A Vine-Copula Model for Time Series of Significant Wave Heights and Mean Zero-Crossing Periods in the North Sea

Jäger, Wiebke; Morales Napoles, Oswaldo

DOI
10.1061/AJRUA6.0000917

Publication date
2017

Document Version
Accepted author manuscript

Published in
ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering

Citation (APA)
Jäger, W., & Morales Napoles, O. (2017). A Vine-Copula Model for Time Series of Significant Wave Heights and Mean Zero-Crossing Periods in the North Sea. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *3*(4), Article 04017014. https://doi.org/10.1061/AJRUA6.0000917

This is an Accepted Manuscript of an article published by ASCE library
ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems,
Part A: Civil Engineering
Vol. 3, Issue 4 (December 2017)
See more at: https://doi.org/10.1061/AJRUA6.0000917

A Vine-Copula Model for Time Series of Significant Wave Heights and Mean Zero-Crossing Periods in the North Sea

W.S. Jäger[1] and O. Morales Nápoles, Ph.D.[2]

[1]Ph.D. Candidate, Department of Hydraulic Engineering, Delft University of Technology, Stevinweg 1, 2600 GA Delft, The Netherlands. (Corresponding author). Email: w.s.jager@tudelft.nl
[2]Assistant Professor, Department of Hydraulic Engineering, Delft University of Technology. Stevinweg 1, 2600 GA Delft, The Netherlands. Email: o.moralesnapoles@tudelft.nl

Keywords: Hydraulic boundary conditions; synthetic storms; coastal risk; bivariate time series

ABSTRACT

Stochastic description and simulation of oceanographic variables are essential for coastal and marine engineering applications. In the past decade, copula-based approaches have become increasingly popular to estimate the multivariate distribution of some variables at the peak of a storm along with its duration. The modeling of the storm shape, which contributes to its impact, is often simplified. This article proposes a vine-copula approach to characterize hourly significant wave heights and corresponding mean zero-crossing periods as a random process in time. The model is applied to a data set in the North Sea and time series with the duration of an oceanographic winter are simulated. The synthetic wave scenarios emulate storms as well as daily conditions. The results are for example useful as input for coastal risk analyses and for planning offshore operations. Nonetheless, selecting a vine structure, finding appropriate copula families and estimating parameters is not straightforward. The validity of the model, as well as the conclusions that can be drawn from it, are sensitive to these choices. A valuable by-product of the proposed vine-copula approach is the bivariate distribution of significant wave heights and corresponding mean zero-crossing periods at the given location. Its dependence structure is approximated by the flexible skew-t copula family and preserves the limiting wave steepness condition.

INTRODUCTION

Wind-induced sea waves affect coastlines, marine structures and offshore operations. Unfavorable conditions can cause significant morphological change, damages or down-time. Intuitively, the higher a wave, the more energy it carries and the more destructive it is. At beaches and dunes, which act as defenses to coastal developments, high waves with long periods lead to higher run-up and intensify erosion rates (van Gent et al. 2008). Wave overtopping is also problematic at non-sandy coasts. Moreover, extreme wave heights and periods can contribute to structural failure of marine structures. In particular offshore, wave periods close to the resonant heave period of a vessel or a structure pose an additional threat (e.g., Faltinsen 1993).

Wave heights and periods are strongly interrelated, usually arising from a common meteorological system. In addition, wave heights are limited by their associated periods in terms of a so-called maximum steepness condition, which postulates that too steep waves break and reduce in height. Understanding the joint behavior of wave heights and periods and being able to estimate possible extreme conditions is important, for instance, to determine design criteria and for risk analyses of marine structures and coastal environments (Hawkes et al. 2002; Salvadori et al. 2014; Gouldby et al. 2014) or for scheduling and budgeting offshore operations (Leontaris et al. 2016). Instead of considering individual waves, one generally uses statistics that describe the sea state under stationary conditions. For example, in this article, we concentrate on the significant wave height, H_{m0}, and the mean zero-crossing period, T_{m02}, which are computed from the zeroth- and second-order moments of the variance density spectrum of a wave record (Holthuijsen 2010).

Multiple studies have been dedicated to modeling the dependencies between oceanographic variables. Most common are analyses of maxima or peak over threshold values to model storms, with variables of interest being, for instance, significant wave heights, peak wave periods and water levels.

A popular approach is based on copulas, which isolate the marginal properties from the dependence structure of random variables (e.g., Genest and Favre 2007 for an introduction). A combination of any copula with any marginal distribution leads to a valid specification of a joint distribution, enabling representations of a wide range of complex multivariate behaviors. In the bivariate case, many different copula families have proven to be useful (e.g., Salvadori et al. 2014; Masina et al. 2015, and references therein). For more than two oceanographic variables, nested (also called hierarchical) Archimedean copulas (Wahl et al. 2012; Corbella and Stretch 2013; Lin-Ye et al. 2016) and elliptical copulas, such as Gaussian or t, (Li et al. 2014a; Wahl et al. 2016; Rueda et al. 2016) have been implemented and found valuable, but also dependence trees (Poelhekke et al. 2016) and vines (De Michele et al. 2007; Montes-Iturrizaga and Heredia-Zavoni 2016), which are a generalization thereof, have been proposed.

Callaghan et al. (2008) and Serafin and Ruggiero (2014) adopt two other approaches to model dependencies. They use a bivariate logistics model (Tawn 1988) and they specify parameters for a conditional distribution of one variable based on the value of the conditioning variable.

Not only the dependencies between variables are important for impact assessment, but also their temporal evolution; impacts amass during long-lasting or recurring extreme conditions (e.g., Karunarathna et al. 2014, and references therein for impacts on a sandy beach). Storm sequences (i.e., time series of storm events) have been modeled as different types of renewal processes (De Michele et al. 2007; Callaghan et al. 2008; Corbella and Stretch 2013; Li et al. 2014b; Wahl et al. 2016) and storm shapes are

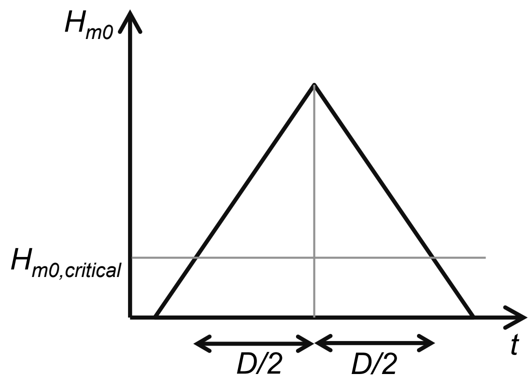


Fig. 1. Illustration of a "triangular equivalent storm" in terms of significant wave height (H_{m0}) and storm duration (D).

often approximated by a triangle (Fig. 1) whose size is determined by a peak value and a critical threshold value of, for example, wave height or water level together with the storm duration (e.g., Boccotti 2000). For instance, Corbella and Stretch (2012) and Poelhekke et al. (2016) have used these triangles and the dependencies between variables at the peak to derive idealized storm time series with high resolution (~1h) to force numerical, physics-based models that compute resulting erosion and flooding. Differently, Wahl et al. (2011) apply linear regression to parameterize and simulate the temporal evolution of total water levels during storm surges.

In this article, we explore a vine-copula approach to represent bivariate time series of significant wave heights and mean zero-crossing periods. By modeling the two variables as a stochastic process in time, we do not need to make explicit assumptions on shape, duration and inter-arrival time of a storm. We model only one time step explicitly, but simulate time series of arbitrary lengths by assuming stationarity and a Markov property of order 1. The time series may, for instance, serve as input for coastal hazard models or offshore operations.

Our aim is three-fold: (1) to present an overview of vine-copula models for time series, (2) to point out useful copula families for the instantaneous relationship of significant wave height and mean zero-crossing period, which safeguard the physical maximum wave steepness, and (3) to set forth a methodology for parsimoniously representing sea waves as a stochastic process.

Vine-copulas have recently been suggested for time series modeling in the financial field, for energy research and in the social sciences (Smith et al. 2012; Smith 2015; Brechmann and Czado 2015; Nai Ruscone and Osmetti 2017). Dependence trees, the simplest form of a vine-copula, have been applied for environmental time series before: in a preliminary study for this article (Jäger and Morales Nápoles 2015) and to compute workability windows for the installation of offshore wind farms based on wind speeds and significant wave heights (Leontaris et al. 2016). The dependencies between wave height and period statistics have also received attention before. Repko et al. (2004) compared a physics-based model to different statistical models for analyses of extremes and Vanem

(2016) proposed the construction of skewed copulas by extra-parameterization.

The remainder of the article is structured as follows. The next section provides an overview on vine-copulas, in general and in particular on time series modeling. Section 3 describes the data set and the preprocessing that has been carried out prior to the analysis. In sections 4 and 5 the results of the marginal and dependence analyses are shown and discussed. Section 6 presents the main conclusions.

METHOD

Preliminaries on Copulas and Vines

A copula is a specific type of joint distribution function that fully characterizes the joint dependence between random variables, separately from their respective marginal behaviors.

Definition 1 *For a n-variate distribution function F with univariate margins* $F_1, \dots, F_n$*, the* **copula** *associated with F is a distribution function* $C : [0,1]^n \to [0,1]$ *with uniform margins on* $[0,1]$ *that satisfies*

$$F(\mathbf{x}) = C(F_1(x_1), \dots, F_n(x_n)), \qquad \mathbf{x} \in \mathbb{R}^n. \tag{1}$$

Theorem 1 *(Sklar 1959)*
If F is a continuous n-variate distribution function with univariate margins $F_1, \dots, F_n$ *and quantile functions* $F_1^{-1}, \dots, F_n^{-1}$*, then the copula*

$$C(\mathbf{u}) = F(F^{-1}(u_1), \dots, F_n^{-1}(u_n)), \qquad \mathbf{u} \in [0,1]^n, \tag{2}$$

is unique.

A valid parametric model for F arises when $F_1, F_2, \dots, F_n$ and C are chosen from appropriate parametric families of distributions. In the bivariate case, $n = 2$, many parametric families have been proposed, covering a wide range of dependence structures. Joe (2014) and Nelsen (2013) provide comprehensive overviews.

Constructing higher dimensional families of copulas has proven to be difficult and existing models, for example, multivariate elliptical or Archimedean copulas, can be too restrictive for many applications. Montes-Iturrizaga and Heredia-Zavoni (2016) discuss this with respect to environmental variables. A more flexible approach to modeling multivariate dependencies can be taken with a graphical model called R-vine (Joe 1996; Cooke 1997; Bedford and Cooke 2002; Aas et al. 2009).

Definition 2 *Vine, R-vine (Kurowicka and Cooke 2006)*
$\mathcal{V}$ *is a* **vine** *on n elements if*

1. $\mathcal{V} = (T_1, \dots, T_{n-1})$
2. T_1 *is a connected tree with nodes* $N_1 = \{1, \dots, n-1\}$*, and edges* E_1*; for* $i = 2, \dots, n-2$*,* T_i *is a connected tree with nodes* $N_i = E_{i-1}$*.*

and $\mathcal{V}$ *is a* **regular vine***, or* **R-vine***, on n elements if additionally*

3 (**proximity**) *For $i = 2, ..., n-2$, if $\{a, b\} \in E_i$, then $\# a \Delta b = 2$, where Δ denotes the symmetric difference. In other words if a and b are nodes of T_i connected by an edge in T_i, where $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i.*

A special class of R-vines that are considered in this article are drawable vines, or D-vines, for which the maximal number of edges attached to any node in the first tree is 2. Loosely stated, the nodes reachable from a given edge in a R-vine are called the *constraint set* of that edge. When two edges are joined by an edge in tree T_i, the intersection of the respective constraint sets form the *conditioning set*. The symmetric difference of the constraint sets form the *conditioned* set. R-vines can be used to specify a joint density through a decomposition into univariate densities and (conditional) bivariate copulas:

Theorem 2 *R-Vine density (Kurowicka and Cooke 2006)*
Let $\mathcal{V} = (T_1, ..., T_{n-1})$ be an R-vine on n elements. For each edge $e(j,k) \in T_i$, $i = 1, ..., n-1$ with conditioned set $\{j, k\}$ and conditioning set D_e, let the conditional copula and copula density be $C_{jk|D_e}$ and $c_{jk|D_e}$. Let the marginal distributions F_i, $i = 1, ..., n$ be given. Then the vine-dependent distribution is uniquely determined and has a density given by

$$f_{1\ldots n} = f_1 \cdots f_n \prod_{i=1}^{n} \prod_{e(j,k) \in E_i} c_{jk|D_e}(F_{j|D_e}, F_{k|D_e}). \tag{3}$$

For a given R-vine this density is unique. The product on the right hand side contains $n(n-1)/2$ copulas and conditional copulas, which is the exact number of ways in which n elements can be coupled. This property is one reason why a vine-copula is more flexible than a fully nested Archimedean copula. Only $n-1$ bivariate margins can be modeled distinctively with the latter, while all others are recurrent (e.g., Serinaldi and Grimaldi 2007, for an exemplification). Furthermore, the families of the bivariate copulas in an R-vine are not restricted to the Archimedean class.

A practical difficulty arises from the many possible different R-vine structures when attempting to estimate a suitable vine-copula, especially if the dimension is high. On n variables there are in total

$$\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}} \tag{4}$$

labeled R-vines (Morales-Nápoles et al. 2010; Morales-Nápoles 2011). We address the issue of selecting a suitable structure for time series in the next section.

Once we selected a structure, we make the simplifying assumption that copulas of conditional distributions do not directly depend on the conditioning variable in order to keep inference and model selection fast and robust. While Haff et al. (2010) showed that a simplified pair copula decomposition can be a good approximation even when the assumption is far from being fulfilled, Acar et al. (2012) illustrated that it can also be misleading. To simulate time series, we sample recursively according to well known algorithms using the inverse conditional copulas corresponding to the R-vine density decomposition in (3) (Kurowicka and Cooke 2006; Aas et al. 2009). In the following section, we describe three vine structures in more detail and we included the accompanying sampling algorithms in Appendix I.

Vines for Time Series

Suppose $\{X_t^{(1)}\}, ..., \{X_t^{(n)}\}$ are n univariate time series jointly observed at time points $t = 1, ..., T$. For $T = 30$ and $n = 2$ this amounts to a vine on 60 nodes. According to equation (4) the number of possible R-vines on 60 nodes is incredibly high, any requiring the specification of 60 marginal distributions and 1770 pair copulas.

A significant simplification of this problem can be achieved by making stationarity and Markov assumptions. For stationary data, the marginal distributions for each variable do not change with time. In the same way, the bivariate copulas (including the conditional ones) joining any two variables are preserved in time. A Markov property of order k implies that all copulas with a lag length greater than k in the conditioned nodes can be set to the independence copula. As a consequence, the number of bivariate copulas that need to be estimated no longer depends on T, but on k and the vine structure. The simplest case is, of course, $k = 1$. When developing the vine-copula model for significant wave heights and mean zero-crossing periods, we limit ourselves to this case.

To achieve a desirable parsimonious model for the time series data, the vine structure should be selected so that (1) the number of unique copulas to be used is minimized, (2) the number of independence copulas in the model is maximized, and (3) the model is easily extendable to arbitrarily many time steps. The last point automatically leads to a structure that represents the flow of time in an intuitive manner. Structures that fulfill these requirements are usually not obtained when using the popular structure selection algorithm by Dissmann et al. (2013), which allocates the strongest pairwise dependencies to the first tree.

Advantageous structures for stationary and Markovian multivariate time series have been proposed by Brechmann and Czado (2015), Smith (2015) and Nai Ruscone and Osmetti (2017). Brechmann and Czado (2015) call their approach COPAR, which stands for COPula AutoRegressive model, and investigate inflation effects on industrial production, stock returns and interest rates. In COPAR serial dependence of $X_t^{(1)}$ is modeled unconditionally of $X_t^{(2)}...X_t^{(d)}$ and the serial dependence of $X_t^{(i)}$, $i \geq 2$, is modeled conditionally on all $X_t^{(j)}$, $j < i$. In two dimensions, this approach is most appropriate when modeling the dependence of one time series onto another. The first three trees of a modified version of the COPAR model for $n = 2$ and $T = 3$ are depicted in Fig. 2a. In this version, serial dependence of $\{X_t^{(1)}\}$ is also modeled conditionally, in this case on $\{X_t^{(2)}\}$. We made the modification for two reasons. On one hand, there is no obvious cause and effect relationship between significant wave heights and mean zero-crossing periods. On the other hand, all non-independence copulas are now included in the first three trees yielding a clearer graphical representation. Just as for the COPAR, 5 unique copulas are needed to quantify this R-vine. These are listed in the third column of Table 4. It is evident that we can sample $X_{t+s}^{(1)}$ and $X_{t+s}^{(2)}$ iteratively for all $s > 1$ in the same way we sampled $X_{t+1}^{(1)}$ and $X_{t+1}^{(2)}$

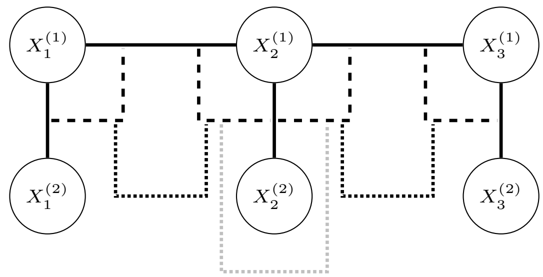
Smith (2015) propose a D-vine structure of a univariate series $\mathbf{Y} = (Y_1, ..., Y_N)$, $N = Tn$, into which the elements of a multivariate time series have been re-ordered. Their article examines five dimensional time series from the Australian electricity market. Fig. 2b shows the first three trees of this model for $n = 2$ and $T = 4$. From now on, we call this model *alternating* D-vine, because $X^{(1)}$ and $X^{(2)}$ alternate in the first tree. Similarly to the modified COPAR example, 5 unique bivariate copulas are needed to quantify it (column 4 in Table 1), although they differ except for $C_{X_t^{(1)},X_t^{(2)}}$. Again, it is sufficient to create an alternating D-vine up to $t = 2$. After that, we can iteratively sample $X_{t+s}^{(1)}$ and $X_{t+s}^{(2)}$ for $s > 1$ (Appendix I).

A third alternative is given in Fig. 2c with unique copulas reported in column 5 of Table 1. In the 2D case it coincides with a proposal by Nai Ruscone and Osmetti (2017). This structure is referred to as *branching* D-vine, as $X^{(1)}$ and $X^{(2)}$ branch out. In higher dimensions, the first tree is star-like. Note that 6 instead of 5 unique copulas are needed, with one of them being in the highest order tree. The main difference with the previous approaches is that serial correlations in each univariate series are modeled unconditionally on the other series. While cross-series dependence is modeled conditionally on previous time points, except of course for $t = 1$. Because of the high number of independence copulas, we can again sample $X_{t+s}^{(1)}$ and $X_{t+s}^{(2)}$ iteratively, this time for $s > 2$ (Appendix I).

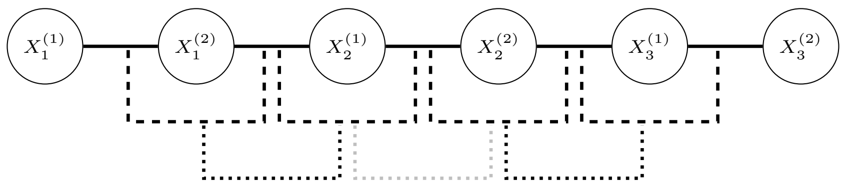
Table 1. Overview of bivariate copulas other than the independence copula that are needed for each of the three vines to specify a stationary bivariate stochastic process $\{X_t^{(1)}, X_t^{(2)}\}_{t=1,2,...}$ with Markov property of order $k = 1$.

Tree	Copula	modified COPAR	Alternating D-Vine	Branching D-Vine
1	$C_{X_t^{(1)},X_{t+1}^{(1)}}$	x		x
	$C_{X_t^{(1)},X_t^{(2)}}$	x	x	x
	$C_{X_t^{(2)},X_{t+1}^{(1)}}$		x	
	$C_{X_t^{(2)},X_{t+1}^{(2)}}$			x
2	$C_{X_t^{(2)},X_{t+1}^{(1)}\|X_t^{(1)}}$	x		x
	$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_{t+1}^{(1)}}$	x		
	$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_t^{(2)}}$			x
	$C_{X_t^{(2)},X_{t+1}^{(2)}\|X_{t+1}^{(1)}}$		x	
	$C_{X_t^{(1)},X_{t+1}^{(1)}\|X_t^{(2)}}$		x	
3 or higher	$C_{X_t^{(2)},X_{t+1}^{(2)}\|X_t^{(1)},X_{t+1}^{(1)}}$	x		
	$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_t^{(2)},X_{t+1}^{(1)}}$		x	
	$C_{X_{t+1}^{(1)},X_{t+1}^{(2)}\|X_t^{(1)},X_t^{(2)}}$			x

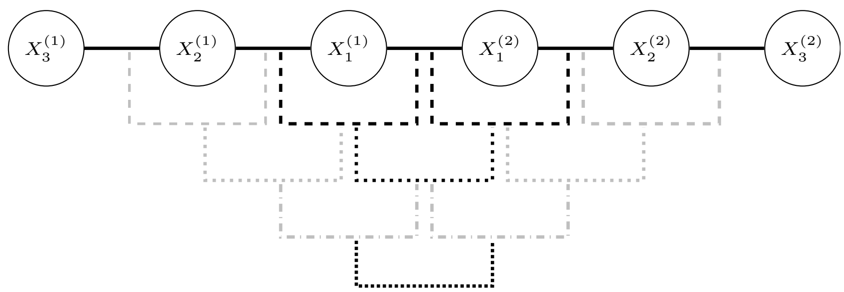
In theory, the three vine structures define different decompositions for the same joint density. In other words, they are equivalent representations of a bivariate stationary stochastic process with Markov property of order 1. However, in practice, it may be easier to find suitable bivariate copula models for one structure than for another. This



(a) First three trees of the modified COPAR model



(b) First three trees of the alternating D-Vine Structure



(c) Branching D-vine Structure

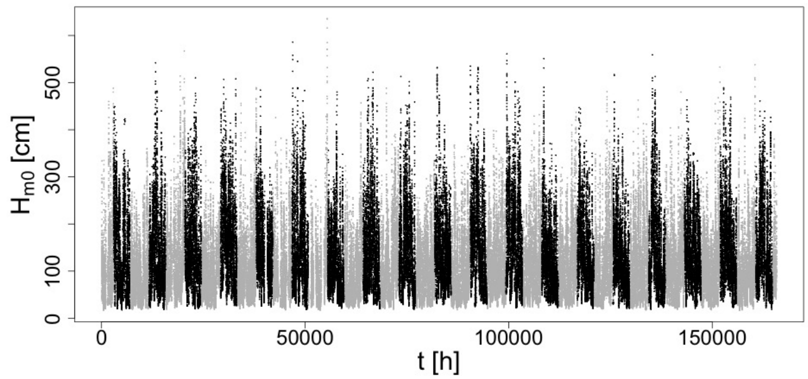
Fig. 2. Examples of (truncated) vines for a bivariate stochastic process $\{X_t^{(1)}, X_t^{(2)}\}_{t=1,2,3}$. Edges of the same tree have the same line type and edges signifying serial dependence with Markov property of order $k > 1$ are depicted in gray.

will affect the overall performance of the vine-copula, which is discussed in section 4.

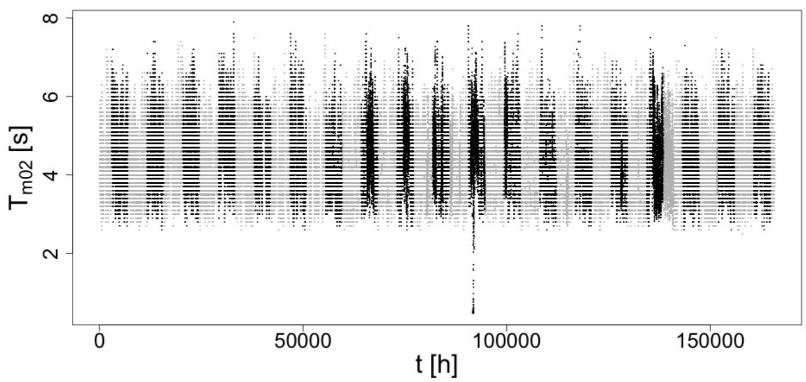
Although Smith (2015) and Brechmann and Czado (2015) propose integral vine-copula estimation methods, the sea waves application relies on a sequential estimation approach, because fast algorithms are not available for some of the copula models we consider. Estimation and simulation efforts are comparable for the three vine-copulas.

DATA

Hourly values of the significant wave height, H_{m0}, and the mean zero-crossing wave period, T_{m02}, are available for a period of 27 years, starting April 1989 and ending in February 2016. The data has been collected at the Europlatform, a Dutch offshore station located 58km off the coast of Rotterdam ($52°00'N$, $03°17'E$) at a water depth of approximately 26.5m, which is operated by Rijkswaterstraat. To bypass seasonal effects, the study is limited to the wave climate during the oceanographic winter period (1 November - 15 April) in which storm events are more frequent and heavier than during the summer period, see Fig. 3. An analysis for the summer season may be conducted similar to the one presented in this paper.



(a)



(b)

Fig. 3. Hourly (a) significant wave heights and (b) mean zero-crossing periods at the Europlatform (April 1989 - February 2016). Winters are depicted in black, summers in gray.

Both time series exhibit strong temporal serial correlation. Fig. 4 shows the serial correlation, in terms of Spearman's rank correlation coefficient, and indicates that it only drops below 0.1 for a time lag of at least 200 hours for significant wave heights. A similar relationship arises for the mean zero-crossing periods (Fig. 4). Here a periodic variation is notable in the order of 12 hours, which could be attributed to the tidal cycle. In the following analysis this variation is neglected to keep the modeling parsimonious.

While we try to model the serial correlation structure, ignoring it in the inference process can cause unwanted bias. In particular, standard statistical inference procedures require independent and identically distributed (i.i.d.) observations. To obtain (approximately) i.i.d. observations, we have sub-sampled the data. More precisely, we extracted a comparatively small number of time series fragments ($N = 506$, $T = 2$). These are sufficient for the inference of the vine-copula models under consideration, because we assume stationarity and a Markov property of order 1, as explained in the

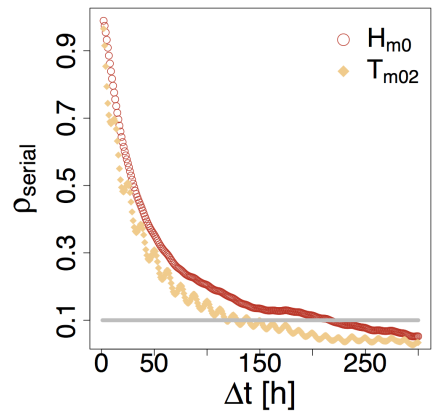


Fig. 4. Serial correlation in terms of Spearman's rank correlation coefficient for measured significant wave heights and mean zero-crossing periods

previous section. In the time series fragments, we denote the respective j-th sample of the t-th time point of significant wave height and mean zero-crossing period as $h_{m0_{j,t}}$ and $t_{m02_{j,t}}$, where $j = 1, ..., N$ and $t = 1, T$. The N joint samples are (approximately) $i.i.d$ and can be used for inference of the marginal distributions and bivariate copulas.

Because formal hypothesis testing for stationarity yielded contradictory results, we followed a diagnostic approach. QQ-plots attest that the univariate distributions do not depend on t (not shown) and the observed differences in rank correlation coefficients for the same variable pairs at different points in time are very small. Table 2 reports all coefficients for a time series fragment of length $T = 4$ to indicate the order of magnitude in the differences. Thus, we concluded that it is reasonable to model the processes, in a first instance, as stationary. Nevertheless, more research in this direction is desirable.

Another aspect that has to be dealt with before the data analysis, is the limited instrumental resolution, which causes the available measurements to be discrete, while

Table 2. Spearman's rank correlation coefficients for time series fragments with $T = 4$ ($N = 506$)

	H_{m0_t}	T_{m02_t}	$H_{m0_{t+1}}$	$T_{m02_{t+1}}$	$H_{m0_{t+2}}$	$T_{m02_{t+2}}$	$H_{m0_{t+3}}$	$T_{m02_{t+3}}$
H_{m0_t}	1	0.80	0.99	0.81	0.98	0.80	0.95	0.79
T_{m02_t}	0.80	1	0.78	0.97	0.76	0.92	0.73	0.85
$H_{m0_{t+1}}$	0.99	0.78	1	0.80	0.99	0.80	0.98	0.79
$T_{m02_{t+1}}$	0.81	0.97	0.80	1	0.78	0.97	0.76	0.90
$H_{m0_{t+2}}$	0.97	0.76	0.99	0.78	1	0.79	0.99	0.79
$T_{m02_{t+2}}$	0.80	0.92	0.80	0.97	0.79	1	0.78	0.96
$H_{m0_{t+3}}$	0.95	0.73	0.98	0.76	0.99	0.78	1	0.79
$T_{m02_{t+3}}$	0.79	0.85	0.79	0.90	0.79	0.96	0.79	1

the physical variables are continuous. Mainly in the case of the mean zero-crossing period this phenomenon causes a high number of ties, imposing additional challenges for the statistical analysis. To overcome these, the data has been randomized following the extensive guidelines in Salvadori et al. (2014): Suppose $T^*_{m02_j}$ is a discrete measurement of T_{m02_j}. It is transformed to

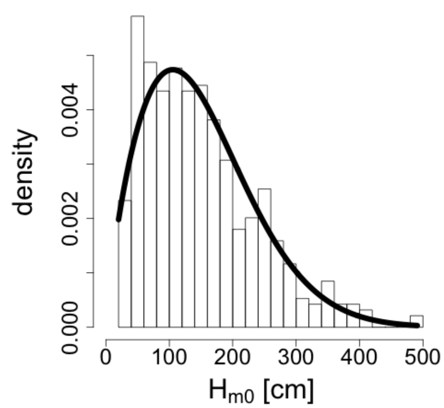
$$\widetilde{T}_{m02_j} = T^*_{m02_j} + \Delta \cdot W_j, \tag{5}$$

where $j = 1, ..., N$, Δ is the approximate resolution of the measurement instrument and W_j has a uniform distribution on $[0, 1]$. Thus, $T^*_{m02_j}$ is re-sampled uniformly over the resolution interval and the resulting $\widetilde{T}_{m02_j}$ is assumed to be statistically equivalent to the real T_{m02_j}. We did not find our results to be sensitive to the choice of Δ.

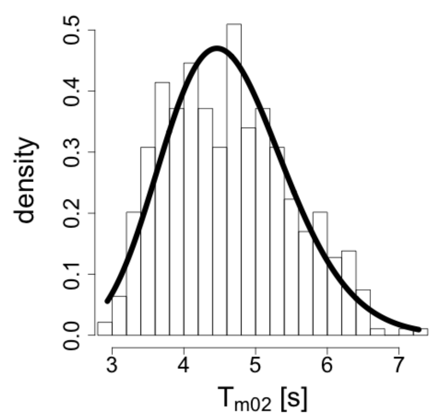
RESULTS

Univariate Margins

The marginal distribution of the significant wave height is assumed to follow a Weibull distribution with shape parameter 1.78 and scale parameter 167.21. The mean zero-crossing period is approximated by a gamma distribution with shape parameter 30.08 and rate parameter 6.51. Both distributions are estimated from sub-sampled data, as described in section 3. Histograms and the maximum likelihood fits are displayed in Fig. 5. The adequate fit of these distributions is attested by the QQ-plots in Fig. 6. Moreover, we sampled 506 random numbers from the specified distributions and performed two-sample Kolmogorov-Smirnov tests. These did not reject the null hypotheses that the observed and sampled data stem from the same distributions. We performed such tests 100 times and obtained p-values ranging from 0.07 to 0.90 for significant wave heights and from 0.08 to 0.97 for mean zero-crossing periods.



(a)



(b)

Fig. 5. Histograms and maximum likelihood fit of (a) significant wave heights and (b) mean zero-crossing periods

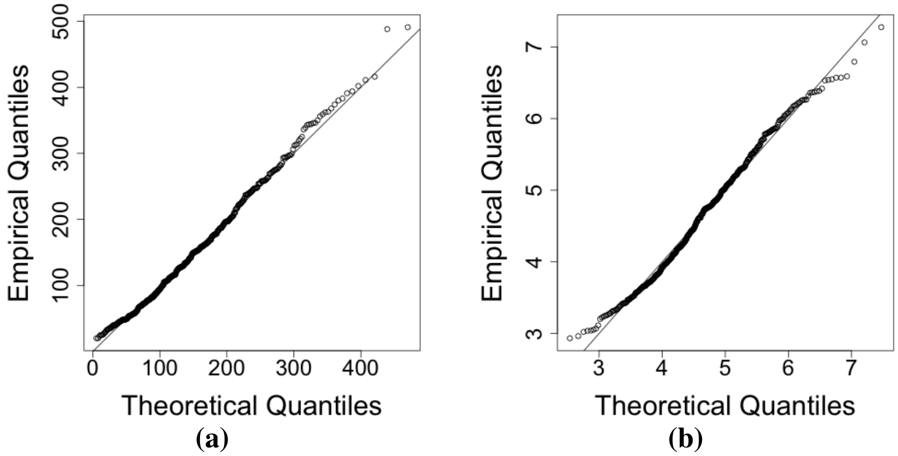


Fig. 6. QQ-plots showing the fit of the univariate models for (a) significant wave heights and (b) mean zero-crossing periods

Table 3. Nomenclature for variables

Variable	Standard name	Name in this article
Significant wave height [cm]	H_{m0}	$X^{(1)}$
Mean period [s]	T_{m02}	$X^{(2)}$

Dependence Structure

To be consistent with the mathematical notation of section 2, we denote significant wave heights and mean zero-crossing periods now as $X^{(1)}$ and $X^{(2)}$ (cf. Table 3).

We estimated the vines sequentially, that is, we estimated the bivariate copulas in lower order trees before the ones in higher order trees. To quantify the three candidate models, 12 unique bivariate copulas need to be estimated. They are listed in the first column of Table 4. Inference on the bivariate copula models is based on ranks and not by using the parametric univariate distributions. They give a faithful representation of dependence, irrespective of the marginal behavior of the variables. The rank of $x_{j,t}^{(i)}$ among $x_{1,t}^{(i)}, ..., x_{N,t}^{(i)}$ divided by $N + 1$ will be denoted as $u_{j,t}^{(i)}$. Before copula selection, we performed an independence test based on Kendall's τ. This test is implemented in the VineCopula package (Schepsmeier et al. 2015) and described in (Genest and Favre 2007). If the null hypothesis of independence could not be rejected, we chose the independence copula. In general, we estimated model parameters by pseudo maximum likelihood and selected families according to the AIC criterion with the VineCopula package. As potential candidate copula models, we considered the following commonly known families, which are included in the package: Gaussian, t, Frank, Gumbel, Clayton, Joe, BB1, BB6, BB7, BB8 families and, if applicable, rotated versions thereof. All selected models as well as their parameters are listed in Table 4. The independence copula is represented by $\prod$.

We made an exception to this procedure for $C_{X_t^{(1)},X_t^{(2)}}$ and $C_{X_t^{(2)},X_{t+1}^{(1)}}$, because the corresponding data exhibits a skewness feature that none of the above models can represent. The observations of $(U_t^{(1)}, U_t^{(2)})$ are shown in Fig. 7a). The variables are positively associated. Very notable is an asymmetry with respect to the main diagonal, or skewness, and a sharp line bounding the maximum significant wave height for given mean zero-crossing period values. Considering the water depth at the measuring location, this feature of the data can be attributed to steepness-induced wave breaking ("white-capping"). The observations of $(U_t^{(2)}, U_{t+1}^{(1)})$ are visually not distinguishable from the ones of $(U_t^{(2)}, U_t^{(1)})$. For $C_{X_t^{(1)},X_t^{(2)}}$ and $C_{X_t^{(2)},X_{t+1}^{(1)}}$, we considered the Tawn, the skew-t and the gamma 1-factor model (GFM) families. Again, we estimated parameters by pseudo maximum likelihood. Appendix II provides more details on these families and shows samples of the respective maximum likelihood fits for $(U_t^{(1)}, U_t^{(2)})$. The Tawn copula does not capture the sharp line which limits the wave height with respect to the period, and indicates the threshold for wave breaking (Fig. 14a). This is an important physical process and, therefore, we did not select this family. Both, the GFM and the skew-t family are able to capture the wave breaking (Fig. 14b and Fig. 14c). Besides visually comparing the simulated points from the three copulas to the data, we computed semi-correlations, in terms of Pearson's product moment correlation, for their normal scores ($N_{sim} = 10^5$). These are listed in Table 5. Semi-correlations are correlations of the lower and upper quadrants, ρ_N^+ and ρ_N^-, respectively. The empirical ρ_N^+ and ρ_N^- are notably different, indicating stronger upper than lower tail-dependence in the data. Because, the GFM has more symmetric semi-correlations than the skew-t copula, the latter has been selected for further analysis. A disadvantage of the skew-t family is that closed form conditional and inverse conditional copula are not available and need to be approximated numerically.

The three vines are different approaches to specifying the dependence structure between significant wave heights and mean zero-crossing periods. We investigated their relative performance on three bivariate margins, namely on $(U_t^{(1)}, U_t^{(2)})$, $(U_t^{(1)}, U_{t+1}^{(1)})$ and $(U_t^{(2)}, U_{t+1}^{(2)})$. To this end, we simulated 506 bivariate time series with $T = 100$. The scatter plots for the empirical pairs of ranks are depicted in Fig. 7. Fig. 7b exposes the dependence between significant wave heights at subsequent hours, while Fig. 7c shows the dependence between mean zero-crossing periods at subsequent hours. In both cases the points are distributed closely around the main diagonal indicating a very strong positive association, which is in agreement with Fig. 4.

Fig. 8 shows the simulated ranks for $t = 1$ as well as $t = 99$. Results of the modified COPAR model are promising (Fig. 8 (a) - (c)). Pairwise scatter plots are very similar to the ones of the data (Fig. 7 (a) - (c)) and the dependence structure does not appear to change significantly over time. Moreover, considering a small time series fragment of $T = 4$, the simulated and observed rank correlation coefficients are comparable (cf. Table 6 and Table 2). The maximum difference we observe is 0.047.

The simulated ranks from the alternating D-vine are plotted in Fig. 8 (d) - (f). This model does not capture the strong serial correlation between mean zero-crossing periods; $\rho_{U_t^{(2)},U_{t+1}^{(2)}}$ of the simulated data is around 0.61, while we estimated a value of 0.97 for the data (Table 2). In the branching D-vine the skewness of the $(U_t^{(1)}, U_t^{(2)})$ pair deteriorates in time (actually already within a few steps), which is not surprising

Table 4. Copula families and parameters for vines

Copula	Family	Parameters
$C_{X_t^{(1)},X_{t+1}^{(1)}}$	t	$\rho = 0.99,\ \nu = 11.18$
$C_{X_t^{(1)},X_t^{(2)}}$	skew-t	$(\rho, \delta_1, \delta_2, \nu) = (0.73, -0.85, -0.27, 30)$
$C_{X_t^{(2)},X_{t+1}^{(1)}}$	skew-t	$(\rho, \delta_1, \delta_2, \nu) = (0.7, -0.85, -0.26, 30)$
$C_{X_t^{(2)},X_{t+1}^{(2)}}$	t	$\rho = 0.96,\ \nu = 12.06$
$C_{X_t^{(2)},X_{t+1}^{(1)}\|X_t^{(1)}}$	Π	–
$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_{t+1}^{(1)}}$	Π	–
$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_t^{(2)}}$	Frank	$\theta = 1.35$
$C_{X_t^{(2)},X_{t+1}^{(2)}\|X_{t+1}^{(1)}}$	Gumbel	$\theta = 3.01$
$C_{X_t^{(1)},X_{t+1}^{(1)}\|X_t^{(2)}}$	Survival Gumbel	$\theta = 5.21$
$C_{X_t^{(2)},X_{t+1}^{(2)}\|X_t^{(1)},X_{t+1}^{(1)}}$	Gumbel	$\theta = 3.17$
$C_{X_t^{(1)},X_{t+1}^{(2)}\|X_t^{(2)},X_{t+1}^{(1)}}$	90° rotated BB8	$(\theta_1, \theta_2) = (-1.47, -0.96)$
$C_{X_{t+1}^{(1)},X_{t+1}^{(2)}\|X_t^{(1)},X_t^{(2)}}$	t	$(\rho, \nu) = (0.36, 17.05)$

Table 5. Lower and upper semi-correlations, ρ_N^+ and ρ_N^-, for the pair $(X_t^{(1)}, X_t^{(2)})$

Copula	ρ_N^+	ρ_N^-
Empirical	0.80	0.45
Tawn type 2	0.79	0.54
GFM	0.66	0.59
Skew-t	0.72	0.56

given the skew-t copula is only used to model the very first point in time (Fig. 8 (d) - (f)).

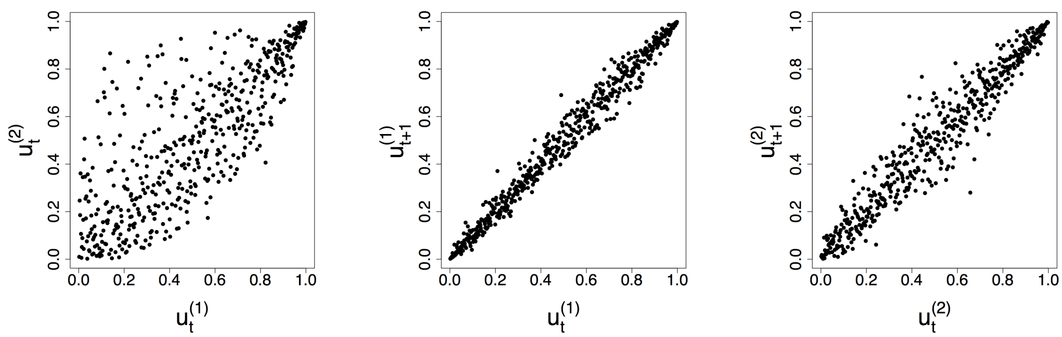
Hence, we chose the modified COPAR structure for time series simulation of significant wave heights and mean zero-crossing periods. Copula samples can be transformed to the original (H_{m0}, T_{m02})-space by applying the respective inverse marginal distribution functions.

Bivariate Distribution of Significant Wave Heights and Mean Zero-Crossing Periods

The specification of the modified COPAR includes a specification of the bivariate distribution of significant wave height and corresponding mean zero-crossing period. Samples of this distribution are shown in Fig. 9. A visual comparison of observed and simulated points indicates that the model is valuable. In particular, it captures the physical limitations on wave steepness due to wave breaking. Wave steepness can be

Table 6. Spearman's rank correlation coefficients for time series fragments simulated with the modified COPAR ($T = 4$, $N_{sim} = 506$)

	$\hat{H}_{m0_t}$	$\hat{T}_{m02_t}$	$\hat{H}_{m0_{t+1}}$	$\hat{T}_{m02_{t+1}}$	$\hat{H}_{m0_{t+2}}$	$\hat{T}_{m02_{t+2}}$	$\hat{H}_{m0_{t+3}}$	$\hat{T}_{m02_{t+3}}$
$\hat{H}_{m0_t}$	1	0.79	0.99	0.76	0.98	0.76	0.97	0.77
$\hat{T}_{m02_t}$	0.79	1	0.79	0.93	0.78	0.88	0.78	0.83
$\hat{H}_{m0_{t+1}}$	0.99	0.79	1	0.78	0.99	0.78	0.98	0.78
$\hat{T}_{m02_{t+1}}$	0.76	0.93	0.78	1	0.77	0.94	0.76	0.88
$\hat{H}_{m0_{t+2}}$	0.98	0.78	0.99	0.77	1	0.78	0.99	0.78
$\hat{T}_{m02_{t+2}}$	0.76	0.88	0.78	0.94	0.78	1	0.78	0.93
$\hat{H}_{m0_{t+3}}$	0.97	0.78	0.98	0.76	0.99	0.78	1	0.79
$\hat{T}_{m02_{t+3}}$	0.77	0.83	0.78	0.88	0.78	0.93	0.79	1



(a) **(b)** **(c)**

Fig. 7. Empirical normalized ranks of (a) significant wave heights at t and mean zero-crossing periods at t, (b) significant wave heights at t and at $t+1$, (c) mean zero-crossing periods at t and at $t+1$.

computed as

$$S_p = \frac{2\pi}{g} \frac{H_{m0}}{T_{m02}^2}. \quad (6)$$

$S_{p_{max}} = 0.07$ is thought to be an upper limit (Holthuijsen 2010), and is represented as a line in Fig. 9. A histogram of wave steepness for observations and simulated points confirms that the distribution of S_p is well described by the model (Fig. 10). The $skew-t$ family is a promising, although computationally more intensive, alternative to the skewed families constructed by extra-parameterization in (Vanem 2016).

DISCUSSION

We simulated 100 time series of 3984 hours, which is equivalent to an oceanographic winter, with the modified COPAR. However, these synthetic series contain notably fewer storm events than observed. An exploratory analysis of the root cause of this

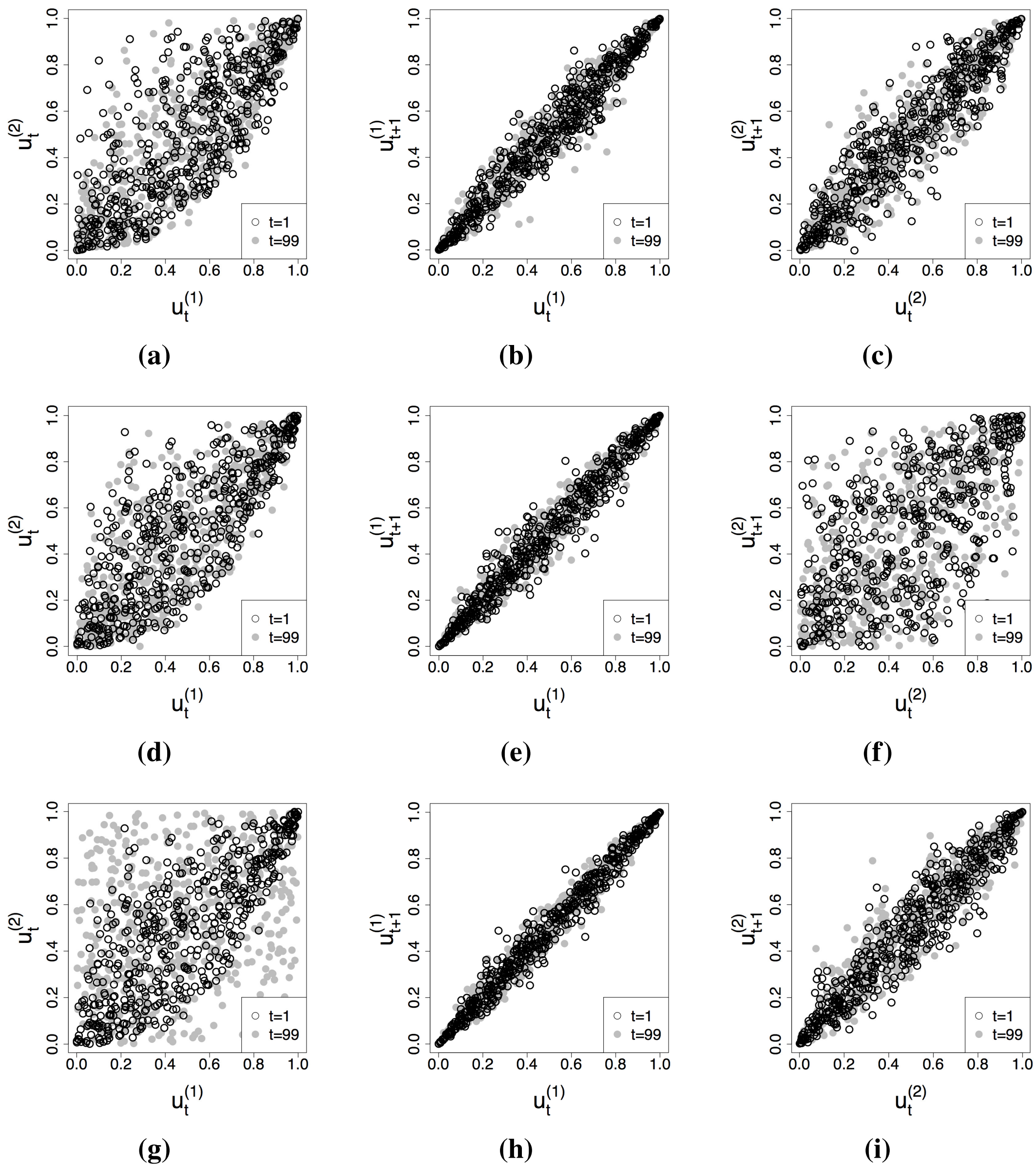


Fig. 8. Simulated points for significant wave heights at t and mean zero-crossing periods at t, significant wave heights at t and at $t + 1$, and mean zero-crossing periods at t and at $t + 1$: (a) - (c) modified COPAR, (d) - (f) alternating D-vine, and (g) - (i) branching D-vine.

phenomenon shows that the long term serial correlation is very sensitive to the choice of the ρ parameter in the $C_{X_t^{(1)},X_{t+1}^{(1)}}$ copula, now denoted by $\rho_{t,t+1}^{(1)}$.

We conducted a sensitivity analysis, assuming values 0.98 and 0.95 besides the estimated 0.99 for $\rho_{t,t+1}^{(1)}$. Fig. 11 shows examples of simulated time series as well as the observations from the winter 2013/2014. The three simulated time series are based on the same random seed. They exhibit the same pattern, however, the spikes in

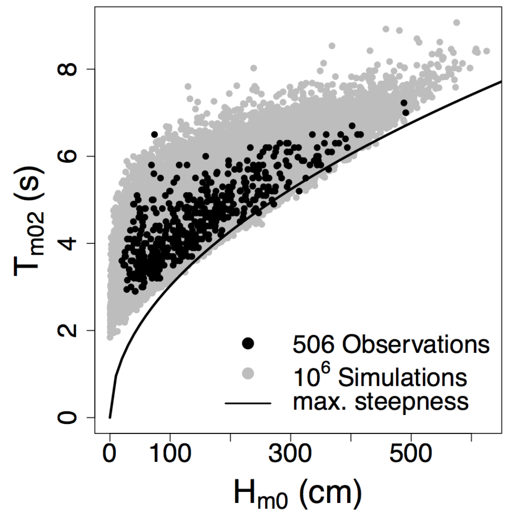


Fig. 9. Simulated and observed significant wave heights and mean zero-crossing periods

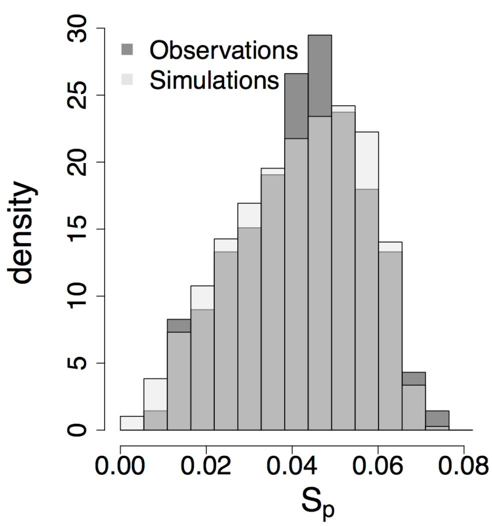


Fig. 10. Histogram of wave steepness for simulated and observed significant wave heights and mean zero-crossing periods

significant wave height and mean zero-crossing period are higher for lower values of $\rho^{(1)}_{t,t+1}$. The boxplot in Fig. 12a provides more insight into the total hours with significant wave heights higher than 400 cm, $N_{H_{m0} \geq 400}$. Recall that $N = 27$ time series have been observed and $N = 100$ have been simulated. For smaller $\rho^{(1)}_{t,t+1}$, the median increases, while the variance decreases. The reason for this model behavior is likely to lie in the faster decay of the serial correlation for smaller $\rho^{(1)}_{t,t+1}$ (Fig. 13), increasing the likelihood of sudden, extreme changes in H_{m0} and T_{m02}. The serial correlation function of both variables is most closely approximated when using $\rho^{(1)}_{t,t+1} = 0.98$. Results for the number of hours with significant wave heights smaller than 50 cm, $N_{H_{m0}\leq 50}$, are different (Fig. 12b). In this case the median is higher in the simulations than in observations.

To investigate whether the observed $N_{H_{m0}\geq 400}$ and $N_{H_{m0}\leq 50}$ are statistically different to the ones simulated, we conducted an analysis of variance (ANOVA) and performed a two-sample Kolmogorov-Smirnov (KS) test. At the 5%-level the Null-hypothesis of the ANOVA, the mean of observed and simulated $N_{H_{m0}\geq 400}$ or $N_{H_{m0}\leq 50}$ are equal, is not rejected for any of the values of $\rho^{(1)}_{t,t+1}$. For the KS-test with the Null-hypothesis that observed and simulated $N_{H_{m0}\geq 400}$ are equal in distribution, $\rho^{(1)}_{t,t+1} = 0.99$ is rejected with a p-value close to 0, but $\rho^{(1)}_{t,t+1} = 0.98$ and $\rho^{(1)}_{t,t+1} = 0.95$ cannot be rejected at the 5%-level. The results are different when testing the Null-hypothesis that observed and simulated $N_{H_{m0}\leq 50}$ are equal in distribution. $\rho^{(1)}_{t,t+1} = 0.99$ cannot be rejected at the 5%-level, $\rho^{(1)}_{t,t+1} = 0.98$ is rejected with a p-value of 0.04 and $\rho^{(1)}_{t,t+1} = 0.95$ is rejected with a p-value close to 0.

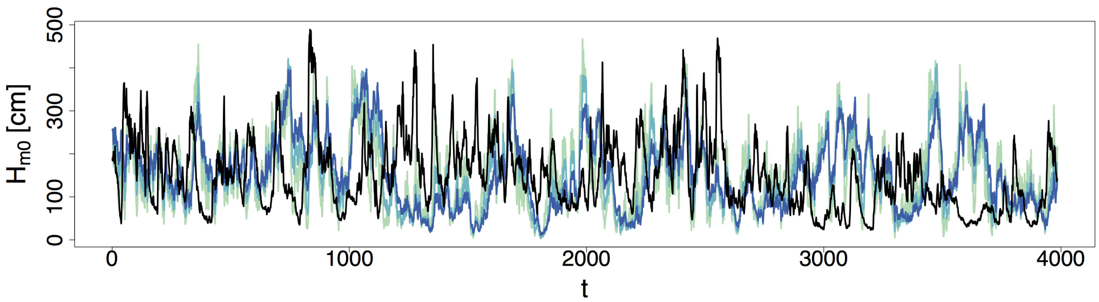
Considering these findings, the modified COPAR is useful for applications that require extreme wave scenarios, such as coastal risk analyses, as well as for applications that need estimations of quiet seas, such as scheduling of offshore operations. It is not surprising that very small differences in $\rho^{(1)}_{t,t+1}$, which could even be sample fluctuations, gain importance when simulating longer time frames, such as $T = 3984$. It is therefore essential to validate the model in view of the application and time frame of interest.

CONCLUSIONS

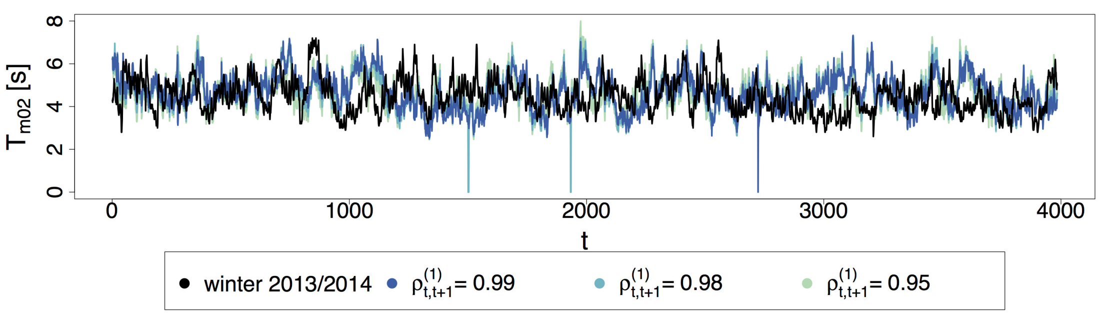
In this article, we presented a vine-copula approach to simulate joint time series of significant wave heights and mean zero-crossing periods. First, we reviewed existing vine-copula models for time series. While many different vine structures are possible, a few have recently been proposed that make efficient use of stationarity and Markov assumptions and yield parsimonious representations of such stochastic processes. In two dimensions, depending on the structure, no more than 5 or 6 unique copulas have to be specified to characterize arbitrarily long, stationary time series with Markov property of order 1 .

Furthermore, as a building block for the vine-copula, we investigated the dependence structure of significant wave heights and mean zero-crossing periods at the same time. We approximated the dependence with the flexible 4-parameter skew-t copula family. It captures asymmetric tail dependence as well as skewness patterns of the data, and, as a result, preserves the limiting wave steepness condition in simulated samples of the two variables.

Finally, we found an R-vine structure similar to the COPAR introduced by Brechmann and Czado (2015) to be most valuable for this application. Based on observations of two successive pairs of significant wave height and mean zero-crossing period, we constructed a vine-copula on four variables. These observations are a small i.i.d. subset of measurements (506 out of 53492), collected during the oceanographic winter period at a single location in the North Sea. We simulated time series with the duration of an oceanographic winter and compared them against all measured winters. While results are sensitive to small changes in the copula parameters, the presented model has potential to provide valuable input for various applications, such as coastal risk analysis,



(a)



(b)

Fig. 11. Hourly measurements of (a) significant wave height and (b) mean zero-crossing period during winter 2013/2014 and examples of simulated time series using the same random seed, but a different parameter in $C_{X_t^{(1)},X_{t+1}^{(1)}}$ denoted as $\rho_{t,t+1}^{(1)}$.

design of marine structures and offshore operations.

ACKNOWLEDGMENT

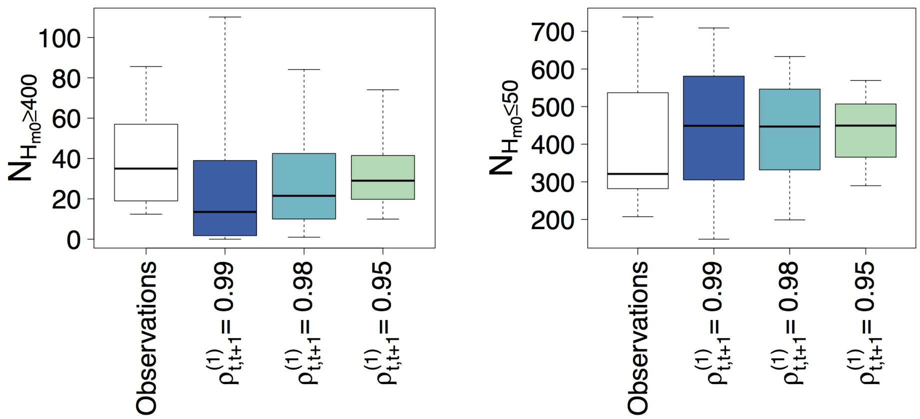
We thank the two anonymous reviewers whose comments and suggestions helped to improve and clarify this manuscript. This work was supported by the European Community's 7th Framework Programme through the grant to RISC-KIT ("Resilience-increasing Strategies for Coasts - Toolkit"), contract no. 603458.

APPENDIX I. SAMPLING ALGORITHMS

This appendix presents the sampling algorithms for the three vine-copulas that have been used in this article. Conditional copulas are denoted by the following *h-function*:

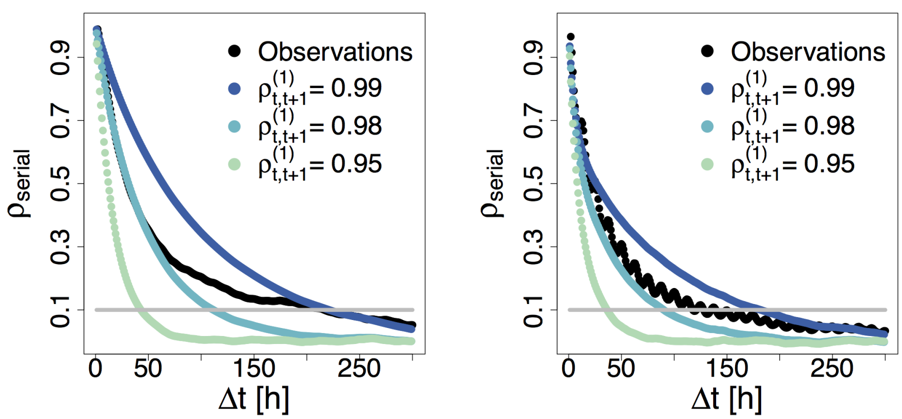
$$h(u, v, \theta) = \frac{\partial C_{u,v}(u, v, \theta)}{\partial v}, \tag{7}$$

where $u, v, \in [0, 1]$ and θ denotes the set of parameters for the copula of the joint distribution function of u and v. Thus, the second parameter of $h(\cdot)$ always corresponds to the conditioning variable. The inverse of the h-function with respect to u is denoted by h^{-1} and corresponds to the inverse conditional copula. We recall that n is the number of variables and k is the order of the Markov property.



(a) **(b)**

Fig. 12. Boxplots for (a) $N_{H_{m0} \geq 400}$ and (b) $N_{H_{m0} \leq 50}$ in the measured and simulated data. $\rho^{(1)}_{t,t+1}$ denotes one of the two parameters of $C_{X_t^{(1)}, X_{t+1}^{(1)}}$. The whiskers represent the 5th percentile and the 95th percentile of the distribution of interest.



(a) **(b)**

Fig. 13. Serial correlation of measured and simulated (a) significant wave heights and (b) mean zero-crossing periods. $\rho^{(1)}_{t,t+1}$ denotes one of the two parameters of $C_{X_t^{(1)}, X_{t+1}^{(1)}}$.

Algorithm 1 (modified COPAR for $n = 2$ **and** $k = 1$**)**

1. *Simulate* $2T$ *random variates* $s_1, ..., s_{2T}$ *from* $U(0, 1)$.
2. *Set* $x_1^{(1)} = s_1$
3. *Set* $x_1^{(2)} = h^{-1}(s_1, x_1^{(1)}, \theta_{X_t^{(1)}, X_t^{(2)}})$.

For $i = 2, ..., T$*:*

4. *Set* $x_i^{(1)} = h^{-1}(h^{-1}(s_{2i-1}, h(x_{i-1}^{(2)}, x_{i-1}^{(1)}), \theta_{X_t^{(1)}, X_{t+1}^{(1)} | X_t^{(1)}}), x_{i-1}^{(1)}, \theta_{X_t^{(1)}, X_t^{(2)}})$
5. *Set* $x_i^{(2)} = h^{-1}(h^{-1}(h^{-1}(s_{2i}, h(h(x_{i-1}^{(2)}, x_{i-1}^{(1)}, \theta_{X_t^{(1)}, X_t^{(2)}}), h(x_i^{(1)}, x_{i-1}^{(1)}, \theta_{X_t^{(1)}, X_{t+1}^{(1)}}),$

$\theta_{X_t^{(2)},X_{t+1}^{(1)}|X_t^{(1)}}), \theta_{X_t^{(2)},X_{t+1}^{(2)}|X_t^{(1)},X_{t+1}^{(1)}})\theta_{X_t^{(1)},X_{t+1}^{(1)}}), x_{i-1}^{(1)}, \theta_{X_{t+1}^{(2)},X_t^{(1)}})$

Algorithm 2 (Alternating D-Vine for $n = 2$ and $k = 1$)

1. *Simulate* $2T$ *random variates* $s_1, ..., s_{2T}$ *from* $U(0, 1)$.
2. *Set* $x_1^{(1)} = s_1$.
3. *Set* $x_1^{(2)} = h^{-1}(s_2, x_1^{(1)}, \theta_{X_t^{(1)},X_t^{(2)}})$.

For $i = 2, ..., T$:

4. *Set* $x_i^{(1)} = h^{-1}(h^{-1}(s_{2i-1}, h(x_{i-1}^{(1)}, x_{i-1}^{(2)}, \theta_{X_t^{(1)},X_t^{(2)}}), \theta_{X_{t+1}^{(1)},X_t^{(1)}|X_t^{(2)}}), x_{i-1}^{(2)}, \theta_{X_{t+1}^{(1)},X_t^{(2)}})$
5. *Set* $x_i^{(2)} = h^{-1}(h^{-1}(h^{-1}(s_{2i-2}, h(h(x_{i-1}^{(1)}, x_{i-1}^{(2)}, \theta_{X_t^{(1)},X_t^{(2)}}), h(x_i^{(1)}, x_{i-1}^{(2)}, \theta_{X_{t+1}^{(1)},X_t^{(2)}}),$
   $\theta_{X_t^{(1)},X_{t+1}^{(1)}|X_t^{(2)}}), \theta_{X_t^{(1)},X_{t+1}^{(2)}|X_t^{(2)},X_{t+1}^{(1)}}), h(x_{i-1}^{(2)}, x_i^{(1)}, \theta_{X_t^{(2)},X_{t+1}^{(1)}}), \theta_{X_t^{(1)},X_{t+1}^{(2)}|X_{t+1}^{(1)}}), x_i^{(1)},$
   $\theta_{X_t^{(2)},X_t^{(1)}})$

Algorithm 3 (Branching D-vine for $n = 2$ and $k = 1$)

1. *Simulate* $2T$ *random variates* $s_1, ..., s_{2T}$ *from* $U(0, 1)$.
2. *Set* $x_1^{(1)} = s_1$.
3. *Set* $x_1^{(2)} = h^{-1}(s_2, x_1^{(1)}, \theta_{X_t^{(1)},X_t^{(2)}})$
4. *Set* $x_2^{(1)} = h^{-1}(h^{-1}(s_3, h(x_1^{(2)}, x_1^{(1)}, \theta_{X_t^{(1)},X_t^{(2)}}), \theta_{X_{t+1}^{(1)},X_t^{(2)}|X_t^{(1)}}), x_1^{(1)}, \theta_{X_t^{(1)},X_{t+1}^{(1)}})$
5. *Set* $x_2^{(1)} = h^{-1}(h^{-1}(h^{-1}(s_4, h(h(x_2^{(1)}, x_1^{(1)}, \theta_{X_{t+1}^{(1)},X_t^{(1)}}), h(x_1^{(2)}, x_1^{(1)}, \theta_{X_t^{(2)},X_t^{(1)}}),$
   $\theta_{X_{t+1}^{(1)},X_t^{(2)}|X_t^{(1)}}), \theta_{X_{t+1}^{(1)},X_{t+1}^{(2)}|X_t^{(1)},X_t^{(2)}}), h(x_1^{(1)}, x_1^{(2)}, \theta_{X_t^{(1)},X_t^{(2)}}), \theta_{X_{t+1}^{(1)},X_{t+1}^{(2)}|X_t^{(2)}}), x_1^{(2)},$
   $\theta_{X_{t+1}^{(2)},X_t^{(2)}})$

For $i = 3, ..., T$:

6. *Set* $x_i^{(1)} = h^{-1}(s_{2i-1}, x_{i-1}^{(1)}, \theta_{X_t^{(1)},X_{t+1}^{(1)}})$
7. *Set* $x_i^{(2)} = h^{-1}(h^{-1}(s_{2i}, h(x_i^{(1)}, x_{i-1}^{(1)}, \theta_{X_{t+1}^{(1)},X_t^{(1)}}), \theta_{X_{t+1}^{(1)},X_{t+1}^{(2)}|X_t^{(1)},X_t^{(2)}}), x_{i-1}^{(2)}, \theta_{X_{t+1}^{(2)},X_t^{(2)}})$

APPENDIX II. EXISTING PARAMETRIC COPULA FAMILIES UNDER CONSIDERATION

Tawn

The Tawn copula (Tawn 1988) is an asymmetric extension of the Gumbel copula and its cdf is given by

$$C(u_1, u_2) = \exp\{\log(u_1 u_2) A\left(\frac{\log(u_1)}{\log(u_1 u_2)}\right)\}, \quad u_1, u_2 \in [0, 1], \tag{8}$$

where $A(t) = 1 - \beta + (\beta - \alpha)t + \{\alpha^r t^r + \beta^r (1-t)^r\}^{\frac{1}{r}}$, $0 \leq \alpha,\ \beta \leq 1$, $1 \leq r < \infty$, $t \in [0, 1]$. Two simplified versions of the Tawn copula with two parameters each are implemented in the R's VineCopula package (Schepsmeier et al. 2015). They are called "Tawn type 1" and "Tawn type 2" and have one of the asymmetry parameters fixed to 1 such that the dependence is either left- or right-skewed.

Gamma 1-Factor Model

A bivariate gamma 1-factor model has the form

$$X_j = Z_0 + Z_j, \qquad j = 1, 2 \tag{9}$$

where Z_0, Z_1 and Z_2 are independent gamma distributed variables with shape parameters θ_1, θ_2 and θ_3 and scale parameters all equal to 1. The copula cdf is (Joe 2014, p.211)

$$C_{\theta_1,\theta_2,\theta_3}(u_1, u_2) = F(F^{-1}_{\theta_0+\theta_1}(u_1), F^{-1}_{\theta_0+\theta_2}(u_2)), \quad u_1, u_2 \in [0, 1] \tag{10}$$

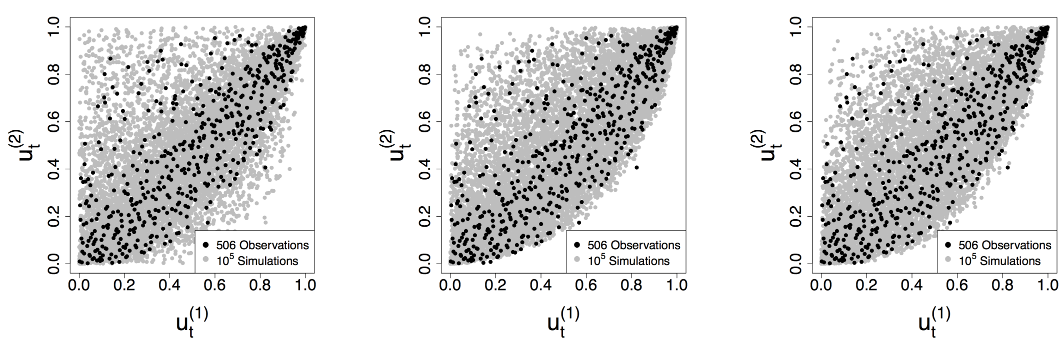
Here F is the joint cdf of X_1 and X_2, while $F_{\theta_0+\theta_1}$ and $F_{\theta_0+\theta_2}$ are the marginal distribution functions of X_1 and X_2 obtained by convoluting the marginal distributions functions of Z_0, Z_1 and Z_2. This model is implemented in the R package "CopulaModel" (Joe and Krupskii 2014).

Skew-t

The skew-t copula derives from the skew-t distribution, which is a normal mean variables mixture. While several formulations exists, only the formulation by (Azzalini and Capitanio 2003) is considered for this article. For this formulation an estimation procedure based on maximum likelihood has been developed by (Yoshiba 2015), who also published the corresponding R code. The copula cdf is given by

$$C_{\rho,\delta_1,\delta_2,\nu}(u_1, u_2) = St(St_1^{-1}(u_1; 0, 1, \delta_1, \nu), St_2^{-1}(u_2; 0, 1, \delta_2, \nu); 0, \rho, \alpha, \nu), \quad u_1, u_2 \in [0, 1], \tag{11}$$

where St is the multivariate skew-t distribution with correlation parameter ρ, a transformed skewness vector α and degrees of freedom ν. St_1 and St_2 are the univariate margins of this distribution and δ_1 and δ_2 are the respective skewness parameters.



(a) (b) (c)

Fig. 14. Comparison of observations $(U_t^{(1)}, U_t^{(2)})$ with simulated points from (a) Tawn type 2 copula, (b) gamma-1-factor model and (c) skew-t copula. Parameters are estimated via maximum likelihood.

REFERENCES

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). "Pair-copula constructions of multiple dependence." *Insurance: Mathematics and economics*, 44(2), 182–198.

Acar, E. F., Genest, C., and NešLehová, J. (2012). "Beyond simplified pair-copula constructions." *Journal of Multivariate Analysis*, 110, 74–90.

Azzalini, A. and Capitanio, A. (2003). "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2), 367–389.

Bedford, T. and Cooke, R. M. (2002). "Vines: A new graphical model for dependent random variables." *Annals of Statistics*, 1031–1068.

Boccotti, P. (2000). *Wave Mechanics for Ocean Engineering*. Elsevier Oceanography Series. Elsevier Science.

Brechmann, E. C. and Czado, C. (2015). "Copar—multivariate time series modeling using the copula autoregressive model." *Applied Stochastic Models in Business and Industry*, 31(4), 495–514.

Callaghan, D., Nielsen, P., Short, A., and Ranasinghe, R. (2008). "Statistical simulation of wave climate and extreme beach erosion." *Coastal Engineering*, 55(5), 375–390.

Cooke, R. M. (1997). "Markov and entropy properties of tree-and vine-dependent variables." *Proceedings of the ASA Section of Bayesian Statistical Science*, Vol. 27.

Corbella, S. and Stretch, D. D. (2012). "Predicting coastal erosion trends using non-stationary statistics and process-based models." *Coastal Engineering*, 70, 40–49.

Corbella, S. and Stretch, D. D. (2013). "Simulating a multivariate sea storm using archimedean copulas." *Coastal Engineering*, 76, 68–78.

De Michele, C., Salvadori, G., Passoni, G., and Vezzoli, R. (2007). "A multivariate model of sea storms using copulas." *Coastal Engineering*, 54(10), 734–751.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). "Selecting and estimating regular vine copulae and application to financial returns." *Computational Statistics & Data Analysis*, 59, 52–69.

Faltinsen, O. (1993). *Sea loads on ships and offshore structures*, Vol. 1. Cambridge University Press.

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

Gouldby, B., Méndez, F., Guanche, Y., Rueda, A., and Mínguez, R. (2014). "A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis." *Coastal Engineering*, 88, 15–26.

Haff, I. H., Aas, K., and Frigessi, A. (2010). "On the simplified pair-copula construction—simply useful or too simplistic?." *Journal of Multivariate Analysis*, 101(5), 1296–1310.

Hawkes, P. J., Gouldby, B. P., Tawn, J. A., and Owen, M. W. (2002). "The joint probability of waves and water levels in coastal engineering design." *Journal of Hydraulic Research*, 40(3), 241–251.

Holthuijsen, L. H. (2010). *Waves in oceanic and coastal waters*. Cambridge University Press.

Jäger, W. and Morales Nápoles, O. (2015). "Sampling joint time series of significant wave heights and periods in the north sea." *Safety and Reliability of Complex Engineered Systems*, L. Podofillini, B. Sudret, B. Stojadinović, E. Zio, and W. Kröger, eds.

Joe, H. (1996). "Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters." *Lecture Notes-Monograph Series*, 120–141.

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

Joe, H. and Krupskii, P. (2014). *CopulaModel: Dependence Modeling with Copulas*. R package version 0.6.

Karunarathna, H., Pender, D., Ranasinghe, R., Short, A. D., and Reeve, D. E. (2014). "The effects of storm clustering on beach profile variability." *Marine Geology*, 348, 103–112.

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

Leontaris, G., Morales-Nápoles, O., and Wolfert, A. R. (2016). "Probabilistic scheduling of offshore operations using copula based environmental time series–an application for cable installation management for offshore wind farms." *Ocean Engineering*, 125, 328–341.

Li, F., Van Gelder, P., Ranasinghe, R., Callaghan, D., and Jongejan, R. (2014a). "Probabilistic modelling of extreme storms along the dutch coast." *Coastal Engineering*, 86, 1–13.

Li, F., van Gelder, P., Vrijling, J., Callaghan, D., Jongejan, R., and Ranasinghe, R. (2014b). "Probabilistic estimation of coastal dune erosion and recession by statistical simulation of storm events." *Applied Ocean Research*, 47, 53–62.

Lin-Ye, J., Garcia-Leon, M., Gracia, V., and Sanchez-Arcilla, A. (2016). "A multivariate statistical model of extreme events: An application to the catalan coast." *Coastal Engineering*, 117, 138–156.

Masina, M., Lamberti, A., and Archetti, R. (2015). "Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves." *Coastal Engineering*, 97, 37–52.

Montes-Iturrizaga, R. and Heredia-Zavoni, E. (2016). "Multivariate environmental contours using c-vine copulas." *Ocean Engineering*, 118, 68–82.

Morales-Nápoles, O. (2011). "Counting vines." *Dependence Modeling: Vine Copula Handbook*, 189–218.

Morales-Nápoles, O., Cooke, R. M., and Kurowicka, D. (2010). "About the number of vines and regular vines on n nodes (working paper).

Nai Ruscone, M. and Osmetti, S. A. (2017). "Modelling the dependence in multivariate longitudinal data by pair copula decomposition." *Soft Methods for Data Science*, Springer, 373–380.

Nelsen, R. B. (2013). *An introduction to copulas*, Vol. 139. Springer Science & Business Media.

Poelhekke, L., Jäger, W. S., van Dongeren, A., Plomaritis, T. A., McCall, R., and Ferreira, Ó. (2016). "Predicting coastal hazards for sandy coasts with a bayesian network." *Coastal Engineering*, 118, 21–34.

Repko, A., Van Gelder, P., Voortman, H., and Vrijling, J. (2004). "Bivariate description of offshore wave conditions with physics-based extreme value statistics." *Applied Ocean Research*, 26(3), 162–170.

Rueda, A., Camus, P., Tomás, A., Vitousek, S., and Méndez, F. (2016). "A multivariate extreme wave and storm surge climate emulator based on weather patterns." *Ocean Modelling*, 104, 242–251.

Salvadori, G., Tomasicchio, G., and D'Alessandro, F. (2014). "Practical guidelines for multivariate analysis and design in coastal and off-shore engineering." *Coastal Engineering*, 88, 1–14.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T. (2015). *VineCopula: Statistical Inference of Vine Copulas*. R package version 1.6-1.

Serafin, K. A. and Ruggiero, P. (2014). "Simulating extreme total water levels using a time-dependent, extreme value approach." *Journal of Geophysical Research: Oceans*, 119(9), 6305–6329.

Serinaldi, F. and Grimaldi, S. (2007). "Fully nested 3-copula: procedure and application on hydrological data." *Journal of Hydrologic Engineering*, 12(4), 420–430.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Publications de l'Institut Statistique de Université de Paris 8.

Smith, M., Min, A., Almeida, C., and Czado, C. (2012). "Modeling longitudinal data using a pair-copula decomposition of serial dependence." *Journal of the American Statistical Association*.

Smith, M. S. (2015). "Copula modelling of dependence in multivariate time series." *International Journal of Forecasting*, 31(3), 815–833.

Tawn, J. A. (1988). "Bivariate extreme value theory: models and estimation." *Biometrika*, 75(3), 397–415.

van Gent, M., van Thiel de Vries, J., Coeveld, E., De Vroeg, J., and Van de Graaff, J. (2008). "Large-scale dune erosion tests to study the influence of wave periods." *Coastal Engineering*, 55(12), 1041–1051.

Vanem, E. (2016). "Joint statistical models for significant wave height and wave period in a changing climate." *Marine Structures*, 49, 180–205.

Wahl, T., Mudersbach, C., and Jensen, J. (2011). "Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a stochastic storm surge model." *Natural Hazards and Earth System Sciences*, 11(11), 2925–2939.

Wahl, T., Mudersbach, C., and Jensen, J. (2012). "Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on copula functions." *Natural Hazards and Earth System Science*, 12(2), 495–510.

Wahl, T., Plant, N. G., and Long, J. W. (2016). "Probabilistic assessment of erosion and flooding risk in the northern gulf of mexico." *Journal of Geophysical Research: Oceans*, 121(5), 3029–3043.

Yoshiba, T. (2015). "Maximum likelihood estimation of skew t-copulas, with its applications to stock returns." *The Institute of Statistical Mathematics Research Memorandum*, (1195).