Asymmetric Copula-Based Distribution Models for met-ocean data in offshore wind engineering applications

T. Fazeres-Ferradosa[1], F. Taveira-Pinto[1], E. Vanem[2], M. T. Reis[3] and L. Das Neves[1,4]

[1] FEUP, Faculty of Engineering of University of Porto, Department of Civil and Environmental Engineering, Hydraulics Water Resources and Environmental Division, Porto, Portugal, e-mails: dec12008@fe.up.pt; fpinto@fe.up.pt; lpneves@fe.up.pt

[2] DNV-GL Strategic Research and Innovation, Høvik, Norway & University of Oslo, Department of Mathematics, Oslo, Norway e-mail: Erik.Vanem@dnvgl.com

[3] LNEC, National Laboratory for Civil Engineering, Hydraulics and Environment Department, Lisbon, Portugal, email: treis@lnec.pt

[4] IMDC International Marine & Dredging Consultants, Antwerp, Belgium, email: luciana.das.neves@imdc.be

Abstract

Joint statistical models for long-term wave climate are a key aspect of offshore wind engineering design. However, to find a joint model for sea-state characteristics is often difficult due to the complex nature of the wave climate and the physical constraints of sea-state phenomena. The available records of wave heights and periods are often very asymmetric in their nature. This paper presents a copula-based approach to obtain the joint cumulative distribution function of significant wave heights and the mean up-crossing periods. This study is based on 124 months hindcast data concerning Horns Rev 3 offshore wind farm. The extra-parametrization technique of symmetric copulas is implemented to account for the asymmetry present in the data. The analysis of the total sea, the wind-sea and primary swell components is performed separately. The results show that the extra-parametrization technique with pairwise copulas consistently provided a better goodness-of-fit when compared to symmetric copulas. Moreover, it is demonstrated that the separation of the total sea into its components does not always improve the extra-parametrized copula's performance. Furthermore, this paper also discusses copula application to offshore wind engineering.

Keywords

Offshore wind, Significant wave height, Mean wave period, Copula, Extra-parametrization, Joint statistical model, Crámer-von Mises, Asymmetry

Introduction

The design of offshore wind structures is a complex task, which requires a detailed knowledge on the environmental conditions during their construction and lifetime. Moreover, the reliability and safety assessment of these structures is very much dependent on the ability to account for the uncertainties related to the environmental loads. When dealing with meta-ocean data, the wave climate, i.e. wave height and wave period, is a crucial component for a safe design. Therefore, several standards for the offshore wind and marine industry require a proper modelling of these random variables and the inherent uncertainties, e.g. [1]-[5].

Also when attempting to estimate extreme events associated with a specific return period, e.g. 100-year significant wave height, one often has to deal with the statistical modelling and inference techniques applied to the long-term sea-state characteristics. However, the complex nature of the met-ocean environment, the physical constraints of wave propagation and steepness and other uncontrolled climate effects, make it practically impossible to perfectly model the wave climate at the desired location. Nevertheless, the statistical description of relevant ocean parameters is a requirement for risk and reliability analysis [6].

Despite the extensive use of numerical models to describe the wave climate and the deployment of monitoring buoys, cost-related problems have contributed to the development of statistical models. Often the time series provided by buoy systems have considerable missing data created by a damage or vagrant buoy or because the observations at a temporary site are not enough to build an accurate model for extreme events. Sometimes, during extreme events, which are very energetic, buoys present malfunctions due to the harsh oceanic environment [7],thus leading to the lack of information during these occurrences. This missing information also affects the input for hindcast numerical models, such as the WAve Modelling (WAM) or the Simulating WAves Nearshore (SWAN). On the other hand, the deployment of buoys and the use of numerical models such as WAM or SWAN are considerably expensive, both in man and simulation hours and computational requirements [7]. In fact the common project of an offshore foundation or a marine structure often implies a balanced mix between buoy observations, numerical and statistical models, whichare further used for design, construction and maintenance operations.

In the majority of structures placed at sea the significant wave height is the parameter adopted to express the severity of sea state [8]. However, the joint modelling of the significant wave height and the mean up-crossing period, or eventually the peak period, enables a more precise description of the sea-state and it is fundamental for several aspects of design, e.g. in the fatigue limit state assessment [9], the scour phenomena at the foundation [10] or the reliability of structural elements as the mooring lines [11].

Joint modelling of long-term sea state has been attempted by means of several statistical approaches, for example: the earlier Peak Over Threshold method and the Annual Block Maxima (e.g. see [12]), the conditional modelling approach, Conditional Extremes models, the bi-variate Maximum Entropy Method, the bi-variate lognormal model and the normality transformation-based models (e.g. see [13]-[16]), among others. The number of statistical approaches to model wave characteristics has increased considerably, not only in quantity but also in complexity. Comprehensive reviews on these and other models are given in [17] and also in [18].

An alternative that has been used recently for this purpose is the Copula approach. The use of copula-based models to build the joint distribution function of the significant wave height and mean up-crossing period, or the spectral peak period, has been newly applied, e.g. [6], [7]. The use of copulas is mainly due to their simplicity of calculation and due to their ability to describe the dependence structure between random variables, regardless of the assumptions made for the marginal distributions [19].Moreover, copula-based models present a straightforward and computationally fast method to simulate the random variables [20].

There are numerous families of copulas (see [21]). Furthermore, if one attends to the techniques used to combine different copulas, this number is practically infinite [6]. The most used families are the elliptical copulas, which are radially symmetric, the Archimedean copulas, which are only able to capture either lower or upper tail dependences, and the extreme value copulas, which according to [6] arise in the limit of component-wise maxima, but are also used to model general dependence structures. However, one of the major problems with these families is the fact that they are not able to account for asymmetry in the data. Asymmetry is often due to physical limitations related to e.g. wave breaking and maximum wave steepness.

Still the above mentioned copula families have been applied with a reasonable degree of success in other studies, namely, when modelling single storms with multivariate Archimedean copulas [22]. Also, [7] successfully used the Gaussian and t-copula, from the elliptical family, to estimate the wave height records through spatial correlation at the south coast of England. Still using the same Elliptical and Archimedean copulas (Clayton, Frank and Gumbel), [23] found that the Gaussian and the Gumbel copulas could accurately fit the empirical densities of the individual wave steepness and height. However, as stated by [6] and re-confirmed in the present study, it is far from straightforward to find a good copula-based model for non-symmetric met-ocean data, such as the significant wave height and the up-crossing mean wave period.

A possible way to solve the asymmetry problem is to use more complex copula-based models. For instance, one can use non-symmetric copulas as the Marshal-Olkin family [21], the dependence trees association [24] or the C- and D-vine copulas [25], which are usually applicable for higher dimensional problems. In the bi-variate cases, a simpler way of combining different copulas is to use the extra-parametrization technique, originally proposed by [26] and later on extended by [27]. This technique was then applied by [28] to develop multivariate extreme models. More recently, by using the same copula construction, [6] concluded that extra-parametrization with an independent or a pairwise copula could be used to model significant wave heights and mean up-crossing periods. Using the root-mean-square error (RMSE), the author concluded that this technique provided a better fit than the bivariate lognormal model, widely used in practice and introduced by [29]. Moreover, this work concluded that extra-parametrization of copulas provided a similar goodness-of-fit when compared with the conditional modelling used in current practice [5], [13]. However, in these works no extra-parametrization was performed with different copulas, e.g. Frank-Clayton or Gumbel-Clayton.

Finally, another important aspect is the fact that the literature does not present many works comparing the copulas' performance and their possible parametrizations to the separated components of wind-sea and swell. The majority of the works performed so far are dedicated to modelling the total sea-state, e.g. [6], [7], [16] or [30]. The separated performance of copula-based models for the total sea and the wind-sea and swell components has not been addressed yet. This aspect is of great importance, since sea components may present less asymmetry than the one showed by the combined sea. This may lead to an improvement of certain copulas' performance, i.e. the ones that are constructed to deal with symmetric cases only, as the Archimedean copulas. Furthermore, the study of wind-sea and swell components may be of importance when dealing with bi-modal sea states or in certain fields of application, e.g. the dynamic analysis of FPSO units or the floating foundations for offshore wind structures [31].

This paper provides inputs to these literature gaps, i.e. the use of copulas to deal with data asymmetry and the wind-sea and swell components. Extra-parametrization of copulas will be applied with an independent copula and pairwise copulas to the total sea (also referred here as the combined sea) and the wind-sea (referred to as the wind component) and the swell. Going one step further, an attempt is made to combine copulas from different families and with different tail dependences, to understand if any improvements are obtained when fitting the hindcast data. The case study corresponds to a record of 124 months of significant wave heights and mean up-crossing periods at the Horns Rev 3 offshore wind farm, located in the North Sea.

This paper is structured in the following manner: in this introduction one has explained the importance of this study and the problem to be solved (i.e. joint modelling of asymmetric combined, wind and swell met-ocean data), the section "Case study and wave data" presents the case study and the data to be modelled after short-term and seasonal dependence are removed. The section "Marginal distributions" the study of the marginal distributions of the significant wave height and the up-crossing mean wave period is performed. The section entitled "Copula models applied to the significant wave height and the mean up-crossing period" introduces the symmetric copula models and the asymmetric extra-parametrized models based on the Independent copula and the respective Pairwise copula. Then, the section called "Copula based models with different copulas" extends the extra-parametrization technique for different copulas and new models are proposed based on the Gumbel copula. Section "Tail dependence analysis" discusses these models' ability to deal with the tail behaviour, while section "Weighted-root-mean-square-error" provides an alternative tail analysis. The section "Discussion and Applications to Wind Engineering" addresses some of the most important aspects outlined by the results previously obtained and the application of copulas to several fields of wind engineering, with particular focus on offshore wind. Finally, one presents the "Conclusions" regarding the tested models for the separated sea components and their ability to deal with asymmetric wave datasets.

Case Study and Wave Data

The case study concerns Horns Rev 3 offshore wind farm [32]. This offshore wind farm is located in the Danish sector of the North Sea, 20-35 km north-west of Blåvands Huk and 45-60 km from the city of Esbjerg [32]. This area is relatively shallow and the water depth ranges closely from 10 m to 20 m. The statistical model of the wave heights and the mean up-crossing periods is based on the hindcast series presented in [32]. The present dataset includes the wind-sea and the primary swell values of significant wave height (H_s) and the mean up-crossing period (T_z), as well as the total sea values.

Detailed information concerning the location of the data sampling and hindcast validation is reported in [32] and corresponds to the following coordinates: Latitude of 55.725ºN and Longitude of 7.750ºE. The available dataset resulted in a total of 90553 pairs of significant wave height and mean up-crossing period, for both the combined sea and its components. The data are obtained from DMI-WAM (wave model) and correspond to an hourly output resolution between 01-01-2003 and 01-05-2013, i.e. 124 months. Table 1 provides the descriptive statistics for both components and the combined seastate. The data are not treated for the incident direction.

Table 1 shows that the wind significant wave height tends to be higher than the swell one, while the swell component tends to present larger mean periods. Positive kurtosis and skewness values indicate, respectively, that fat and long upper tails are expected for the present data. This is important, as the upper tail is the region of interest when dealing with the failure of marine systems and structures, because it is related with the largest wave heights.

Legend table 1: Descriptive statistics for the combined sea state, the wind-sea and swell components (hindcast data of 124 months).

Descriptive	Combined		Wind		Swell	
	H_s (m)	T_z (s)	H_s (m)	T_z (s)	H_s (m)	T_z (s)
Mean	1.458	5.9	1.033	3.9	0.883	7.2
Median	1.220	5.6	0.790	3.6	0.760	6.9
Std.deviation	0.932	1.7	0.938	1.9	0.517	2.0
Max	6.110	14.0	6.040	13.8	5.120	17.8
Min	0.140	2.2	0.000	1.0	0.110	2.6
Skewness	1.581	0.8	1.644	0.9	1.555	0.7
Kurtosis	6.161	3.6	6.426	4.0	6.590	3.6
Percentile 25%	0.800	4.6	0.360	2.5	0.510	5.8
Percentile 50%	1.220	5.6	0.790	3.6	0.760	6.9
Percentile 75%	1.850	6.9	1.420	4.9	1.120	8.4
Percentile 90%	2.690	8.2	2.270	6.4	1.560	9.8
Percentile 95%	3.350	9.0	2.940	7.5	1.900	10.8
Percentile 99%	4.780	10.7	4.400	9.5	2.640	12.7
Percentile 99.5%	5.320	11.3	4.950	10.2	2.957	13.4
Percentile 99.9%	5.890	12.7	5.700	11.7	3.640	15.3
N	90553					

When dealing with the combined results from wind and swell components the asymmetry between H_s and T_z is considerable. This may lead to difficulties when attempting to fit a statistical model to the overall data. [6] points out that fitting a model to the separate components may contribute to better fittings and to reduce the asymmetries in the data. The present research presents several copula-based models and analyses their comparative performance to understand if treating wind and swell components separately leads to a better goodness-of-fit compared to the overall model.

1. Data Pre-processing

Copula-based models are typically built on measures of dependence which are rank-based, e.g. the Kendall's tau (τ_K) or the Spearman's rho (ρ_S) [21]. Therefore, the data had to be treated for the existence of ties. Moreover, serial dependence needs to be removed in order to fulfil the underlying assumptions for the joint models application ([33], [6]).

The R package RANKS was used to randomly break ties present in the data, when constructing the pseudo-observations used to fit the copulas. Similarly to the data presented by [6], it was found that this did not affect the marginal distributions of the significant wave height and mean period, nor their parameters, which are estimated based on the Maximum Likelihood Estimation method (MLE).

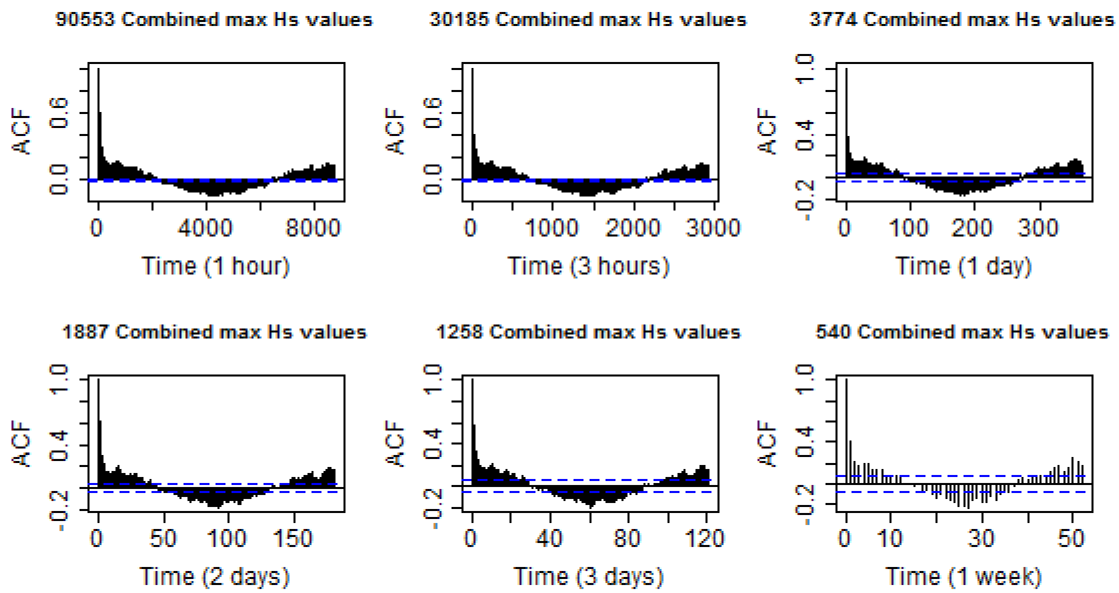
The short-term dependence was removed by subsampling the data. The underlying question is "which time interval should be used to subsample the data?" The auto-correlation function (ACF) was computed for several intervals, namely 1 hour, 3 hours, 1 day, 2 days, 3 days and 1 week. For each interval the maximum value of the significant wave height and the respective mean up-crossing period was selected as the subsample object. In the present study, the maxima are selected instead of the means in order to conserve the extremes information. It was concluded that, for both the significant wave height and mean up-crossing period no substantial reductions of short-term dependence were obtained from considering intervals larger than 2 days, e.g. 3 days or 1 week, as shown in figure 1.

Moreover, this interval complies with the value used in [32] to define a storm event for the Peak over Threshold application. One should note that the definition of storm duration and storm threshold of H_s is always questionable. [34] provides further details on storm characterization for different climate conditions, which is not the main focus of this work. However, according to the auto-correlation analysis (figure 1), subsampling the maximum significant wave height and the associated mean up-crossing period for a 2-day interval still shows a bit of short dependence. This can be due to the fact that the maximum significant wave height of a block of 2 days can be chronologically close to the maximum of the next block of 2 days. Nevertheless, in the present case, it was considered that the bi-daily maxima was a suitable subsampling.

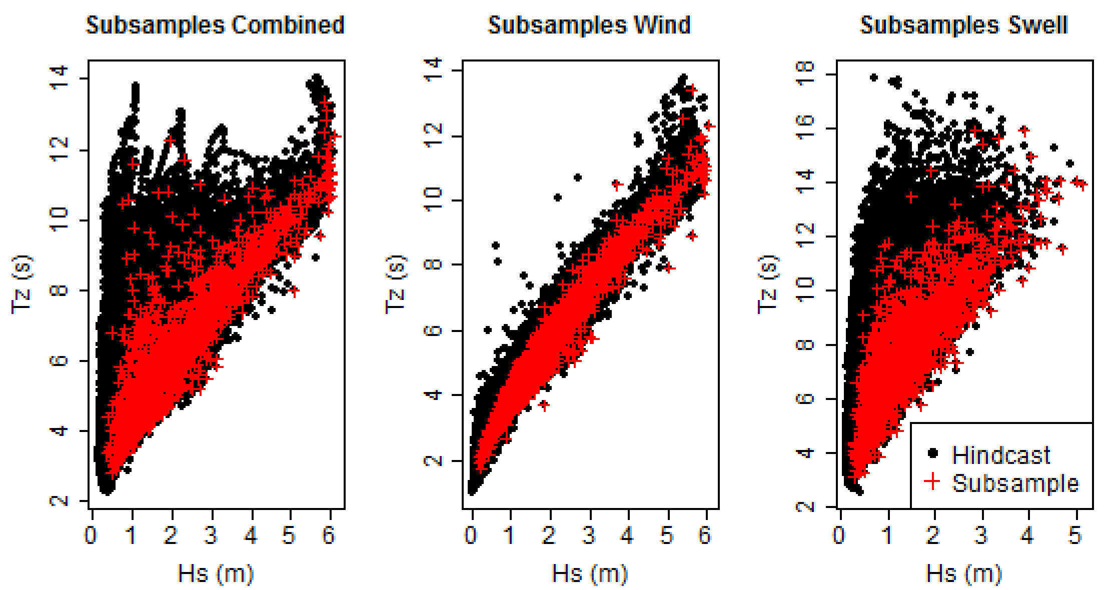
This subsampling led to samples of 1887 pairs of H_s and T_z for the combined sea-state, the wind and the swell components. Figure 2 provides the subsampling before seasonality is removed from the data.

According to [6], Seasonal effects can be removed by calculating the seasonal mean and the standard deviation for each annual cycle. The weekly data is then normalized by subtracting the seasonal mean and then dividing by the standard deviation for each week of the annual cycle. Then the overall mean is added.

This procedure may lead to pre-processed data that has negative values. This may pose a problem when dealing with the domain of certain distributions, e.g. the lognormal or the Weibull distributions. A possible way to deal with negative values is to work with the log of the data. If the data are indeed lognormal, their log should lead to new variables that follow a Gaussian (Normal) marginal distribution.



Legend figure 1: Autocorrelation function for the maximum value selected for different time intervals.



Legend figure 2: Subsamples obtained from choosing the maximum H_s for intervals of 2 days.

However, for the present case, the Shapiro-Wilk test for normality showed that neither the logarithmic transformation of the significant wave height or the mean period followed a Gaussian distribution, as shown in table 2. One should also note that for such large samples as these ones (n=1887 pairs), the Shapiro-Wilk test tends to be very sensitive to any departures from normality. An ad hoc solution was adopted by adding to the sum of the overall mean (as explained before) with the minimum integer number of standard deviations necessary to turn negative values into positive ones. This does not affect the dependence structure of the pre-processed data, since it corresponds to a location shift solely.

Moreover, after the copula-based models are fitted to the data the seasonality is added back again. Therefore, adding the overall mean alone or adding this value plus the necessary standard deviations, does not lead to differences in the models' outcome.

Legend table 2: Shapiro-Wilk test for normality of the log(H_s) and log(T_z) applied to each subsample. Confidence interval of 90%. Rejection of null hypothesis H0 for p<0.1.

H0: X follows normal distribution; X is either log(H_s) or log(T_z)						
Subsample	**Combined**		**Wind**		**Swell**	
CI=90%	H_s	T_z	H_s	T_z	H_s	T_z
Shapiro-Wilk	0.9933	0.9931	0.9911	0.9948	0.9962	0.9988
p value	1.36E-7	9.66E-8	2.55E-9	3.64E-6	1.1E-4	2.44E-1
Evaluation	Reject	Reject	Reject	Reject	Reject	Reject

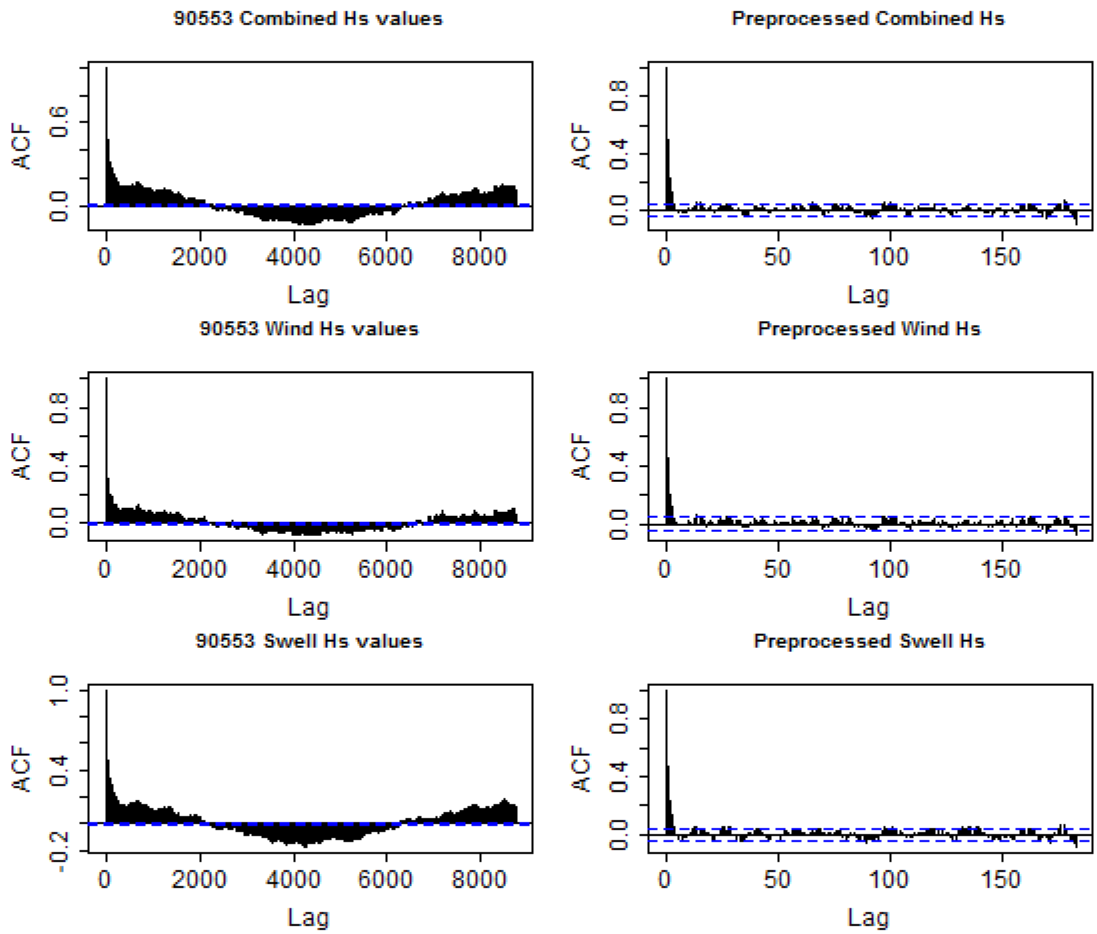
The seasonality is removed according to Eq. (1), where Y_i is the pre-processed data, X_i is the subsampled value of H_s or T_z, and the μ_j and σ_j are respectively the mean and standard deviation of the significant wave height and the mean period of the week j, with j=[1;52]. M stands for the overall mean of the subsampled data and w represents the minimum integer number of overall standard deviations of the subsampled data (Sd) necessary to make all Y_i positive.

$$Y_i = \frac{X_i - \mu_j}{\sigma_j} + M + w \cdot Sd \tag{1}$$

For the present dataset all values of T_z in the pre-processed data were already positive. However, for the significant wave height, it was found that the required w was equal to 0 for the combined sea and 1 for the wind and swell components. Note that the Sd of the subsampled wind and swell components is different, respectively, equal to 1.221 m and 0.78 m. Note that with such transformation, the pre-processed data of the combined sea does not necessarily have the highest H_s when compared with the wind and swell components. The same is valid for the percentiles of H_s (also see table 6), as mentioned when the seasonality effects are added back to the pre-processed data and this somehow counter-intuitive aspect gets dissipated.

Figure 3 shows that the subsampling and seasonality treatment led to a considerable reduction in the autocorrelation functions for the significant wave heights in the wind-sea, the swell and the combined sea-state.

The same was concluded for the mean up-crossing period, thus leading to pre-processed data that can be used to obtain the pseudo-observations for copula fitting. The left-most images of figure 3 concern the original hindcast data, while the right-most ones concern the pre-processed data, i.e. 2-day maximum H_s and respective T_z with weekly seasonal effects removed according to Eq. 1. Figure 3 also shows that the seasonality effect is more evident in the swell component than in the wind-sea component.

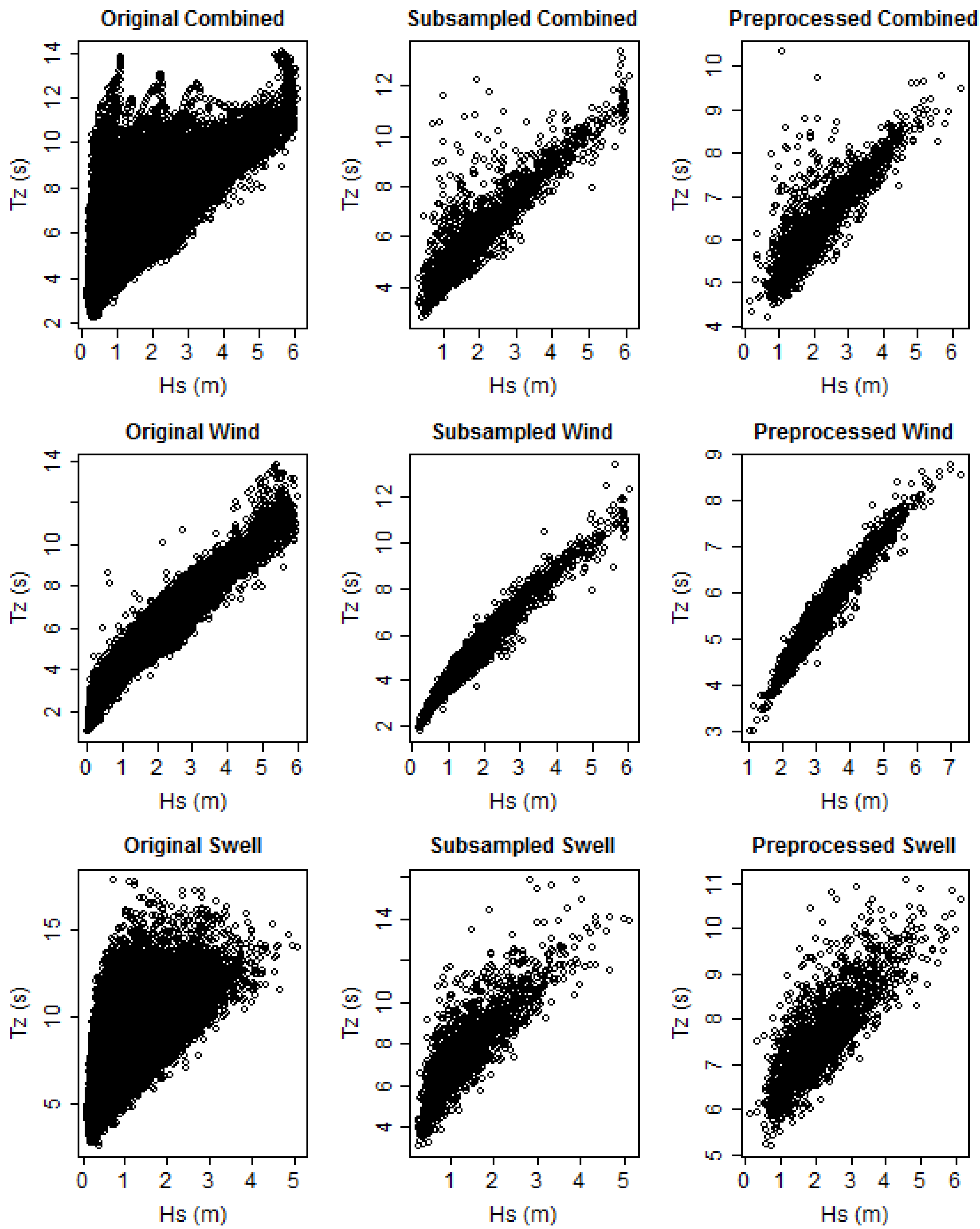


Legend figure 3: Autocorrelation functions for the original hindcast data (left) and the pre-processed data (right). Values presented for H_s. T_z has similar plots.

Figure 4 presents the scatterplots for the original data, the subsampled and the pre-processed data. When performing the subsampling procedure, one is able to see that the dependence between H_s and T_z changes considerably. This occurs because only the maximum data for the wave heights is being selected along with the associated up-crossing periods. Once again, since the season effect will be added back again, one expects that this change is dissipated after the final generation process. In the present case, the generated data will of course correspond to a model that expresses the 2 days maxima for the significant wave heights and the associated values of the mean wave up-crossing periods. The straight comparison between the original hindcast data and the pre-processed data becomes less relevant, because the first concerns to an hourly output, while the latter concerns the bi-daily maxima. Therefore, further comparisons are to be performed between the subsample and the pre-processed data. The model's output will correspond to maximum values per each two days, which are useful for offshore wind engineering design, namely in loads calculation. However, for reliability assessment purposes, the fact that the models refer to local maxima of H_s and T_z may lead to an overestimation of a system's probability of failure. Although this may result in a conservative assessment of an offshore system's safety, this option should be the aim of further detailed research.

One should note that the dependence measures have changed. In this work the measure of dependence, τ_K, was corrected for ties existence and obtained for all datasets. These values are summarized in table 3.

From table 3 it is also possible to understand that for the wind-sea component the positive dependence between the H_s and T_z is more evident than the one showed by the swell component. This is maintained after the subsampling and the removal of seasonality.



Legend figure 4: Comparison of H_s and T_z before and after the subsampling and the pre-processing.

Legend table 3: Kendall's tau (τ_K) for the original data, the subsamples and the pre-processed data.

τ_K	Original	~~Sub sample~~SubsampleSubsample	Pre-processed
Combined	0.534	0.719	0.688
Wind	0.893	0.875	0.864
Swell	0.539	0.648	0.621

At a first glance one should expect that the best fit is not provided by the same copula for the wind-sea, swell and the combined data, not only because the asymmetry shown by the data varies (figure 2 and figure 3), but also because the measure of dependence τ_K is different (table 3). Also note that for the same copula, the estimation of the copula parameter depends on τ_K. The values obtained in table 3 are similar to the ones presented in the literature for other locations; e.g. [6], [7], [11], [15] and [22] reported values of τ_K for several locations, ranging from 0.21 to 0.8.

An important aspect to be noted is that τ_K may not be a suitable analysis parameter when the behaviour of the tails is being analysed. For this matter the asymptotic dependence should be looked into in further detail. The dependence in the tails region considerably affects the choices on the possible models to be tested, because different copulas display different dependences at the lower or upper tails. Next section discusses the tail dependences with further detail. For now one should bear in mind that the values obtained for the overall dependence measure τ_K seem to be reasonable, in the sense that due to the wave steepness it is physically impossible to have very large wave heights with very short periods. In general, as the significant wave height increases the mean up-crossing period is also expected to increase.

Marginal Distributions

Before applying the copula-based models to the pre-processed data, an assessment of the goodness-of-fit of several marginal distributions was performed (see table 4). The Kolmogorov-Smirnov distance (KS) and the Wasserstein (W) distance were calculated between each tested marginal and the empirical cumulative distribution function of the pre-processed data. Table 4 provides the distances. One is able to conclude that for both measures the lognormal distribution function (table 4 in italics) was the one that provided a closer fit to the significant wave heights and the mean up-crossing periods. In fact, this distribution was also the best candidate for the original and the subsampled data.

The MLE method was then used to estimate the lognormal distribution parameters associated with the pre-processed data. In table 5 these parameters and the 95% confidence interval are shown. Table 6 gives the descriptive statistics of the pre-processed data.

Legend table 4: Kolmogorov-Smirnov (KS) and Wasserstein (WS) distances between theoretical marginal distribution and the empirical cumulative distribution function of the Pre-processed data.

Distribution	**Combined**		**Wind**		**Swell**	
	KS H_s	**KS T_z**	**KS H_s**	**KS T_z**	**KS H_s**	**KS T_z**
Normal	0.0972	0.0633	0.1004	0.0767	0.0684	0.0612
Exponential	0.3129	0.5125	0.4163	0.4956	0.2863	0.5273
Rayleigh	0.1068	0.4091	0.2461	0.3816	0.0789	0.4338
GEV	0.1523	0.1167	0.1554	0.1196	0.1507	0.1102
GP	0.2542	0.4451	0.3538	0.4202	0.2257	0.4535
Lognormal	*0.0242*	*0.0419*	*0.0507*	*0.0457*	*0.0353*	*0.0353*
Weibull	0.0711	0.0795	0.0916	0.0846	0.0406	0.0877
Weibull 3p	0.0711	0.0474	0.0916	0.0756	0.0406	0.0462
Distribution	**Combined**		**Wind**		**Swell**	
	WS H_s	**WS T_z**	**WS H_s**	**WS T_z**	**WS H_s**	**WS T_z**
Normal	0.1856	0.1151	0.1912	0.1338	0.1420	0.1141
Exponential	0.7098	1.5276	1.0961	1.4760	0.7150	1.6213
Rayleigh	0.1244	1.0733	0.4553	1.0025	0.1128	1.2460
GEV	0.3411	0.2517	0.3577	0.2983	0.3168	0.2913
GP	0.4414	1.0856	0.7421	1.0467	0.4446	1.1824
Lognormal	*0.0592*	*0.0756*	*0.0810*	*0.0751*	*0.0688*	*0.0612*
Weibull	0.1254	0.1731	0.1955	0.1951	0.0807	0.2113
Weib 3p	0.1620	0.0997	0.1755	0.1472	0.1178	0.0976

Legend table 5: Parameters and 95% Confidence interval of the lognormal distribution fitted to the pre-processed data. Log-mean and log-st.dev are the parameters of the distribution and CI is the respective 95% confidence interval.

	Combined		Wind		Swell	
	H_s	T_z	H_s	T_z	H_s	T_z
Log-mean	0.691	1.840	1.119	1.693	0.679	2.0186
CImean 95%	[0.669; 0.711]	[1.833; 1.847]	[1.106; 1.133]	[1.685; 1.701]	[0.657; 0.700]	[2.013; 2.024]
Log-st.dev.	0.454	0.152	0.296	0.175	0.467	0.128
CI 95%	[0.440; 0.469]	[0.147; 0.157]	[0.286; 0.305]	[0.170; 0.181]	[0.454; 0.484]	[0.123; 0.132]

Legend table 6: Descriptive statistics of the pre-processed data.

Descriptive stats	Combined		Wind		Swell	
	H_s (m)	T_z (s)	H_s (m)	T_z (s)	H_s (m)	T_z (s)
Mean	2.203	6.4	3.202	5.5	2.188	7.6
Median	1.976	6.2	2.961	5.4	2.027	7.5
Std. Deviation	0.986	1.0	0.986	1.0	0.986	1.0
Max	6.261	10.3	7.294	8.8	6.138	11.1
Min	0.136	4.2	1.035	3.0	0.107	5.2
Percentile 25%	1.460	5.6	2.477	4.8	1.426	6.9
Percentile 50%	1.976	6.2	2.961	5.4	2.027	7.5
Percentile 75%	2.771	7.1	3.750	6.2	2.795	8.2
Percentile 90%	3.685	7.7	4.653	6.9	3.541	9.0
Percentile 99%	4.985	8.8	5.971	8.1	5.040	10.3
Percentile 99.5%	5.303	9.0	6.286	8.4	5.577	10.5
Percentile 99.9%	5.872	9.7	6.983	8.7	5.964	10.9

Copula Models Applied to the Significant Wave Height and the Mean Up-Crossing Period

In this section several copula-based models will be fitted to the pre-processed data. The goodness-of-fit is assessed by means of the Crámer-von Mises distance (s) [35], which according to [33] corresponds to a more formal goodness-of-fit test. Other metrics used for model selection, as the well- known AIC and BIC criteria, can be seen in [36]-[39], Moreover, an interesting and broad review on this matter is provided by [40]. The present analysis is mainly focused on the Crámer-von Mises distance, as it enables a straightforward comparison of each model with the empirical copula of the pre-processed data. In this section, an attempt to fit the data is made with a set of symmetric copulas. This trial will be followed by the application of the extra-parametrization procedure with an independent copula and pairwise copulas. Going one step further and based on the best scores of the Crámer-von Mises distance, pairs of different copulas will be tested.

1. Introduction to copula based models

In this section only the fundamentals of copula theory are introduced. Comprehensive insights on copula's theory and a large spectrum of copula's families is provided by [21] and [41].

A copula is a function that couples multivariate distribution functions to their marginal distributions [11]. These functions have uniform one-dimensional margins on the interval [0; 1] and are invariant under monotone increasing transformations of the marginal distributions [21]. The main advantage of copulas is that they enable one to separate the marginal behaviour and the dependence structure of the variables from their joint distribution function [19]. Copulas present a simple way to build the joint distribution function in multivariate problems, taking into consideration the dependence structure between the considered random variables. The representation of the dependence structure is vital when dealing with reliability problems that somehow imply a random variables generation process. For instance, in maritime and offshore engineering problems one often needs to simulate pairs of wave heights and periods, which must be in agreement with the location's characteristics.

If one considers X as a vector of random variables (x_i) with the marginal distribution functions defined by $F(x_i)$, with i=1…d. The transformation $U_i=F(x_i)$ is a dependent uniformly distributed vector of random variables, with $U=(U_1,\ldots,U_d)$ on the space $[0,1]^d$. If $F(x_i)$ are continuous, the joint distribution function of X can be expressed as in Eq. 2.

$$F(x) = C(F_1(x_1), \ldots, F_d(x_d) = C(u_1, \ldots, u_d) \quad (2)$$

Where C(u) is the copula of the distribution, $C:[0,1]^d \rightarrow [0,1]$ and $u=(u_1,\ldots,u_d)$. The Eq. 2 was originally introduced as the Sklar's theorem [21]. The copula C(u) and the correspondent copula density c(u) can also be defined as in Eqs. 3 and 4, thus leading to the joint density of X given by Eq. 5:

$$C(u) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \quad (3)$$

$$c(u) = \frac{\partial^d(C(u_1,\ldots,u_d))}{\partial_{u1},\ldots,\partial_{ud}} \quad (4)$$

$$f(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{i=1}^{d} f_i\,(x_i) \quad (5)$$

Note that $f_i(x_i)$ is the marginal of x_i. Therefore, the joint distribution function of X corresponds to the combination of each marginal in X and the information on the dependence structure, which is retained by the copula function [6].

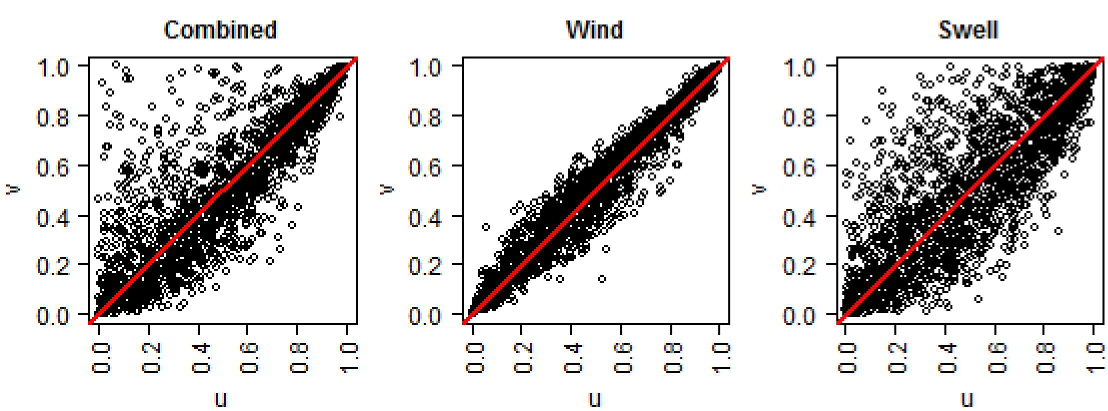
2. Empirical copula

As reference to the models one can use the empirical copula, which is built on the empirical distribution function $F_n(X_i)$ of each random variable (X_i). The empirical copula can be defined as in Eq. 6 [42]:

$$C_n(u) = F_n\left(F_{n1}^{-1}(u_1), \dots, F_{nd}^{-1}(u_d)\right) \tag{6}$$

where F_{np} and F_{np}^{-1} denote the p-th marginal empirical cdf and its generalized inverse, for p=1,…,d and u is in the interval $[0;1]^d$.

Figure 5 provides the empirical copula for the pre-processed data in the (u,v)-space, i.e. ($F^{-1}(H_s)$; $F^{-1}(T_p)$). From the left to the right one has the combined sea, the wind-sea and swell components. One can confirm that the data has an asymmetric behaviour in the (u,v)-space, i.e. generated pseudo-observations, thus $C(u,v) \neq C(v,u)$. As in other works and datasets, e.g. [6], [7] and [24], the asymmetry in the data was already expected. This can be explained by the physical limitations of the wave steepness, i.e. due to wave breaking after a certain limit, it is not possible to have very high waves with very short periods. The evident asymmetry indicates that symmetric copulas will struggle when fitting with quality the present data. One expects that a straightforward application of such copulas does not perform well under the Crámer-von Mises evaluation. A solution for this is proposed further on with the extra-parametrization technique based on the Khoudraji algorithm. As it will be demonstrated, the goodness-of-fit will be improved in the copula constructions with extra parameters. Figure 4 also indicates that the wind-sea component is the dataset with the least degree of asymmetry, while the combined sea seems to be the most asymmetric one. With the available dataset from [32] it was not possible to assess why the wind component was less asymmetric than the swell. Due to Horns Rev 3 location, it is possible that this fact might be related to shallow water depth effects. The improvements obtained in the Crámer-von Mises distance are expected to be more pronounced in the swell component and the combined sea, because symmetric models may fit the wind component better.

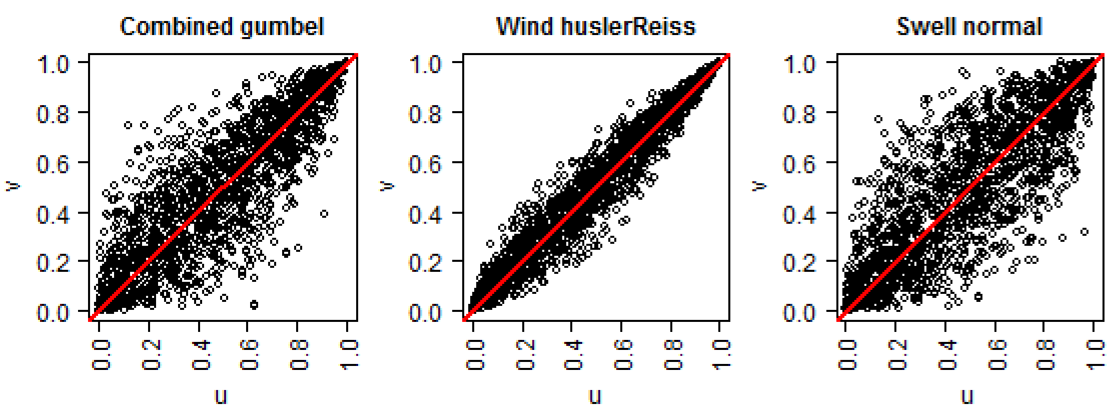


Legend figure 5: Empirical copulas of the pre-processed data in the (u,v)-space.

3. Symmetric copula-based models

To confirm the suspicion~~s~~ based on the asymmetry shown by the empirical copulas, a first attempt was performed with the following list of copulas: the Gumbel, the Frank, the Clayton, the Galambos, the Hüssler-Reiss (HR), the Joe, the Normal (Gaussian), the Tawn~~;~~ the Plackett,~~;~~ the Ali-Mikhail-Haq (AMH) and the Farlie-Gumbel-Morgenstern (FGM).

As expected the tested copulas failed at capturing the data's asymmetry. In table 7 the estimated parameter and the Crámer-von Mises distance (see [33]) of the tested copulas are presented. The mentioned distance is quite large when compared with the extra-parametrized copulas. The copulas with the lowest score in the Crámer-von Mises distances were the Gumbel copula for the combined dataset, the Hüssler-Reiss for the wind-~~-~~sea component and the Normal copula for the swell component. These values are presented in italics format in table 7. When comparing figure 6 with figure 5 (empirical copulas) it becomes obvious that the best option among the proposed copulas is not able to accurately reproduce the asymmetry shown by the empirical copulas. However, as referred before, the wind-sea component, which was the less asymmetric one seems to be the case where the proposed copulas are able to provide the lowest value of the Crámer-von Mises distance. In the combined dataset, the distances are evidently larger (table 7) when compared with the components' distances (wind-sea and swell). The asymmetry of the swell component is not as evident as in the combined dataset but it is still present (see figure 4). Still, the copulas shown in figure 5 reproduced symmetric data which, do not comply with the knowledge obtained by the empirical copulas in figure 4. This is the case where extra-parametrization could be applied to improve the goodness-of-fit.



Legend figure 6: Symmetric copulas without extra-parametrization with the lowest Crámer-von Mises distances, (u,v)-space per dataset.

An aspect worth mention in the present analysis is the fact that separating the total seainto its swell and wind-sea components does not necessarily lead to an improvement of a certain copula's performance. For example, in Joe copula one obtains s=0.5857, a lower Crámer-von Mises distance than the same copula applied to the swell component, which yields s=0.9948. On the other hand, in the wind-sea component, the Joe copula presents an s=0.1250, which reflects the fact that the copula's goodness-of fit is very much dependent on the symmetry/asymmetry present in the data.

As a result, when fitting copula models the separation of the combined sea into its components only results in a better goodness-of-fit if the actual components yield a degree of symmetry in the (u,v)-space, which symmetric copulas are able to retain.

According to Table 7, the AMH, the FGM and the Tawn copula (with one parameter, i.e. symmetric version) provided the worst (i.e. highest) values of the Crámer-von Mises distance. The parameter estimation leading to the minimization of s yields $\theta=1$, which is in the limit of the parameters' interval for the three cases ([21], [43]). This enables one to conclude that these families of copulas are not suitable for the present case study. The extra-parametrization of these copulas is not expected to present any benefit for the present analysis, as it will be confirmed further ahead. This is related with the specific nature of these copulas, for example, FGM copula is designed to hold a quadratic section in the (u,v)-space. The quadratic section, say in u, implies that $C(u,v)=a(v)^2+b(v)u+c(v)^2$. The pre-processed data do not hold a quadratic section within FGM's θ limit and therefore this copula is not suitable for these data. The Tawn copula tested in this case only has one parameter. Other versions, as the Tawn with three parameters, could lead to better results. This asymmetric version of Tawn copula was not tested in this case, but details on its application are provided in the R package VineCopula [44].

In the other copulas presented in table 7, only the wind-sea component presents Crámer-von Mises distances that are in the order of 10^{-2}. Since these copulas imply different tail dependences, this is also a reason that contributes to the disparity in the goodness-of-fit of the tested copula-based models. As an overall remark, it should be noted that the asymmetry in the data is a key factor when using copulas to model the significant wave heights and the mean up-crossing periods. Therefore, the statistical modelling of the sea components per se may not contribute to a better goodness-of-fit of given copulas. Nevertheless, it should be recognised that if the components have a noticeable symmetry, simple copulas could indeed be used to model the data. This can be perceived by the low values of s obtained for some copulas applied to the wind-sea component, e.g. the Galambos, the Gumbel or the HR copulas.

Some of the copulas tested without extra-parametrization seem to compare well, in terms of the order of magnitude obtained for the Crámer-von Mises distances, with the values obtained with extra-parametrization for the combined sea analysed by [6]. This could be due to the characteristics of the present dataset and also due to the accuracy of the estimation of the parameters and the Crámer-von Mises distances. Although not specified in [6], in this is study, the estimation of the copulas' parameters and the values of s was performed with an iterative procedure, until the Crámer-von Mises distance is stabilized at the fourth decimal place. Table 7 also shows that for well-known simple copulas, as the Frank, Clayton, Gumbel and Gaussian ones, the Crámer-von Mises distance is not worse than the ones presented, for example, by the Plackett or the Joe copulas which are not so common in the literature. However, these distances can be improved with the extra-parametrization technique. Thus several spectral parameters modelling and reliability problems studied under the use of simple Archimedean or elliptical copulas, e.g. [11], [23], [45] or [46], can improve their accuracy with the use of the same copula with extra-parameters. In the next section, the extra-parametrization of copulas is introduced and compared with the results obtained in the symmetric models (table 7).

Legend table 7: Estimated symmetric copulas, without extra-parametrization and their Crámer-von Mises distances (lowest in italics).

Copula	Combined		Wind		Swell	
	θ	*s*	θ	*s*	θ	*s*
AMH	1	19.733	1	32.7288	1	14.4045
Clayton	5.105	2.2590	17.655	0.2697	3.517	1.7917
FGM	1	30.9645	1	48.5962	1	24.7091
Frank	9.779	0.6682	24.909	0.0944	7.910	0.3478
Galambos	2.451	0.3913	6.438	0.0171	1.965	0.2364
Gumbel	3.161	*0.3840*	7.151	0.0172	2.675	0.2347
HR	3.149	0.4159	7.661	*0.0169*	2.592	0.2451
Joe	5.252	0.5857	14.682	0.1250	4.267	0.9948
Normal	0.879	0.6536	0.976	0.0376	0.829	*0.2258*
Plackett	36.288	0.5463	259.693	0.0615	23.181	0.2845
Tawn	1	9.8314	1	20.8852	1	6.4406

4. Copula-based models with extra-parametrization

In order to build asymmetric copula-based models the extra-parametrization technique [26] can be applied to combine different copulas. This algorithm results in a new copula, with extra-parameters α and β, which retain the information regarding the asymmetry present in the (u,v)-space. If one considers the symmetric copulas $C_1(u,v)$ and $C_2(u,v)$, the new asymmetric copula C(u,v) is obtained as in Eq. 7 [6]:

$$C(u,v) = C_1\left(u^{\alpha}, v^{\beta}\right) \cdot C_2\left(u^{1-\alpha}, v^{1-\beta}\right) \tag{7}$$

Note that if C_1 and C_2 have the parameters θ_1 and θ_2, the new copula C will have four parameters, the ones that come from each parametric copula plus the pair (α;β). The parameters α and β may vary between 0 and 1. If α is different from β, C(u,v) corresponds to an asymmetric copula [28]. Note that the resulting new copula can be combined with another one, say A(u,v), which leads to a new set of parameters. Thus the number of copula combinations is almost infinite [6]. However, there is no guaranty that the goodness-of-fit is improved with the number of copulas combined. Hence, using the Crámer-von Mises distance, the AIC and BIC or other goodness-of-fit criteria should always be performed to understand if the complexity of the combinations is actually leading to better copula-based models.

In the present study, the estimation of α, β and θ is performed as in [6], which seeks for the minimum value of the Crámer-von Mises distance (s), which is defined as the sum of the distances between the empirical copula and the parametric copula distribution function, over a grid of [0;1]×[0;1] as in Eq. 8:

$$s = \sum_{i=1}^{N}\left[C^{empirical}(u_i^*, v_i^*) - C_{\alpha;\beta;\theta}(u_i^*, v_i^*)\right]^2 \tag{8}$$

In this case, and for the sake of comparison with Vanem (2016) the calculation of s is performed for the referred grid with 100×100=10 000 points, where the empirical and the tested copulas are evaluated.

5. Extra-parametrization with an Independent copula

Assume that C_1 is one of the copulas proposed in symmetric set of copulas introduced beforeand that C_2 is an independent copula, referred to as $I(u^{1-\alpha},v^{1-\beta})$. The independent copula yields $I(u^{1-\alpha},v^{1-\beta}) = u^{1-\alpha}v^{1-\beta}$, which leads from Eq. 7 to Eq. 9:

$$C(u,v) = C_1\left(u^{\alpha}, v^{\beta}\right) \cdot I\left(u^{1-\alpha}, v^{1-\beta}\right) = C_1\left(u^{\alpha}, v^{\beta}\right) \cdot u^{1-\alpha} \cdot v^{1-\beta} \tag{9}$$

This procedure corresponds to the definition of a new asymmetric copula, which is extra-parametrized with an independent copula. For the tested copula-based models, one obtains a set of 3 parameters and the associated Crámer-von Mises distance. These values are summarized in Table 8. One is able to see that the AMH, the FGM and the Tawn copulas, were not able to catch the asymmetry of the data. Note that these copulas yield $\alpha=\beta=1$. This was already expected, since in the application without extra-parametrization, the copula's parameter θ had reach its limit and still the values of s were very high when compared with the other models.

Without reaching its θ limit, the Plackett model also led to $\alpha=\beta=1$ for the wind-sea component. Hence, this model did not catch the asymmetry of this dataset, which was the one closest to symmetry. One is also able to note that the estimation of θ and the Crámer-von Mises distance obtained in table 8 is the same as in table 7 (s= 0.0615). However, note that the symmetric Plackett copula applied to the wind-sea component still provided a lower distance than the asymmetric models obtained with the Frank, the Joe and the Clayton copulas. This interesting result is justified by the fact that the present data do not have a much remarked asymmetry for this sea component, as it is presented for the combined sea and the swell component. This emphasizes the fact that copula's performance is very much dependent on the asymmetry of the data. Also, the tested models account for different tail dependences which may also contribute for the disparity of the Crámer-von Mises distance between models.

From table 8 one can also conclude that the pre-processed data for the combined sea was best fitted by extra-parametrizing with an independent copula with the Gumbel copula, which was closely followed by extra-parametrizing with the Galambos and the HR copulas. The swell component was best fitted by extra-parametrizing an independent copula with the Normal copula, which was followed by extra-parametrizing with the Gumbel and the HR ones. The HR copula provided the lowest Crámer-von Mises distance for the wind-sea component, which also presented very similar values of s, when modelled with the Galambos and the Gumbel copulas. The lowest Crámer-von Mises distances appear in italics in table 8.

In general, the Gumbel and the HR copulas, with extra-parametrization with an independent copula, were the models that tended to provide the lowest Crámer-von Mises distances. If one compares the distances obtained between the models with the same copula, with and without extra-parametrization with an independent copula, i.e. (table 7 and table 8), it is possible to conclude that the construction proposed by Eq. 9 leads to reductions in the s values, thus, approximating the proposed models to the empirical copulas of the pre-processed data. The exception occurs for the AMH, the FGM and the Tawn copulas, which remain the same as the symmetric versions explained before. For the other models, the improvements obtained with the extra-parametrization are more noticeable for the combined and the swell pre-processed datasets. This occurs because the asymmetry in these datasets contributed to worse estimations with the symmetric set of copulas. This emphasizes the notion that the added complexity of the extra-parametrization procedure is more valuable if the asymmetry in the data is more evident. The improvement of the Crámer-von Mises distances agrees with the results reported in [6] and [28].

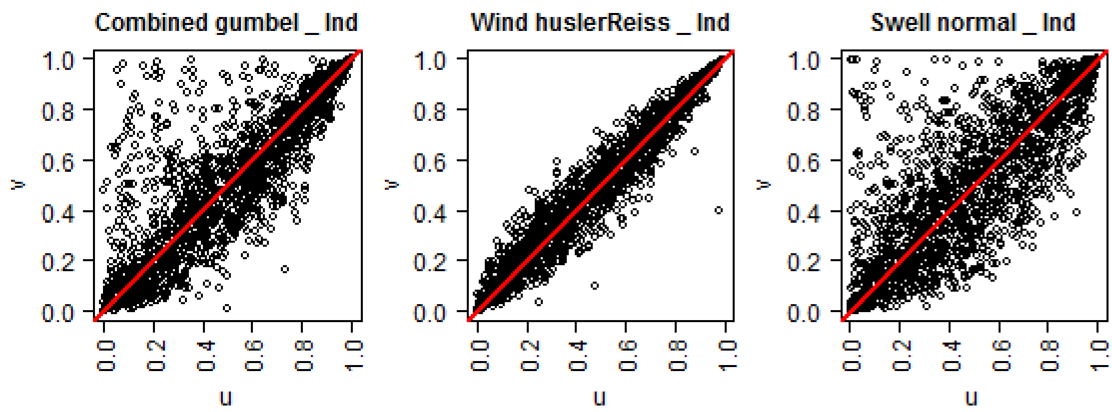
Figure 7 provides the extra-parametrized copulas that led to the lowest Crámer-von Mises distances. One can visually confirm that these copulas provide asymmetric results, which are closer to the empirical copulas shown in figure 5 than the results provided by the set of symmetric copulas from figure 6.

One may also want to look at the original (H_s; T_z)-space. A series of 1887 points generated in this original space are presented in figure 8. These values result from the inverse of the marginal distributions proposed in section 3 (table 5) applied to the series of generated values in (u,v)-space presented in figure 7. In order to obtain the values in the original (H_s; T_z)-space the seasonality is added back to the pre-processed data, by inverting Eq. 1. Once the seasonality is added, the autocorrelation function becomes similar to the subsampled data. The autocorrelation function for H_s in the generated series of 1887 pairs of (H_s [m]; T_z [s]) is shown in figure 9. The analysis also showed that the autocorrelation of the mean up-crossing period agreed with the subsampled data.

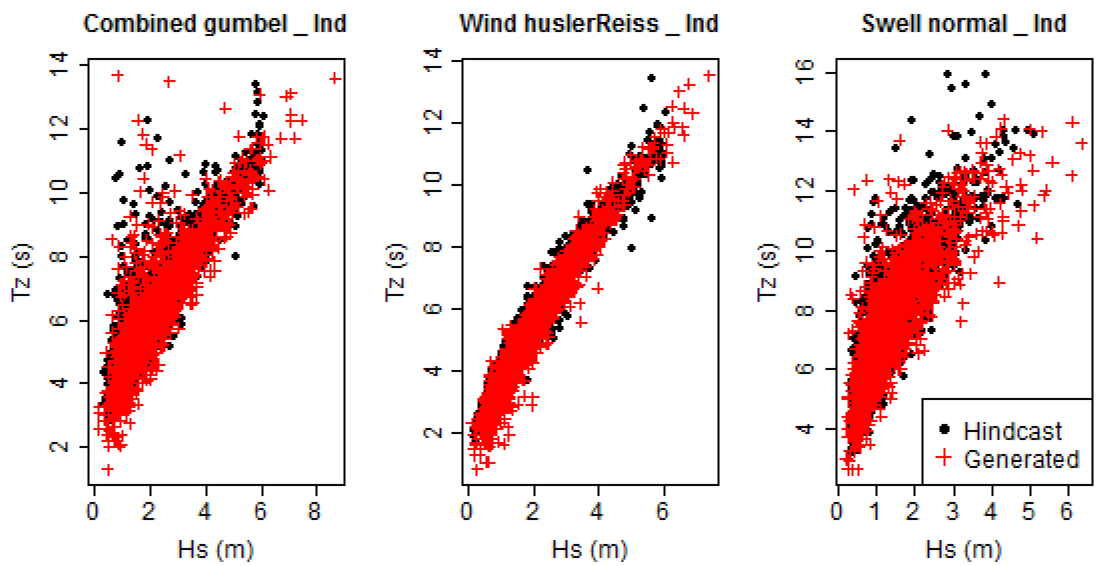
In figure 8 one is able to note that the generated series include values that can exceed the maximum of the hindcast significant wave heights. Sometimes the exceedances may be up to 2 m, which is quite striking. Despite the reasonable agreement between the generated pairs of (H_s; T_z), the models tended to provide a worse fit when dealing with the upper tail of the distributions. At a first look, it seems that the marginal distributions previously defined are too heavy tailed for the present data. It should be noted that this might be a problem of the marginal modelling and not necessarily due to the employed copula models. This also occurred with the extra-parametrization with pairwise and the Gumbel copulas. This emphasizes the importance of a good marginal definition, regardless of the copula model employed for the joint distribution of H_s and T_z. There are also large uncertainties in the fitting of the marginal models and this would particularly affect the tails of the distributions. The authors recognize that this aspect should be investigated in further research. However, the same marginal distributions applied to the pre-processed data are used when converting the data to the original (H_s;T_z)-space. Therefore, relative comparisons between models are still valid. The overestimation of significant wave heights may lead to a conservative safety and reliability assessment of offshore wind or marine structures, because the large significant wave heights are the most energetic phenomena and the most likely to lead to the failure of a structure. However, when dealing with the probabilities of failure of a certain maritime system, one should note that such models may provide an overestimation of these probabilities, which does not contribute to the optimization of the structures' design.

Legend table 8: Estimated asymmetric copulas, with extra-parametrization with an independent copula and their Crámer-von Mises distances (lowest in italics).

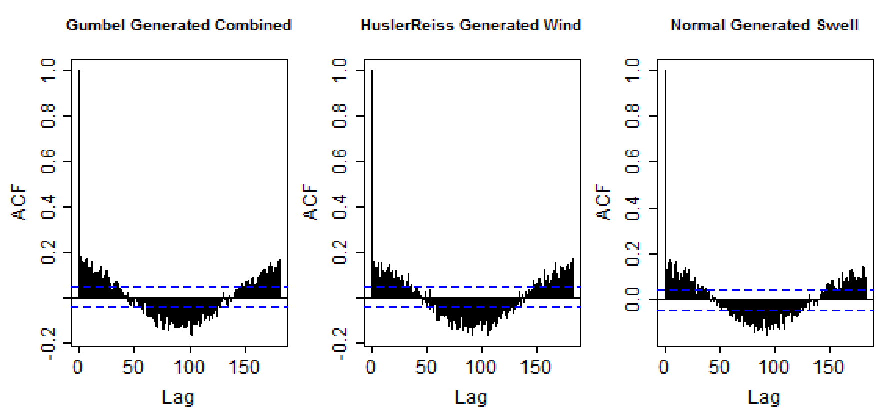
Extra-parametrization with an Independent copula						
	Combined		**Wind**		**Swell**	
Copula	**α; β; θ**	**s**	**α; β; θ**	**s**	**α; β; θ**	**s**
AMH	1; 1; 1	19.7326	1; 1; 1	32.7288	1; 1; 1	14.4045
Clayton	0.936; 0.812; 26.776	0.4084	0.971; 0.978; 28.699	0.1707	0.912; 0.804; 10.759	0.5466
FGM	1; 1; 1	30.9645	1; 1; 1	48.5963	1; 1; 1	24.7091
Frank	0.977; 0.846; 18.203	0.2207	0.986; 0.992; 28.867	0.0836	0.985; 0.869; 10.838	0.1923
Galambos	1; 0.868; 3.729	0.0667	0.997; 1; 6.572	0.0165	0.965; 0.845; 3.048	0.1649
Gumbel	1; 0.868; 4.445	*0.0665*	0.997; 1; 7.293	0.0166	1; 0.911; 3.067	0.1171
HR	0.992; 0.864; 4.785	0.0696	0.995; 1; 7.887	*0.0159*	0.985; 0.865; 3.462	0.12
Joe	0.985; 0.856; 8.147	0.0869	0.982; 0.989; 17.809	0.1022	0.943; 0.827; 6.983	0.4705
Normal	0.983; 0.849; 0.956	0.1674	0.992; 0.997; 0.979	0.0348	1; 0.888; 0.882	*0.0728*
Plackett	1; 0.868; 93.048	0.2055	1; 1; 261.051	0.0615	1; 0.905; 35.11	0.1799
Tawn	1; 1; 1	9.8314	1; 1; 1	20.8852	1; 1; 1	6.4406



Legend figure 7: Asymmetric copulas, with extra-parametrization with an independent copula, with the lowest Crámer-von Mises distance, (u,v)-space per dataset.



Legend figure 8: Generated series (+) of 1887 pairs of (H_s;T_z) over the subsampled data (o) based on the best copulas extra-parametrized with and independent copula.



Legend figure 9: Autocorrelation function for H_s in the generated series of 1887 pairs of (H_s;T_z) based on the best copulas extra-parametrized with an independent copula.

6. Extra-parametrization with a Pairwise copula

If one recalls Eq. 7 and assumes that $C_1(u,v)$ is coupled with a copula $C_2(u,v)$ of the same family, the Khoudraji algorithm can be applied to perform an extra-parametrization with pairwise copulas, e.g. Gumbel-Gumbel, Clayton-Clayton or Plackett-Plackett. In this case, the procedure is similar, but one needs to estimate the parameters of the first and the second copulas, respectively θ_1 and θ_2. This leads to Eq. 10.

$$C(u,v) = C_{1,\theta 1}\left(u^{\alpha}, v^{\beta}\right) \cdot C_{1,\theta 2}\left(u^{1-\alpha}, v^{1-\beta}\right) \tag{10}$$

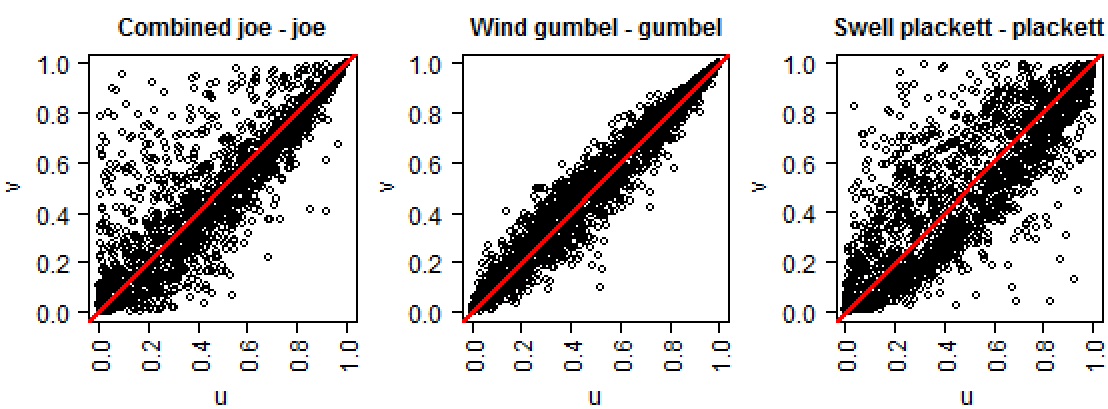
In this section, only a set of the copulas with the lowest Crámer-von Mises distances obtained in the previous section will be tested. Therefore, the models based on the Tawn copula, the FGM and the AMH copulas are excluded from the analysis. The estimated parameters for the pairwise copulas and the Crámer-von Mises distances are provided in table 9. Comparing table 9 with tables 7 and 8, one concludes that the extra-parametrization with a pairwise copula leads to improvements in the Crámer-von Mises distances. The improvements are more noticeable in some copulas than in others. Generally, the models that had already presented low values of s are the ones where the improvement is less noted, e.g. the Gumbel-Gumbel or the HR-HR copulas in the wind component. However, the improvements are quite remarkable in some other models, for example in the swell component, the Plackett model extra-parametrized with an independent copula (table 8) has s=0.1799, while the Plackett-Plackett model is improved to s=0.0238.

Table 9 shows that the pairwise models with lowest Crámer-von Mises distances were the Joe-Joe copula for the combined sea, the Gumbel-Gumbel copula for the wind-sea component and the Plackett-Plackett copula for the swell component. One should also note that for the three datasets, the Galambos-Galambos and the HR-HR copulas still present low Crámer-von Mises distances. Although their distances are not the lowest ones, the differences for the best models are not so large. Due to the existence of several models with similar Crámer-von Mises distances, it is important to understand if these models are actually very different when dealing with generated series. Note that the Crámer-von Mises distance provides a measure on the overall goodness-of-fit. Therefore, two copulas with the same Crámer-von Mises distance, may fit the extreme events, i.e. upper tail, with different localized goodness-of-fit. Hence the tail behaviour of the models should be analysed depending on the final objective or intended use of the proposed model. For instance. if one wants to predict the extreme significant wave height and the mean up-crossing period for a specific return period, one may chose the model that fits the upper tail the best, although not being the best model in terms of the Crámer-von Mises distance. On the other hand, if one wishes to deal with probabilities of failure of a marine structure, it is important to reach a balance between the overall goodness-of-fit and the upper-tail fit, mainly, because the probability of failure depends on the proportion between the extreme events that can cause failure and the common events that are not expected to lead to the collapse of the system.

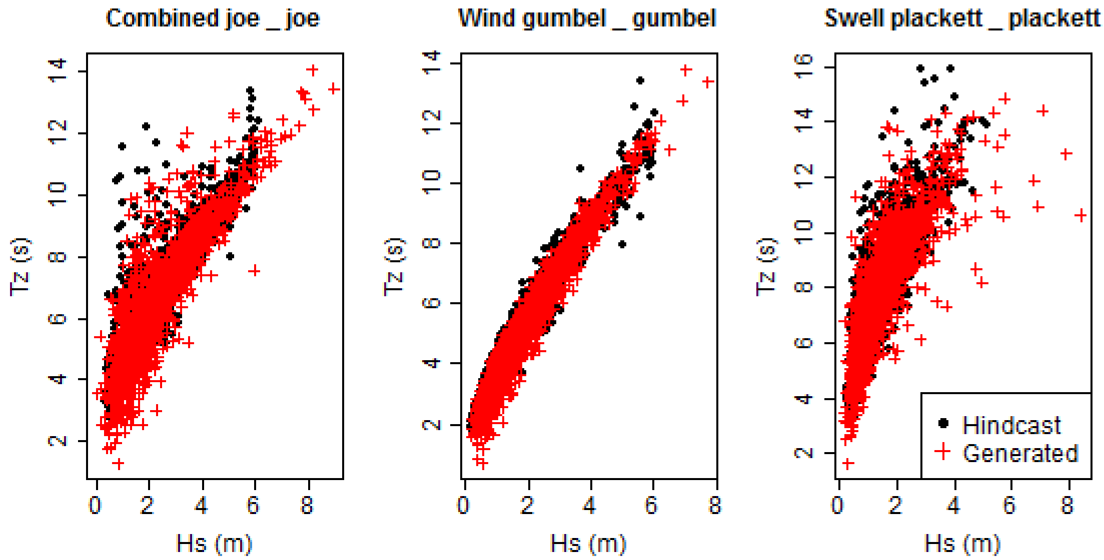
Before diving into the tail dependence analysis, figure 10 provides the data for the pairwise copulas that provided the lowest Crámer-von Mises distances. It is possible to see that the asymmetry present in the data is also captured with the pairwise copulas. The generated series of 1887 points in the original (H_s; T_z)-space is provided in figure 11, which compares with figure 8. One is able to see that the extra-parametrization with a pairwise copula enables one to generate the values which are within the range of the subsampled data. Once again the effect of subsampling the maximum values in each block of 2 days from the original hindcast data should be the aim of further research, as well as the effect of the selected marginal distributions. In figure 12 it is presented the autocorrelation function of the significant wave heights for the generated series with the pairwise copulas. The autocorrelation compares well with the one presented by the subsampled data. Also in this case, the analysis showed that the autocorrelation of the mean up-crossing period also agreed with the subsampled data. The analysis also showed that the autocorrelation is also preserved for the correspondent generation of T_z.

Legend table 9: Estimated asymmetric copulas, with pairwise extra-parametrization, and their Crámer-von Mises distances (lowest s in italics).

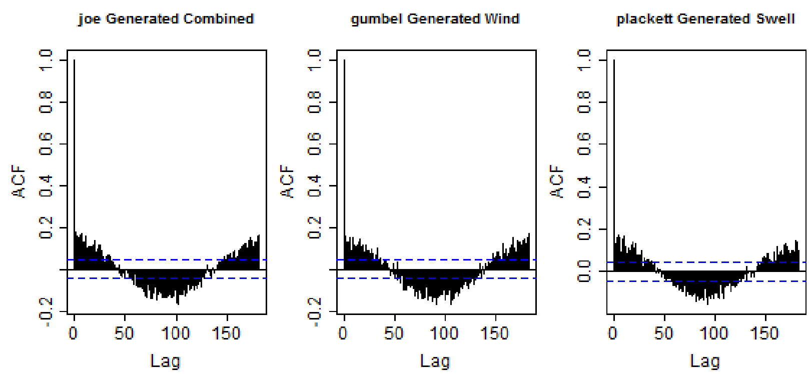
Copula	Combined		Wind		Swell	
	α; β; θ1; θ2	**s**	**α; β; θ1; θ2**	**s**	**α; β; θ1; θ2**	**s**
Clayton-Clayton	0.124; 0.373; 31.824; 31.824	0.2937	0.491; 0.605; 82.324; 100	0.0380	0.252; 0.558; 24.872; 14.832	0.1586
Frank-Frank	0.047; 0.248; 20.992; 64.998	0.1984	0.481; 0.370; 84.998; 99.978	0.0344	0.147; 0.422; 17.492; 28.602	0.0984
Galambos - Galambos	0.029; 0.181; 4.077; 0.934	0.0622	0.581; 0.515; 9.592; 6.579	0.0152	0.243; 0.455; 3.568; 1.564	0.0901
Gumbel-Gumbel	0.041; 0.195; 4.926; 1.651	0.0624	0.382; 0.447; 7.317; 10.770	*0.0151*	0.241; 0.475; 4.467; 2.452	0.0926
HR-HR	0.119; 0.275; 6.065; 1.503	0.0653	0.435; 0.476; 6.849; 10.710	0.0152	0.181; 0.377; 4.030; 1.923	0.0883
Joe-Joe	0.105; 0.260; 9.056; 1.806	*0.0529*	0.472; 0.369; 18.906; 15.475	0.0233	0.212; 0.502; 7.795; 3.944	0.2059
Normal-Normal	0.922; 0.745; 0.852; 0.951	0.1583	0.115; 0.173; 0.984; 1	0.0280	0.115; 0.346; 0.934; 0.934	0.0492
Plackett-Plackett	0.064; 0.255; 162.688; 209.984	0.1634	0.463; 0.563; 998.976; 2474.976	0.0238	0.289; 0.543; 185.984; 46.413	*0.0455*



Legend figure 10: Asymmetric copulas, with extra-parametrization with a pairwise copula, with the lowest Crámer-von Mises distance, (u,v)-space per dataset



Legend figure 11: Generated series (+) of 1887 pairs of (H_s;T_z) over the original hindcast data (o) based on the best copulas extra-parametrized with a pairwise copula.



Legend figure 12: Autocorrelation function for H_s in the generated series of 1887 pairs of (H_s;T_z) based on the best copulas extra-parametrized with a pairwise copula.

Copula Based Models with Different Copulas

7. Extra-parametrization with a Gumbel copula

From the extra-parametrized models one is able to understand that the introduction of new parameters leads to more flexible models. However, the non-parametric estimations of the tail dependence and the WRMSE show that the tails' behaviour is not automatically improved. This aspect is discussed in the next section. Moreover, a model that performs well in the asymmetric combined dataset, may perform worse under the asymmetry showed by the wind-sea and the swell components. In this sense, one may want to try an extra-parametrization performed with different copulas. As the number of possible combinations is almost infinite, as explained before, in this section, an extra-parametrization with a Gumbel copula will be tested for the symmetric models initially presented (table 7). Note that other combinations could be tested but this was not performed for the present case.

Further research should extend the combinations hereby presented for a deeper assessment on the quality of possible fittings.

The parameters obtained and the optimised Crámer-von Mises distance for each model are provided in table 10. The Gumbel copula as $C_1(u,v)$ is not tested because it corresponds to the pairwise model obtained from the Gumbel-Gumbel model.

Table 10 shows that extra-parametrization enabled one to build copula-based models based on the AHM, the FGM and the Tawn copulas. One is also able to note that the Crámer-von Mises distances did not improve for all models.

Comparing table 10 with table 9, one can see that, for the combined dataset the Joe-Joe model provides a slightly lower Crámer-von Mises distance than the Joe-Gumbel model. All the Crámer-von Mises distances obtained in table 10 are improved when compared with the symmetric models from table 7. These improvements also occur in the wind-sea and swell components.

The wind-sea component, which was the most symmetric dataset, provided low values of the Crámer-von Mises distance for models which only had an upper-tail dependence in the symmetric version. Note that the Hüssler-Reiss and the Galambos copulas extra-parametrized with a Gumbel copula give similar distances when compared with their extra-parametrization with a pairwise copula. In the swell component the Plackett copula (extra-parametrized with a Gumbel copula) remained as the model with the lowest Crámer-von Mises distance. In this case, the distance is slightly above the one obtained with the pairwise extra-parametrized Plackett copula.

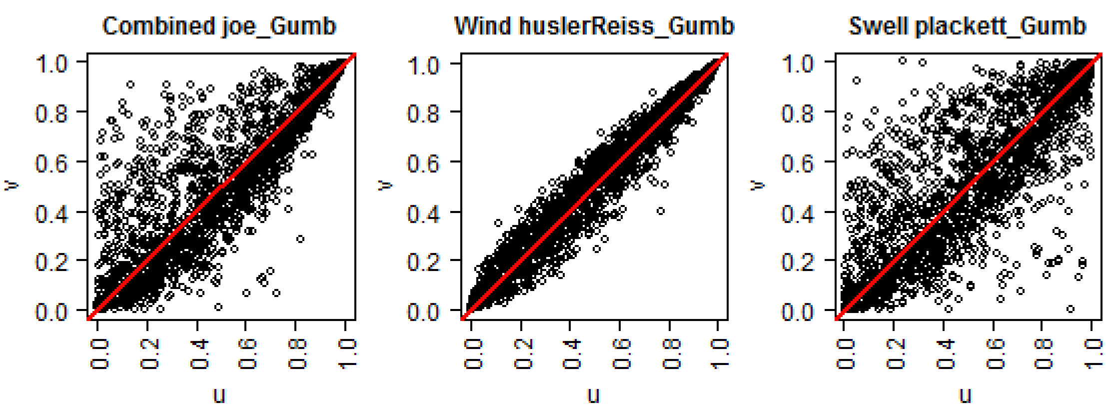
For some copulas, e.g. the Frank or the Clayton copulas in the combined dataset, the extra-parametrization with a Gumbel copula gave lower Crámer-von Mises distances when compared with the extra-parametrization made with an independent copula. However, in other models this did not occurr, e.g. the Galambos-Gumbel copula did not provide a lower distance than the Galambos-Independent copula. From the tested extra-parametrizations, the pairwise copulas were the ones that provided the best goodness-of-fit in terms of the Crámer-von Mises distances. Nevertheless, it is concluded that testing the Khoudraji algorithm with different copulas may lead to some improvements. The Crámer-von Mises distances obtained in the present study seem to compare well with the ones presented for a different dataset by [6].

The results from table 10 also show that the separation between wind-sea and primary swell components does not necessarily lead to a better goodness-of-fit when compared with the combined dataset. For instance, the Hüssler-Reiss copula extra-parametrized with a Gumbel copula presents a lower Crámer-von Mises for the combined sea than for the wind-sea and swell components.

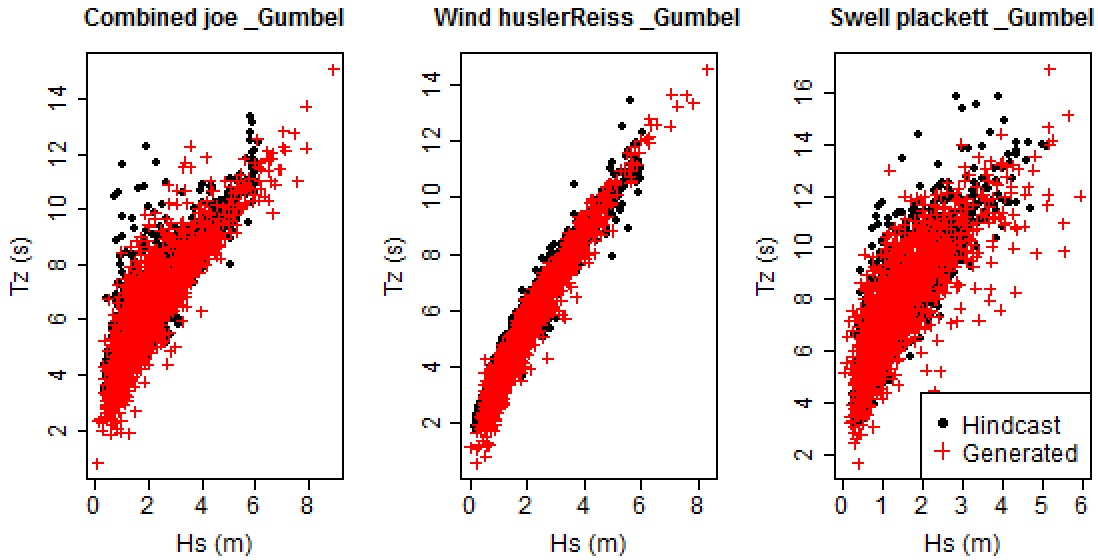
In figure 13, simulated data from the copulas extra-parametrized with a Gumbel copula also show an asymmetry, which is in agreement with the one obtained for the pre-processed data. In figure 14, the generated values show a visually good agreement with the subsampled data, as it occurred for the copulas extra-parametrized with independent or pairwise copula.

Legend table 10: Estimated asymmetric copulas, with extra-parametrization with a Gumbel copula and their Crámer-von Mises distances.

Extra-parametrization with a Gumbel copula						
	Combined		**Wind**		**Swell**	
	α; β; θ_1; θ_2	**s**	**α; β; θ_1; θ_2**	**s**	**α; β; θ_1; θ_2**	**s**
AMH	0.046; 0.156; 6.049; -0.933	0.1387	0.018; 0.013; 8.694;-0.704	0.0350	0.117; 0.226; 5.112; 1.000	0.3141
Clayton	0.115; 0.343; 6.483; 27.833	0.1381	0.456 ;0.351; 12.843; 79.998	0.0242	0.264; 0.548; 5.365; 11.573	0.0731
FGM	0.046; 0.143; 6.055; -0.90	0.1386	0.018; 0.011; 8.654; -0.704	0.0356	0.112; 0.217; 4.985; 1.000	0.2976
Frank	0.887; 0.691; 2.010; 25.004	0.1660	0.433; 0.535; 10.016; 109.648	0.0243	0.804; 0.532; 3.030; 17.960	0.0883
Galambos	0.902; 0.696; 2.464; 5.306	0.0859	0.547; 0.465; 10.937; 7.445	0.0154	0.199; 0.480; 4.926; 2.469	0.1143
HR	0.861; 0.657; 2.315; 6.803	0.0849	0.448; 0.529; 8.002; 12.145	*0.0153*	0.792; 0.509; 3.128; 5.268	0.1129
Joe	0.155; 0.360; 6.704; 2.578	*0.0799*	0.473; 0.564; 8.771; 16.468	0.0175	0.234; 0.520; 5.203; 3.841	0.1563
Normal	0.967; 0.770; 50.333; 0.967	0.1561	0.679; 0.609 ;33.889; 0.984	0.0263	0.712; 0.835; 33.889; 0.835	0.0744
Plackett	0.898; 0.693; 2.260; 333.342	0.1619	0.517; 0.610; 10.459;995.730	0.0222	0.900; 0.685; 3.576; 60.302	*0.0698*
Tawn	0.189; 0.327; 7.310; 1.000	0.0814	0.042; 0.035; 8.840; 1.000	0.0280	0.334; 0.452; 6.206; 1.000	0.1558



Legend figure 13: Asymmetric copulas, with extra-parametrization with a Gumbel copula, with the lowest Crámer-von Mises distance, (u,v)-space per dataset.



Legend figure 14: Generated series (+) of 1887 pairs of (H_s;T_z) over the original hindcast data (o) based on the best copulas extra-parametrized with a Gumbel copula.

Tail Dependence Analysis

As in other statistical models, the behaviour of the tails is important to understand how well the proposed model is able to account for extreme events.

A possible way to measure the tail dependence (lower or upper tails) is to compute the tail dependence coefficients. Further details on these dependence measures can be seen in [21] and [47]. Regarding other tail dependence functions, [48] also provides an extensive review with an application case study. Assume that X and Y are continuous random variables with marginal distributions F(X) and G(Y). The upper (or lower) tail dependence is the limit (if it exists) of the conditional probability that Y is greater (or lower) than the u-th percentile of G given that X is greater (or lower) than the same u-th percentile of F, as u approaches 1 (or 0 for the lower tail). The upper (λ_U) and lower tail dependence (λ_L) coefficients can be defined, respectively, by Eq. 11 and Eq. 12:

$$\lambda_U = \lim_{u \to 1^-} \left[P\left(Y > G^{-1}(u) | X > F^{-1}(u)\right)\right] \tag{11}$$

$$\lambda_L = \lim_{u \to 0^+} \left[P\left(Y \leq G^{-1}(u) | X \leq F^{-1}(u)\right)\right] \tag{12}$$

These limits can be expressed in terms of the copula function by Eq. 13 and Eq. 14, as follows [21]:

$$\lambda_U = \lim_{u \to 1^-} \left[\frac{1-C(u,)}{1-u}\right] \tag{13}$$

$$\lambda_U = \lim_{u \to 0^+} \left[\frac{C(u,)}{u}\right] \tag{14}$$

X and Y are considered to be upper (or lower) asymptotically dependent if λ_U (or λ_L) belongs to the interval]0;1]. If λ_U (or λ_L)=0 the variables are considered to be upper (or lower) asymptotically independent [21]. Note that if the tails are asymptotically independent that does not mean that the variables are actually independent. It solely means that as one moves to the upper or lower tails of the data, the probability that Y exceeds a certain quantile is independent of the probability that X exceeds the same quantile. For instance, X may represent the significant wave height (H_s) in the pre-processed data and Y may represent the mean up-crossing period (T_z), also in the pre-processed data.

In the present study the non-parametric estimation originally introduced by [49] is implemented. Some of the symmetric copulas do not have tail dependence, e.g. the Clayton copula does not have upper tail dependence, the Gumbel copula does not have lower one and the Frank copula has both upper and lower zero tail dependency. As it turns out, the non-parametric estimation method proposed in [49] converges very slowly to these dependences. Therefore, in limited samples, the obtained values of the tail coefficients may not correspond to the most realistic evaluation. For the purpose of illustration, assume a Clayton copula with the copula parameter θ=2. One expects that, theoretically, the λ_U=0, while $\lambda_L=2^{-1/\theta}$=0.7071 [50]. Table 11 provides the Schmid & Schmidt estimations for both coefficients, in the percentile u=0.01, for different sample sizes (n).

Legend table 11: Example of [49] non-parametric estimation of the upper and lower~~-~~-tail dependence coefficients for a Clayton copula with a ~~copula~~ parameter θ=2.

n	**1000**	**10 000**	**50 000**	**100000**	**500000**	**1E6**
$\boldsymbol{\lambda_U}$	0.0198	0.0168	0.0112	0.0212	0.0180	0.0141
$\boldsymbol{\lambda_L}$	0.6611	0.7690	0.8468	0.8397	0.8325	0.8264

Table 11 shows that the lower tail dependence is clearly higher than the upper tail dependency. However, the values obtained for several sample sizes are not that close to the theoretical values. Still, this non-parametric estimation of finite tail dependence is widely referred to and used in the copula's literature, e.g. [50] or [51]. Moreover, it is a straightforward method that can easily be implemented for the more complex copulas built with the Khoudraji algorithm, as the ones implemented in this paper. Since the same method is used for all the copula-based models presented in this work, the use of [49] estimator introduces a systematic bias, which is not relevant for the relative comparison of the tested copulas. However, one should keep in mind that the estimator is useful for relative comparison but it might not lead to the most accurate assessment of the tail dependence coefficients, as was illustrated by table 11.

In order to have a better perception of the differences in the tail dependences, the lower and upper tail coefficients were computed, respectively at the 5% and the 95% quantiles, i.e. λ_L(u=5%) and λ_U(u=95%). Note that these values correspond to the finite dependence estimation provided by [49]. Therefore, they do not represent the actual asymptotic dependence of the pre-processed data, although they still provide a notion on the possible dependence or independence. These values are summarized in table 12 and they were computed for 10 random samples of 10 000 pairs (u,v), which lead to the mean values presented.

One is able to note that the models that provided the best Crámer-von Mises distance may not be the ones that provided the closest non-parametric estimation of λ_L(u=5%) and

λ_U(u=95%). For example, in the extra-parametrization with a pairwise copula, the combined sea fitted with a Joe-Joe copula had provided the best Crámer-von Mises distance (see table 9). However, the values of λ_L(u=5%) and λ_U(u=95%) in this case are better approximated by the Frank and the Gumbel copula respectively. Despite this occurrences, it is fair to note that the best models in terms of the Crámer-von Mises distance were not the worse models in terms of the non-parametric estimation of λ_L(u=5%) and λ_U(u=95%). This could somehow be expected because, as mentioned before, the Crámer-von Mises distance provides a measure of the overall goodness-of-fit thus not ensuring the best fit on the tails, but still including them in the estimation process of α, β and θ.

An interesting aspect from table 12 is that the pairwise models Gumbel-Gumbel, the HR-HR, Joe-Joe and Galambos-Galambos, assume a perfect asymptotic dependence at the 95% quantile in the wind-sea component. The same occurs for the Gumbel model extra-parametrized with an Independent copula. Although for the pre-processed data this does not occur, it seems reasonable to admit that the data has a strong upper-tail dependence for the wind-sea component. This occurs more frequently in the extra-parametrization with the Gumbel copula, which is expected since the Gumbel copula introduces a strong upper-tail dependence.

Regarding the swell data, the pairwise Plackett-Placket and the Gumbel-Gumbel copulas provided the best estimation for the upper and lower–tail, respectevely. The Plackett extra-parametrized with an Independent copula yield good estimations for both the 5% and the 95% percentiles. In the combined sea, the Gumbel-Independent and the Frank-Frank models yield the best estimations for the lower tail dependence, while the Normal-Independent and the Gumbel-Gumbel copulas provide the best non-parametric estimation of the upper-tail dependence at the 95% percentile. Regarding the copulas extra-parametrized with a Gumbel copula close estimations are obtained for some models. The estimations for the 5% quantile in the combined, wind and swell datasets are obtained by the Clayton-Gumbel, the Plackett-Gumbel and again the Clayton-Gumbel copulas, respectively. For the 95% quantile, the Galambos-Gumbel, Normal-Gumbel and Plackett-Gumbel, respectively yield the best estimations for the combined, wind-sea and swell datasets.

The flexibility introduced with the extra-parametrization enables one to obtain the optimized model parameters. Nevertheless, the results from table 12 emphasize the need for specific models when dealing with the tails' region. It is also interesting to note that, in general, the lower deviations for λ_L(u=5%) and λ_U(u=95%) seem to occur in the models extra-parametrized with a Gumbel copula. From the analysis of table 12 it is not clear which model might be the best to fit the asymmetric data. Nevertheless, one can conclude that the choice of the extra-parametrization with a Gumbel copula is a valid attempt to deal with the present data. However, it is important to note that often the reliability and risk analysis imply the simulation of the overall population of random variables. In this sense, the present technique seems to achieve an interesting balance for the overall goodness-of-fit. Therefore, a corollary from this observation is that when using copula models applied to the significant wave height and a characteristic wave period, the extra-parametrization is a reasonable way to try to improve the results obtained from symmetric copulas, which are still the main copulas applied in the ocean modelling literature, e.g. [7], [45] or [46].

Legend table 12: Non-parametric estimation of the tail dependence at the 5% and the 95% quantiles for extra-parametrized copulas (closest values to empirical estimation in italics- bold).

Empirical copula of the pre-processed data						
	Combined		Wind		Swell	
	λ(u=5%)	λ(u=95%)	λ(u=5%)	λ(u=95%)	λ(u=5%)	λ(u=95%)
Pre-processed	0.450	0.775	0.887	0.890	0.393	0.509
Extra-parametrization with an independent copula						
Copula	Combined		Wind		Swell	
	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)
Clayton	0.740	0.420	0.886	0.568	0.553	0.226
Frank	0.349	0.373	0.620	0.484	0.289	0.218
Galambos	0.525	0.954	0.800	0.989	0.375	0.854
Gumbel	***0.451***	0.961	***0.880***	1.000	0.387	0.879
HR	0.561	0.744	0.680	0.947	0.430	0.888
Joe	0.198	0.846	0.432	0.959	0.206	0.830
Normal	0.647	***0.758***	0.939	0.946	0.561	0.578
Plackett	0.565	0.658	0.856	***0.890***	***0.390***	***0.463***
Extra-parametrization with a Pairwise copula						
	Combined		Wind		Swell	
	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)
Clayton	0.512	0.417	0.798	0.729	0.376	0.242
Frank	***0.446***	0.373	0.762	0.674	0.363	0.269
Galambos	0.421	0.920	0.757	1.000	0.433	0.912
Gumbel	0.532	***0.784***	***0.897***	1.000	***0.386***	0.888
HR	0.538	0.866	0.707	1.000	0.382	0.818
Joe	0.232	0.860	0.802	1.000	0.240	0.809
Normal	0.594	0.734	0.863	***0.873***	0.448	0.675
Plackett	0.509	0.644	0.865	0.940	0.441	***0.590***
Extra-parametrization with a Gumbel copula						
	Combined	Wind	Swell	Combined	Wind	Swell
	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)	λ_L(u=5%)	λ_U(u=95%)
Clayton	***0.444***	0.746	0.812	0.766	***0.390***	0.613
Frank	0.413	0.423	0.842	0.812	0.341	0.383
Galambos	0.479	***0.781***	0.823	1.000	0.437	0.751
HR	0.504	1.000	0.696	1.000	0.353	0.799
Joe	0.500	0.840	0.826	1.000	0.324	0.787
Normal	0.588	0.709	0.842	***0.911***	0.510	0.704
Plackett	0.579	0.727	***0.887***	0.755	0.368	***0.521***
Tawn	0.467	0.865	0.750	1.000	0.364	0.800
FGM	0.474	0.840	0.799	1.000	0.501	0.759
AMH	0.494	0.897	0.844	0.997	0.464	0.794

Of course, when simulating random sea-state parameters for reliability analysis, one may adopt optimised sampling techniques based solely on the tail behaviour. These are not approached in the present research. An extensive review on several optimised simulation techniques is provided in [52]. However, optimised sampling implies a correction on the systems' probabilities of failure, at least in the Monte-Carlo based methods, because one is solely sampling values in the near-failure region. In the present case, extra-parametrized copulas are a simple alternative to optimised sampling, as the simulation-efforts with copulas are reasonably low. The authors also recognise that the present technique should also be compared with other copula constructions, e.g. for a better assessment of the results hereby presented.

As noted in [6] and confirmed in the present study, it is often difficult to visually assess which copula-based model provides the best fit, particularly in (u,v)-space but also in the original (H_s;T_z)-space. Moreover, the results obtained from the Crámer-von Mises distance and the non-parametric estimations of tail dependence are sometimes very close between models. Nevertheless, some of the copula models can be disregarded based on the Crámer-von Mises distance and the tail dependences criteria, which is the case of several symmetric copulas and some of the extra-parametrized ones, for example the ones that use the Clayton copula. In the next section, one introduces the weighted version of the Root-Mean-Square Error (RMSE), adapted to account for the departures at the upper tail of the significant wave heights and their joint mean up-crossing wave periods.

Weighted-Root-Mean-Squared Error (WRMSE)

8. WRMSE applied to the extra-parametrization with Independent or Pairwise copulas

As a straightforward way to analyse these models' performance in the original (H_s;T_z)-space, one may compute the Weighted-Root-Mean-Square Error [6]. A discussion on the advantages and pitfalls of the evaluation of significant wave height models based on the root-mean-square error can be seen in [53]. If one assumes that the significant wave height is the dominant variable, in terms of reliability interest one can also assume that the errors performed on the upper tail of H_s should have a stronger penalty when compared with the errors related to the central part of the joint distribution of H_s and T_z.

In this research, a similar procedure to the one adopted in [6] is implemented. First, a set of N= 100 000 pairs (H_s; T_z) is obtained from each copula model, with the seasonality effects added back in. Then the original space is divided into bins of size 0.1 m × 0.1 s and the points falling in each bin are computed for both the empirical subsampled data and the simulated ones. The sum of the squared difference between fractions is calculated, thus providing a measure on the goodness-of-fit of the models.

Consider B_{ij} as the i-th bin in the H_s direction and the j-th in the T_z direction. Also consider that X_{ij} is the fraction of points from the empirical dataset that fall into B_{ij} and that Y_{ij} is the fraction of simulated points that fall into the same bin (B_{ij}). Then X_{ij} and Y_{ij} are obtained as in Eq. 15 and 16, while the WRSME is obtained from Eq. 17, where w_{ij} are the weights attributed to the errors . Note that w_{ij} varies between 0 and 1.

$$X_{ij} = \frac{1}{N_X}\sum_{k=1}^{N_X} I((H_s;T_z) \in B_{ij}) \tag{15}$$

$$Y_{ij} = \frac{1}{N_Y} \sum_{k=1}^{N_Y} I((H_s; T_z) \in B_{ij}) \quad (16)$$

$$WRMSE = \sqrt{\sum_{ij} w_{ij} \cdot \left(X_{ij} - Y_{ij}\right)^2} \quad (17)$$

There is no criteria on how to define the weights applied to the WRMSE. As mentioned in this case one is interested in penalizing the errors made on the upper tail of the joint distribution. The present dataset comes from a location for which the water depth ranges from 10 m to 20 m, at the Horns Rev 3 offshore wind farm [32]. In fact, the coordinates given for the case study correspond to a water depth of 18 m. Therefore, one expects that the wave heights are depth limited. In [32] it is assumed that the maximum wave height at the location is two times the significant one. According to the Coastal Engineering Manual [54] the breaker index, i.e. H/d, which defines the maximum non-breaking wave height (H) in water depth (d) limited situations, a theoretical value of 0.78 is used. However, a value of 0.6 is applied for irregular sea states. Therefore the maximum expected non-breaking wave height for d=18 m should roughly be H=0.6×18=10.8 m. In the present case, one is concerned with the significant wave height, which means that wave heights over H_s=10.8 m will be breaking. However, assuming that the maximum H_s is equal to 10.8 m, it is indeed a conservative limit when modelling the significant wave height for reliability assessment purposes at the case study location. Note that for H_s=10.8 m, the maximum wave height proposed in [32] would be around 21.6 m, which is an absurd value, given the water depth of 18 m. The limits proposed for the breaker index can be seen as reference levels and for practical situations one could definitely use a lower value than the one proposed for the maximum significant wave height. Note that several works have been performed on the discussion of the breaker index variations. Here, the limit of H/d=0.6 is assumed as a simplistic approach. For further details on this matter one can see [55].

A possible way to account for these physical limitations, which are not perceived in the above mentioned probabilistic models, is to truncate the marginal distribution used to obtain the random values of H_s. This was performed in the present case, i.e. maximum wave height is limited by H/d=0.6. However, one should again note that marginal modelling is an important and inherent part of the joint modelling. In this case, the lognormal distribution was applied as mentioned in section “Marginal Distributions”, but a more extensive analysis of other distributions can contribute to build similar copula-based models that are able to account for the data’s characteristics in a more efficient way. The present paper was focused on the joint asymmetric copula models. However, a comprehensive discussion on the uncertainties related to the marginal distributions is provided in [56]. Taking into consideration the maximum values of the subsamples, which are slightly above 6 m, the scale of weights applied to the WRMSE is defined as in Eq.18. Table 13 provides the WRMSE values for the extra-parametrized copulas. The AMH, the Tawn and the FGM copulas are not presented, because they were already excluded based on the parameter’s values and the respective Crámer-von Mises distance obtained for the extra-parametrized models with independent or pairwise copulas.

$$w_{ij} = \begin{cases} 0 & 0\,m < H_s \le 2\,m \\ 0.33 & 2\,m < H_s \le 4\,m \\ 0.67 & 4\,m < H_s \le 6\,m \\ 1 & 6\,m > H_s \end{cases} \quad (18)$$

Legend table 13: WRMSE for the extra-parametrization with Independent and Pairwise copulas. Lowest WRMSE in bold italics.

Copula	**Extra-parametrization with an Independent copula**		
	Combined	**Wind**	**Swell**
Clayton	0.000253	0.000327	0.000246
Frank	0.000233	0.000275	0.000226
Galambos	0.000236	0.000276	***0.000221***
Gumbel	***0.000226***	***0.000272***	0.000228
HR	0.000239	0.000273	0.000228
Joe	0.000238	0.000306	0.000235
Normal	0.000227	0.000283	0.000227
Plackett	0.000228	0.000288	0.000235
Copula	**Extra-parametrization with Pairwise copula**		
	Combined	**Wind**	**Swell**
Clayton	0.000241	0.000273	0.000219
Frank	***0.000225***	***0.000262***	0.000218
Galambos	0.000228	0.000272	0.000224
Gumbel	0.000227	0.000266	0.000224
HR	0.000231	0.000274	0.000223
Joe	0.000231	0.000268	0.000221
Normal	0.000231	0.000275	0.000229
Plackett	0.000228	0.000271	***0.000216***

Since table 13 refers to the original (H_s;T_z)-space, one is able to see~~,~~ which models hold the highest and the lowest weighted errors, with higher penalties directly given to departures on the upper tail. The models that provided the lowest errors in the extra-parametrization with an independent copula, are not the same as the ones that gave the lowest Crámer-von Mises distance. In this case the combined sea is best approximated by the Gumbel-Independent copula, while the wind-sea and swell components are approximated the best by the Gumbel-Independent and the Galambos-Independent copulas, respectively. Regarding the models extra-parametrized with pairwise copulas, one is able to see that the models that provide the lowest errors are the Frank-Frank copula for the combined sea and for the wind-sea and the Plackett-Plackett copula for the primary swell component.

The different results in terms of the Crámer-von Mises and the WRSME were already expected, mainly because the first measure concerns the overall fit and the second one deals with the upper-tail fit considering increased penalties with the increasing wave height. The differences between criteria were already noted for a different dataset in [6]. This emphasizes the need of analysing several goodness-of-fit criteria when dealing with copula-based models.

Table 13 also shows that, unlike in the Crámer-von Mises distance, the extra-parametrization with a pairwise copula did not lead to a generalized reduction of the WRMSE, when compared with the extra-parametrization with an Independent copula. In some models, e.g. inthe Joe copula-based ones the error is reduced, in the combined sea, while for others, e.g. the Normal-Normal model for the combined sea, the error increases when compared with the Normal-Independent model. Such observations confirm that the minimization process of the Crámer-von Mises distance does not necessarily lead to a minimization of the weighted root-mean-square error, for the same copula-based model.

Also the separation in the wind-sea and the primary swell components did not hold an improvement of the WRMSE when compared with the combined sea. This can be noted for example in the Plackett model extra-parametrized with an Independent copula and a pairwise copula. This occurs due to the fact that a model that presents a certain tail dependence may not be suitable for a specific dataset that does not yield a similar tail dependence.

The dependence structures of the combined sea and its components are different, therefore the performance of a specific copula-based model does not remain the same.

9. WRMSE applied to the extra-parametrization with a Gumbel copula

Table 14 provides the WRSME values for the copulas extra-parametrized with a Gumbel copula. One can see that the Galambos-Gumbel copula provided the lowest errors for the combined sea, while the HR-Gumbel copula provided the lowest ones for both the wind-sea and the swell components.

No systematic improvements are registered in these models' WRMSE when compared with the extra-parametrization technique with an Independent or a pairwise copula (table 13). However, one cannot guarantee that models based on other copulas could not result in lower values of WRMSE than the ones presented in table 14.

The WRMSE presented in both tables 13 and 14 are considerably low when compared with the ones obtained by [6]. This could be due to the dataset used, which is different, but also due to the number of simulations performed to obtain the model's parameters.

Further research should be carried out to clarify this aspect. When dealing with new datasets and based on the lack of improvements in both the Crámer-von Mises distance and the WRMSE, it seems reasonable to implement the extra-parametrized copulas with an Independent or a pairwise copula, before moving on to models with different copulas, which automatically increase the number of possible combinations to be tested.

Legend table 14: WRMSE for the extra-parametrization with a Gumbel copula (lowest WRMSE in italics).

Extra-parametrization with a Gumbel copula			
	Combined	**Wind**	**Swell**
AMH	0.000231	0.000275	0.000238
Clayton	0.000231	0.000269	0.000223
FGM	0.000231	0.000288	0.000240
Frank	0.000236	0.000277	0.000217
Galambos	*0.000230*	0.000272	0.000225
HR	0.000228	*0.00026*	*0.000210*
Joe	0.000233	0.000263	0.000220
Normal	0.000228	0.000276	0.000221
Plackett	0.000235	0.000281	0.000226
Tawn	0.000240	0.000282	0.000227

Discussion and Applications to Wind Engineering

The present section outlines and discusses the results obtained and the pitfalls encountered in the asymmetric copula based models proposed. Furthermore, some insights on possible applications to offshore wind engineering are provided.

This research was mainly focused on the joint modelling of the significant wave height and the mean up-crossing period. However, the results showed that further research should be performed to improve the univariate modelling of the marginal distributions. The main reason for this is the fact that using the log-normal distribution revealed itself as a very heavy tail distribution, which contributed for a worse fit in the upper tail region of the datasets. However, this distribution provided the lowest Kolmogorov-Smirnov and Wasserstein distances, thus being the best choice among the tested marginal distributions (table 4). Nevertheless, this work covered a wide range of distributions, often used in the offshore wind engineering, e.g. [4], [5] or [13].

Other aspects, such as the choices made on the breaker-index, the truncation of the marginal distributions, the selection of subsampled data based on the 2day maximum significant wave height and the associated mean up-crossing period or the weights defined for the WRMSE, affect the model's choice and model's output. Although further research should be carried out to quantify the influence of these choices, at the end of the day, it is reccommended that the designer tests several possibilities, before making a decision. Since each dataset may present its own particularities, the experience of the designer is also a key aspect to assess the quality of the joint model proposed to deal with met-ocean data.

Other procedures can be used to remove the data's seazonality, e.g. see [57] and [58] . The method used, based on [6], proved to be suitable. Although these copula models are built to add back in the seasonality effects to the generated values of (H_s;T_z), some practical issues may appear when dealing with the pre-processed data, namely the occurrence of negative pre-processed values. Negative values of pre-processed data may pose some difficulties when trying to fit marginal distributions that only have a positive domain, e.g.

lognormal or Weibul distributions. Although this problem can be solved by performing a shift location, either on H_s or T_z direction, this contributes to obtain a set of pre-processed data that is not directly comparable with the original or the subsampled data. The application of the proposed models remains valid, but the designer should pay attention to the possibility of having some conter-inuitive information, as it occurred in table 6, which presented a maximum pre-processed H_s that was higher than the maximum presented for the total sea.

Also, another aspect that can be further improved is the non-parametric estimation of tail dependence. In this study the estimation based on [49] was applied. However, this estimation does not really enable one to understand the actual assymptotic behaviour of the data, as showed in table 11 for the Clayton copula. A somehow *ad hoc* solution consisted in the application of the non-parametric estimation of the finite quantile dependence at the 5% and the 95% quantiles. Tail dependence analysis can be crucial to rule out some of the inittially proposed models. Therefore, other tail and assymptotic dependence analysis, e.g. [33] or [59], should be further implemented to improve the present research.

Moreover, in order to obtain a better assessment of these copula-based models, one can complement the analysis made with the tail dependence, the Crámer-von Mises distance or the WRMSE, with a response-based analysis, as mentioned in [59] . Several models can be employed to model the met-ocean data and then used e.g. for load estimations at an offshore wind turbine or for dragging forces acting on the soil-structure interaction, and then the behaviour of the response, say the loads in the first case, is further analysed. For example, when dealing with scour protections for offshore wind turbines, [60] compute the bed shear-stress response for several distributions of the environmental variables. By performing this analysis one is able to compare the responses obtained from different modelling assumptions. One is also able to statistically study the response variables along with the environmental ones. In this matter, obtaining an non-parametric estimation of the response is also useful to validate the statistical models.

This paper also showed that opting for a joint modelling (based on copulas) of the total sea, might be a good option, because the asymmetry of combined data is not necessarily reduced for the separate wind-sea and primary swell components. Additionally, in current practice, the separation of the components might be problematic, namely, if one is dealing with real observations instead of the hindcast data. However, being able to statistically describe the sea components might be of great importance when dealing with specific matters of wind engineering design. For example, in offshore wind turbines, the failure caused by waves from the wind-sea component, which is related with the wind speed, can also be related to the operation mode of the turbine, which depends on the cut off and cut in wind speed, e.g. [61]. On the other hand, the swell component can also be important, for example, when dealing with the dynamic behaviour of offshore wind floating foundations, e.g. [62], [63].

The results obtained from the extra-parametrization technique led to lower values than the ones presented in [6] for both the Crámer-von Mises distances and the WRMSE. Also the estimated copula parameters are not only within the theoretical copula's range (see [21]) but seem to compare reasonably well with the ones used in copulas application to met-ocean data, e.g. [11], [45] or [46]. However, these works only include symmetric models, which leaves space for the present technique to be tested, possibly leading to fitting improvements in the datasets related to the mentioned works.

In this case the significant wave height and the mean up-crossing period are analyzed. Nevertheless, these asymmetric copula-based distribution models can also be extended to the analysis of peak periods and extreme wave heights [15], hydrodynamic loads [15], wind speed analysis [64] among other environmental variables needed for design. Note that wind speed is often very asymmetric in terms of directional distribution and spatial correlation. Copula applications as the ones presented here, and for instance in [7], can be used to model this variable. This type of modelling is not only useful for design purposes but it also yields potential applicability to broader aspects of wind engineering, e.g. to weather prediction for offshore maintenance operations [65], to the estimation of seasonal energy production [66], to the lifetime extension of wind turbines [3] or to scour protections design [60].

The majority of the works performed with copulas for met-ocean data is typically applied to a bi-variate case. However, it should be recognized that the proposed models can be applied to problems with several dimensions. This means that one is able to create joint models, for example, for wave heights, peak periods and incident directions, wind speed and ocean current velocity, among several other variables of interest. In this sense, the the extra-parametrization technique poses a straightforward alternative to nested copulas and C,D-Vine copulas, applied for example in wind resource estimation by [24] by [67] or [24] to model spectral sea-state parameters and [68] to perform spatiotemporal modeling of wind generation power for storage sizing.

A final remark should be made to the fact that the popularity of asymmetric copula models is increasing. This occurs not only due the models flexibility but also due to their ability to tackle the complexity and computational burden of modeling high-dimensional data. This advantage coupled with its simple application and the promising results, which often confirm that these models compare reasonably well with current practice methods, e.g. see [69] and [6], make them an interesting alternative for several wind engineering applications, with a special emphasis on the ones related to the offshore wind industry.

Conclusions

The present research proposed several extra-parametrized copula-based models which were applied to a dataset of hindcast significant wave height and up-crossing mean wave period, which refer to the location of the Horns Rev 3 offshore wind farm. An extensive application was performed for the total (combined) sea and its respective components of wind-sea and primary swell. With this research it was possible to conclude that the separate modelling of the wind-sea and primary swell components does not always lead to a better goodness-of-fit of the copula models. Therefore, it seems a reasonable choice to model the total sea, without separation of its components. Moreover, using the extra-parametrization technique it becomes easier to catch the asymmetry of the total sea.

It was noted that the copula's performance is very much dependent on the asymmetry of the data itself. Therefore, if the transformed significant wave heights and the mean up-crossing periods, from primary swell and wind-sea, are asymmetric in the (u,v)-space, it is recommended to use asymmetric copulas to improve the goodness-of-fit. It is concluded that problems caused by asymmetries in the data, can be reduced by the flexibility introduced with the extra-parametrization of copulas.

The extra-parametrization led to significant improvements of the Crámer-von Mises distance, between the models and the empirical data. The improvements were noticed for the extra-parametrization with both an independent copula and with the pairwise copulas. The flexibility introduced enables one to easily build joint models for the significant wave height and up-cross mean period, based on simple symmetric copulas, such as the Archimedean or the Elliptical ones.

Using this technique to combine different copulas is also possible. In this paper, an example was shown for the extra-parametrization with a Gumbel copula. These copula constructions also improved the goodness-of-fit when compared with the symmetric copulas. However, they still presented higher Crámer-von Mises distances when compared with the pairwise construction and some of the independent copula based models.

It was also confirmed that a model that gives the lowest Crámer-von Mises distance, may not always provide the lowest WRMSE adapted to penalize more the departures on the upper tail of the significant wave heights. Therefore, the Crámer-von Mises distance must be seen as an overall goodness-of-fit measure, while the proposed loss function based on WRMSE is more suitable to deal with errors on the tails of the distributions.

Acknowledgements

T. Fazeres-Ferradosa is funded by the Portuguese Foundation for Science and Technology (FCT) under the Ph.D. scholarship PB/BD/113454/2015 – Doctoral Program INFRARISK. The authors also acknowledge to Dr. Mario Lopez (University of Oviedo) and MSc. Bruno Oliveira (University of Porto) and Diogo Mendes (Hydrographic Institute) for the helpful discussions regarding the hindcast data and the models' selection. T. Fazeres-Ferradosa acknowledges Dr. Francisco Q. Fazeres (ULSAM) on the enlightening discussions on survival and reliability analysis.

References

[1] API. (2011). API - RP 2FPS *Planning, Designing and Constructing Floating Production Systems*. USA: American Petroleum Institute.

[2] API. (2014). API RP 2A-WSD - *Planning, Designing, and Constructing Fixed Offshore* - Working Stress Design. United States of America: American Petroleum Institute.

[3] DNVGL. (2017a). DNVGL - ST- 0262 - *Lifetime extension of wind turbines*. Norway: Det Norske Veritas AS.

[4] DNVGL. (2017b). DNV OS J101 - *Design of Offshore Wind Turbine Structures*. Norway: Det Norsk Veritas AS.

[5] DNVGL. (2017c, August). DNVGL-RP-C205 *Environmental conditions and Environmental loads*. Norway: Det Norske Veritas AS.

[6] Vanem, E. (2016). Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*. vol. 49, 180-205.

[7] Jane, R., Valle, L. D., Simmonds, D., & Rabya, A. (2016). A copula-based approach for the estimation of wave height records through spatial correlation. *Coastal Engineering. vol. 117*, 1-18.

[8] Ferreira, J., & Guedes Soares, C. (2002). Modelling bivariate distributions of significant wave height and mean wave period. *Applied Ocean Research. vol. 24*, 31-45.

[9] Ziegler, L., Voormeeren, S., Schafhirt, S., & Muskulus, M. (2015). Sensitivity of Wave Fatigue Loads on Offshore Wind Turbines under Varying Site Conditions. *Energy Procedia. vol. 80*, 193-200.

[10] Whitehouse, R., Brown, A., Audenaert, S., Bolle, A., Schoesitter, P., Haerens, P., Baelus, L., Troch, P., das Neves, L., Ferradosa. T., Taveira-Pinto, F. (2014). Optimising scour protection stability at offshore foundations. *Proceedings of the 7th International Conference on Scour and Erosion* (pp. 593-600). Pearth, Australia: CRC Press.

[11] Montes-Iturrizaga, R., & Heredia-Zavoni, E. (2016). Reliability analysis of mooring lines using copulas to model statistical dependence of environmental variables. *Applied Ocean Research. vol. 59*, 564-576.

[12] Holthuijsen, L. (2007). *Waves in Oceanic and Coastal Waters*. New York: Cambridge University Press.

[13] Bitner-Gregersen, E. (2015). Joint met-ocean description for design and operation of marine structures. *Applied Ocean Research*, 279-292.

[14] Jonathan, P., & Ewans, K. (2013). Statistical Modelling of extreme ocean environments for marine design: A review. *Ocean Engineering. vol. 62*, 91-109.

[15] Dong, S., Wang, N., Liu, W., & Soares, C. G. (2013). Bivariate maximum entropy distribution of significant wave height and peak period. *Ocean Engineering. vol. 59*, 86-99.

[16] Lucas, C., & Guedes-Soares, C. (2015). Bivariate distributions of significant wave height and mean wave period of combined sea states. *Ocean Engineering. vol. 106*, 341-353.

[17] Guedes-Soares, C., & Scotto, M. (2011). Long term and extreme value models of wave data. In G. Soares, *Marine Techonology and Engineering* (pp. 97-108). London: Taylor & Francis Group.

[18] Ewans, K., & Jonathan, P. (2014). Evaluating environmental joint extremes for the offshore industry using conditional extremes model. *Journal of Marine Systems. vol. 130*, 124-130.

[19] Mahfoud, M. (2012). Bivariate Archimedean copulas: an application to two stock market indices. Retrieved from *https://beta.vu.nl/nl/Images/werkstuk-mahfoud_tcm235-277460.pdf*

[20] Zhang, Y., Quck, S., & Beer, M. (2013). Long-term reliability assessement of offshore structures in a complex environment. In G. Deodatis, B. Ellingwood, & D. *Frangopol, Safety, Reliability, Risk and Lyfe-Cycle Performance of Structures & Infrastructures* (p. 2209). London: Taylor & Francis Group.

[21] Nelsen, R. (2006). An Introduction to Copulas. New York: Springer.

[22] Corbella, S., & Stretch, D. (2013). Simulating a multivariate sea storm using Archimedean copulas. *Coastal Engineering. vol.76*, 68-78.

[23] Antão, E., & Soares, C. G. (2016). Approximation of the joint probability density of wave steepness and height with a bivariate gamma distribution. *Ocean Engineering. vol 126 (1)*, 402-410.

[24] Jäger, W., & Napoles, O. M. (2015). Sampling joint time series of significant wave heights and periods in the North Sea. *Safety and Reliability of Complex Engineered Systems*, 4287-4294.

[25] Brechmann, E., & Schepsmeier, U. (2013). Modeling Dependence with C- and D-Vine Copulas: The R Package CD Vine. *Journal of Statistical Software. vol. 52*, issue 3, 1-27.
[26] Khoudraji, A. (1995). *Contributions à l'étude des copules et àla modélisation des valeurs extrêmes bivariées - PhD thesis*. Québec, Canada: Université Laval.
[27] Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis. vol. 99*, 2234-2250.
[28] Durante, F., & Salvadori, G. (2010). On the construction of multivariate extreme models via copulas . *Environmetrics* 21, 143-161.
[29] Ochi, M. (1978). Wave statistics for the design of ships and ocean structures. *SNAME Trans (86)*, 47-76.
[30] Vanem, E. (2017). A regional extreme value analysis of ocean waves in a changing climate. *Ocean Engineering*, 277-295.
[31] Bitner-Gregersen, E., & Toffoli, A. (2009). Uncertainties of Wind Sea and Swell Prediction From the Torsethaugen Spectrum. *ASME 29th International Conference on Ocean, Offshore and Arctic Engineering* (pp. 1-8). Hawai - Honolulu: American Society of Mechanical Engineers.
[32] DMI. (2013). *Horns Rev 3 Offshore Wind Farm - Metaocean*. Copenhagen, Denmark: DMI - Danish Metereological Institute & Orbicon A/S.
[33] Genest, C., & Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*. Vol 12, issue 4, 347-368.
[34] Soldevilla, M. J., Martín-Hidalgo, M., Negro, V., López-Gutiérrez, J. S., & Aberturas, P. (2015). Improvement of theoretical storm characterization for different climate conditions. *Coastal Engineering. vol. 96*, 71-80.
[35] Wang, W., & Wells, M. (2000). "Model selection and semiparametric inference for bivariate failure-time data (with discussion). *Journal of the American Statistical Association. vol. 96 (2)*, 62-76.
[36] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. vol. 19, 716-723.
[37] Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics. vol 6. No. 2*, 461-464.
[38] Shumway, R., & Stoffer, D. (2011). *Time Series Analysis and its Applications with R* examples. London: Springer.
[39] Huang, W., & Prokhorov, A. (2014). A goodnesss-of-fit Test for Copulas. *Econometric Reviews. vol. 33*, issue 7, 751-771.
[40] Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics. vol. 44*, issue 2, 199-213.
[41] Durante, F., & Sempi, C. (2015). Principles of Copula Theory. London: CRC Press, Taylor & Francis Group.
[42] Bücher, A., & Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis. vol. 119*, 61-70.
[43] Kumar, P. (2010). Probability Distributions and Estimation of Ali-Mikhail-Haq Copula. *Applied Mathematical Sciences. vol. 4. No. 14*, 657-666.

[44] Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., & Erhardt, T. (2017). R package 'VineCopula' - Statistical Inference of Vine Copulas. Retrieved from cran.r-project.org: *https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf*

[45] Montes-Iturrizaga, R., & Heredia-Zavoni, E. (2015). Environmental contours using copulas. *Applied Ocean Research, vol. 52*, 125-139.

[46] Redón-Conde, C., & Heredia-Zavoni, E. (2015). Reliability assessment of mooring lines for floating structures considering statistical parameter uncertainties. *Applied Ocean Research. vol. 52*, 295-308.

[47] Ferreira, H., & Ferreira, M. (2018). Multidimensional extremal dependence coefficients. *Statistics & Probability Letters. vol. 133*, 1-8.

[48] Lekina, A., Chebana, F., & Ouarda, T. (2015). On the tail dependence with bivariate hydrological frequency analysis. *Dependence Modeling*, 203-227.

[49] Schmid, F., & Schmidt, R. (2007). Multivariate conditional versions of spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis. vol 98. issue 6*, 1123-1140.

[50] Durante, F., Fernández-Sanchez, J., & Pappadá, R. (2015). Copulas, diagonals and tail dependence. *Fuzzy, Sets and Systems. vol. 264*, 22-41.

[51] Pérez, A., & Prieto-Alaiz, M. (2016). A note on nonparametric estimation of copula-based multivariate extensions of Spearman’s rho. *Statistics & Probability Letters*, 41-50.

[52] Fajardo, F., Perez, J., Alsina, M., & Marques, J. R. (2010). *Simulation Methods for Reliability and Availability of Complex Systems*. New York: Springer.

[53] Mentaschi, L., Besio, G., Cassola, F., & Mazzino, A. (2013). Problems in RMSE-based wave model validation. *Ocean Modelling*, 53-58.

[54] US Army Corps Engineers. (2002). *Coastal Engineering Manual*. USA: USACE - Army Publications.

[55] Shand, T., Peirson, W., & Cox, R. (2010). Engineering Design in the Presence of Wave Groups. *Proceedings of 32nd Conference on Coastal Engineering*, Shanghai, China (pp. 1-10). Shangai, China: ICCE.

[56] Vanem, E. (2015). Uncertainties in extreme value modelling of wave data in a climate change. *Journal of Ocean Engineering and Marine Energy*, 339-359.

[57] Menéndez, M., J.Méndez, F., Izaguirre, C., Alberto Luceño, J.,Losada, I. (2009). The influence of seasonality on estimating return values of significant wave height. *Coastal Engineering. vol. 56, issue 3*, 211-219.

[58] Sartini, L., Cassola, & Besio, G. (2015). Extreme waves seasonality analysis: An application in the Mediterranean Sea. Journal of Geophysical Research: Oceans, 6266-6288.

[59] Su, J., & Lei. (2017). A general approach to full-range tail dependence copula. *Insurance: Mathematics and Economics. vol. 77,* 49-64.

[60] Fazeres-Ferradosa, T., Taveira-Pinto, F., Neves, L. d., & Reis, M. T. (2016). Design of scour protections and structural reliability techniques. In Erpicum, *Sustainable Hydraulics in the Era of Global Chang*e (pp. 527-531). London: Taylor & Francis Group.

[61] Dowell, J., Zitrou, A., Walls, L., Bedford, T., Infield, D. (2013). Analysis of Wind and Wave Data to Assess Maintenance Access to Offshore Wind Farms. *European Safety and Reliability Association Conference (ESREL)*, (pp. 1-8). Amesterdam.

[62] Brommundt, M., Ludwig, K., Merz, K., & Muskulus, M. (2012). Mooring System Optimisation for Floating Wind Turbines using Frequency Domain Analysis. *Energy Procedia. vol. 24*, 289-296.

[63] Campanile, A., Piscopo, V., & Scamardella, A. (2018). Mooring design and selection for floating offshore wind turbines on intermediate and deep water depths. *Ocean Engineering*. vol. 148, 349-360.

[64] Wang, Y., Ma, H., Wang, D., Wang, G., Wu, J., Bian, J., & Liu, J. (2018). A new method for wind speed forecasting based on copula theory. *Environmental Research. vol. 160*, 365-371.

[65] Taylor, J., Jeon, J. (2017). Probabilistic forecasting of wave heights for offshore wind turbines maintenance. *European Journal of Operational Research*, (In Press, Accepted Manuscript).

[66] Elginoz, N., & Bas, B. (2017). Life Cycle Assessment of a multi-use offshore platform: Combining wind and wave energy production. *Ocean Engineering. vol. 145*, 430-443.

[67] Veeramachaneni, K., Cuesta-Infante, A., & O'Reilly, U.-M. (2015). Copula graphical models for wind resource estimation. *IJCAI'15 Proceedings of the 24th International Conference on Artificial (pp. 2646-2654).* Buenos Aires: AAAI Press.

[68] Haghj, H., Lotifard, S. (2015), Spatiotemporal Modeling of Wind Generation for Optimal Energy Storage Sizing. *IEEE Transactions on Sustainable Energy. vol. 6 No. 1*, 113-121.

[69] Jonathan, P., & Ewans, K. (2013). Statistical Modelling of extreme ocean environments for marine design: A review. *Ocean Engineering. vol. 62*, 91-109.