A SIMULATION STUDY ON THE USEFULNESS OF THE BERNSTEIN COPULA FOR STATISTICAL MODELING OF METOCEAN VARIABLES

Erik Vanem\textsuperscript{1,2,∗}, Øystein Lande\textsuperscript{3,2}, Elias Fekhari\textsuperscript{4}

\textsuperscript{1}DNV Group Research & Development, Høvik, Norway
\textsuperscript{2}University of Oslo, Oslo, Norway
\textsuperscript{3}Ocean Oasis, Oslo, Norway
\textsuperscript{4}EDF R&D, Chatou, France

ABSTRACT

Probabilistic modelling of relevant environmental variables are crucial for the safe design and operation of marine structures. Using metocean data, a joint model of several variables can be estimated, including their dependence structure. Often, a conditional model is assumed for this, but recently the non-parametric Bernstein copula has been suggested as an alternative tool to model such dependencies. As a non-parametric technique, it is very flexible and often provides excellent goodness-of-fit to data with different dependencies. However, non-parametric techniques are prone to over-fitting and generalizability might be challenging. Moreover, care should be taken when using such models for extrapolation. In this paper, a simple simulation study will be presented that has investigated the usefulness of the Bernstein copula in modeling joint metocean variables. First, data have been generated from a known parametric joint distribution model. Then, a joint model based on the Bernstein copula is fitted to a subset of these data. Data simulated from the Bernstein-based models are then compared to data from the initial model. A particular focus will be put on how the model captures the dependencies in the extremes.

Keywords: Multivariate analysis, Wind-wave environment, Offshore wind, Renewable energy, Environmental loads, Non-parametric copula

1. INTRODUCTION

Probabilistic descriptions of extreme wave and wind conditions are essential for safe and reliable design and operation of marine structures. Accumulated environmental loads over time may lead to structural fatigue and must be accounted for in the design. In addition, the most extreme environmental conditions the structure is expected to experience during its lifetime must be accounted for in ultimate limit state assessment during design. Hence, both a proper statistical description of the most likely environmental conditions as well as extreme value analysis of relevant metocean variables are crucial. Two of the most relevant wave parameters for ocean engineering applications is the significant wave height, $H_s$, and peak wave period, $T_p$. For offshore wind installations, both bottom-fixed and floating, the mean wind speed, $U$, and turbulence, $\sigma_U$, are crucial wind parameters. Hence, extreme wind and wave conditions are of great interest to the ocean engineering community [1].

The structural response to environmental loads depends on the joint occurrence of both wind and wave conditions. The extreme structural response might not necessarily occur when either environmental variable is extreme, but could just as well happen at certain combinations of conditions that are not marginally extreme. Therefore, it is important to establish multivariate probabilistic models that accounts for both the marginal behavior of each individual variable as well as the dependence between them. This is described by joint distribution models which may typically be fitted to location specific wind-wave data.

There are different ways of constructing joint distribution models for multivariate wave and wind data, and several models have been proposed previously for ocean engineering applications. In [2], a conditional modeling approach was adopted, where a multivariate joint model is constructed as a product of marginal and conditional distribution models for the different variables. This is in line with established practice in ocean engineering, and different conditional models have previously been proposed for various combinations of metocean variables, see e.g. [3–7]. Other approaches to multivariate modeling include the so-called Nataf model, i.e., to find transformations of the data and fit multivariate Gaussian distributions to the transformed data [8–10], or to fit other parametric multivariate distributions such as the multivariate log-normal [11] or Weibull [12] distributions. For multivariate extremes, the conditional extremes model [13]...
is an alternative that has also been used in ocean engineering applications [14].

Another approach is to use copulas, which allows for the marginal distributions to be fitted independently and the dependence to be modeled by way of copulas; a joint distribution can be constructed based on the marginals and the dependence structure, see e.g. [15]. Several copula-based joint distributions for metocean variables have been proposed in the literature, see e.g. [16–19]. Copula-based methods were compared to conditional models for sea state variables in [17], and were generally found to perform well. However, it was also emphasized that any conditional-based joint model can be expressed as a copula-based joint model with a particular specification of the copula function. Hence, both the conditional modeling approach and the copula approach can in principle be used to describe the exact same joint distribution. The challenge is to find suitable parametric models.

Even though there are many methods for establishing parametric joint distribution models for wind and wave conditions for design and assessment of marine structures, finding good models remains a challenge and joint distribution fitting remains an active area of research. In particular, as the dimension grows, it will become increasingly challenging to define accurate multivariate distributions, and some simplifications may be needed in the modeling. A useful approach for establishing parametric copula-based models in high dimensions is to combine pairwise copula constructions into high-dimensional multivariate distributions in so-called vine copulas [20], see also e.g. [21, 22]. However, defining an appropriate tree structure remains a challenge for high-dimensional problems. For the conditional modeling approach, assumptions on conditional independence between some of the variables similarly simplifies the modeling.

As an alternative to parametric models, non-parametric approaches such as kernel density-based methods have been proposed to model multivariate environmental data, see e.g. [23, 24]. Such non-parametric models would typically give very good fit to the data, but it is questionable how good such models would generalize and extrapolate outside the support of the data. Hence, such methods should be used with particular care if the interest is in extremes or rare environmental conditions. Another drawback of multivariate kernel density estimators is that they generally don’t scale well, especially for higher dimensions. A semi-parametric approach is to combine parametric (or non-parametric) distributions for the marginals with a non-parametric copula such as the Bernstein copula [25]. Environmental contours based on non-parametric copulas are presented in [26].

In this paper, a simple simulation study that investigates the usefulness of the non-parametric Bernstein copula for modeling joint distributions of metocean variables is reported. Based on data simulated from a known parametric joint distribution -- the one presented in [2] -- a semi-parametric model assuming parametric models for each marginal and a Bernstein copula for the dependencies will be established. This semi-parametric model is then compared to the initial joint distribution and in particular data simulated from both models may be compared. In particular, how well the non-parametric Bernstein copula captures the tail dependencies between the variables will be explored.

2. THE BERNSTEIN COPULA

2.1 Joint modeling with copulas

A d-dimensional copula \( C : [0, 1]^d \mapsto [0, 1] \) is a multivariate cumulative distribution function whose variables are each uniformly distributed on \([0, 1]\). This function is often used to model the dependence between random variables. In fact, any multivariate joint distribution can be written as an association of univariate marginal distribution functions and a d-dimensional copula describing the dependence structure between the variables, according to Sklar’s theorem [27]. If \( H \) is a joint distribution for a random vector \( X \in \mathbb{R}^d \) with marginals \( \{F_j\}_{j=1}^d \), then there exist a copula \( C \) such that, for all \( x \in \mathbb{R}^d \),

\[
H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).
\]

If the marginals CDF \( \{F_j\}_{j=1}^d \) are continuous, then \( C \) is unique. Conversely, if \( C \) is a copula and \( \{F_j\}_{j=1}^d \) are the marginal CDFs, then the function \( H \) defined by eq. (1) is a joint distribution function with marginals \( \{F_j\}_{j=1}^d \).

One important consequence of Sklar’s theorem is that any joint probability density can be written as the product of the marginal probability densities and the copula density, \( c \), i.e.,

\[
h((x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d) c(F_1(x_1), \ldots, F_d(x_d))
\]

Hence, if the copula is differentiable, the joint density function can be estimated from estimates of the marginal distributions and the copula. It can easily be seen that if the copula density is 1, then the variables are independent. Otherwise, the dependence is described by the copula density which is also sometimes referred to as the dependence function. There is a plethora of parametric copula-models that can be used to fit joint distribution models to multivariate data by estimating the copula parameters. For a more thorough introduction to copulas reference is given to textbooks such as [15, 28]; see also e.g. [29].

2.2 Empirical Bernstein copula

When inferring a multivariate distribution from a dataset, Sklar’s theorem allows to infer the marginals and the copula separately. Even if a wide catalogue of parametric copulas exists for parametric inference, they offer a limited flexibility. Alternatively, nonparametric copula are built as a functional approximation of an empirical copula \( C_n \) associated with a dataset \( X_n = \{x^{(1)}, \ldots, x^{(n)}\} \sim X \). Note that the empirical copula density \( c_n \) associated with the sample \( X_n \) is given by the componentwise normalized ranks \( R_n = \{r^{(1)}, \ldots, r^{(n)}\} \) of \( X_n \). Applying strictly increasing transforms does not modify copulas, which is the case of the rank transform (see e.g., [30] p.57).

Bernstein polynomials have been introduced to prove that any continuous and real-valued function defined on a compact set can be uniformly approximated as closely as desired by a polynomial (Weierstrass theorem). Therefore, they were good candidates to approximate unknown copulas on the unit hypercube. This concept was introduced as the empirical Bernstein copula (EBC) by [25].

Let us define a regular discretization of the unit hypercube

\[
G = \left\{ \frac{0}{m}, \ldots, \frac{m}{m} \right\} \times \cdots \times \left\{ \frac{0}{m}, \ldots, \frac{m}{m} \right\}, \quad m \in \mathbb{N}.
\]

The Bernstein
approximation $B_m(\cdot)$ of the empirical copula $C_n$ is then given by:

$$B_m(C_n)(\mathbf{u}) = \sum_{t_1=0}^{m} \cdots \sum_{t_d=0}^{m} C_n\left(\frac{t_1}{m}, \ldots, \frac{t_d}{m}\right) \prod_{j=1}^{d} b_{m,t_j}(u_j),$$

$$\mathbf{u} = (u_1, \ldots, u_d) \in [0, 1]^d, \quad (3)$$

where $b_{m,t}(u) = \binom{m}{t} u^t (1 - u)^{m-t}$, $t \in \{0, \ldots, m\}$ is a Bernstein polynomials of order $m$. The evaluations of the empirical copula over the regular grid $G$ are smoothed by the Bernstein polynomials. Interestingly, [31] remarks that this method can also be seen as a mixture of Beta distributions. An EBC is a genuine copula if and only if $m$ is a divisor of $n$ (see the demonstration in [31]).

Asymptotically, the EBC $B_m(C_n)$ converges towards the copula $C$ under some conditions over $m$ (see [25]). However, the choice of the polynomial order leads to a bias-variance trade-off. For example, increasing the polynomial order can cause an overfit of the empirical copula which induces a bias. Theoretical tuning rules were proposed by [32–34] and more recently for small samples by [31] (setting $m = n$).

The empirical Bernstein copula (EBC) is implemented in OpenTURNS, a Python package for the treatment of uncertainties [35], which has been used to fit the Bernstein copula and subsequently simulate from it in this study. To ensure that the fitted copula is genuine, this implementation drops samples until $m$ becomes a divisor of $n$.

3. THE METOEAN DATA

This study assumes a known parametric joint distribution for several metocean variables. Considering a known model rather than simply raw data means that one knows the true underlying distribution and is therefore allowed to extrapolate. One may simulate data corresponding to a certain time period that may be construed as actual observations, and that will be used to fit the semi-parametric models. In addition, one may simulate an extended period of time for comparison with the fitted models. For the purpose of this study, the joint distribution model proposed in [2] will be assumed, see also [36] for more details. This model was initially fitted to 32 years of data For a location outside South Brittany from the ANEMOC atlas\footnote{URL: http://anemoc.cetmef.developpement-durable.gouv.fr/}. The following metocean variables were included in the joint model, at hourly resolution:

- Wind direction ($\theta$)
- Mean wind speed ($U$)
- Turbulence ($\sigma_U$)
- Significant wave height ($H_s$)
- Wave direction ($\beta$)
- Peak wave period ($T_p$)

An initial dataset corresponding to 1000 years of data are simulated from the known joint distribution. Pairwise scatter plots of the simulated data (all 1000 years of data) are shown in Figure 1. In addition, pairwise environmental contours are shown in the lower panels, corresponding to 1-, 10- and 50-year extreme conditions. These contours are based on the direct sampling approach [37, 38]. Together, these scatter plots and environmental contours are assumed to describe the true data well, and in the following it will be explored how well a semi-parametric Bernstein-based joint model are able to describe this. Some summary statistics and the correlation matrix are presented in table 1.

![FIGURE 1: PAIRWISE SCATTER PlOTS AND ENVIRONMENTAL CONTOURS OF THE WIND AND WAVE DATA](image)

| TABLE 1: SUMMARY STATISTICS OF THE WIND AND WAVE DATA |
|---------------------------------|---------------------------------|----------------|----------------|----------------|----------------|
|                                 | WSP (m/s) | WaveDir (°) | Turbulence (m/s) | $H_s$ (m) | $T_p$ (s) | WindDir (°) |
| min                             | 0.004     | 0           | 0.147            | 0.345     | 2.786       | 0             |
| max                             | 37.98     | 360         | 4.738            | 16.87     | 58.52       | 360           |
| mean                            | 10.17     | 116.6       | 1.599            | 2.346     | 13.28       | 184.4         |
| sd                              | 4.843     | 132.3       | 0.545            | 1.343     | 3.221       | 114.1         |
| skewness                        | 0.635     | 0.936       | 0.535            | 1.420     | 0.875       | -0.124        |

| TABLE 1: SUMMARY STATISTICS OF THE WIND AND WAVE DATA |
|---------------------------------|---------------------------------|----------------|----------------|----------------|----------------|
|                                 | WSP (m/s) | WaveDir (°) | Turbulence (m/s) | $H_s$ (m) | $T_p$ (s) | WindDir (°) |
| min                             | 0.004     | 0           | 0.147            | 0.345     | 2.786       | 0             |
| max                             | 37.98     | 360         | 4.738            | 16.87     | 58.52       | 360           |
| mean                            | 10.17     | 116.6       | 1.599            | 2.346     | 13.28       | 184.4         |
| sd                              | 4.843     | 132.3       | 0.545            | 1.343     | 3.221       | 114.1         |
| skewness                        | 0.635     | 0.936       | 0.535            | 1.420     | 0.875       | -0.124        |

3.1 Fitted Marginal Distributions

Data corresponding to 1 year of data are assumed to constitute the observations and will be used to fit the Bernstein copula in
this study. Hence, data corresponding to 1 year are extracted from the 100 years of simulated data and used for model fitting. This is done in order to take sampling variability and the fact that only finite amount of data will be available to fit distributions in actual applications. However, in this study, the interest is in how well the Bernstein copula describes the dependencies in the data, and the marginals are not of equal interest. Thus, a set of marginal models are fitted to the data for all 1000 years and all subsequent comparisons are shown in the original parameter space to ease the interpretation of results. The fitted marginal models for each variable are shown in Figure 2.

It is noted that the following distribution families are assumed for the various marginal distributions: 3-parameter Weibull for $H_s$, Weibull for $U$, log-normal for $\sigma_U$ and $T_F$ and mixtures of von Mises distributions for $\beta$ and $\theta$. Not also that the effect of misspecifying the marginals on the subsequent estimation of copulas is discussed in [39].

4. FITTING THE BERNSTEIN COPULA TO THE METEOCEAN DATA

Having established the Bernstein-based joint distributions, three different Bernstein-copula models (or rather, models with three different tunings) are fitted to these pseudo-observations in this study, i.e., with hyper-parameters (number of bins) $N_b = 50, 500$ and 5000, respectively. These are henceforth referred to as Bernstein model 1 ($N_b = 50$), model 2 ($N_b = 500$) and model 3 ($N_b = 5000$), respectively. Intuitively, increasing the number of bins will give a better fit to the sample (with the risk of over-fitting) and will also be more computationally heavy.

5. EVALUATING THE FITTED DISTRIBUTIONS

Having established the Bernstein-based joint distributions, one wants to compare how well they describe the initial distribution. In this simulation exercise, this will be done by comparing data simulated from the known true distribution and from the different Bernstein-based models. 1000 years of data will be simulated from each distribution model and compared. Note that the Bernstein-based models were fitted to 1 year of data, meaning that the models needs to be extrapolated. Note also that the different models assume the same marginal distributions, so the differences will stem from the different models for the dependencies in the data.

5.1 Comparing Simulated Data

Having fitted the semi-parametric distributions consisting of the marginal fits and the non-parametric Bernstein copulas, one may simulate as much data as desired from them. Hence, the 3 Bernstein-based models are used to simulate 1000 years of data from each of the models. Pairwise scatter plots and environmental contours based on these simulated data are shown in Figures 3 – 5. These are directly comparable to similar plots for the simulated observations in Figure 1.

It is observed that the simulated data somewhat resembles the initial observations, and that increasing the number of bins to use when fitting the Bernstein copula appears to improve the fit. However, comparison is not straightforward from simply considering the scatter plots and environmental contours. Another visual technique to compare multivariate distributions is to plot scale curves. These are based on data depth and visualize the volume of convex hulls containing subsequent central regions, see [40] for details. Such scale curves are shown in Figure 6 for data from the known model and the three alternative Bernstein-based models. Interestingly, these scale curves suggest that the second Bernstein model fits best to the data, indicating that perhaps the most complicated Bernstein model is over-fitting to the sample. This information may also be illustrated in scale-scale plots [41], as shown in Figure 7, where a perfect fit would correspond to the black dotted line. It is observed that all models deviate somewhat from this line, but that the best fit is the second Bernstein-based model.

There are several more formal ways to compare multivariate distributions, e.g. based on e.g. kernel distances or permutation tests [42–46], but these may not work well if the interest is on the tail behavior. Non-parametric models will typically fit very well to the data where there is much data, but in order to explore how well such models generalize and extrapolate, one needs to look specifically at the tails and tail-behavior. In the following,
FIGURE 3: SIMULATED DATA FROM BERNSTEIN MODEL 1 (WITH $N_b = 50$); PAIRWISE SCATTER PLOTS AND ENVIRONMENTAL CONTOURS

FIGURE 4: SIMULATED DATA FROM BERNSTEIN MODEL 2 (WITH $N_b = 500$); PAIRWISE SCATTER PLOTS AND ENVIRONMENTAL CONTOURS

FIGURE 5: SIMULATED DATA FROM BERNSTEIN MODEL 3 (WITH $N_b = 5000$); PAIRWISE SCATTER PLOTS AND ENVIRONMENTAL CONTOURS

FIGURE 6: COMPARING SCALE CURVES FOR THE DIFFERENT MODELS
a closer look at the tail dependencies from the different models will be made.

5.2 Tail Dependencies

In this study we compare the pairwise strong lower and upper tail dependence indices for all pairs of variables in order to investigate how well non-parametric distributions based on the Bernstein copula can reproduce observed data. That is, we compare different bivariate subsets of the multivariate data. Note that tail behavior is not very meaningful for circular variables, so the variables wind and wave directions are not included in these comparisons.

The bivariate lower and upper strong tail dependence index for variables \((X, Y)\), with cumulative distribution functions \(F_X(x)\) and \(F_Y(y)\), are defined as follows (see e.g. [28, 47]):

\[
\begin{aligned}
\lambda_l &= \lim_{u \to 0} \frac{P(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(u))}{P(X \leq F_X^{-1}(u))} \\
\lambda_r &= \lim_{u \to 1} \frac{P(X > F_X^{-1}(u), Y > F_Y^{-1}(u))}{P(X > F_X^{-1}(u))} \\
\end{aligned}
\]

In order to visualize and compare this limiting behavior, one may plot these as a function of \(u\), i.e. to plot the functions \(L(u)\) and \(R(u)\) for different distributions in order to visually compare the tail dependencies, where

\[
\begin{aligned}
L(u) &= \frac{P(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(u))}{P(X \leq F_X^{-1}(u))} \\
R(u) &= \frac{P(X > F_X^{-1}(u), Y > F_Y^{-1}(u))}{P(X > F_X^{-1}(u))}.
\end{aligned}
\]

The empirical versions of these can be estimated based on the empirical distribution function, \(\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \leq x)\), where \(\mathbb{1}\) is the indicator function and \(n\) is the sample size, and hence be calculated from a dataset (see also [48]). Hence, the following empirical dependence functions will be calculated and visualized based on simulated data from the known true model and the different models based on the Bernstein copula,

\[
\begin{aligned}
\bar{L}_n(u) &= \frac{\sum_{i=1}^{n} \mathbb{1}(x_i \leq \hat{F}_X^{-1}(u), y_i \leq \hat{F}_Y^{-1}(u))}{\sum_{i=1}^{n} \mathbb{1}(x_i \leq \hat{F}_X^{-1}(u))} \\
\bar{R}_n(u) &= \frac{\sum_{i=1}^{n} \mathbb{1}(x_i > \hat{F}_X^{-1}(u), y_i > \hat{F}_Y^{-1}(u))}{\sum_{i=1}^{n} \mathbb{1}(x_i > \hat{F}_X^{-1}(u))}.
\end{aligned}
\]

The lower tail will be plotted for \(u \in (0, 0.5)\) and the upper tail for \(u \in (0.5, 1)\).

An example of such dependence functions is shown in Figure 8 for the variable-pair wind speed and turbulence. The black line corresponds to the known true distribution, colored lines correspond to alternative fitted models based on Bernstein copula; red = Bernstein copula 1 \((N_b = 50)\), green = Bernstein copula 2 \((N_b = 500)\) and blue = Bernstein copula 3 \((N_b = 5000)\). From this plot it is observed that the fitted Bernstein-based models capture the sub-asymptotic tail dependence reasonably well, although the simplest model (with \(N_b = 50\)) significantly underestimates the tail dependence. Similar pairwise plots are shown for all the linear variables in Figure 9. Again, it appears that the tail dependencies are reasonably well captured by the Bernstein-copula, and with increasingly better fit for increased bin size.

However, in order to have a closer look at the tail dependencies towards the far tails, similar plots can be made on logarithmic scale. This is shown in Figure 10. Now, it can clearly be seen...
that for some of the variables, the Bernstein-based joint models are not able to fully capture the dependencies in the far tails of the distributions. Most notable are the deviations in both upper and lower tail for the wind speed and turbulence variables and in the upper tails for the wind speed, turbulence and significant wave height variables. For the spectral wave period variable, $T_p$, it appears that this is asymptotically independent with the other variables, with negligible tail dependence for $u > 0.9$, and in this case, the Bernstein copulas work well.

It is also easily seen from these plots that as the number of bins used to estimate the Bernstein copula increases, the tail dependence is better described by the Bernstein copula. However, the extreme tail dependence is under-estimated, and this can be a reason for concern if the distribution model is used for multivariate extreme value analysis and assessment of extreme responses of marine structures.

An alternative and related approach to evaluate the usefulness of the empirical Bernstein copula could be to consider the pairwise conditional expected value of one variable, conditioned on the other being large, i.e. $E[Y|(X > u)]$ for $u$ large, see also [39].

6. DISCUSSION

The simple simulation study presented above has illustrated that 1) the Bernstein copula is generally able to capture dependence structures in metocean data quite well, but that 2) it may fail to capture the extreme tail behavior. In this study, the lower and upper tail dependencies have been studied, and it is observed that for sufficiently high quantiles, the tail dependencies might not be captured. This is because the Bernstein copula in particular, and non-parametric approaches in general, fail to generalize and extrapolate well. Hence this is as expected. It is observed that a better fit can generally be obtained by increasing the number of bins used for fitting the Bernstein copula. In this study, three alternative models with $N_b = 50, 500$ and $5000$, respectively, is tested resulting in models with quite large number of bins. However, using a too high bin number might lead to over-fitting, and in this exercise, it appears that the Bernstein copula with the second highest bin number fit better than the one with the highest number of bins, according to scale curves. The optimal choice of this hyper-parameter would be a trade-off between goodness-of-fit to the sample, generalizability and computational efficiency. The different models fail to capture the tail dependencies at different quantile levels, with the simplest models failing first. At what quantile level this occurs will obviously also depend on the amount of data available and, in principle, one way to remedy this could be to collect more data so the need for extrapolation is minimized. However, in practice this is most often not possible.

In this study, only the lower and upper tail dependencies have been studied. However, in a multivariate setting, there will be omnidirectional tails, and there will e.g. be tail dependencies when one variable is large and another is small. This has not been investigated in this study, but it is assumed that the results for the lower and upper tails are representative for any directional tails. How well the Bernstein copula captures tail dependence in arbitrary directions could be further investigated in future research. It is also noted that in this study, only pairwise tail dependencies have been compared, and the picture might be even more complicated if the multivariate tails are considered jointly. Nevertheless, it is believed that the pairwise comparisons yields a good indication of the way the Bernstein copula performs in the extremes.

The plots in Figure 10 indicate that all pairwise variables might be asymptotically independent, but with varying degree of dependence in sub-asymptotic regions. It is not entirely clear if
this is a property of the joint distribution model or if it is a consequence of a finite sample and limited data in the tails. This is an interesting topic for further research, and it could be explored how well the empirical Bernstein copula is able to capture tail dependencies for distributions that are truly asymptotically dependent. Similar simulation studies could be carried out with simulated data from models with known strong asymptotic dependencies in order to investigate this.

The joint model and multivariate data analyzed in this paper, contains a combination of circular and linear variables. It should be noted that circular variables will not have any upper and lower tails, and therefore would presumably be described very well by non-parametric techniques such as the Bernstein copula. Hence, it could be recommended that Bernstein copula is a good alternative for modelling circular variables. If, for example, a joint model is made based on a pairwise copula construction [20], then it could be appropriate to assume a Bernstein copula for all cyclic variables, but to seek other models for the linear variables where there is a need for extrapolation.

[49] commented on the use of empirical Bernstein copulas for multivariate extremes, and suggested that bivariate linear B-splines might be more appropriate and that it might be better to transform to a different scale before evaluating the empirical Bernstein copula. These are interesting topics for further research.

It is also noted that the appropriateness of the Bernstein copula is assumed to be heavily application-dependent. For example, in fatigue assessment or wind potential estimation, where it is more important to describe frequent non-extreme conditions well than very rare extreme conditions, Bernstein copula based models may be an excellent choice. On the other hand, for extreme value analysis and ultimate limit state assessment, where the main concern is on extreme conditions and rare events, then Bernstein copula based models might not be adequate. Hence, it is recommended to carefully assess this on a case-by-case basis. In either case, it is believed that the simulation study presented in this paper provides some useful insight on the usefulness of the Bernstein copula for modeling metocean data.

7. SUMMARY AND CONCLUSIONS

This paper has presented a simple simulation exercise to evaluate how appropriate the non-parametric Bernstein copula is for modeling the joint behavior of several metocean variables often used in design and assessment of marine structures. Starting with data from a known multivariate distribution, joint models are established based on the Bernstein copula and compared to the true model. The results indicate that the Bernstein copula might give excellent fit to the data. However, if the model should be used for extrapolation and for extreme value analysis where the tail dependencies are important, such models should be used with caution. Although the dependencies are well captured for moderately high quantile levels, the models might fail to capture the tail dependencies in the far tails of the distributions. Hence, the usefulness of the Bernstein copula will depend on the application and should be assessed on a case-by-case basis. For applications where the far tails are not very important, e.g. in fatigue assessment, the Bernstein copula might be a very useful tool for describing the joint probabilistic behavior of several metocean variables. In applications where such tail dependencies are important, however, it is recommended to use Bernstein copulas with care.

Acknowledgement

The work presented in this paper has been carried out within the research project HIPERWIND, with support from EU H2020 (grant agreement number 101006689). The data has been obtained from the ANEMOC Digital Atlas of Ocean and Coastal Sea States.

REFERENCES


