

PAPER • OPEN ACCESS

## Spatial met-ocean data analysis for the North Sea using copulas: application in lumping of offshore wind turbine fatigue load cases

To cite this article: Alahyar Koochekali and Michael Muskulus 2019 *J. Phys.: Conf. Ser.* **1356** 012042

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

# Spatial met-ocean data analysis for the North Sea using copulas: application in lumping of offshore wind turbine fatigue load cases

Alahyar Koochekali<sup>1</sup>, Michael Muskulus<sup>1</sup>

Department of Civil and Environmental Engineering, NTNU, Høgskoleringen 7A,  
7491 Trondheim, Norway

[akoochekali@gmail.com](mailto:akoochekali@gmail.com)

**Abstract.** Knowledge about stochastic environmental loads and the response of the substructure are essential in fatigue damage calculations for offshore wind turbines. Site-specific and high-quality met-ocean data that contains joint measurements of wind and waves however, are not always available. An approach based on copulas was used to generate lumped joint met-ocean data in a location where only marginals (i.e., individual, univariate probability distributions) are known. Empirical copulas were calculated for different stations in the North Sea for pairs of wind speed and wave height variables. These copulas are then combined with marginals from another station to generate lumped data. The generated data is compared with the known lumped joint data, and the difference of two methods is measured. The lumped values are used to estimate fatigue damage based on simple formulas. The results show that the joint behaviour of wind speed and wave height follows a similar pattern around the North Sea. In addition, the mean difference of generated lumped wind speeds based on this technique from lumped wind speeds based on real measurements is less than 5%. The mean difference of the fatigue damage caused by the generated lumped wind speeds relative to lumped wind speeds based on real measurements is less than 5% at two locations, and less than 15% in another location. This suggests that the approach is promising for estimating joint met-ocean data at new locations in the North Sea where only marginals are available, which is of major interest for wind farm planning and design.

## 1. Introduction

An adequate representation of site-specific met-ocean data has several important applications in marine and coastal industry. Cost-efficient and reliable design of offshore wind turbines especially for fatigue damage are highly dependent on high-quality, joint met-ocean data. In addition, reducing uncertainties in environmental data can effectively improve the planning of installation and maintenance-operation of marine structures [1].

Offshore wind turbines are ubiquitous. The main drawback of offshore wind power is its comparatively high capital cost. This high cost can be attributed to more expensive support structures, grid connections and offshore installations [2]. However, it is expected that the wind industry will reduce its cost by 40% until 2020. Application of large offshore wind farms and large-sized monopiles in water depths of 25-



40 meters may be a solution to reduce the costs. The design of these monopiles is typically governed by fatigue damage, with significant contributions from wave excitation [3].

Integrated time-domain analysis that can capture the dynamic behaviour of the support structure is an accurate tool for fatigue design of offshore wind turbines [4]. However, to avoid excessive computational time of these simulations, input reduction is imperative [5]. Lumping is a way to reduce a full-sea-state to a limited number of load cases by weighting environmental parameters by a certain criterion. In terms of fatigue damage, this criterion can be the desire to achieve approximately the same total damage, when all load cases considered are assessed and correctly weighted.

A practical challenge is that joint measurements of site-specific met-ocean data, e.g. correlated wind speed and wave heights, are not always available, especially in early wind farm project phases. The pairwise dependence between two random variables (e.g. wave height and wind speed) can be described using classical families of bivariate distributions such as the normal, Weibull, or Gumbel distributions. However, one drawback of using these joint distributions is that their univariate distributions (marginals) must then be characterized by the same parametric family of distributions [6], which can be too limiting. Another, more flexible way of modelling dependence between random variables is to use a copula<sup>1</sup>. This is a multivariate probability distribution whose individual variables have uniform marginals [7] and a standard statistical tool to capture the dependence structure of two or more random variables [8].

Accurate lumping of site-specific joint met-ocean data is necessary for fatigue damage estimation, but, it is possible that joint measurements of data are not available. For example, wind speed measurements for a given offshore location may start some years after wave measurements.

This work develops a new method based on estimating copulas at one location in the North Sea with joint measurements, and then combining it with marginals measured at another location, to generate (lumped) joint data for the new location. The hypothesis here is that although the scale of environmental parameters of interest will vary between the locations, the correlations between those parameters might be quite similar. The effectiveness of this method to estimate fatigue damage caused by lumped loads is investigated.

In order to address these objectives, long-term met-ocean data were gathered from four locations around the North Sea. Corrupted data and uncorrelated measurements were omitted. Empirical marginal distributions of wave height and wind speed, and the empirical copula describing their relationship for the four sites were determined. The copula domain was discretised on a fine grid. A finite difference method was used to determine copula density (i.e., joint probability of occurrence) in the copula domain. The marginal domains were discretised into bins at a coarser resolution typical for applications. These bins were transferred to the copula domain using the empirical cumulative distribution functions of the marginals. A wave height representative for each wind speed bin was then calculated by weighted averaging of wave heights from all copula grid cells inside each wind speed bin. Similarly, a wind speed for each wave height bin was calculated by weighted averaging of wind speeds that occur inside it. The root mean square error between the generated lumped conditional wind speeds and lumped wind speeds from measurements was quantified. Finally, a simplified formula was used to compare fatigue damage caused by these lumped data.

## 2. Theory

### 2.1. Empirical copula

---

<sup>1</sup> If one random variable is a linear function of another random variable, then linear correlation completely describes their relationship. A copula is a generalization of this idea that provides a complete description of the relationship between random variables that are correlated in an arbitrary (non-linear) manner. A copula can be considered a version of the joint distribution of two (or more) random variables that is “normalized” with respect to the (marginal) distributions of the individual random variables: The dependence is not described in absolute terms between values that the random variables take, but in terms of which quantiles occur jointly with which other quantiles.

Copulas isolate the marginal statistical properties from the dependence structure of random variables. Combining copulas with marginal distributions generates a joint distribution which represents multivariate behaviours [9].

The copula  $C$  of two random variables,  $X, Y$  with marginals  $F_X(x) = P(X \leq x) = U$  and  $F_Y(y) = P(Y < y) = V$  is defined as the joint cumulative distribution of the pair  $(F_X(x), F_Y(y))$ :

$$F_{X,Y}(u, v) = C(u, v) = \text{prob}[U \leq u, V \leq v], \quad u, v \in [0,1] \quad (1)$$

Here  $U$  and  $V$  are the cumulative distribution functions and are uniformly distributed.

Based on Sklar's theorem there exists a unique copula  $C$  for which equation (1) holds, and the joint dependence of  $X$  and  $Y$  is characterized fully and uniquely by  $C$  [6].

It can be mathematically proven that a copula associated with  $(X, Y)$  is invariant under monotone increasing transformations of the marginals  $F_X(x), F_Y(y)$ . One of the functions that meets this requirement is the rank function, therefore, pairs of ranks of the variables  $X, Y$  have the same copula as the original variables [6].

A realization of  $n$  joint measurements  $(X_i, Y_i)$  and their joint ranks are shown in equation (2)

$$(R_1, S_1), \dots, (R_n, S_n) \quad (2)$$

where  $R_i$  is the rank of  $X_i$  among  $X_1, \dots, X_n$  and  $S_i$  is the rank of  $Y_i$  among  $Y_1, \dots, Y_n$ . By dividing both  $R$  and  $S$  by  $\frac{1}{n+1}$ , the domain is rescaled to the copula domain  $[0,1]^2$ . An *empirical copula* can therefore be defined as:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1 \left( \frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right) \quad (3)$$

where  $1(\cdot)$  is an indicator function. This is an unbiased estimate of the true copula  $C(u, v)$  underlying the data.

The copula density  $c(u, v)$  equals the derivative of  $C(u, v)$  relative to its arguments, and based on Sklar's theorem it can be written as

$$f(u, v) = f(u)g(v)c(U(u), V(v)) \quad (4)$$

in terms of the probability densities  $f, g$  of  $X, Y$ . If the copula density is diagonal (with values of 1 along the diagonal), then the variables are independent [7].

## 2.2. Lumping of environmental conditions

To reduce the number of load cases for fatigue-damage design of an offshore wind turbine, a scatter plot of joint sea-state is typically reduced to a number of lumped environmental conditions (e.g. sea-states and associated wind speeds) in such a manner that the set of all such load cases represents the total expected damage that accumulates in a structure as accurately as possible [10].

Some of the variables that play a role in determining wind and wave load combinations for fatigue design are mean wind speed at hub height, wind direction, wave direction, and the sea state [11].

Most of the traditional lumping methods aim at lumping sea-states (conditional on the wind speed). The estimation of such lumped loads for an offshore wind turbine is typically based on assuming a quasi-static response. The lumped loads can be refined through an iterative process using a correction factor [10]. The traditional lumping methods for lumping sea-states work in two ways: preserving the wave period distribution and averaging the (significant) wave height for each wave period considered [12], or the other way around, preserving the wave height distribution and averaging the wave period [13]. However, these methods cannot provide accurate lumped load cases for larger offshore wind turbines established on monopiles where fatigue damage has to be checked at different places along the structure

[11]. Neither of the traditional methods takes the dynamics of the structures into account. The dynamics of the support structure can be captured using frequency domain analysis. By assuming a narrow-banded response spectrum for a monopile, lumping of sea-states was performed based on the square root of spectral wave energy at the structure's first natural frequency for each sea-state [14]. Damage equivalent wave height and wave periods were introduced to lump sea-states based on preserving fatigue damage, neglecting wind speed [11]. The structure's response was summarized in a damage matrix. A more recent wind-wave correlation method combined sea-state lumping and wind speeds by considering the dynamic response of offshore wind turbines [4]. The goal of this new method was to establish wind-wave correlations in terms of a unique, damage equivalent  $H_S - T_p$  combination per wind speed bin, wind direction, and wave direction.

In this article a simple lumping method is used to simplify a scatter diagram of wave height versus wind speed. The lumping is performed over wave height, i.e., by preserving the wave height distribution and lumping the wind speed, resulting in one representative wind speed associated with each wave height bin. A weighted average is used to obtain the conditional wind speed:

$$W_i = \frac{\sum_{j=1}^m P_{i,j} \cdot W_{i,j}}{\sum_{j=1}^m P_{i,j}} \quad (5)$$

Here  $i$  is the index of the wave height bins,  $j$  is the index of the wind speed bins,  $m$  is the number of wind speed bins, and  $P_{i,j}$  is the probability of occurrence of the combination  $[W_{i,j}, H_{i,j}]$ .

### 2.3. Fatigue damage

To calculate the cumulative fatigue damage the Palmgren-Miner rule is used. The accumulated fatigue damage ( $D$ ) is given by:

$$D = \sum_{i=1}^k \frac{n_i}{N_i} \quad (6)$$

where  $n_i$  is the total number of stress cycles of stress range  $S_i$ , taking into account environmental loading, which results in an expected failure at  $N_i$  cycles, and  $k$  is the number of different stress ranges. However, in this work a more simplistic approach is used. In this method the damage is proportional to key characteristics of the environmental conditions, as follows [10]:

$$\begin{aligned} D &\propto \Delta\sigma \propto H_S^\mu \\ D &\propto n_{total} \propto \frac{1}{T_z} \\ D &\propto \Delta\sigma^\mu \propto W^\mu \end{aligned} \quad (7)$$

where  $D$  is the damage,  $W$  is the mean wind velocity at hub height,  $\mu$  is the slope of the S-N curve (i.e., the Wöhler exponent),  $H_S$  is significant wave height, and  $T_z$  is the mean zero crossing period. This formulation is based on quasi static analysis of an offshore wind turbine [10].

## 3. Methods

In this section the theory of copulas is used to capture the dependence of pairs of wave height and wind speed measured together at one location in the North Sea. The dependence between these two variables, as captured by the copula, is transferred to another location in the North Sea and is combined with marginal data. The resulting environmental conditions are lumped and used to estimate fatigue damage for an offshore wind turbine.

### 3.1. Data gathering

The data considered in the present study are collected at four different locations around the North Sea, shown in Figure 1. Each dataset includes average wind speed (m/s), measured at 10 m above the sea level, wind direction, significant wave height, wave direction, and wave period, that are measured in time intervals of 10 minutes and 3 hours. The first two data-sets include 16 years (2002-2018) and 24

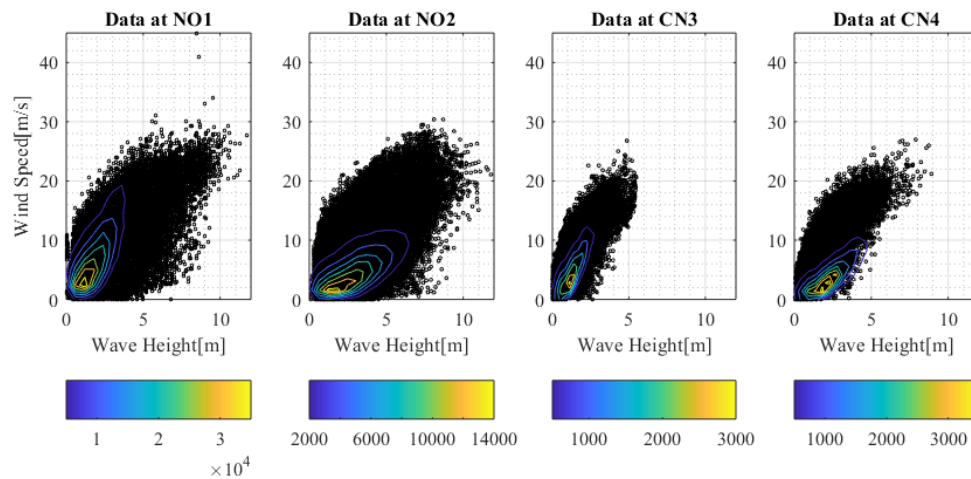
years (1994-2018) of data measured in ten-minute intervals from two offshore stations, located in the north-east part of the North Sea, NO<sub>1</sub> and NO<sub>2</sub>, available from the Norwegian Meteorological Institute website ([www.eklima.no](http://www.eklima.no)). The third and fourth environmental data sets include 24 years (1992-2016) of data measured in three-hour intervals from two offshore stations located in the centre of the North Sea, CN<sub>3</sub> and CN<sub>4</sub>, available from ([www.waveclimate.com](http://www.waveclimate.com)).

**Table 1.** Data source

Name	Source	Duration
NO.1	<a href="http://www.eklima.no">www.eklima.no</a>	2002_2018
NO.2	<a href="http://www.eklima.no">www.eklima.no</a>	1994_2018
CN.3	<a href="http://www.waveclimate.com">www.waveclimate.com</a>	1992_2016
CN.4	<a href="http://www.waveclimate.com">www.waveclimate.com</a>	1992_2016



**Figure 1.** Data source locations



**Figure 2.** Scatter plots of raw data at the four North Sea locations studied.

### 3.2. Excluding unrealistic values

Missing data and defect measurements exist in all of the data sets. To keep only correlated data, records with missing values in one or more variables were excluded from the data. The number of excluded data points are relatively small, compared to the size of the remaining data set.

For purposes of the analysis it is assumed that the measured met-ocean data are realizations of independent and identically distributed random variables (iid). However, the original met-ocean data may not be iid, as they may contain ties and serial dependence [7]. This is not a serious limitation, as serial dependence of the data does not influence the results. The wind speed and wave height are the two variables that are treated in this analysis. The pre-processed data at the four locations are plotted in Figure 2.

### 3.3. Applying copulas

#### Copula generation

The empirical copula at each station was calculated. To simplify the calculations, small random values between  $[1e^{-9} - 1e^{-5}]$  were added to the data to prevent the existence of completely similar values. This makes tracking and ranking joint measurements easier. The pairs of ranks  $[R, S]$  were determined and the copula was calculated using equation (3). The copula domain  $[0,1]^2$  was discretised into a

100x100 mesh grid using equally spaced intervals of size 0.01 in each variable. The calculated copulas are shown in Figure 3.

The wave height inside each cell of the copula grid was calculated by averaging the inverse CDF of the cell boundaries,

$$H_{cell,i,j} = \frac{1}{2} (ICDF(U(i)) + ICDF(U(i+1))) \tag{8}$$

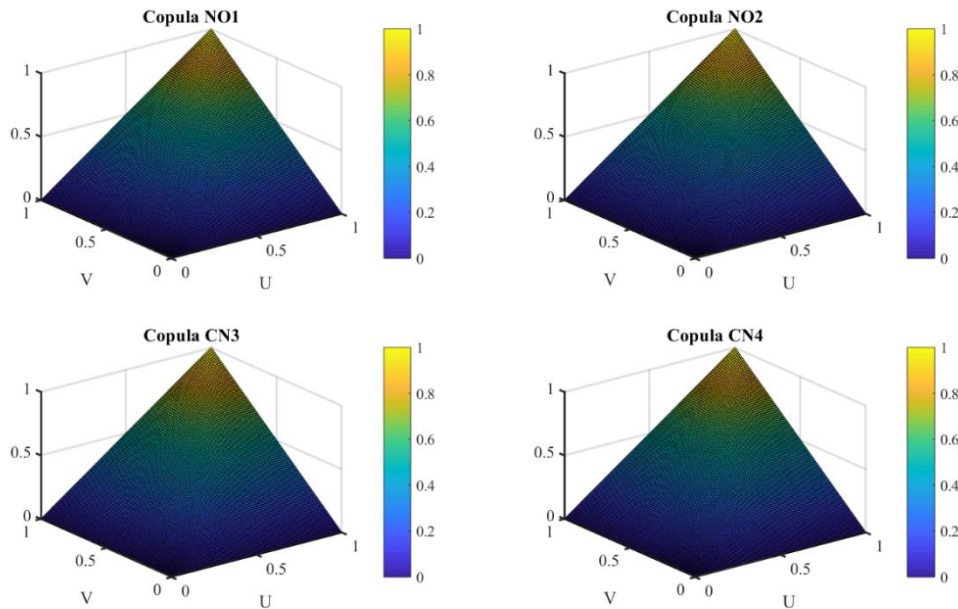


Figure 3. Empirical copulas at the four different locations studied.

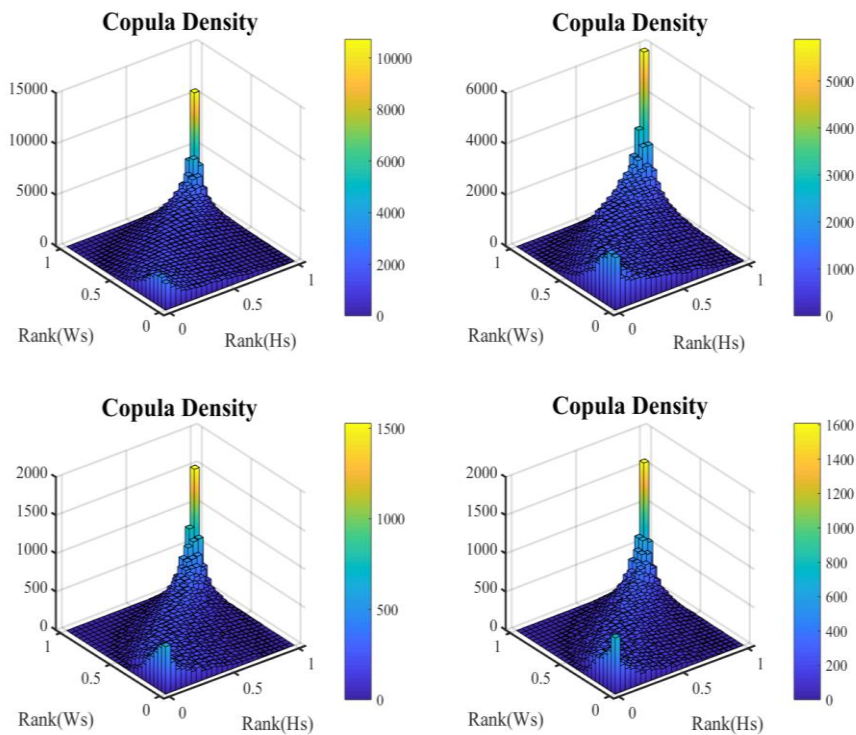


Figure 4. Copula densities at the four different locations studied

the ICDF is the inverse cumulative distribution function, and  $U$  is the cumulative distribution function of (significant) wave height  $H_s$ , and  $i, j$  are the indices of  $U$  and  $V$  (running from 1 to 100 each) in the copula domain.

### *Copula density*

The copula density inside each cell of the copula domain was estimated by a finite difference,

$$c_{i,j}(U, V) = \frac{d^2 C}{dU dV} \approx \frac{C_{i+1,j+1} - C_{i+1,j} - C_{i,j+1} + C_{i,j}}{dU \cdot dV} \quad (9)$$

The resulting densities are shown in Figure 4.

### *3.4. Univariate data discretization and bin formation*

The lumping of scatter diagrams was performed over wave heights, and the wave heights were divided into bins, with the index  $k$  representing the bin number. Bin sizes are based on two criteria:

- A bin size of 0.5[m] was chosen for wave heights that are located in the range between the mean wave height minus one standard deviation and mean wave height plus one standard deviation.
- A bin size of 1[m] was chosen for the rest of the wave heights.

The number of wave height bins varied, depending on the site (see Figure 6).

### *3.5. Using copula to generate lumped data at different locations*

A copula calculated at location NO<sub>1</sub> was combined with marginal of wave height and wind speed, measured at another location, to calculate lumped pairs of wave height and wind speed.

#### *Transfer marginal grid to copula mesh grid*

Marginal wave height bins are transferred to the [0,1] domain using the empirical cumulative distribution function. As the copula mesh grid is at a finer resolution, for the  $k$ -th bin of the marginal wave height there exist  $m(k)$  copula cells with values of the wave height inside it. A wave height representing each bin was calculated by a weighted averaging of wave heights in the cells inside each bin,

$$H_{k,j} = \frac{\left( \sum_{i=1}^{m(k)} c_{i,j} H_{cell,i,j} \right)}{\sum_{i=1}^{m(k)} c_{i,j}} \quad (10)$$

Here  $k$  is the marginal bin number,  $i$  is the cell number of copula grid in  $U$  direction,  $j$  is the cell number if copula grid in  $V$  direction, and  $m(k)$  is the number of cells inside each bin

#### *Lump wave height over wind speed*

The wind speed inside each cell of the copula grid was calculated using another finite difference approximation,

$$W_j \approx \frac{1}{2} (ICDF V(j) + ICDF V(j + 1)) \quad (11)$$

The ICDF of wind speed was based on the wind speed measurement data. The density for each marginal bin is the sum of the copula densities in the cells inside each bin

$$c_{k,j} = \sum_{i=1}^{m(k)} c_{i,j} \quad (12)$$



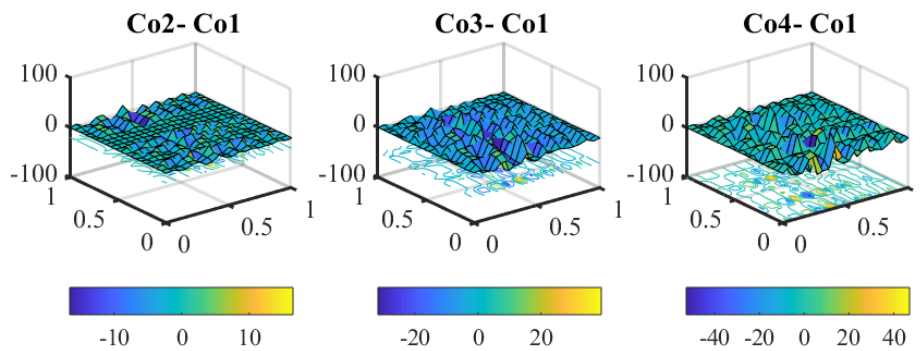


Figure 5. Copula subtraction

Finally, the lumped (conditional) wave height and wind speed for each marginal bin is calculated using weighted averages:

$$H_k = \frac{\sum_{j=1}^{100} c_{k,j} H_{k,j}}{\sum_{j=1}^{100} c_{k,j}} \quad (13)$$

$$W_k = \frac{\sum_{j=1}^{100} c_{k,j} W_k}{\sum_{j=1}^{100} c_{k,j}} \quad (14)$$

The difference between lumped real data and lumped data using the above method was quantified by the root mean square error:

$$\text{RMS} = \sqrt{\sum_k (W_{k\text{real}} - W_k)^2} \quad (15)$$

## 4. Results and discussion

### 4.1. Differences in copulas

The empirical copula that was calculated at location NO<sub>1</sub> was subtracted from the copula estimated at other locations (Figure 5). This difference between copulas quantifies the similarity of wave height and wind speed dependence at two different locations. The percentile difference of the two copulas varies spatially for different cells. The maximum difference locally reached up to 40 percent, however the average copula difference was less than 15%. The sites that were close to each other show somewhat more similar copulas than sites that were far from each other, which is attributed to being subject to a more similar, joint mechanism of wave and wind generation.

### 4.2. Lumped data using copulas

Lumped met-ocean data was generated at different sites using the empirical copula calculated at NO<sub>1</sub>. In Figure 6 the copula from NO<sub>1</sub> was combined with marginals measured at each site. The generated joint lumped data are plotted, together with the real lumped data (using the available joint measurement data). The difference between red dots and circles illustrates how well the copula from NO<sub>1</sub> can predict the joint environmental conditions at other locations of the North Sea. The blue line represents the wave height for which 99% of measurements were smaller. In other words, this line represents the upper tail of the joint data with a very low probability of occurrence. The number of circles in this region is high (relative to only 1% probability of occurrence) because the bin size is defined in uniform steps of 1[m]. Regarding the calculations inside this upper tail, the results are not as accurate, as the averages use information from only a few cells in the copula domain. Moreover, only a few values in the data determine the values of the copula. Therefore, the determination of the copula in this cell is sensitive.

The RMSE at the four stations are shown in Figure 7. In this graph the wave heights are the same for both real data and generated data. The wind speed was lumped and the difference in the lumped wind speed was calculated. The figure shows that the mean difference of the estimated lumped wind speeds with lumped real data is less than 5%, which shows a very good agreement. It should be noted that the comparison is done only for lumped data with probability of occurrence less than 99%. In other words, the upper tail is not included in this calculation as it does not have a big influence on the fatigue damage. In addition, the results suggest that the joint behaviour of wave height and wind speed can be predicted from joint data at different locations around the North Sea.

#### 4.3. Fatigue damage difference.

The damage caused by the lumped loads were calculated using equation ). The results (Figure 8) show that for three stations this method results in a mean difference of less than 5%, while for one location the mean difference is larger, with a value around 12%. The fatigue damage is highly dependent on the number of cycles, however the calculated values only reflect the relative damage for single wave and wind.

### 5. Conclusion and recommendations

High-quality met-ocean data that contains joint measurements of random variables at every location is often not available for a specific site. Having access to such data, though, is necessary for obtaining accurate designs, especially with regard to fatigue damage or when planning operation and maintenance of offshore structures. This paper uses a novel method to predict lumped wave height and wind speed distributions at different locations around the North Sea, assuming only marginal distributions are

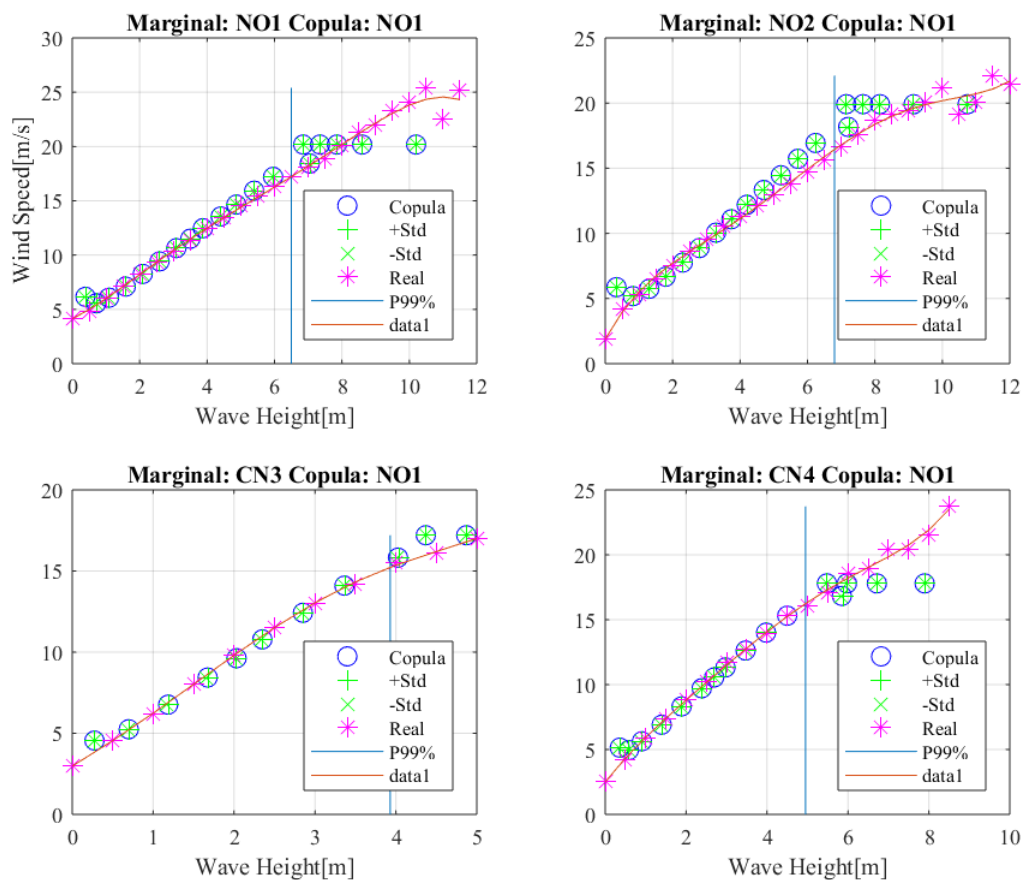


Figure 6. Copula generated lumped data and real lumped data

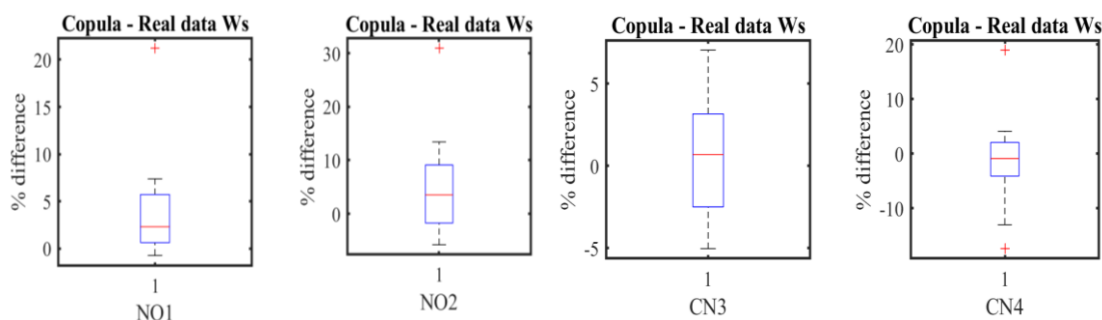


Figure 7. Percentile difference of calculated and real wind speeds

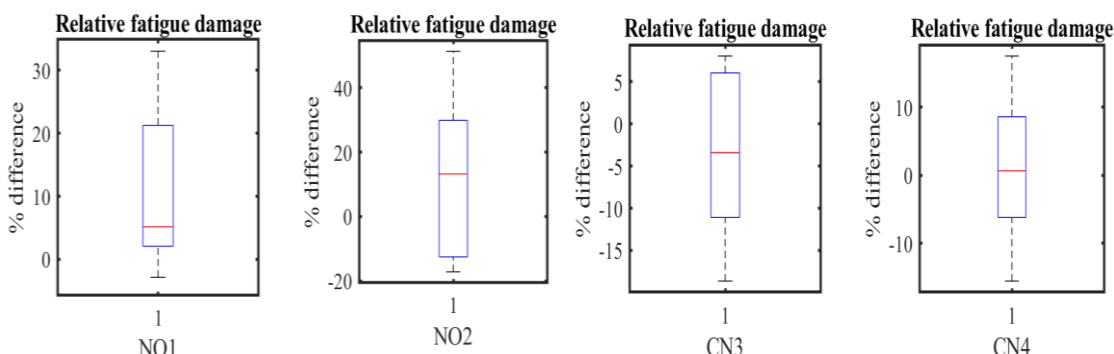


Figure 8. Fatigue damage difference between real and generated lumped data cause by wind speed

known in these locations. These marginals are combined with an empirical copula calculated at another site in the North Sea, in order to generate lumped wave heights and wind speeds. The fatigue damage of offshore wind turbines caused by these generated lumped data was estimated using a simple formula. The copula density calculated in the four sites shows that conditional wind speed and conditional wave height show different tail behaviours. In addition, stations that are located relatively close to each other have relatively similar copulas with average differences of less than 10% in copula density. An increase in the distance between measurements shows that the average copula difference is increased up to 15%. The generated lumped data was compared with lumped data from measurements. The average root mean square error was less than 5%. In this comparison the upper tail, with a probability of occurrence of less than 1%, was excluded. The fatigue damage shows an average difference of less than 5% for three locations, and an average difference of less than 15 % for another site.

The similarity of the copulas at different locations around the North Sea suggests that the joint behaviour of wind speed and wave height in the North Sea is predictable to a remarkable extent. It is recommended to conduct further studies, in order to find a family of empirical copulas that can model the joint behaviour of wind speeds and wave heights in the North Sea, possibly separately for a few different regions.

## 6. References

- [1] Leontaris G, Morales-Nápoles O, Wolfert ARM. 2016 *Ocean Engineering*; **125**: 328-41.
- [2] Bilgili M, Yasar A, Simsek E. 2011 *Renewable and Sustainable Energy Reviews*; **15(2)**: 905-15.
- [3] Ziegler L. Probabilistic estimation of fatigue loads on monopile-based offshore wind turbines: Technical University of Delft, NTNU; 2015.
- [4] Passon P. 2015 *Renewable Energy*; **81**: 723-36.
- [5] Walstra DJR, Hoekstra R, Tonnon PK, Ruessink BG. 2013 *Coastal Engineering*; **77**: 57-70.
- [6] Genest C, Favre A-C. 2007 *Journal of Hydrologic Engineering*; **12(4)**: 347-68.
- [7] Vanem E. 2016 (49941): V003T02A33.
- [8] De Michele C, Salvadori G, Passoni G, Vezzoli R. 2007 *Coastal Engineering*; **54(10)**: 734-51.

- [9] Jäger WS, Nápoles OM. 2017 *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*; **3(4)**: 04017014.
- [10] Kühn M. Dynamics and design optimisation of offshore wind energy conversion systems 2001.
- [11] Passon P, Branner K. Condensation of long-term wave climates for the fatigue design of hydrodynamically sensitive offshore wind turbine support structures 2014. 1-25 p.
- [12] Mittendorf K. Joint Description Methods of Wind and Waves for the Design of Offshore Wind Turbines 2009. 23-33 p.
- [13] Tony Burton NJ, David Sharpe, Ervin Bossanyi. Wind Energy Handbook
- [14] Seidel M. 2014 *Stahlbau*; **83(8)**: 535-41.