak period ($H_s$, $T_p$). This method enables us to compute a large number of many-years-realisations of the process ($H_s$, $T_p$) and then to deduce accurate estimations of duration statistics. Sea-state parameters are known as non stationary and non Gaussian processes. In the case of sea-state parameters, non stationarity is induced by season. During autumn and winter, there are more observed storms and with a higher level of severity than those observed during summer. The non Gaussian property is due to the definition of the parameters ($H_s$, $T_p$). If the artificially generated sea-state histories exhibit statistical properties substantially the same as those of the observed data, they can be used as a substitute for real long-term observations for estimating the feasibility of offshore operations in areas where no sufficiently long experience exists.

Several authors have been worked on the problem of the modelisation and simulation of processes $H_s$ and $T_p$. For instance, Stefanokos [8] proposes a model based on time series. Some authors have developed simulation methods based on Markov chains models particularly for the study of mean storm durations [1 and 4]. Scheffner and Borgman [7] and Walton and Borgman [7] describe a simulation method (cited in this paper as Borgman's method) that uses the empirical marginal distribution of probability of the observed parameters. Derval [3] has adapted Borgman's method to bivariate simulation and has implemented it by replacing the empirical distributions of probability by parametrical estimations. The large number of observations and the good fitting of the sea-state parameters to Gamma distribution may justify her choice of Gamma distributions. However, Derval shows, by studying the significant steepness of the sea-state, that the method she has used induces a bad approximation of the instantaneous joint distribution of ($H_s$, $T_p$).

The aim of the study presented here is to propose an improvement of the empirical simulation method of Borgman. The simulation method described hereafter will enable a better approximation of the joint instantaneous distribution of ($H_s$, $T_p$) and it will respect the physical constraints existing between variables $H_s$ and $T_p$ (e.g. due to wave breaking, it is not possible to observe any sea-states with high $H_s$ and low $T_p$). The method we propose gives a good enough approximation of the joint distribution of probability of the time process $(H_s(t),T_p(t))_t$ to retrieve some statistical properties of persistence parameters.

The data which we will use consist of 8 years of measurements of $H_s$ and $T_p$ on the Frigg field in the North Sea. The significant wave height $H_s$ and the peak period $T_p$ are computed from 20 min time series of sea surface elevation. Between each 20 min of time series, we have 3 h gap. The time series of $H_s$ and $T_p$ obtained from these measurements contains a large number of missing data that we will deal with. Fig. 1 presents the records of the significant wave height and the peak period for the fourth year. All the blanks correspond to missing data.

## 2. Borgman's method

The method which we will describe here was first proposed by Walton and Borgman [9] for one dimensional non stationary and non Gaussian processes and it has been adapted by Derval [3] for bivariate processes in a simple way. The method consists of two parts: the analysis part where the statistical properties of the time processes are estimated and the synthesis part where one simulates new time series having the same statistical properties as the observed one. The main steps of both parts of Derval's version of Borgman's method are described hereafter.

### 2.1. Analysis part

The data are first lowpass filtered to determine a time-varying mean and a time-varying variance. We then attempt to stationarise the data by removing the mean and variance trends. Let us denote $H_{ij}$ as the observations of significant wave height, where $j$ denotes the year of observation and $i$ the observation number in year $j$. We suppose that $H_{ij}=\sigma_i^{(l)}H_{ij}^{st}+m_i^{(l)}$ where $m^{(l)}{}_i$ represents the time-varied mean, $\sigma^{(l)}{}_i$ the time-varied standard deviation and $H^{st}$ the stationarised significant wave height. $m^{(l)}{}_i$ and >σ(l)i (Fig. 1) are smoothed versions of $m_i$ and $\sigma_i$. The mean $m_i$ is computed as follows for all $i$:

(1)

$$m_i = \frac{1}{n_{\text{years}}(i)} \sum_{j=1}^{n_{\text{years}}} H_{ij}$$

where $n_{\text{years}}$ is the number of years of observations and $n_{\text{years}}(i)$ the number of years for which the observation $i$ is not missing. Missing values are replaced by zeros. The variance $\sigma^2{}_i$ is defined by:

(2)

$$\sigma_i^2 = \frac{1}{n_{\text{years}}(i)} \sum_{j=1}^{n_{\text{years}}} (H_{ij} - m_i)^2$$

The lowpass filtered mean and variance are defined in the continuous time domain as a convolution of the signal (mean or variance) with the standard Gaussian kernel. In practice, it is usual to revert to discrete Fourier transform tools. The series of transformed data $H^{st}{}_{ij}$ is a series of dependent variables which is supposed to be stationary. A similar transformation is computed on the peak periods to obtain the process $T^{st}$. In the second step, the stationarised observations are transformed by a normal score transformation.

For a stationary process $X$ with marginal distribution function $F_X$, the process $Z$ with standard Gaussian marginal distribution function is constructed as follows:

(3)

$$Z_i=\Phi^{-1}[F_X(X_i)]\forall_i$$

where $\Phi$ denotes the standard normal distribution function. In practice, an empirical estimation of the distribution function $F_X$ is used. Let us denote $Z_H$ and $Z_T$ as the series obtained by the normal score transformation respectively for the significant wave height and the peak period. The pair ($Z_H$, $Z_T$) is supposed to be Gaussian and the bivariate process $(Z_H(t),Z_T(t))_t$ is supposed to be stationary. Then, a smoothed version of bivariate spectrum of ($Z_H$,$Z_T$) is estimated.

### 2.2. Synthesis part

The smoothed bivariate spectrum of the Gaussian process ($Z_H$, $Z_T$) is used to simulate a new version $(\tilde{Z}_H, \tilde{Z}_T)$ of the Gaussian process. In the following, $\tilde{X}$ defines a simulated version of $X$. The inverse normal score transformation is applied to $(\tilde{Z}_H, \tilde{Z}_T)$ using the following relations to obtain $(\tilde{H}^{st}, \tilde{T}^{st})$.

$$\tilde{H}^{st}(i,j) = F_{H^{st}}^{-1}(\Phi(\tilde{Z}_H(i,j)))\,\forall(i,j)$$

(4)

$$\tilde{T}^{st}(i,j) = F_{T^{st}}^{-1}(\Phi(\tilde{Z}_T(i,j)))\,\forall(i,j)$$

where $F_H^{st-1}$ and $F_T^{st-1}$ denote respectively the empirical estimations of the inverse distribution functions of $H^{st}$ and $T^{st}$.

Finally, the mean and variance trends are restored to simulated $\tilde{H}^{st}$ and $\tilde{T}^{st}$ to obtain the simulated bivariate process $(\tilde{H}_s, \tilde{T}_p)$.

### 2.3. Results and discussion

Now, we will compare several statistical properties of simulations and observations. In Fig. 2 we present the marginal distribution of observed and simulated $H_s$ and $T_p$. We note that the simulated marginal distributions closely fit to the observed ones.

here are essentially three problems in the method described above for bivariate simulation.

1. The pair obtained after normal score transformation of the stationarised data is clearly not Gaussian. We have used a test of normality to verify that the couple ($Z_H$, $Z_T$) obtained after elimination of the mean and variance trends and after normal score transformation is not a Gaussian pair. The normality test of Lin and Mudholkar [5] has been adapted to two random variables. We may observe by simulations that the more the parameter observations Hs and Tp are smoothed, the further the couple (ZH, ZT) is from normality. Furthermore, Fig. 5 which presents the joint distribution of (ZH, ZT) confirms the non normality: the distribution is not symmetric. This is due to the physical relationship between $H_s$ and $T_p$, which prevents the definition domain of ($H_s$, $T_p$) being whole $R^+ \times R^+$.

2. The inverse normal score transformation does not use the information which may be given by the joint distribution of ($H_s$, $T_p$). In Borgman's method, the inverse normal score transformation is done marginally and instantaneously, although the random variables $H^{st}(t)$ and $T^{st}(t)$ are dependent. Hence, the information on the joint distribution of the process (Hs, Tp) is only brought by the simulated bivariate Gaussian process. In Appendix A, we show that the joint distribution of significant wave height and peak period is better restored if the Gaussian random variables are instantaneous independent.

3. The hypothesis of stationarity done after the removal of mean and variance trends is not valid here. The processes obtained by removing the mean and variance trends are supposed to be stationary, although the operation of removing the trends changes the definition domain of the variables. For instance, after removing the trend, $H_s \in [0,+\infty[$ and after $H^{st}_i \in [m_i/\sigma_i,+\infty[$ and this interval depends on time.

In what follows, we propose to adapt Borgman's method to improve the statistical properties of the simulated processes.

## 3. Conditional simulation method

We have shown in Borgman's method that the instantaneous marginal distributions of $H_s$ and $T_p$ do not give enough information to simulate the pair ($H_s$, $T_p$) by failing to respect the physical

relationship between both processes. Secondly, in Borgman's method the assumption of stationarity of the unseasonal processes generates non physical situations in the simulation. In the conditional simulation method proposed below, we will provide solutions to these two points.
The analysis part of the conditional simulation method (stationarisation, normal score transformation, spectral estimation) is the same as in Borgman's method. Then, we simulate a bivariate Gaussian process which has the same spectrum as the series of Gaussian variables obtained in the analysis part. This process is simulated by applying the method proposed in Ref. [9].
Our idea to improve the dependence properties of the simulated bivariate process ($H_s$, $T_p$) is to use the bivariate distribution function of ($H_s$, $T_p$) in the inverse normal score transformation instead of the product of the marginal distribution function as done by Derval [3]. The simulation scheme is given by the Bayes formula as follows:
(6)

$$\tilde{H}_s^{st}(i,j) = F_{H^{st}}^{(i)-1}(\Phi(\tilde{Z}_H(i,j)))$$

$$\tilde{T}_p^{st}(i,j) = F_{T^{st}|H^{st}}^{(i)-1}(\Phi(\tilde{Z}_T(i,j))\,|\,H_s^{st} = \tilde{H}_s^{st}(i,j))$$

As mentioned earlier, we suppose in Borgman's method that the processes obtained after removing the mean and variance trends are stationary, although this assumption is not valid. The seasonal parameters being smoothed, they may be approximated by constants on a small interval. Furthermore, the definition domain is only depending on the mean and variance trends and it can be seen in Fig. 1 that these trends are slow-varying and may be supposed to be constant on a small time interval. So that we can assume that the probability law of (Hs(t), Tp(t)) is stationary on a small time interval around t. Following this hypothesis, we will use a small set of observation points to estimate the conditional distribution functions of Eqs. (6). For instance, $F_H^{st(i)-1}$(.) will be estimated using the points $\{H_s^{st}(k,j), k=i-1,\ldots,i+l, h=1,\ldots,n_{years}\}$. Numerical tests show that we can choose a time interval of 120 h, which corresponds to $l$=20 observations before and after $i$, for each year.
Recently, Rychlik et al. [6] have developed a procedure to transform stationary non Gaussian data to Gaussian data for univariate processes. We can verify that the inverse normal score transformation proposed here for $H_s$ is approximately equal to the transformation $g$ of Rychlik et al., before smoothing, based on the level up-crossings of $H_s$. On the contrary, the transformation for the peak period $T_p$ is really irregular, and only the trend of the transformation corresponds to the transformation of Rychlik et al.

## 3.1. Results and discussion

Let us compare the simulated processes $\tilde{H}_s$ and $\tilde{T}_p$ with the observed ones. We have simulated 4 years of data, with one point every 3 h. The data files contain a large number of missing data. In the cases where only one successive observation is missing, it is replaced by a linear interpolation of the neighbour points. In other cases, we have to take care of the missing data in the different steps of the method.
We first compare the marginal properties of $H_s$ and $T_p$. The histograms obtained from observations and simulations (Fig. 6) are sufficiently close that we can say that the simulations have the same instantaneous and marginal probability distributions as the observations. The mean of up-crossings per unit of time of $H_s$ gives some information about the time dependence of the $H_s$ process. Indeed, the frequency of up-crossings depends on $H_s$ and its time derivative. We note (Fig. 7) that the mean of up-crossings of levels higher than 4 m are quite well fitted, but we observe that the $H_s$ simulated process crosses the low levels between 1 and 3 m too often. The same phenomenon is observed in winter and summer. It is due to high frequency oscillations in the simulated process. Cutting the very high frequency of the spectra of the Gaussian process used for the simulation may reduce these oscillations.

One of our main objectives is to predict the duration of sea-states of given levels of significant wave height. Fig. 8 gives the probability that a sea-state with $H_s$ higher than 2, 3, 4 and 5 m lasts more than time $t$ (in hours). The curves show that the duration properties of the observations are well restored in the simulations. For the observations, the results are computed only for the completely observed storms (without missing data).
The simulations must also restore the joint distribution of the pair ($H_s$, $T_p$). We compare the probability density functions of observed and simulated significant steepness. Fig. 9 presents the kernel density estimate of the significant steepness. It is clear that the conditional method that we propose improves the results of Borgman's method. Although the observed probability of the high levels of steepness is overestimated by the simulations, the definition domains of simulations and observations are the same. It means that the physical constraints between the significant wave height and the peak period are respected. Finally, Fig. 10 shows that the instantaneous joint probability density function of simulated ($H_s$, $T_p$) is a good approximation of the observed one. Furthermore, we see that the simulations respect the physical constraint between $H_s$ and $T_p$. The whole joint probability is over the breaking-waves limit. 3.2. Example of application
Let us now consider again the practical application mentioned in the Introduction. Suppose that a company has to install an underwater pipeline and that this operation needs 10 days of effective work and 2 h for the installation at each arrival on the site. Furthermore, the operation has to be stopped if $H_s$>2 m or 9.5 s<$T_s$<13 s because this period matches with the heave resonance frequency of the boat. After each stop, the installation has to be restarted. Now, the company responsible for the operation has to evaluate how much time it will spend to perform it with respect to the date of the beginning of the operation. Fig. 11 shows the prediction of the mean number of days necessary to complete the mission computed from observations and from simulation with respect to the month the operation begins. It is supposed that the operation always starts the first day of the month. Observed data contain 9 years of measures recorded on the area of Statfjord (Norway) and simulations represent 35 sets of 9 years of the same data. Fig. 11 shows that it is more convenient to begin the operation during end of winter or spring. The results obtained from simulations permit us to associate a confidence interval to the prediction computed from the observed data. For instance, we note that the prediction is less precise in August than in March.

## 4. Conclusion

In this paper, we propose a non parametric method to simulate bivariate non stationary and non Gaussian processes. This method is applied to generate new time series of the pair ($H_s$, $T_p$) for a given area where observations have been recorded during several years. The generated series do not present the same successions of storm and calm conditions, but they have globally the same statistical properties as the observed time series. Then, this method enables us to test the influence of different scenarios of the time evolution of the sea-state for many applications. An example is proposed.
The described method has been tested on only one record of ($H_s$, $T_p$), but it has been verified that the results are similar on other series of observations.
Several statistical properties of simulated and observed processes are compared and it is shown that the conditional method of simulation clearly improves its precedent version called here Borgman's method. In particular, it is important to note that the conditional method of simulation enables us to well restore the instantaneous joint distribution of the significant wave height and the peak period by respecting the physical relationship existing between these parameters. Furthermore, we show that the simulated processes give a good approximation of statistics of duration such as the persistence of storm of given severity.
The conditional method is a non parametrical method of simulation. It means that the number of the observations, used as ‘model’ of simulation, should be large enough. One of the deficiencies of this method is that it is difficult to include long-term trends such as the elevation of the mean sea level in the simulations. To take into account such trends, we should write a semi-parametric model.

## Acknowledgements

The authors are much obliged to the STATOIL Company, which has supplied the data used in this study.

## References

[1] Anastasiou K, Tsekos C. Persistence statistics of marine environmental parameters from Markov theory, Part I: analysis in discrete time. Applied Ocean Research 1996;18:187-99.
[2] Bosq D, Lecoutre JP. Théorie de l'estimation fonctionnelle. Economica, 1987.
[3] Derval C. Processus d'état de mer. DEA Génie Civil, Ifremer, 1998.
[4] Haring RE, Richardson JR, Russell LR. Environmental simulation in offshore petroleum operations. Preprint, 1975.
[5] Lecoutre JP, Tassi P. Statistique non paramétrique et robustesse. Economica, 1987.
[6] Rychlik I, Johannesson P, Leadbetter MR. Modelling and statistical analysis of ocean-wave data using transformed Gaussian processes. Marine Structures 1997;10:13-47.
[7] Scheffner NW, Borgman LE. Stochastic time-series representation of wave data. Journal of Waterway, Port, Coastal, and Ocean Engineering 1992;118(4):337-51.
[8] Stefanakos CN. Nonstationary stochastic modelling of time series with application to environmental data. PhD of National Technical University of Athens, 1999.
[9] Walton TL, Borgman LE. Simulation of non stationary, non Gaussian water levels on great lakes. Journal of Waterway, Port, Coastal, and Ocean Engineering 1990;116(6):664-85.

## Appendix A. Inverse normal score transformation

Let ($U$, $V$) denote a pair of standard Gaussian random variables and ($X$, $Y$) a couple obtained by inverse normal score transformation of ($U$, $V$). Let $p_{X,Y}$ denote a given probability density function and $F_{X,Y}$ the associated distribution function. The Bayes formula gives:

$$p_{X,Y}(x,y)=p_X(x)p_{Y|X}(y,x)$$

$$F_{X,Y}(x,y)=F_X(x)F_{Y|X}(y,x)$$

where $p_{Y|X}$ and $F_{Y|X}$, respectively, denote the conditional probability density and distribution functions. We use standard notation $\Phi$ for the normal distribution function.
The inverse normal score transformation is equivalent to the following transformation:

$$(X,Y)=(F_X^{-1}(\Phi(U)),F_{Y|X}^{-1}(\Phi(V),F_X^{-1}(\Phi(U))))$$

For this ($X$, $Y$), we have for all integrable functions $h$:

$$E[h(X,Y)]=\int h(F_X^{-1}(\Phi(u)),F_{Y|X}^{-1}(\Phi(v),(F_X^{-1}(\Phi(u))))p_{U,V}(u,v)\mathrm{d}u\mathrm{d}v=\int h(x,y)p_{U,V}(\Phi^{-1}(F_X(x)),\Phi^{-1}(F_{Y|X}(y,x)))(\Phi^{-1})'(F_X(x))(\Phi^{-1})'F_{Y|X}(y,x)p_X(x)p_{Y|X}(y,x)\mathrm{d}x\mathrm{d}y$$

If $U$ and $V$ are independent random variables, we can verify that

$$p_{U,V}(\Phi^{-1}(F_X(x)),\Phi^{-1}(F_{Y|X}(y,x)))(\Phi^{-1})'(F_X(x))(\Phi^{-1})'(F_{Y|X}(y,x))=\frac{(\Phi\Phi^{-1}F_X(x))'}{(F_X(x))'}\frac{(\Phi\Phi^{-1}F_{Y|X}(y,x))'}{(F_{Y|X}(y,x))'}=1$$

We deduce that

$$E[h(X,Y)]=\int h(x,y)p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

such that the probability law of ($X$, $Y$) has $p_{X,Y}$ for probability density function.

## Figures

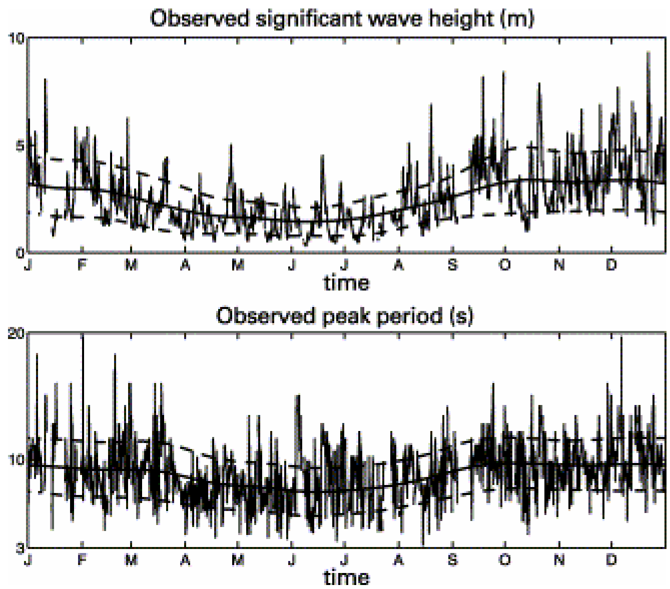


Fig. 1. Observations of $H_s$ and $T_p$ for the fourth year of measure with the smoothed mean trend and the smoothed mean trend plus and minus the R.M.S. trend.

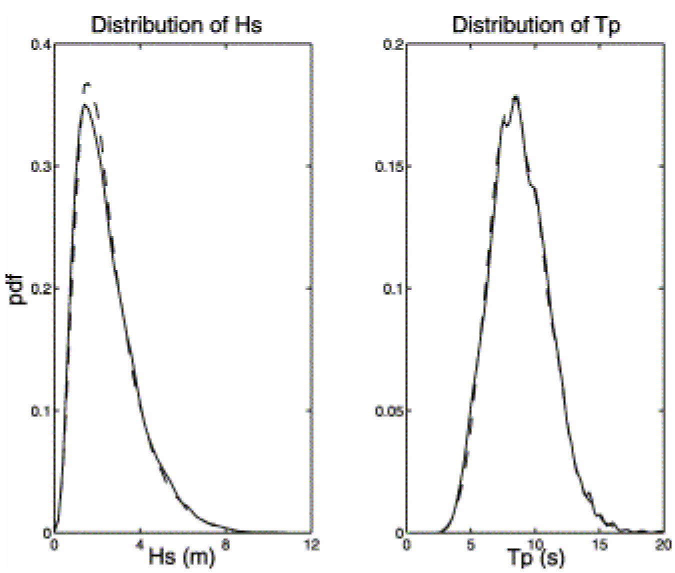


Fig. 2. Probability density functions of $H_s$ and $T_p$ for Borgman's method. Observations (continuous line) and simulations (dashed line).

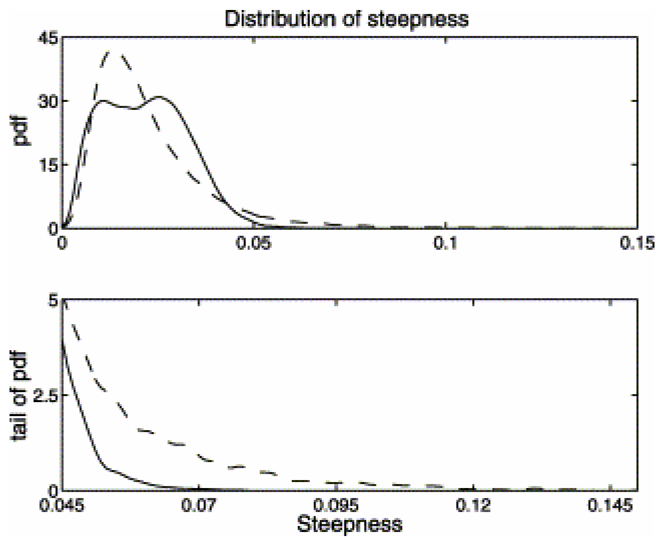


Fig. 3. Probability density functions of significant steepness for Borgman's method. Observations (continuous line) and simulations (dashed line).

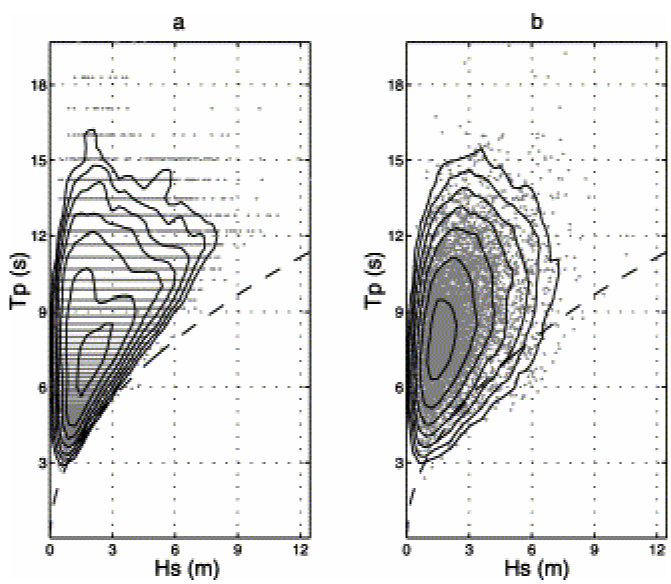


Fig. 4. Joint distribution functions of $T_p$ with respect to $H_s$ for Borgman's method: (a) observations; (b) simulations. The dashed line represents the limit of breaking-waves.

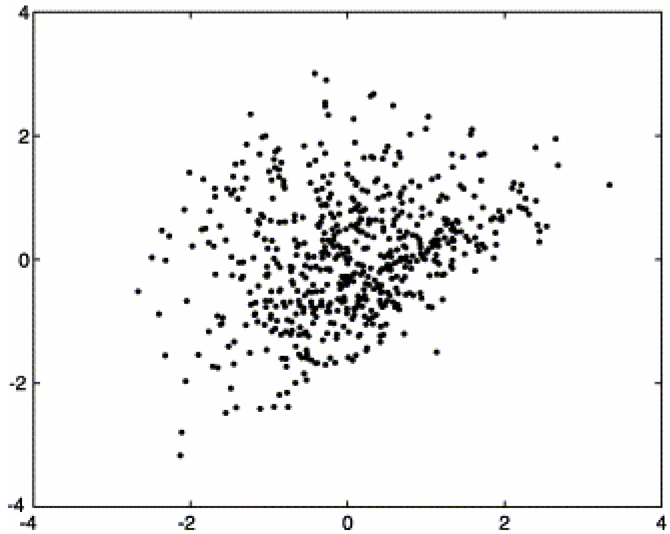


Fig. 5. Joint distribution of pair observed by normal score transformation in Borgman's method.

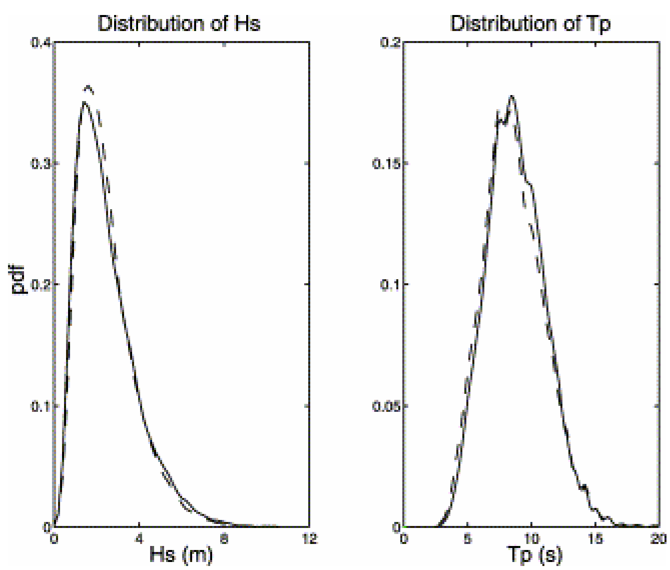


Fig. 6. Probability density functions of $H_s$ and $T_p$ for conditional method observations (continuous line) and simulations (dashed line).

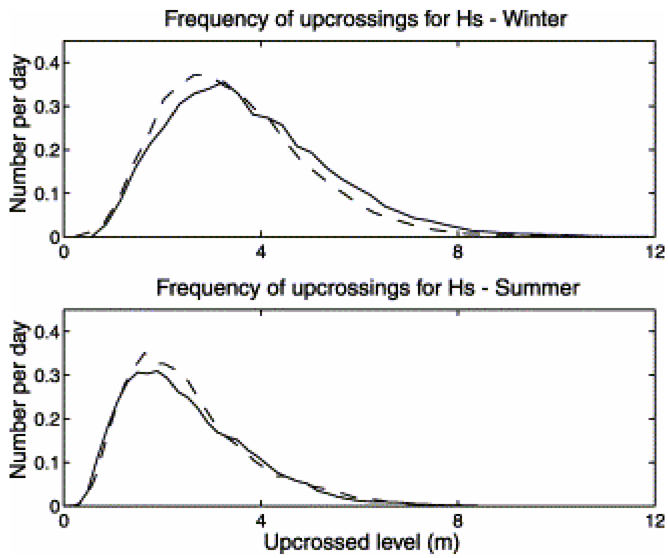


Fig. 7. Frequency of up-crossings for $H_s$ in winter and in summer. Observations (continuous line) and simulations (dashed line).

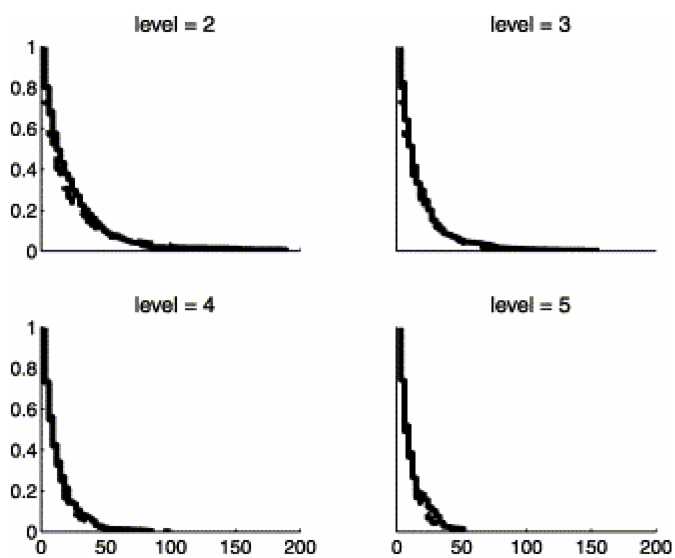


Fig. 8. Probability that a the duration of a storm of given severity level is more than time *t*. Observations (continuous line) and simulations (dashed line).

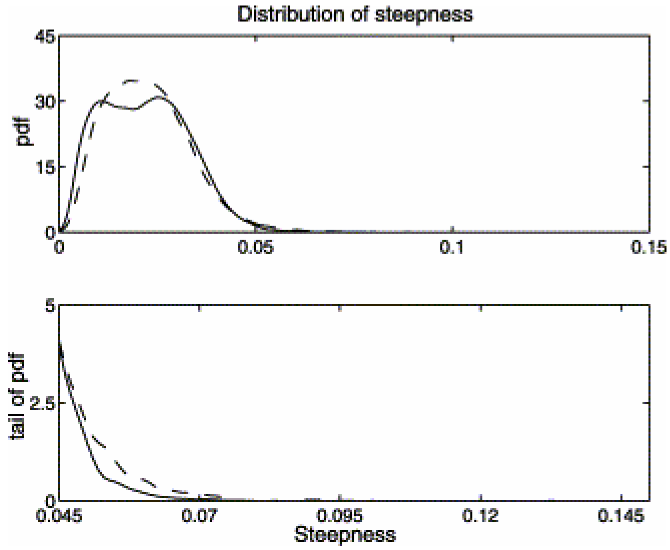


Fig. 9. Distribution of significant steepness for conditional method. Observations (continuous line) and simulations (dashed line).

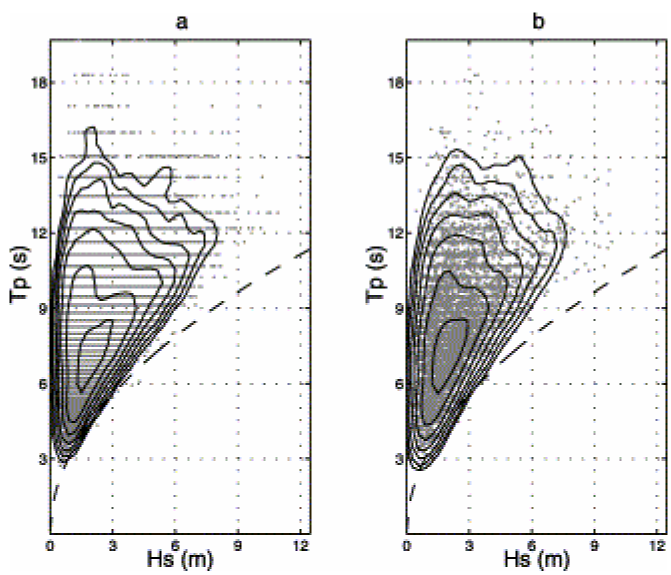


Fig. 10. Joint distribution functions of $T_p$ with respect to $H_s$ for conditional method: (a) observations; (b) simulations. The dashed line represents the limit of breaking-waves.

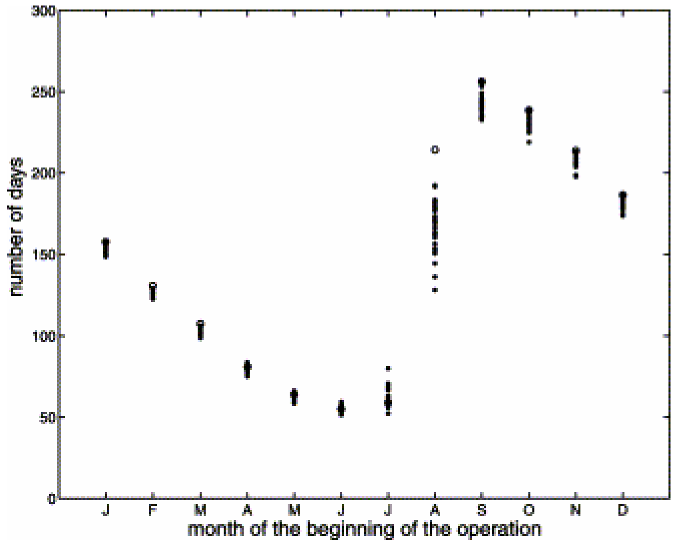


Fig. 11. Mean number of days necessary to perform an offshore operation of 10 days, supposing that $H_s$ has to be lower than 2 m and $T_p$ greater than 13 s or lower than 9.5 s. Prediction computed from observed data (circles), prediction computed from simulation (points).