Probabilistic scheduling of offshore operations using copula based environmental time series

An application for cable installation management for offshore wind farms

Leontaris, George; Morales-Nápoles, Oswaldo; Wolfert, A. R M (Rogier)

DOI
10.1016/j.oceaneng.2016.08.029

Publication date
2016

Document Version
Accepted author manuscript

Published in
Ocean Engineering

Probabilistic scheduling of offshore operations using copula based environmental time series

- An application for cable installation management for offshore wind farms -

Georgios Leontaris[a], Oswaldo Morales-Nápoles[b], A.R.M (Rogier) Wolfert[a]

a: Engineering Assets Management group, fac. Civil Engineering and Geosciences, TU Delft

b: Probabilistic Design group, fac. Civil Engineering and Geosciences, TU Delft

Abstract

There are numerous uncertainties that impact offshore operations. However, environmental uncertainties concerning variables such as wave height and wind speed are crucial because these may affect installation and maintenance operations with potential delays and financial consequences. In order to include these uncertainties into the duration estimation, adequate tools should be developed to simulate an installation scenario for a large number of historical environmental data. Data regarding environmental time series are usually scarce and limited, therefore they should be modelled. Since the environmental variables are in reality dependent, we propose a probabilistic method for their construction using copulas. To demonstrate the effectiveness of this method compared to the cases where observed or independently constructed environmental time series are used, a realistic cable installation scenario for an offshore wind farm was simulated. It was found that the proposed method should be followed to acquire more reliable and accurate estimation of the installation's duration.

Keywords: Copulas, Environmental time series, Stochastic model applications, Offshore wind farms, Simulation, Project management

1. Introduction

During the past years a lot of attention has been drawn towards the development of renewable energy technologies due to the expected depletion of fossil fuels and respective consequences (EWEA 2012; Dickel et al. 2014). Offshore wind energy is considered one of the most promising renewable energy sources and it is expected to grow even more during the upcoming years, because of better quality of wind far from shore, more space available and less noise and visual impact compared to onshore wind farms (Esteban et al. 2011). However, the high costs of offshore wind farms (OWF) make them non-competitive compared to the conventional energy sources. Especially installation costs highly contribute on the total cost of OWF, which "add up to nearly one quarter of the entire project value" (Wüstemeyer et al, 2015).

The installation of OWF as every offshore operation, is subject to a variety of uncertainties such as environmental conditions, failure of vessels and/or equipment, variation in the duration of operations, availability of the required components etc. However one of the main cause of miss-estimations of project duration and delays is the miss-estimation of environmental parameters, such as the wind speed and the significant wave height, which are difficult to predict in the planning phase. For those reasons, project schedulers may use buffers in the planning phase which can lead to overestimation of the duration of a project and subsequently the cost of the installation. Therefore it is essential to find a method which will assist schedulers in acquiring more accurate and reliable estimates of the duration of offshore installation operations by incorporating these uncertainties.

A lot of research has been conducted in the past regarding forecasting of environmental time series. Zounemat-Kermani et al (2015) mention the following methods to model wind-wave characteristics: discrete spectral approach, stochastic simulation, numerical methods and data driven non-statistical models (such as artificial neural networks, fuzzy wavelet model, fuzzy logic and chaos theory). Moreover, a survey regarding the stochastic models for wind and wave state time series was conducted by Monbet et al. (2007) and categorizes these models into: non-parametric models, models based on Gaussian approximations and other parametric models. These methods however do not always explain the underlying physical properties which should be captured by the joint probability distribution. In particular, nothing or little may be said in terms of joint probabilities of environmental random variables that are described by a non-normal joint distribution. For example, as will be seen later on in this paper, high values of wave height may be more correlated with high values of wind speed than low values for each of the two variables.

Univariate distributions are used frequently in order to estimate the design parameters of wind speed and wave characteristics without considering their dependence (Yang and Zhang 2013). Some studies were focussed on estimating the joint distribution of wave characteristics such as significant wave height and wave period. Particularly, Salvadori et al (2013) used Copulas, Athanassoulis et al (1994) used applications of Placket model and Galiatsatou and Prinos (2007) investigated different bivariate distributions, in order to find the dependence between significant wave height and wave period. Another example can be found in Memos and Tzanis (2000). The authors propose a model to represent the joint probability distribution of wave heights and wave periods, both in deep and shallow waters employing a wave-by-wave transformation. In Athanassoulis and Belibassakis (2002), the authors proposed a kernel density model to obtain an analytic representation of univariate or multivariate empirical distributions of metocean parameters (e.g. significant wave height, mean wave period and wave direction). This approach is particularly useful when the environmental data are only available as histograms and not historical time series. However, the aforementioned approaches were not used to produce synthetic time series. Furthermore, only a few studies investigate the joint distribution of the wind speed and the significant wave height. Particularly, Fouques et al (2004) propose one method using only the correlation matrix and another method based on multivariate Hermite polynomials expansion of the multinormal distribution, in order to model the joint occurrence of those variables including the wave period. Moreover Bitner-Gregensen and Haver (1989, 1991) developed a joint environmental model which is based on conditional modelling approach (CMA) and concerns wind, waves, current and sea water level. This model was also applied for design and operations of marine structures by calculating the joint distribution based on parametric fits for each one dimensional marginal (Bitner-Gregensen 2015). Also the Nataf model (Nataf 1962) is used in many applications in literature for modelling metocean variables. Nevertheless, in Bitner-Gregensen et al. (2014), it is noted that Nataf model may lead to bias results, when the transformation to standard normal variates deviates from a multi normal distribution. Finally, Yang and Zhang (2013) followed a similar approach as the one described in this article, using Copulas to estimate the joint distribution of wind speed and significant wave height without taking into account the autocorrelation which is essential when time series are required.

The main goal of this article is to propose an alternative method to produce large number of realistic time series of wind speed and significant wave height, which can be valuable for planning and scheduling more efficiently offshore installation operations. In order to plan the sequence of complex offshore installation operations and decide the optimal combination of vessels and equipment required for a particular operation, different scenarios should be simulated and compared. Therefore, large number of environmental time series is needed to account for uncertainties regarding the environmental conditions that limit the operations. Usually it is difficult, expensive and sometimes impossible to acquire a large data set of environmental time series and when it is possible there are often missing values due to failures in the measuring equipment (Monbet et al. 2007), which can influence the estimation of the duration of offshore operations. For these reasons it is important to create realistic environmental time series by taking into account the dependence between the environmental characteristics. In this paper a method using copulas is proposed in order to produce time series of environmental characteristics that limit the operations (i.e. wind speed and significant wave height) by taking into account their dependence and the observed autocorrelation. Copulas are a way of studying scale free measures of dependence and a starting point for constructing families of bivariate distribution (Nelsen 2006). Copulas allow us to construct models which go beyond the standard ones at the level of dependence (Embrechts et al. 2003) and they avoid the restriction that presents the traditionally used method which describes the pairwise dependence using families of bivariate distribution characterized by the same parametric family of univariate distributions (Genest and Favre 2007). Following the copula approach, it is possible in many cases to construct the joint distribution requiring only the marginal distributions of the variables and measures of their dependence (Clemen and Reilly 1999). Also, in our case, the characterization of the joint distribution of the environmental variables of interest is semi-parametric. In other words, the one dimensional margins are modelled by non-parametric estimators while the underlying dependence structure are described by one parameter copulas. Moreover the use of copulas has made the investigation of asymmetries in the joint distribution relatively easier since they satisfy different types of tail behaviour (Joe 2014). These asymmetries are, as it shall be demonstrated in this paper, crucial for offshore operations which are mainly influenced by extreme environmental conditions. Finally, in order to investigate the effect of this approach, an application of the proposed method concerning the estimation of the duration of the cable installation of an offshore wind farm was conducted.

2. Preliminary concepts and definitions of copulas

Before continuing to the method proposed for the construction of time series for significant wave height and wind speed, the main concepts and definitions to be used in the remainder of this paper are introduced. Copulas are defined as functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distribution functions. In particular, they are multivariate distribution functions whose one-dimensional margins are uniform on the interval [0,1] (Nelsen 2006). The most important theorem of copulas theory is Sklar's theorem (Sklar 1959) which states that any multivariate joint distribution can be written in terms of the univariate marginal distribution functions and a copula which describes the dependence between the random variables. For the two dimensional case let $H_{XY}(x,y)$ be a joint distribution function with marginal distribution $F_X(x)$ and $G_Y(y)$ which lie in the interval [0,1]. Then there is a copula C on the unit square I^2 such that for all x,y satisfies the following (Genest and Favre 2007):

$$H_{XY}(x,y) = C\{F_X(x),\ G_Y(y)\}\ \ x,y\ \in \mathbb{R} \qquad \text{(eq. 1)}$$

There is a large variety of copulas which can be used to model joint distributions with different characteristics. For the purpose of this paper three of the most common families of copulas are investigated: the Gaussian, Gumbel and Clayton copulas. These copulas can model different tail asymmetries of the joint distributions and have been used in many financial applications (e.g. see Aas et al. 2009).

The Gaussian copula is given by:

$$C(u,v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \qquad \text{(eq. 2)}$$

where Φ denotes the standard normal distribution function and Φ_ρ the standard bivariate normal distribution function with linear correlation coefficient ρ.

The Gumbel and Clayton copulas are two of the most used one-parameter Archimedean copulas. For the bivariate case, Archimedean copulas are defined as $C(u,v) = \varphi^{-1}\big(\varphi(u) + \varphi(v)\big)$. The generator function of Gumbel copula is $\varphi(u) = (-\ln(u))^{\theta}$, $\theta \in [1, \infty)$, while the generator of Clayton copula is $\varphi(u) = (u^{-\beta} - 1)/\beta$, $\beta \in [-1, \infty)$ (Nelsen, 2003).

So the Gumbel copula is defined as:

$$C(u,v;\theta) = \exp\{-\left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{\frac{1}{\theta}}\} \qquad \text{(eq. 3)}$$

and the Clayton copula is defined as:

$$C(u,v;\beta) = (u^{-\beta} + v^{-\beta} - 1)^{-\frac{1}{\beta}} \qquad \text{(eq. 4)}$$

A way of fitting copulas to data concerns the use of correlation estimators (or measures of dependence) such as Spearman's rho and/or Kendall's tau. These important measures of dependence refer to the ranks of the data achieving scale-invariant estimates (Schmidt 2006). In eq. 5, Spearman's rho r_S is presented in terms of copulas.

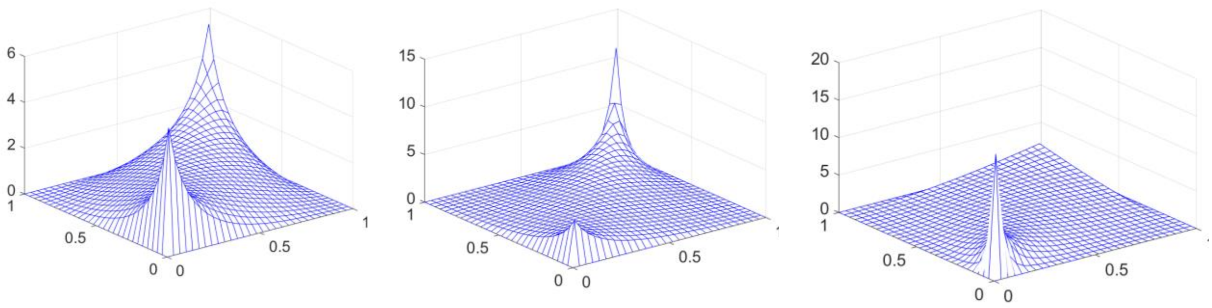
$$r_S(X,Y) = 12\iint_{I^2} uv dC(u,v) - 3 = 12\iint_{I^2} C(u,v) dudv - 3 \qquad \text{(eq. 5)}$$

Another important concept that should be introduced for our analysis, is tail dependence. Tail dependence allows the study of dependence between extreme values, because (for positive dependence) shows the amount of dependence in the upper right quadrant tail or lower left quadrant tail of a bivariate distributions (Embrechts et al 2003). The upper tail dependence coefficient is defined in eq. 6.

$$\lambda_U = \lim_{q \to 1} P\{Y > G^{-1}(q) | X > F^{-1}(q)\} \qquad \text{(eq.6)}$$

One can characterize X and Y as asymptotically dependent in the upper tail when $\lambda_U \in (0,1]$ or as asymptotically independent in the upper tail if $\lambda_U = 0$. The coefficient of lower tail dependence can be defined analogously for the lower tail.

The three parametric models of interest have been selected precisely because they capture lower, upper or no-tail dependence. Usually the density of bivariate copulas defined as: $c(\boldsymbol{u}) = \frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2}$ is used to illustrate copulas’ distributions. Plots of the density function of the three copulas of interest in this paper (for Spearman’s rank correlation equal to 0.7, which is descriptive of the level of correlation of the available data) are presented in Figure 1. Intuitively the reader may see that for positive correlation the mass in the upper tail of the Gumbel copula is significantly larger than that in the lower tail, which is indication of upper tail dependence. Analogously for the Clayton copula the mass in the lower tail is larger than that in the upper tail, which indicates lower tail dependence. Having described briefly the main concepts to be used in the rest of the article, we proceed to describe the data of interest.



Gaussian copula ($\rho = 0.7$) Gumbel copula ($\theta = 2, r_s = 0.7$) Clayton copula ($\beta = 2.1, r_s = 0.7$)

Figure 1: Density functions of Gaussian, Gumbel and Clayton copulas.

3. Environmental data analysis

Two different environmental data sets concerning average wind speed (m/s), that is the average of the wind speeds observed in the time interval of interest, and significant wave height (m) which is the mean values over the upper third of the observed wave heights during the time interval (Bauer and Staabs 1998), were analysed in order to find which of the three copulas families introduced in section 2 describes best the dependence between those two variables. The first environmental data set concerns 21 years (1990-2011) of measurements in one-hour intervals from an offshore station (41010) located east of Cape Canaveral in Florida, available on National Oceanic and Atmospheric Administration (NOAA) web site (www.noaa.gov). The second environmental data set concerns 3 years (2010-2013) of observations from an offshore station located in the North Sea and it was provided by Deltares. Deltares' environmental data were provided in 10-min intervals and they were transformed into one-hour intervals by taking the maximum value observed during one hour. It was decided to take the maximum values of the means observed during one hour; however it must be noted that the results would not be different concerning which copula fits better the data if the mean of the mean values was used. Both available environmental data sets were analysed following the same procedure, which consists of the following steps:

Excluding unrealistic values

In both environmental data sets there were unrealistic values that probably occurred due to failures in the measuring equipment. Although the number of these values was comparatively small to the size of the set, it was decided to exclude these values. Excluding these obvious measurement errors helped to avoid the influence of those values to the estimation of the best fitting copula and its parameter and subsequently to the synthetic time series.

Transforming observations into pseudo-observations

Because the marginal distributions of random variables are usually unknown, it is often recommended to estimate the parameters of the copula of interest using pseudo-observations. These may be interpreted as a sample of the underlying copula (Genest et al. 2009; Charpentier et al. 2007). The underlying copula of a random vector is invariant by continuous, strictly increasing transformations. Therefore the observations (X_j), when j refers to the random variable, can be safely transformed to pseudo-observations, using the ranks (R_j). The pseudo-observations are defined as (Genest et al. 2009): $U_j = \frac{R_j}{n+1} = n\hat{F}_j(X_j)/(n+1)$, where n refers to the number of the observations and $\hat{F}_j$ the empirical cumulative distribution function defined as: $\hat{F}_j(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_j \leq t)$, where $\mathbf{1}(\,)$ is the indicator function, which is defined as follows for a set E: $\mathbf{1}_E(\omega) = \begin{cases} 1, \omega \in E \\ 0, \omega \notin E \end{cases}$.

Performing three different statistical tests

i. Sum of square differences based on Cramer von Mises statistics

In order to find which copula fits the data best, "blanket tests" are usually used. The "blanket tests" (i.e. “those whose implementation requires neither an arbitrary categorization of data nor any strategic choice of smoothing parameter, weight function, kernel, window, etc.”), are favoured compared to other methodologies due to the fact that they do not involve parameter tuning or other strategic choices (Genest et al. 2009). There are various types of "blanket tests" nevertheless in our study only the test statistic that concerns the calculation of the sum of square differences between the empirical C_n and the parametric copula C_{θ_n}, based on Cramer-von Mises statistic was performed. The empirical copula is a non-parametric estimator of the true copula and it summarizes the information of pseudo-observations. For the bivariate case with two random variables (u_1, u_2) the empirical copula is defined as: $C_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_1 \leq u_1, U_2 \leq u_2)\,,\ \boldsymbol{u} = (u_1, u_2) \in [0,1]^2$. Moreover, the sum of square differences based on Cramer - von Mises statistic for an empirical process $A_n = \sqrt{n}(C_n - C_{\theta_n})$, is defined as (Remillard 2010): $S_n = \int_{[0,1]^d} A_n^2(\boldsymbol{u}) dC_n(\boldsymbol{u}) = \sum\{C_n(\boldsymbol{u}) - C_{\theta_n}(\boldsymbol{u})\}^2$

The sum of the square difference between the empirical and the parametric copula is calculated for every copula under consideration (i.e. S_N for Gaussian, S_{Gum} for Gumbel and S_{Cl} for Clayton) and the copula for which the smallest value is obtained should be preferred.

ii. Calculation of semi-correlations

Another approach to investigate which copula describes better the dependence between significant wave height and wind speed, concerns the calculation of Pearson correlation for upper and lower quadrant of the actual observations transformed to standard normal N(0,1) margins. Let Φ denote the standard normal cumulative distribution function, then $Z_j = \Phi^{-1}(U_j)$, for $j = 1 \ldots, d$ are the standard normal transforms of the pseudo-observations (Joe 2014).

After dividing the standard normal transforms of observations into four quadrants, for positive correlation the upper semi-correlation is defined as: $\rho_{ne} = \rho(Z_1, Z_2 \mid Z_1 > 0,\ Z_2 > 0)$ and the lower semi-correlation is defined as: $\rho_{sw} = \rho(Z_1, Z_2 \mid Z_1 < 0,\ Z_2 < 0)$. The upper and lower quadrant correlations indicate whether or not there is tail asymmetry. When there is tail asymmetry, the two semi correlations present an obvious difference (Joe 2014). Also, these values could be compared to the product moment correlation of all quadrants ρ. This procedure has been exemplified in the context of traffic load measurements before, for example in Morales-Nápoles and Steenbergen (2014). If the values of semi-correlation are larger than the overall Pearson correlation or there is big difference between the upper and lower semi correlation, then there is indication of tail dependence.

iii. Calculation of exceedance probabilities for different percentiles

The third test concerns the calculation of conditional exceedance probability for different percentiles concerning the observations as well as the investigated copulas. The calculation of the joint exceedance probabilities for each copula under consideration is described by the following formula $P(U > u_p, V > u_p) = 1 - 2u_p + C(u_p, u_p)$ where u_p is the percentile of interest. Therefore the calculation of the joint conditional exceedance probabilities for each copula under consideration is:

$$P(U > u_p | V > u_p) = \frac{P(U > u_p, V > u_p)}{P(V > u_p)} = \frac{1 - 2u_p + C(u_p, u_p)}{1 - u_p} \qquad \text{(eq. 7)}$$

The calculated exceedance probabilities of the observations, which were computed based on the empirical version of (eq. 7), were plotted along with the conditional exceedance probabilities for each different copula under investigation, in order to evaluate which copula describes better the extreme cases of large wind speeds occurring together with large wave heights. The purpose of this study is to find the right copula that will be used to produce environmental time series which will show whether or not certain operations can be performed during the time intervals. Hence, the percentiles u_p that are investigated should correspond to values close to the environmental limits of the operations. For that reason, it was decided to conduct this analysis for values of u_p larger than 80th percentile. Following this approach it is safe to assume that these percentiles include the values that address the limits of the vessels.

4. Results of tests

The described environmental data analysis was conducted for every month in order to ensure that seasonality will be taken into account in the estimation of the copula parameter and subsequently in the produced time series. Concerning data from Deltares and NOAA, Tables 1 and 2 present the calculated values of the sum of square differences and the semi-correlations for February, June and for the entire data sets. For both environmental data sets the Gumbel copula had the smaller square difference compared to Gaussian and Clayton copula; meaning that the Gumbel copula has the smallest "distance" between empirical copula and the estimate represented by the parametric copula. Moreover, the calculated values of semi-correlations clearly show that there is tail asymmetry. It can be seen that the upper quadrant semi-correlation regarding Deltares’ data is larger than the overall correlation while the upper quadrant semi-correlation regarding NOAA data are very close. This result suggests that a model with upper tail dependence is preferable. Considering the three copulas under investigation, only Gumbel copula has upper tail dependence. The results of these tests evidently indicated that the Gumbel copula is the copula that fits the data best among the copulas under consideration.

In Figures 2 and 3, the values of exceedance probability for percentiles larger than 80^{th} are presented for Gaussian, Gumbel and Clayton copula, while the dots indicate the exceedance probability of observations. Based on the presented plots, it was found that Gumbel copula underestimates the exceedance probability less than the other investigated families, as far as percentiles smaller than 90^{th} and 96^{th} are concerned for NOAA and Deltares entire data sets respectively. For higher percentiles, the size of the sample is smaller and therefore the calculated exceedance probabilities of the observations tend to have larger distance from those of the Gumbel copula. However it is obvious that Gaussian copula, which was the second best among the investigated families, underestimates the probability of extreme environmental conditions in all cases. This is crucial in our case where the quality of the estimated duration of the operations is influenced by the quality of estimation regarding extreme environmental conditions that hinder the operability of the vessels and equipment. Therefore, based on the conducted tests on both available environmental data sets, one can safely conclude that among the one parameter copula families investigated Gumbel is the most appropriate to model the dependence between wind speed and significant wave height.

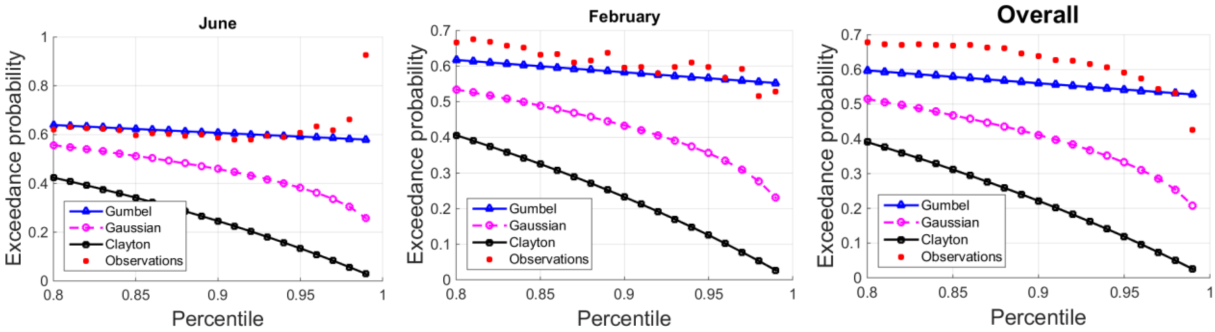


Figure 2: Conditional exceedance probabilities $P(U > u_p | V > u_p)$ for different percentiles concerning Deltares' environmental data

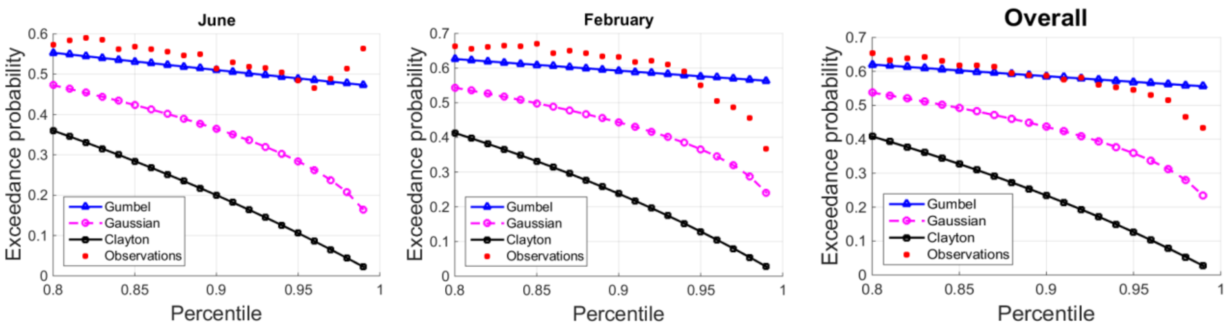


Figure 3: Conditional exceedance probabilities $P(U > u_p | V > u_p)$ for different percentiles concerning NOAA environmental data

Table 1: Semi correlation and square differences of Deltares environmental data

	ρ	ρ_{ne}	ρ_{sw}	S_N	S_{Gum}	S_{Cl}
Overall	0.6123	0.7092	0.1278	2.1420	0.9587	11.2876
February	0.6486	0.6479	0.2536	2.5932	1.3222	12.3801
June	0.6750	0.6431	0.3244	0.8740	0.3189	9.3321

Table 2: Semi correlation and square differences of NOAA environmental data

	ρ	ρ_{ne}	ρ_{sw}	S_N	S_{Gum}	S_{Cl}
Overall	0.6412	0.6322	0.1531	1.2848	0.5084	9.7843
February	0.6542	0.6496	0.1442	1.1522	0.4223	9.7625
June	0.5595	0.5692	0.1232	1.0443	0.4731	6.8582

5. Method of generating time series

Since our goal is to create environmental time series to simulate offshore installation operations in the near future, we can safely assume that using extreme value theory is not necessary. In this application the interest is on exceeding a certain percentile which is usually observed in the empirical margin. Hence, the proposed method uses copulas modelling and provides realistic time series of wind speed and wave height that do not exceed the maximum observed values. The methods presented here can be extended also by fitting a parametric distribution to the one dimensional margins.

By knowing the copula among the investigated families, that describes best the dependence between wind speed and significant wave height, it is possible to produce couples that take into account the dependence between the variables by using the estimated parameter for each month. However, in order to produce realistic environmental time series the autocorrelation should also be taken into account. Similar analysis as the one presented in section 3 was performed in order to investigate which copula describes the dependence of the wind speed with lagged versions of itself and it was found that the Gaussian copula describes it best.

After the copulas that describe the dependence of the environmental variables and the autocorrelation are both known, random environmental (synthetic) time series can be produced. In our case, the procedure of random time series generation was conducted for each month separately in order to include seasonality. The generation of time series for each month could be represented by simple vine or a dependence tree, as it is defined by Kurowicka and Cooke (2006), consisting of three nodes and two edges. The nodes are associated with marginal densities while the first edge specifies the autocorrelation of wind speed and the second edge specifies the dependence between wind speed and wave height, using the copula families that were determined by the analysis. The procedure consists of the following steps:

- Generate the first wind speed value u_t in $[0,1]$ using a uniform random number generator

- Calculate the values of wind speed in [0,1] based on the previous value (u_t) by solving the inverse conditional Gaussian copula (Joe 2014): $C^{-1}(u_{t+1}|u_t;\rho) = \Phi\left\{\Phi^{-1}(u_{t+1})\sqrt{1-\rho^2} + \rho\Phi^{-1}(u_t)\right\}$, where Pearson correlation $\rho = 2\sin\left(\frac{\pi}{6}r_s\right)$ and r_s is the Spearman's rank correlation coefficient.
- Next, the inverse conditional Gumbel copula function written in Matlab by Patton (2012) provides the value of wave height (v_t) in [0,1] for each of the generated wind speed values (u_t). The conditional Gumbel copula is described by the following relation (Joe 2014): $C(v_t \,|\, u_t;\theta) = {u_t}^{-1}\exp\left\{-\left[x^\theta + y^\theta\right]^{\frac{1}{\theta}}\right\}\cdot\left[1+(y/x)^\theta\right]^{\frac{1}{\theta}-1}$, where $x = -\ln u_t$ and $y = -\ln v_t$. Using the calculated parameter θ of the Gumbel copula the inverse conditional Gumbel copula is found numerically using a bisection method.
- The values of wind speed and wave height are transformed back to the original units through the inverse cumulative distribution function of each separate variable.
- Combine the generated time series of each month to acquire time series for the whole year.

It must be mentioned that following the proposed method, there will be a discontinuity between the values occurring at the last hour of the month and those at the first hour of the following month. However, this discontinuity is considered acceptable since our main focus lies on the environmental behaviour over a long period of time. Also, based on the method proposed by Joe (2014) time series were also generated by taking into account more than one lag (by using a D-Vine for the time series process of the wind speed $\{U_t\}$). However it was found that this approach does not improve the persistence of the synthetic time series and therefore it was decided to proceed with the proposed method based on a first order lag.

6. Application of synthetic environmental time series

A simulation algorithm, which performs Monte Carlo simulations concerning the infield cable installation of an offshore wind farm, was developed in order to identify the influence of using synthetic time series instead of observations. This algorithm (whose flowchart can be found in the

Appendix) along with the copulas algorithm were combined into a decision support tool (Figure 4) which can be used by concept engineers to compare different cable installation scenarios. The designed tool works as follows: first historical environmental data observed in the installation site are fed into the copula analysis algorithm, which performs the presented statistical tests and calculate the parameters regarding the dependence of the wind speed and the significant wave height as well as the autocorrelation for each month. Then, as many random environmental time series as needed are produced. Through testing, it was found that 1000 randomly generated annual time series are sufficient since the resulted CDF curves of installation's duration do not present important differences with a larger number of time series. Following, the produced time series, along with the cable installation scenario details are fed into the cable installation algorithm which simulates the proposed installation scenario for every different set of time series. Finally a CDF curve of the duration of the cable installation is obtained as output and the user is able to estimate the duration of a cable installation scenario within a confidence level.

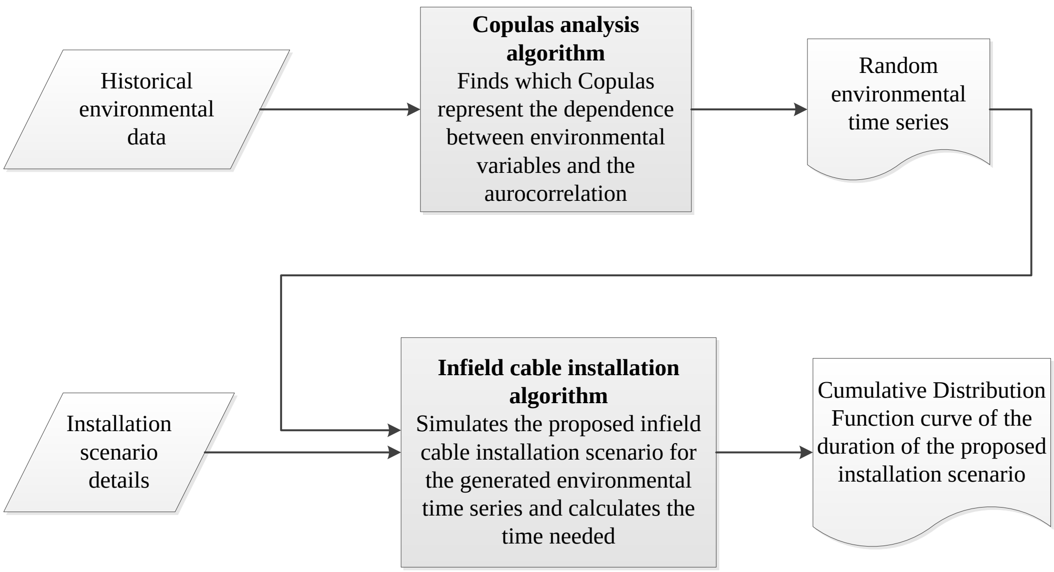


Figure 4: Overview of the designed tool

6.1 Test case description

A test case provided by Van Oord which concerns the cable installation of an OWF consisting of 55 wind turbines in the North Sea was simulated. The layout of the OWF, consisting of nodes (i.e. wind turbines and ports) and edges (i.e. lines that connect different nodes and represent the cables) is presented in Figure 5. The infield cable installation of an offshore wind farm is a complex process consisting of different operations performed by different types of vessels. The cycle of installation operations that take place in every edge of the OWF when post lay burial (PLB) of the cable is concerned, are presented in a Gantt chart (Figure 6). Also, the required performances of the vessels and the environmental limits (Table 3) for the different operations of the cable installation, were provided. To clarify, it must be mentioned that the performance concerns the duration of each operation, and that the limiting environmental conditions may differ for various operations.

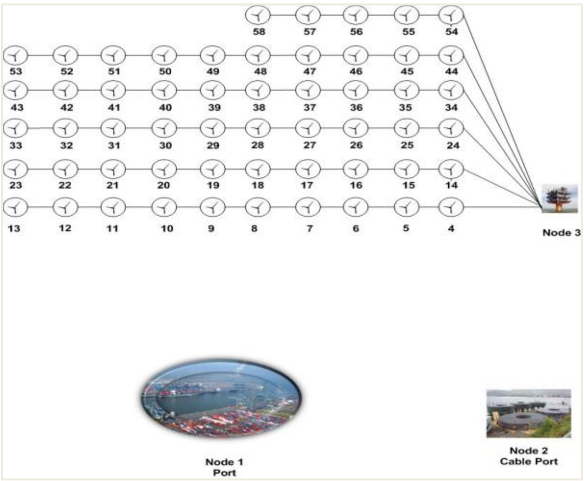


Figure 5: Offshore wind farm layout

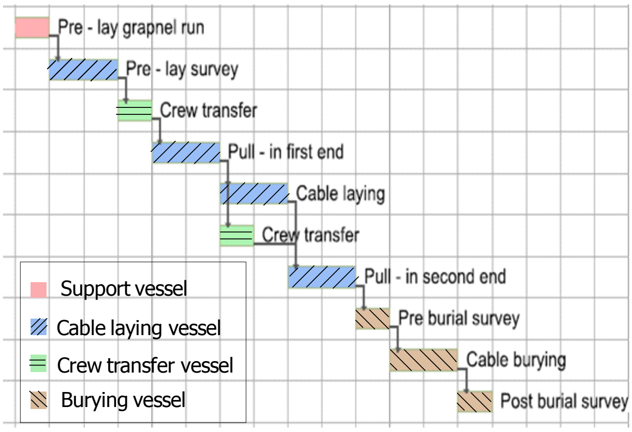


Figure 6: Gantt chart of the infield cable installation

Table 3. Operational weather limits.

Operation	Wind	Wave
Pre-lay grapnel run	-	1.5 m
Crew transfer	12 m/s	1.25 m
Pull-in	12 m/s	1.25 m
Pre-lay survey	-	1.75 m
Cable laying	12 m/s	1.75 m
Pre-burial survey	-	1.5 m
Burying cable	-	1.5 m
Post-burial survey	-	1.5 m

6.2 Synthetic time series validation

Based on 10 years of corrected 6 hours intervals measurements for a location in the North Sea close to the site of the installation, the statistical analysis mentioned in section 3 was performed. It was found that the Gumbel copula describes the dependence of the wind speed and the significant wave height best for all months except November and December when the Gaussian copula was preferred. Therefore, the time series for these two months were produced, as it was described in section 5, by using the inverse h-function of Gaussian copula, instead of the inverse conditional Gumbel copula. Applying the proposed method, 1000 random annual time series of wind speed and wave height, considering 6h intervals, were constructed and their scatter plot is presented in Figure 7. While Figure 8 shows the scatter plot of the observed time series. It can be seen that the plots present similarities in terms of shape, marginal distributions and extreme values.

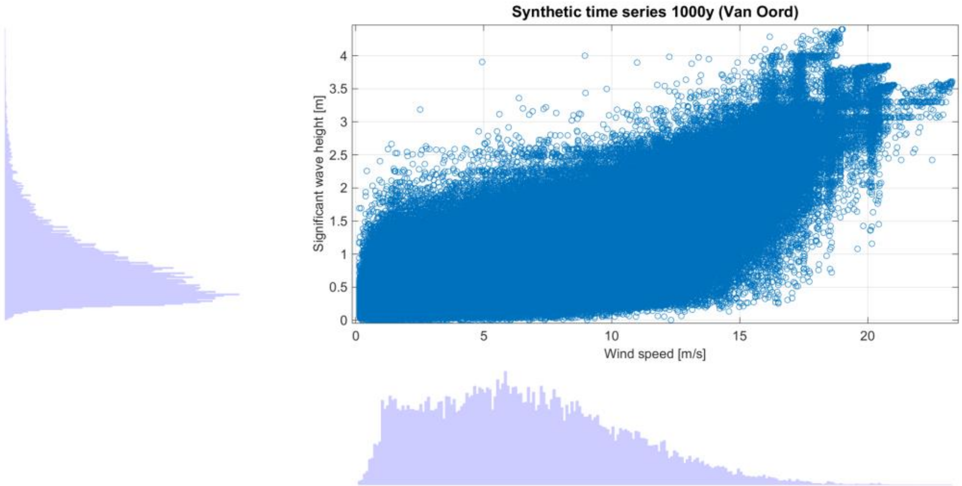


Figure 7: Scatter plot of 1000 constructed annual time series using copulas based on Van Oords environmental data .

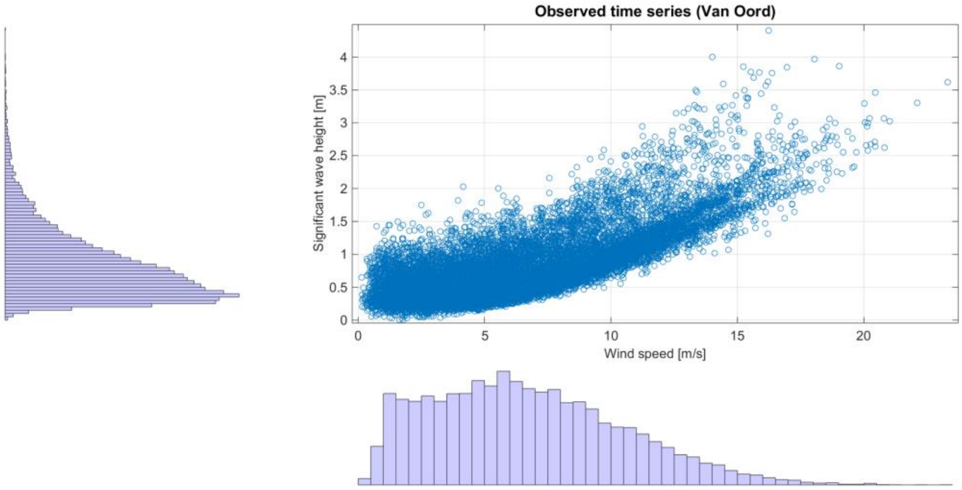


Figure 8: Scatter plot of 10 years of observed time series provided by Van Oord

Besides the visual comparison between the scatterplots of the observed and generated time series, two additional characteristics of the time series were compared in order to validate the synthetic time series. These are the monthly workability and the persistence of weather windows. The monthly workability concerns the percentage of the time steps during which an operation, limited by certain environmental thresholds, can be performed for every month. “The persistence of an environmental parameter above (below) some threshold level is defined as the time interval between an up-crossing (down-crossing) of that threshold level and the first subsequent level down crossing (up-crossing) of the same level” (Anastasiou and Tsekos 1996). Similarly, the persistence of weather windows can be defined as the amount of hours that the environmental parameters (i.e. wind speed and wave height) do not exceed the environmental thresholds (or limits) of an operation. In order to validate the synthetic time series, the monthly workability and the persistence of weather windows were calculated for both synthetic and observed time series, considering an operation limited by values of significant wave height (X_2) larger than $1.5\ m$ and values of wind speed (X_1) larger than $12\ m/s$. Moreover, different realistic environmental limits were tested, without resulting in significant differences from the presented case.

In Figure 9, it can be seen that the mean workability of 1000 synthetic time series is very similar to the mean workability for 10 years of observed time series, for every month. This indicates that the proposed method captures sufficiently the dependence between wind speed and wave height, since it is possible to produce synthetic time series with similar workability to the observed. However in order to ensure that the proposed method produces realistic time series that take into account the time dependence of the environmental variables, the persistence of the weather windows was also tested. The CDF of the persistence of the observed time series is compared to the CDF of the persistence of the synthetic time series and to every year’s persistence of the synthetic time series, in Figure 10 and 11 respectively. As it can be seen in Figure 10, the CDFs of the persistence of weather windows are very similar for both observed and synthetic time series. In Figure 11, one may observe that the CDF of the persistence varies for different years of the generated time series. These results show that using the proposed method, it is possible to produce realistic time series that present similar characteristics to the observations avoiding overfitting of the model.

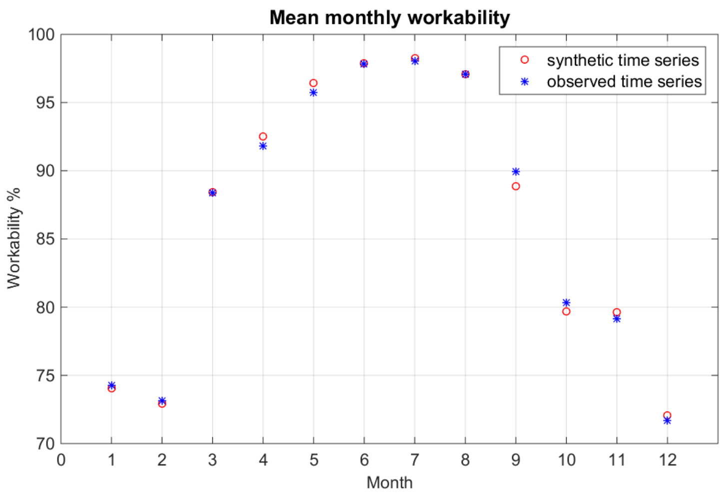


Figure 9: Mean monthly workability of observed and synthetic time series.

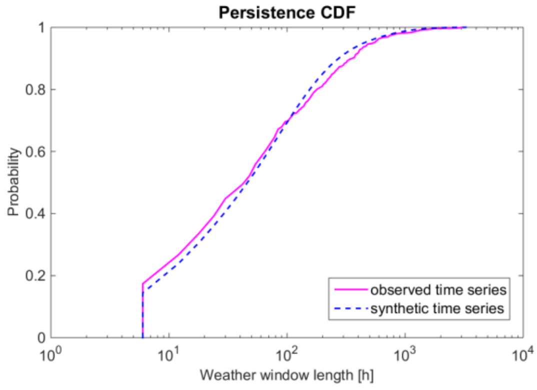


Figure 10: Comparison between CDFs of the persistence for observed and synthetic time series.

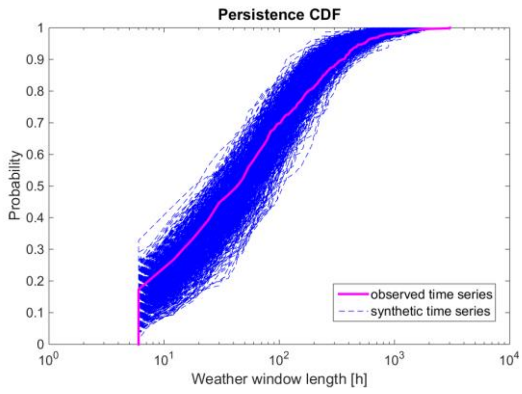


Figure 11: Comparison between CDF of the observed persistence and CDF of persistence for 1000 possible realizations.

6.3 Transforming time series

As it was mentioned before, cable installation consists of different sub-operations performed by various vessels with different operational limits. However some of these sub-operations may have durations smaller than six hours. Therefore it is needed to transform the time series from 6h intervals to 1h intervals. In this research we calculate the values for 1h interval using linear interpolation. After transforming both, observed and synthetic time series, to time series with 1h intervals, the workability and the weather windows' persistence was tested and ensured that the transformed time series (1h) have the same characteristics as their original (6h) time series. The workability plot concerning the interpolated synthetic time series was identical to the one concerning 6h synthetic time series (Figure 9). The plot showing the cumulative distribution of the persistence concerning the observed time series and the interpolated synthetic time series can be found in the Appendix (Figure 18). The resulting plot shows that the persistence of interpolated synthetic time series is similar to the persistence of observations. The main difference with Figure 11 is that there is a probability of having smaller weather windows than 6 hours since the resolution of the synthetic time series is 1 hour. However it must be mentioned that the probability of these cases is lower than the one of the 6h observations. This result supports the statement that it is possible to obtain higher resolution time series maintaining the characteristics of the observations.

To conclude this section, we call the reader's attention to two different cases in which 1h time series may be produced using the proposed method and 6h data available: i) use linearly interpolated 1h historical data for the construction of synthetic time series ii) produce 1h synthetic time series directly from 6h available historical time series. It must be noted that both cases will lead to a wrong process as judged by persistence statistics. Hence, it is important to first produce synthetic time series with the same resolution as the available observed time series and then use interpolation to increase the resolution.

6.4 Results of simulations

The infield cable installation scenario under consideration was simulated for different sets of time series (1h), considering 1st of June as starting date of the operations and the obtained CDF curves were compared. Firstly, it is important to compare the CDF curves of the cases where the observed time series and the randomly constructed time series were considered for the simulation. Secondly, it is of interest to compare the same CDF curves when other types of uncertainties are also included. Thirdly, we compare the CDF curves of the cases where dependently and independently constructed time series are used. The last comparison is for illustration purposes, in order to emphasize why offshore operators should never consider environmental variables as independent.

Observed versus Synthetic time series

i. Only environmental uncertainties considered

One could say that it is sufficient to simulate the installation scenario for the available observed time series in order to acquire a good estimate of the duration. This statement was investigated by simulating the same cable installation scenario (PLB) for observed and synthetic time series without taking into account any other uncertainties (i.e. the durations of the operations were assumed constant and risks of failures were not considered). In Figure 12, the CDF curves of these cases concerning the environmental data provided by Van Oord are presented. Usually project managers use CDF curves to estimate the duration of a project within a confidence interval. In practice experts often base their decisions on the 70th or 80th percentile (or P70, P80 value) of the CDF curve of the duration.

Regarding the 70th percentile, it can be seen that when synthetic time series are used for the simulation, the estimated duration equals 1220 hours. However, the P70 value of the CDF of the duration when observed time series are considered, may range from 1190 – 1240 hours. Hence, when the scenario is simulated concerning dependently constructed time series instead of a limited number of observed time series, it is possible to acquire more accurate estimates of the total duration.

In general it can be said that the estimate of the duration using synthetic time series is similar to the estimate when observed time series are used. However, constructing a large number of time series includes more possible environmental realizations to the estimation of the duration. Thus, the obtained CDF curve presents a bigger range (from 1125 to 1390 hours), incorporating more environmental uncertainties into the estimation of the duration and can be used to acquire a more precise estimate of the distribution of duration.

ii. Performance and failure uncertainties included

Besides environmental conditions there are also other uncertainties that influence the total duration of an offshore operation. Some of these uncertainties could be potential failures of equipment or variation of performance value from the deterministic value that has been assigned. Selected uncertainties were also included in the developed tool after consultation with cable installation experts, using common features of Monte Carlo simulation models (Hopkinson 2011). In particular it was decided to calculate the value of the most uncertain operation (i.e. crew transfer) from a triangular distribution and assign a failure probability equal to 2.5% and its impact in time regarding the pull-in operations. The results of the simulations for the same sets of observed and synthetic time series are presented in Figure 13. It is interesting to see that seven different runs of simulations concerning the observed time series present significant variations in the CDF of the duration of the installation. However, this is not observed for the CDF of the duration, concerning 1000 synthetic time series, which was identical for different runs. Hence, it can be stated that this outcome shows clearly the importance of having a large number of realistic time series in order to acquire a reliable estimate of the duration of the installation, including uncertainties regarding the environment, the performance of the equipment and possible failures.

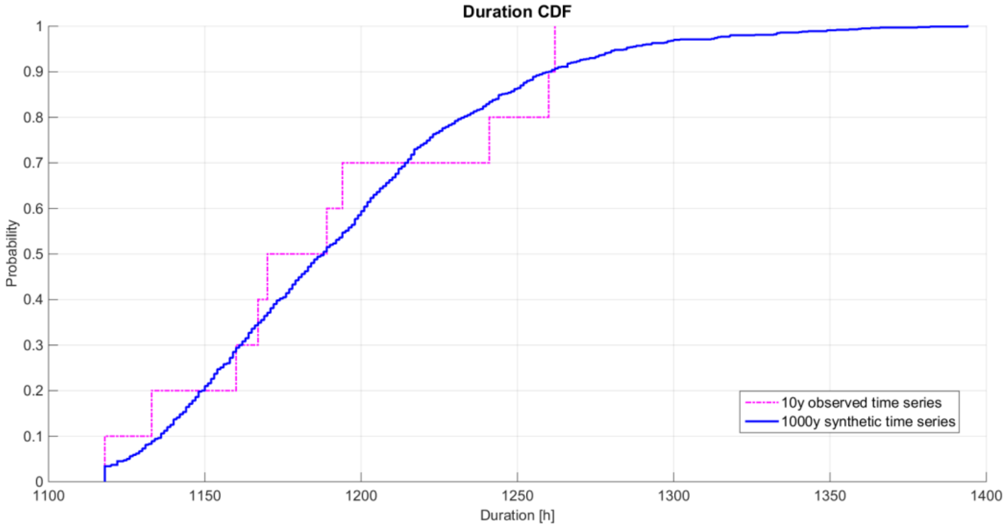


Figure 12: Comparison between observed and constructed time series based on Van Oord's environmental data without taking into account other uncertainties.

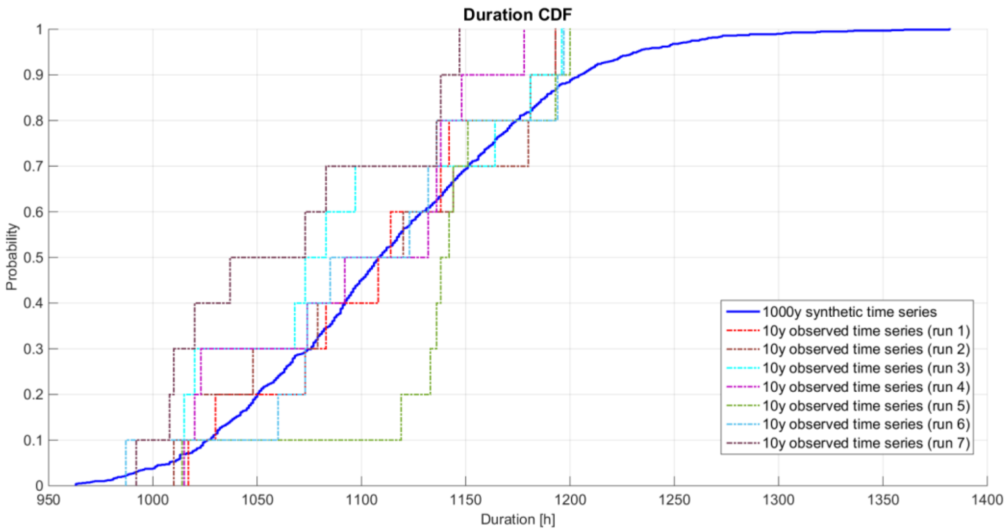


Figure 13: Comparison between CDF curves concerning synthetic time series and multiple runs for observed time series including performance and failure uncertainties.

Dependently versus independently constructed time series

For illustration purposes we present the case where environmental time series are constructed independently by calculating random numbers in [0,1] and using the inverse empirical cumulative distribution of every month to transform them in the appropriate range. As it was expected, the scatter plot of the independently constructed time series (Appendix) seems unrealistic compared to the observed time series. As it can be seen in Figure 14, the duration CDF curve of the independent case has larger values and bigger range than that of the dependent case. This result was anticipated for the independent case, due to the fact that the cases where at least one of the two values exceeds the environmental limits of an operation are more often, resulting in shorter weather windows and subsequently larger total duration. Therefore, the P70 value of the independent case overestimates the estimated duration in an order of 600 hours (i.e. 25 days) compared to the case of dependently constructed time series. An overestimation of that scale may lead to false decisions regarding the scheduling of the operations and the installation components (i.e. vessels and equipment), resulting in increase of the cost of the cable installation. All the above reasons indicate that constructed time series that take into account the dependence of the wind speed and the wave height should be used in order to safely estimate the time of the cable installation.

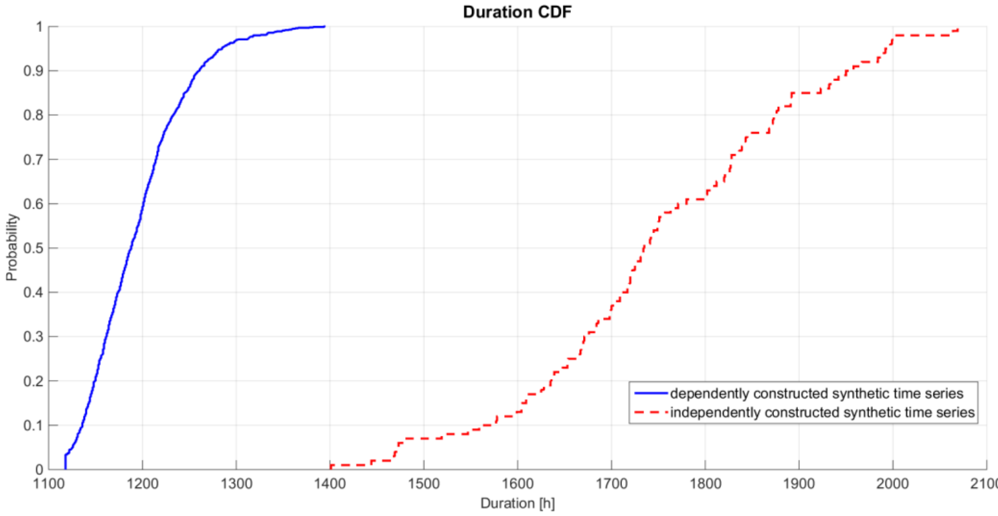


Figure 14: Comparison between dependently and independently constructed time series based on Van Oord's environmental data without taking into account other uncertainties.

6.5 Comparison of different installation scenarios

The tool developed can help project schedulers and researchers in comparing different installation scenarios based on the estimation of their duration, including environmental, performance and failure uncertainties. For that reason a different cable installation scenario was built and simulated. This scenario concerns the case of simultaneously burying (SB) which does not require a separate vessel for the burying of the cable but considers smaller performance regarding the cable laying/burying operation. The installation cycle of simultaneously burying scenario for every edge of the OWF, is presented in a Gantt chart (Figure 15). After simulating both cable installation scenarios, it can be seen from the obtained CDF curves (Figure 16) that it is more probable that the PLB scenario will result in smaller duration of the installation than the SB scenario. Hence, it is obvious that SB scenario would need more time to complete the installation and it should not be preferred over the PLB scenario. However, it must be mentioned that it could be possible that the SB scenario would be preferable in terms of costs, since it concerns three instead of four vessels, whose day rates are very expensive. Therefore, if one is interested in finding the overall optimal scenario, it is recommended to also investigate the cost.

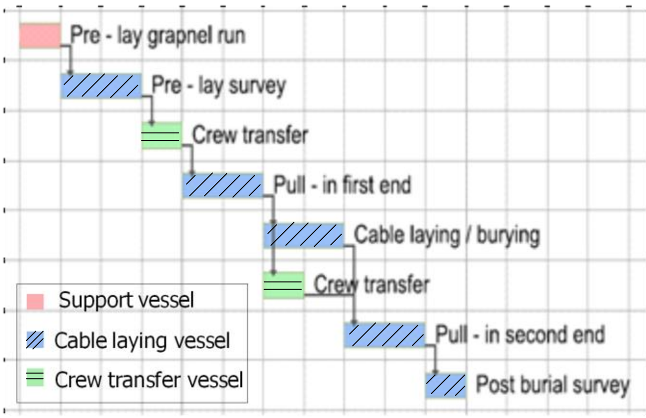


Figure 15: Gantt chart of SB scenario

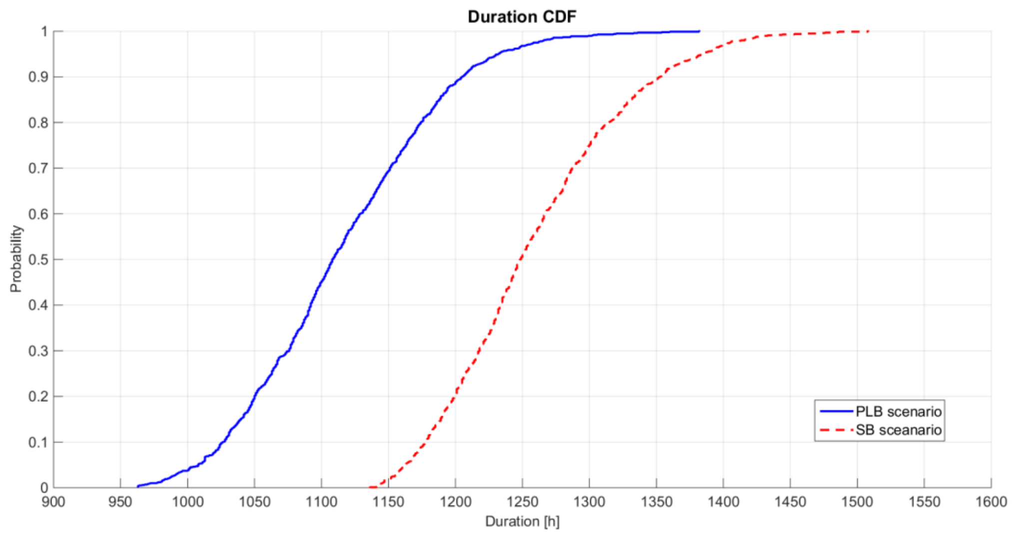


Figure 16: CDF curve of duration concerning PLB and SB scenarios including performance and failure uncertainties

7. Conclusions and recommendations

In this paper a method is proposed in order to produce realistic random time series of the environmental conditions, such as wind speed and significant wave height, that usually hinder the operability of vessels for offshore operations. The proposed method uses copulas to take into account the autocorrelation and the dependence of wind speed and significant wave height. Usually observed time series are not available for many years. Therefore, being able to synthesize large number of realistic time series of wind speed and wave height can help project schedulers or researchers in simulating offshore operations including environmental uncertainties.

In order to evaluate the importance of using the proposed method in the estimation of the duration of offshore operations a decision support tool for the cable installation of offshore wind farm, was developed. A cable installation scenario as well as time series measured wind speed and wave height close to the installation site, were provided by Van Oord. Using the proposed method, 1000 synthetic time series were constructed and validated by comparing important characteristics such as workability and persistence with those of the observed time series. Then, the cable installation scenario was simulated for different sets of time series.

It was found that dependently constructed synthetic time series provide a better insight into the duration of the cable installation compared to the case where only observed time series are used. The main advantage has to do with the fact that more possible realizations of environmental conditions are taken into account. Concerning the case of independently constructed time series, the results show that not including the dependence between wind speed and significant wave height will lead to unrealistic time series which will result in miss-estimation of the duration of the offshore operations which are limited by these environmental characteristics. The presented method can help professionals and researchers who are interested in including uncertainties, concerning environmental conditions, performance of the equipment and possible failures, in the scheduling of offshore operations and the comparison of different scenarios, by acquiring more accurate and reliable estimations of their duration.

Besides wind speed and wave height, there are also other environmental variables such as wave period, wind direction, current speed etc., which limit or influence offshore operations. Extending the models which represent the bivariate joint probability distributions to models which represent multivariate joint probability distributions, adds significantly to the complexity of the models. As it was already mentioned in section 1, many studies have focused on describing the bivariate or multivariate joint distributions of metocean variables without aiming in constructing synthetic time series. Since nowadays it is possible to acquire historical environmental time series, it would be possible to extend the method proposed in this article in order to construct sufficient number of synthetic time series concerning more metocean variables. These synthetic time series can be used as input to stochastic simulation models in order to acquire better estimates of the duration of offshore construction activities. The use of Vine copulas, which have been used for time series in financial applications (Brechmann and Czado 2015; Smith 2015), is suggested for future study. It is expected that different families of copulas with more than one parameter would be needed in order to describe the multivariate distributions. Attention should be paid though, because as the number of metocean variables increases, the dependence of these variables with each other and in time will become more complex. A good way to validate results is through persistence and workability as presented in section

6.2. How much the persistence and workability computed from synthetic time series would differ from those computed from the observations, when more environmental variables are added, is to be investigated.

Finally, for future work, it is suggested to expand the features of the developed tool in order to be able to simulate the entire installation of an OWF including also uncertainties concerning the cost, the availability of components, the duration of operations and failure events. Incorporating these uncertainties will contribute significantly towards the reduction of their current high costs.

Appendix

Scatter plots of independently constructed time series

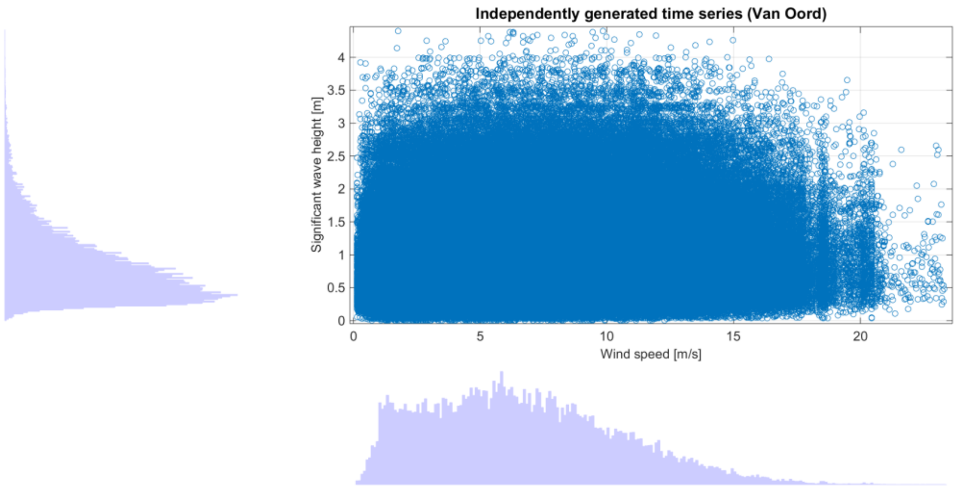


Figure 17 Scatter plot of independently constructed annual time series based on Van Oord's environmental data.

Persistence plot of1h interpolated synthetic time series

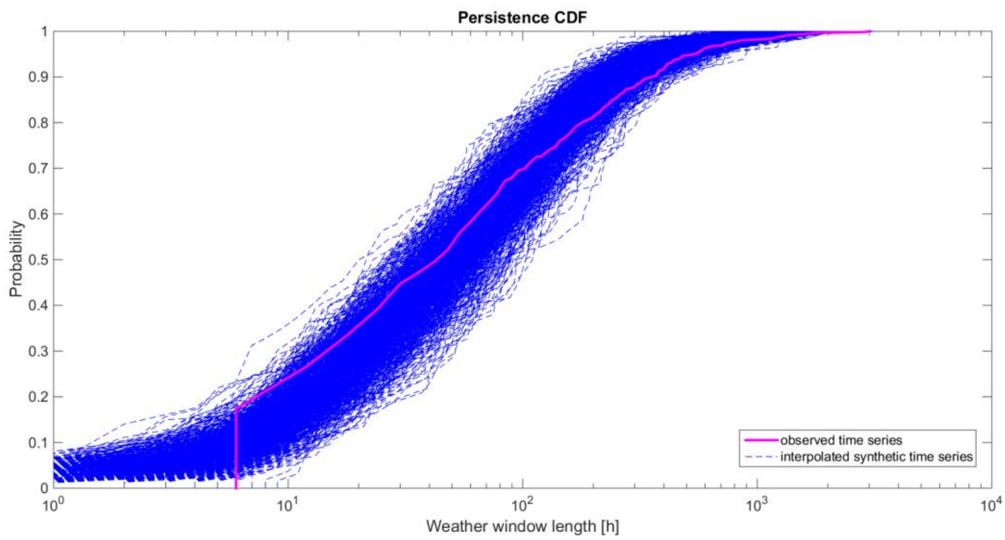
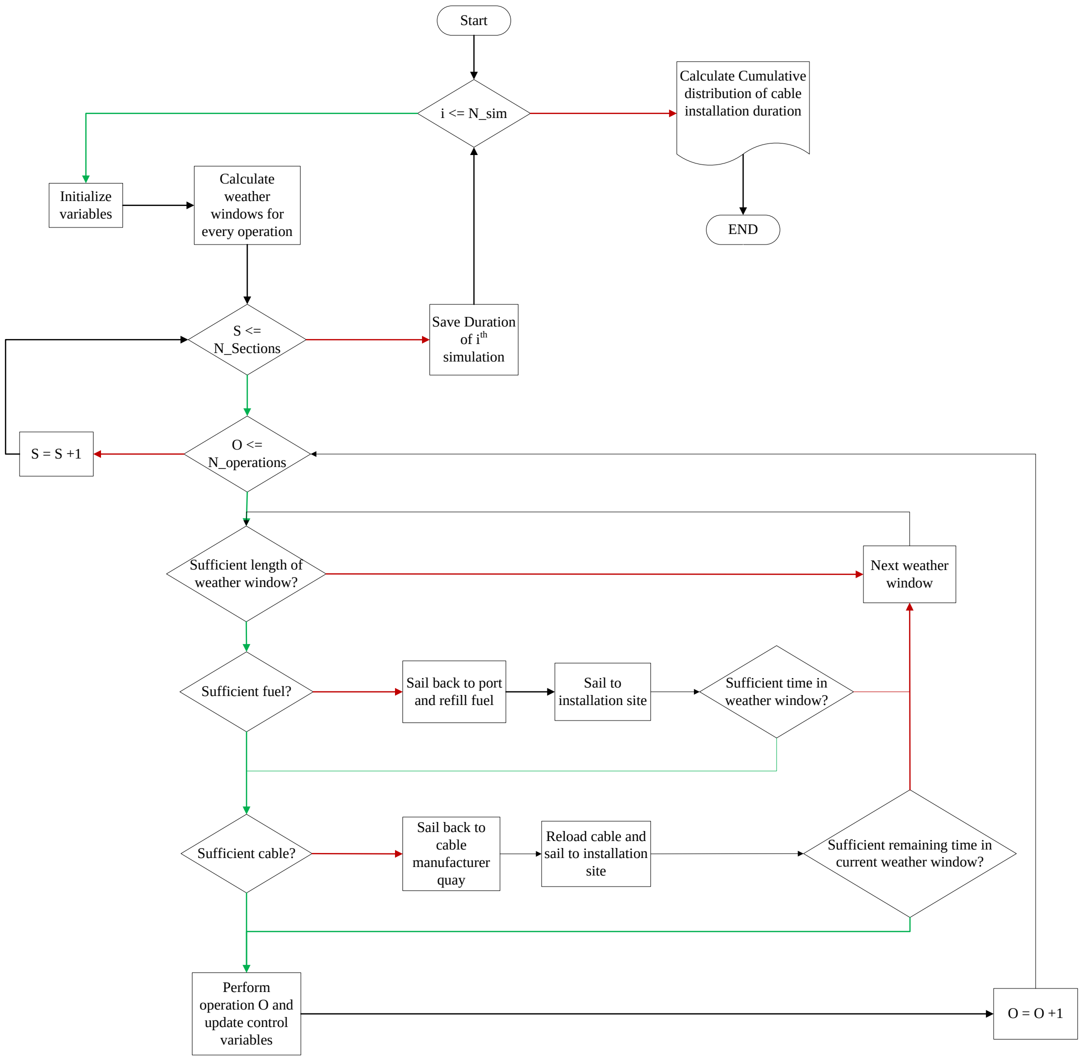


Figure 18: Comparison between CDF of the observed persistence and CDF of persistence for 1000 interpolated synthetic time series (1h).

Cable installation Flowchart



Flowchart nomenclature	
i	Index indicating the number of simulation.
N_sim	Number of simulations. This is equal to the number of synthetic time series.
S	Index indicating the section. Every section consists of two nodes (i.e. wind turbines) and an edge (the line that connects the two nodes), as it can be seen in Figure 5.
$N\_Sections$	The total number of sections.
O	Index indicating the operation of the cable installation cycle (Figures 6 and 15).
$N\_Operations$	Number of total operations of the cable installation cycle.

Acknowledgements

This study was conducted in terms of FLOW and EUROS programmes. The authors want to thank Royal IHC, Van Oord and Deltares for providing the required information and supporting financially this study.

References

Aas K., Czado C. Frigessi A. Bakken H. (2009). Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics, 44, pp. 182-198

Anastasiou K., Tsekos C. (1996). Persistence statistics of marine environmental parameters from Markov theory, Part 1: analysis in discrete time. Applied Ocean Research, 18, pp. 187-199

Athanasoulis G. A, Skarsoulis E. K. Belibassakis K. A. (1994). Bivariate distributions with given marginal with an application to wave climate description. Applied Ocean Research, 16, pp. 1-17

Athanassoulis G.A., Belibassakis K.A. (2002). Probabilistic description of metocean parameters by means of kernel density models 1. Theoretical background and first results. Applied Ocean Research, 24 (1), pp. 1-20

Bauer E., Staabs C. (1998). Statistical properties of global significant wave heights and their use for validation. Journal of Geophysical Research, Vol. 103, NO. C1. pp. 1153-1166, January 15, 1998

Bitner-Gregersen E. M., Haver S. (1989). Joint Long Term Description of Environmental Parameters for Structural Response Calculation. 2nd International Workshop on Wave Hindcasting and Forecasting.

Bitner-Gregersen E.M., Haver S.(1991). Joint Environmental Model for Reliability Calculation. Proc. of the 1st International Offshore and Polar Engineering Conference, Edinburgh, United Kingdom, August 11-15, 1991, ISOPE

Bitner-Gregersen E. M., Bhattacharya S. K., Chatjigeorgiou I. K., Eames I., Ellermann K., Ewans K., Hermanski G., Johnson M. C., Ma N., Maisondieu C., Nilva A., Rychlik I. Waseda T. (2014). Recent developments of ocean environmental description with focus on uncertainties. Ocean Engineering, 86 pp. 26-46

Bitner-Gregersen E. M. (2015). Joint met-ocean description for design and operations of marine structures. Applied Ocean research, 51, pp. 279-292

Brechmann E. C., Czado C. (2015). COPAR-multivariate time series modeling using the copula autoregressive model. Applied Stochastic Models in Business and Industry, 31, pp. 495–514

Charpentier A., Fermanian J. D., Scailler O. (2007). The estimation of Copulas: Theory and practice, Copulas: from theory to applications in finance. Risk Books, pp. 35-62

Clemen T., Reilly T. (1999). Correlations and Copulas for Decision and Risk Analysis, Management Science, 45(2), pp. 208-224. http://dx.doi.org/10.1287/mnsc.45.2.208

Dickel R., Hassanzadeh E., Henderson J., Honoré A., El-Katiri L., Pirani S., Rogers H., Stern J., Yafimava K. (2014). Reducing European Dependence on Russian Gass: distinguishing natural gas security from geopolitics. Oxford Institute for Energy Studies

Embrechts P., Lindskog F., Mc Neil A. (2003). Modelling Dependence with Copulas and Applications to Risk Management, in Handbook of Heavy Tailed distributions in Finance, ed. Rachev S. T, (Amsterdam: Elsevier/North Holland), pp. 329-384

Esteban M. D., Diez J. J., Lopez J. S., Negro V. (2011). Why offshore wind energy?. Renewable Energy, Vol. 36, pp. 444-450

EWEA, (2012). The European Offshore Wind Industry Key 2011 Trends and Statistics, available at: http://www.ewea.org/

Fouques S., Myrhaug D., Nielsen F. G. (2004). Seasonal Modeling of Multivariate distribution of Metocean Parameters with Application to Marine Operations, Journal of Offshore Mechanics and Arctic Engineering, 126, pp. 202-212

Galiatsatou, P., Prinos, P. (2007). Bivariate models for extremes of significant wave height and period. An application to the Dutch coast. Proc. 2nd IMA International Conference on Flood Risk Assessment, University of Plymouth, UK

Genest C., Favre A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 12 (4), pp. 347-368

Genest C., Remillard B, Beaudoin D. (2009). Goodness-of-fit tests for copulas: A review and a power study". Insurance: Mathematics and Economics, 44 (2), pp. 199-213

Hopkinson M. (2011). Monte Carlo Schedule Risk Analysis - a process for developing rational and realistic risk models. Risk Management Capability

Joe H. (2014). Dependence Modeling with Copulas. CRC Press, Taylor & Francis group

Kurowicka D., Cooke R. (2006). Uncertainty analysis with high dimensional dependence modelling. John Wiley & Sons, Ltd

Memos, C.D. and Tzanis, K. (2000). Joint Distribution of Wave Heights and Periods in Waters of any Depth. Journal of Waterway, Port, Coastal, and Ocean Engineering, ASCE, May/June, 126 (3), pp. 162-172

Monbet V., Ailliot P., Prevosto M. (2007). Survey of stochastic models for wind and sea state time series. Probabilistic Engineering Mechanics, 22, pp. 113-126

Morales-Napoles O., Steenbergen R. D. J. M. (2014). Analysis of axle and vehicle load properties through Bayesian Networks based on Weigh-in-Motion data. Reliability Engineering and System safety, 125, pp. 153-164

Nataf A. (1962). Determination des Distribution don’t les Marges sont Donnees. Paris, France: Comptes Rendus l’ Academiedes Sciences, 225, pp. 42-43

Nelsen R. B. (2003). Properties and applications of copulas: A brief survey. Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance, University of São Paulo

Nelsen R. B. (2006). An introduction to Copulas. Springer Series in Statistics, second ed., New York

Patton (2012). Copulas Methods for Forecasting Multivariate Time series. Duke University

Remillard B. (2010). Goodness-of-fit test for copulas of multivariate time series. SSRN Electronic Journal, available at: http://ssrn.com/abstract=1729982

Salvadori G, Tomasicchio G. R., D'Alessandro F. (2013). Multivariate approach to design coastal and off-shore structures, In: Conley, D.C., Masselink, G., Russell, P.E. and O'Hare, T.J. (eds.). Proceedings 12th International Coastal Symposium (Plymouth, England), Journal of Coastal Research, Special Issue No. 65, pp. 386-391,

Schmidt T. (2006). Coping with copulas. In Rank, J. (ed.), Copulas- From Theory to Applications in Finance. London, UK: Incisive Media Risk Books, pp. 3-34

Sklar A. (1959). Fonctions de Réparition à n Dimensions et Leurs Marges. Publications de l' Institut statistique de l' Université de Paris, 8, pp. 229-231

Smith M. S. (2015). Copula modelling of dependence in multivariate time series. International Journal of forecasting, 31, pp. 815-833

Wüstemeyer, C., Madlener, R., Bunn, D.W. 2015. A stake-holder analysis of divergent supply-chain trends for the European onshore and offshore wind installations. Energy Policy 80, 36-44.

Yang X., Zhang Q. (2013). Joint probability distribution of winds and waves from wave simulation of 20 years (1989-2008) in Bohai Bay. Water Science and Engineering, 6 (3), pp. 296-307

Zounemat-Kermani M., Kisi O. (2015). Time series analysis using chaos theory. Ocean Engineering, 100, pp. 46-53