Title	Modeling multivariate ocean data using asymmetric copulas
Author(s)	Zhang, Yi; Kim, Chul-Woo; Beer, Michael; Dai, Huliang; Soares, Carlos Guedes
Citation	Coastal Engineering (2018), 135: 91-111
Issue Date	2018-05
URL	http://hdl.handle.net/2433/235494
Right	© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/; The full-text file will be made open to the public on 01 May 2020 in accordance with publisher's 'Terms and Conditions for Self-Archiving'.; This is not the published version. Please cite only the published version. この論文は出版社版でありません。引用の際には出版社版をご確認ご利用ください。
Type	Journal Article
Textversion	author

Coastal Engineering 135 (2018) 91–111
https://doi.org/10.1016/j.coastaleng.2018.01.008

Modeling multivariate ocean data using asymmetric copulas

Yi Zhang[a,*], *Chul-Woo Kim*[a], *Michael Beer*[b], *Huliang Dai*[c], *Carlos Guedes Soares*[d]

[a] Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Japan

[b] Institute for Risk and Reliability, Leibniz Universität Hannover, Germany / Institute for Risk and Uncertainty, University of Liverpool, UK / Shanghai Institute of Disaster Prevention and Relief, Tongji University, China

[c] School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, China

[d]Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract

Multivariate descriptions of ocean parameters are quite important for the design and risk assessment of offshore engineering applications. A reliable and realistic statistical multivariate model is essential to produce a representative estimate of the sea state for understanding the ocean conditions. Therefore, an advanced modeling of ocean parameters helps towards improving ocean and coastal engineering practices. In this paper, we introduce the concepts of asymmetric copulas for the modeling of multivariate ocean data. In contrast to extensive previous research on the modeling of symmetric ocean data, this study is focused on capturing asymmetric dependencies among the environmental parameters, which are critical for a realistic description of ocean conditions. This involves particular attention to both nonlinear and asymmetrically dependent variates, which are quite common for the ocean variables. Several asymmetric copula functions, capable of modeling both linear and nonlinear asymmetric dependence structures, are examined in detail. Information on tail dependencies and measures of asymmetric dependencies are

[*] Corresponding author. Tel.: +81-075-383-3421
E-mail address: zhang_yi87@163.com; (Y. Zhang).

exploited. To demonstrate the advantages of asymmetric copulas, the asymmetric copula concept is compared with the traditional copula approaches from the literature using actual environmental data. Each of the introduced copula models is fitted to a set of ocean data collected from a buoy at the US coast. The performance of these asymmetric copulas is discussed and compared based on data fitting and tail dependency characterizations. The accuracy of asymmetric copulas in predicting the extreme value contours is discussed.

Keywords: ocean engineering, joint distribution, multivariate analysis, asymmetric copula

1. Introduction

Offshore and coastal structures facilitate the exploitation of the vast ocean resource, which contributes significantly to technological and economic development. Compared with normal structures on land, offshore structures are bulky, expensive and in most cases constructed in a complex marine environment. The marine environment for offshore structures can be severe, adverse, varying and uncertain. It covers a broad area of climatic factors, which generally include tide, current, wind, waves, ice and other sometimes devastating events such as hurricanes. When addressing different environmental risks for the offshore engineering applications designers are usually required to estimate the environmental conditions at the ocean site, and usually a multivariate analysis is performed (Zhang and Cheng, 2016). For example, the environmental contour method developed by Winterstein et al. (1993) is popular for this purpose. As a basis to produce realistic results it requires a reliable multivariate environmental model for finding the maximum system response associated with a given exceedance probability. In this context the interaction among various environmental influences plays an important role. In practical applications, offshore and coastal structures can suffer from severe damages because of the occurrence of critical combinations of the ocean environmental variables which coexist in extreme weather events such as sea storms (Zhang and Lam, 2014; 2015). In turn, deficiencies in modeling their joint statistics may severely overestimate the safety and effectiveness of coastal and offshore structures, hence lead to unsafe design and consequently lead to expensive and unexpected catastrophes (Bitner-Gregersen, 2015; Zhang et al. 2017a,b). Particularly, the modeling of the joint distribution of wave height and wave period is normally a must in marine engineering applications since the sea state at a specific location primarily depends on these two ocean parameters simultaneously (DNV, 2014). However, the ocean climate system is an extremely complex system that contains many more natural factors from both the ocean and the atmosphere. There are various sources of uncertainties and potential biases that influence the ocean conditions (Zhang and Cao, 2015). Specifically, the uncertainties related to the parameter dependencies are one of the most influencing factors. It was recognized that the understanding of the nonlinear dependency between ocean parameters remains one of the most difficult tasks, and the statistical modeling of the multivariate ocean data remains challenging due to their complicated relationships (Ewans and Jonathan, 2014).

Lots of attempts to cope with multivariate statistical analysis can be found in the literature, involving the use of a conditional distribution model (Lucas and Guedes Soares, 2015), a bivariate logistic model (Morton and Bowers, 1996), a Pareto distribution model (Muraleedharan et al., 2015) and so on. Clearly, the traditional conditional joint distribution model is only applicable to the multivariate problem to a certain extent. If the variables contain slightly more complex relationships such as non-constant correlation between pairs of random variables (e.g. the correlation coefficient value may change at extremes), the traditional joint statistical model is not appropriate any more. Therefore, many more advanced techniques have been employed in the multivariate analysis. Among all the developments, the application of copulas has been increasingly popular in recent years. Many initial studies have illustrated that application of copula theory can produce more realistic joint models for the ocean multivariate data. De Michele et al. (2007) have utilized copulas to characterize sea storms in terms of significant wave height, sea storm duration, sea storm inter-arrival (waiting) time, and sea storm direction. Antào and Guedes Soares (2014) have established copula based bivariate models for individual wave steepness and wave height. A similar copula model for description of water levels and waves are also presented by Masina et al. (2015). Montes-Iturrizaga and Heredia-Zavoni (2015) have proposed a formulation for expressing the environmental contours as functions of copulas and show that the dependence structure of sea state parameters can be well presented in this manner. Until recently, Jane et al. (2016) have employed the copula model to predict the wave height at a given location by considering the spatial dependence of the wave height at nearby locations. In contrast to the traditional joint model, a copula approach has the advantage that the dependency structure between the variables can be defined independently of the choice of the marginal distribution. This flexibility is highly desirable in modeling the environmental parameters as the natural factors often exhibit non-obvious dependencies. Moreover, it was also found that the copula model can save numerical effort when it is utilized to characterize the environmental loading in the offshore structural analysis. For instance, Zhang et al. (2015) have demonstrated an approach of using a copula model to characterize the sea load for the reliability analysis of a real jacket structure, which reduces the numerical effort by a factor of five. A practical guideline for using a copula in the design of coastal and offshore engineering applications can be found in Salvadori et al. (2014). Thorough guidelines involving the use of copulas in a structural approach are given in Salvadori et al. (2015). In general, from the recent advances in coastal engineering, it is now widely recognized that a copula approach is very efficient and powerful to model the statistical behavior of ocean dependent variables.

As exciting as the copula approach is, there are some obvious issues, which need to be addressed for a successful application. In former studies it was criticized that most families of parametric copulas can only model data having symmetric dependency (Genest and Favre, 2007). For example, the well established Archimedean copula families are all symmetric. If the data dependencies exhibit asymmetric behavior, the traditional copula model may no longer be adequate. Asymmetric Archimedean copulas are discussed in Grimaldi and Serinaldi (2006). Unfortunately, ocean data, fall into this category; they have been found as asymmetrically dependent in various previous studies. This is especially obvious for the sea state parameters, which are important for in engineering design (deWaal and van Gelder, 2005). Ignoring the asymmetric effects in the modeling of ocean data can be quite critical as it affects the estimates of the response statistics and eventually compromises the quality of the structural reliability assessment. A reason explaining the frequent (possibly unjustified) usage of symmetric (Archimedean) copulas might be that these are the ones provided by the Matlab package, the one traditionally used by maritime engineers. However,, asymmetric copulas can remedy this problem. Asymmetric copulas can be constructed based on the families of symmetric copulas. This compounded procedure can significantly improve the fit (Jondeau, 2016). The modeling of the ocean data utilizing the asymmetric copulas has received much attention recently (Vanem, 2016). The well known Khoudraji-Liebscher family, introduced in (Durante and Salvadori, 2010; Salvadori and De Michele, 2010), gives the possibility to construct asymmetric copulas. The application of this family in a maritime context has been mentioned in Salvadori et al. (2014, 2015). De Michele et al. (2013) have also used it for the modeling of drought. Besides, the conditional mixture construction (Vine copulas), first introduced in maritime engineering by De Michele et al. (2007), also provides the possibility to construct asymmetric copulas starting from symmetric ones. However, the theoretical concepts and procedures of constructing an asymmetric copula have not yet been studied in detail. Despite this, it is recognized that there are many candidate asymmetric copulas in theory. These choices provide potent features and practical meaning in ocean and coastal engineering applications. This potential can readily be utilized once the applicability of asymmetric copulas for the modeling of ocean data has been verified and demonstrated. We aim to contribute to this development with the present real case study for demonstrating and highlighting the features, merits as well as limitations associated with asymmetric copulas.

The remainder of this paper is organized as follows. Section 2 presents a general literature review of the existing techniques in modeling multivariate ocean data. Section 3 presents the fundamental knowledge of copula theory and the basic dependence measure concepts. Basic concepts of asymmetry measure as well as the procedures of constructing asymmetric copula models are explained in detail in Section 4. Specific asymmetric copula models for ocean data are developed in Section 5 and compared against traditional parametric copula models based on collected, preconditioned ocean data. To understand the features of using asymmetric copulas in the ocean data modeling, a comparative study between symmetric and asymmetric copula models is presented in Section 6. The concluding remarks of this paper form Section 7.

2. Joint statistical models for ocean data

Among the probabilistic models available in the literature, the most commonly recommended model adopted in offshore engineering design codes is the conditional joint distribution model, which is widely applied to various kinds of ocean data (Burton et al. 2001; Jonathan and Ewns et al. 2011; Ernst and Seume 2012). The most pertinent joint distribution model that is applied in ocean engineering is for the significant wave height and peak period, which characterize the spectrum of a sea state. For instance, Guedes Soares et al. (1988) and Bitner-Gregersen and Haver (1989) have demonstrated the use of a joint environmental model, which was constructed based on the combination of the marginal distribution of wave height and conditional distribution of the wave period. Later on, Ochi (1992) introduced a bivariate log-normal distribution in the modeling of the significant wave height and peak period. Generally, these conditional bivariate distribution models assume significant wave height follows a Weibull distribution while wave peak period follows a log-normal distribution whose model parameters are conditional on significant wave height. The primary reason for using such conditional distribution model is generally that the significant wave height is the most important parameter, which affects design conditions of ocean structures whereas other parameters have less influence. This also agrees well with the practice design code (DNV 2010), which utilizes a bivariate conditional model for the wave height and wave period. This concept of conditional distribution models can also be extended to modeling other ocean parameters.

Beyond these fundamental developments, many of the current studies demonstrate that the joint models could be further developed. Prince-Wright (1995) showed that a multivariate model using the Box and Cox transformation can model the collected ocean data well, especially under the presence of non-stationarities. This has been proven through a comparative study done by Bitner-Gregersen et al. (1998), in which the analysis has been applied to a dataset containing both wave and wind. Repko et al. (2004) developed a bivariate model of significant wave height and peak period based on a given independent value of wave steepness. A comparison of several of these approaches with maximum entropy principle, including clarifying the differences and performance in estimating a return value, by utilizing a specific data set can also be found in (Dong et al. 2013). Ewans and Jonathan (2014) have incorporated the offshore structural response properties in the multivariate sea state parameter modeling and utilized the reliability concepts to derive the response based environment contours. The conditional multivariate extreme models for ocean parameters considering covariate effects in directionalities are also discussed in Jones et al. (2016). An overview of different methods for multivariate modeling of ocean data such as the conditional modeling approach can be found from Ferreira and Guedes Soares (2002) and Jonathan et al. (2010). Although the use of a conditional joint distribution model is quite convenient, its drawback is also very obvious. That is, the marginal distributions and the dependence structure are both defined within one bivariate model, which reduces the degree of freedom of the model. In fact, it should be noticed that conditional models are simple special cases of copula models. As explained later via the Sklar's Theorem representation (see Section 3.1 and Eq. (3)), the bivariate joint density f_{XY} is simply given by

$$f_{XY}(x,y) = c_{XY}\left(F_X\left(x\right), F_Y\left(y\right)\right) \cdot f_X\left(x\right) \cdot f_Y\left(y\right), \tag{1}$$

where c_{XY} (.) is the copula density, F_X, F_Y are the univariate distribution functions, and f_X, f_Y are the univariate marginal densities. In turn, the conditional density used in conditional bivariate models is a function of the copula at play, i.e.

$$f_{X|Y}(x\mid y) = \frac{f_{XY}(x,y)}{f_Y\left(y\right)} = c_{XY}\left(F_X\left(x\right), F_Y\left(y\right)\right) \cdot f_X\left(x\right). \tag{2}$$

Here, the point is that, in general, it is easier to identify/construct a bivariate copula model than a conditional model, especially when the sample size is scarce. In addition, changing the marginals in a copula model is easy, whereas it may be awkward in a conditional model. This point has also been mentioned by many other studies, which also suggested further development of the conditional bivariate models.

Besides the traditional joint models, some researchers have dedicated their efforts to the study of establishing multivariate models using transformation approaches. Quite popular is the Nataf transformation approach to construct joint probability models in offshore engineering applications (Nataf, 1962). Wist et al. (2004) have applied a Nataf model to capture successive wave heights and found that they can be well approximated by this joint bivariate model. A more complicated methodology has been presented by Sagrilo et al. (2011) for creating a Nataf model which includes the wave, wind and water current parameters. It is also applied in a structural reliability analysis where an environmental contour is estimated from the Nataf model (Silva-González et al. 2013). More recently, the maximum entropy distributions, which are developed based on Nataf transformation, have been utilized in the modeling of wave height and wind speed (Dong et al. 2015). The features of the approximation for the distribution of the physical variables depend on whether the vector of the transformed standard normal variables is close to being multi-normal. Otherwise, it is criticized that certain transformation procedures might not be necessary (Huseby et al., 2013). For comparison, a 4-dimensional model was easily constructed in De Michele et al. (2007) via conditional mixtures of copulas (now generalized by the so-called Vine copulas).

Generally, under certain conditions each of the provided models has its own advantages. Quite a few of these models are flexible enough to provide a realistic characterization of ocean parameter dependencies under various conditions. With the aim of advancing the field of offshore reliability engineering, there is a strong need for establishing a multivariate model that can handle nonlinear dependencies. We found that only little attention has been devoted to the research of non-symmetric multivariate models for ocean parameters. Studies in this direction with greater depth seem useful in order to improve the modeling of dependencies in ocean data.

3. Copula theory and dependence measures

An alternative modeling approach to the multivariate ocean data is to use copulas for constructing multivariate data. Copulas provide a powerful tool for modeling multivariate data, and are widely used in Finance and Economics (see, e.g., Cherubini et al. 2004; McNeil et al. 2005), as well as in Hydrology and Environmental Sciences. In this latter instance, as seminal references, the following ones provide a thorough survey: for a theoretical introduction see Nelsen (2006); Joe (2014); Durante and Sempi (2015), for a practical engineering approach see Genest and Favre (2007); Salvadori et al. (2007); Salvadori and De Michele (2007).

3.1 Definition and basic properties

Copula is a model, which “couples” univariate marginal distributions to form a multivariate distribution. In theory, a copula model is constructed by combining the marginal distributions of variables and a specific dependence structure. The formal definition of a copula as a multivariate distribution with specified marginal distributions is originally introduced in Sklar’s theorem (Sklar, 1959):

Sklar's Theorem: Let F be an n-dimensional distribution function with marginal distributions $F_1, \ldots, F_n$. There exists an n-dimensional copula C such that for all $x \in R^n$

$$F\left(x_1,\ldots,x_n\right)=C\left(F_1\left(x_1\right),\ldots,F_n\left(x_n\right)\right) \tag{3}$$

If $F_1, \ldots, F_n$ are all continuous, then C is unique. Conversely, if C is a copula and $F_1, \ldots, F_n$ are distribution functions, then the function defined in Eq. (3) is a multivariate distribution function with marginal distributions $F_1, \ldots, F_n$.

On account of Sklar's theorem, it is easy to see that the copula model does not need to consider the characteristics of the individual random variables in the multivariate problem. This results from the probability integral transform which states that the random variables $U_i=F_i(X_i)$ are uniformly distributed on [0, 1]. Note that the probability integral transform works for continuous random variables, since F_i should be invertible. In other words, the copula model is a multivariate model for all the variables after their transformation through the cumulative distribution function. Copula is a multivariate cumulative distribution function with uniform marginals. Hence, the domain and the range values for an n-dimensional copula function are

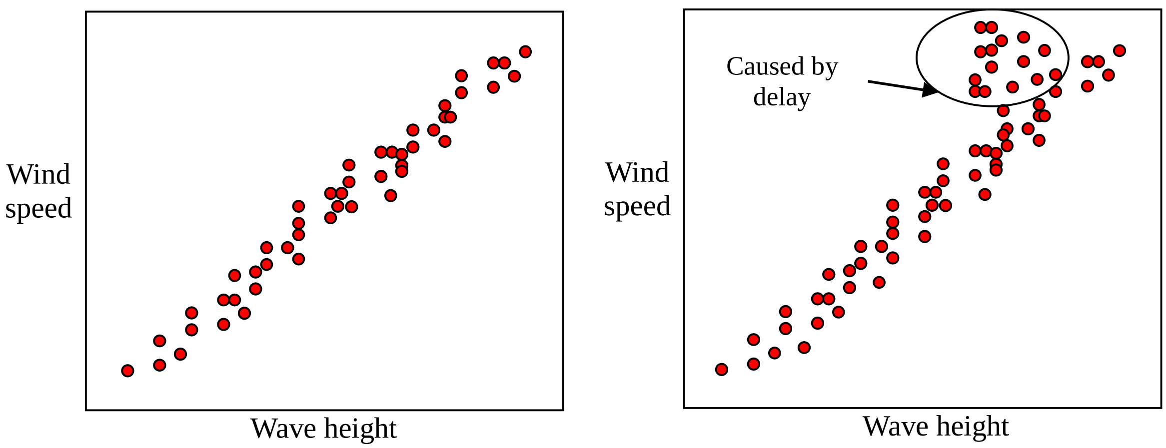
$$C:\left[0,1\right]^n \rightarrow \left[0,1\right].$$

Compared to the other joint distribution models, the copula approach has the advantage that the dependency structure between the variables can be defined independently of the choice of the marginal distributions. The freedom of selecting any marginal distributions for the variables in a copula makes this approach quite flexible in characterizing individual variable's behaviors. This flexibility is highly desirable in the modeling of ocean parameters as the environmental factors often show non-obvious dependencies. Various kinds of parametric copula families and classes can be found in the literature, see e.g. (Hutchinson and Lai 1990; Nelsen 2006; Salvadori et al. 2007; Joe 2014). Each family or class of copulas can characterize a certain kind of dependence in multivariate data. Most of these copulas are used for bivariate data. However, they can be expanded to a multivariate model through straightforward transformations. Examples of copulas, like Archimedean copulas, are presented in Appendix B.

3.2 Dependence measures

To highlight the significance of the copula approach in offshore engineering applications, a detailed interpretation of dependence concepts is provided herein. The dependence structure is generally the most important characteristic in a copula model. The most common and convenient way for measuring the data dependence is using the Pearson's correlation coefficient ρ. It is widely applied in many statistical approaches because of its simplicity and ease of handling. However, the weakness of ρ is also obvious as it can only represent linear dependencies. Therefore, other concepts of dependencies such as Kendall's τ_k and Spearman's ρ_s were introduced in the literature (Joe, 2014). Compared to Pearson's coefficients, these two dependency measure concepts are much more robust. Kendall's τ_k measures the possible excess of concordance/discordance in the sample, while Spearman's ρ_s is a measure of the "distance" (in the L^1 integral sense) between the chosen copula and the one modeling independent variables (see Salvadori et al. (2007)). Pearson coefficient requires the existence of the second order moments, and may depend on the marginals, whereas Kendall and Spearman ones are nonparametric measures of association and do always exist. Furthermore, the influence of unequal variances, outliers and non-linearity, which could cause distortions in Pearson's correlation coefficient, are greatly minimized in Kendall's τ_k and Spearman's ρ_s. In other words, from engineering point of view, Kendall's and Spearman's dependencies are more focusing on the concordance of the ranking whereas Pearson's dependence is focusing on the value. Copula model can describe various kinds of dependencies which include association concepts such as concordance, linear correlation and other related measures. A copula is thus much more flexible than traditional concepts for characterizing dependencies and includes these concepts. However, there are a few issues associated with traditional copulas (e.g. Archimedean copulas) when they are applied to ocean data. A key drawback is that some copulas are symmetric while most ocean data display non-symmetric dependencies. The reason for these asymmetric dependencies can be summarized as follows:

- Different ocean variables respond differently to the same environment conditions. For example, when there is a hurricane, the wind speed is the most directly affected variable. A sudden increase of wind speed is expected to be observed. The value of wind speed has a very quick response to the hurricane. However, this effect may not be reflected in the wave height values instantaneously. There is normally a delay in the observed wave height due to the change of wind speed. Such a delay causes some deviations in the co-movement of wave height and wind speed time series data. Thus, a nonlinear dependency is observed in the data and asymmetric dependencies evolve. This effect can be illustrated by means of scatter plots as shown in Fig. **1**.



Perfect linear relationship

Nonlinear relationship and asymmetric dependency

Figure **1** Asymmetric dependency of ocean data caused by delay.

- The feasible domain of parameters restricted by the physical phenomenon is another significant reason for asymmetric dependencies. For example, a large value of wave height is unlikely to be accompanied by a small wave period because of the breaking wave limit. In other words, the observation of some variable combinations is physically impossible. This effect is illustrated in Fig. 2. As can be seen, there is no observation in the right-lower region (marked with a cross), while observations can be seen in the left-upper region (marked with a tick). That is, implicit physical phenomena can limit the possibility of occurrence for some data combinations. The feasible domain is therefore reduced and becomes asymmetric. Implementing such physical feature in the copula model is not straightforward and is still in its infancy stage of development.

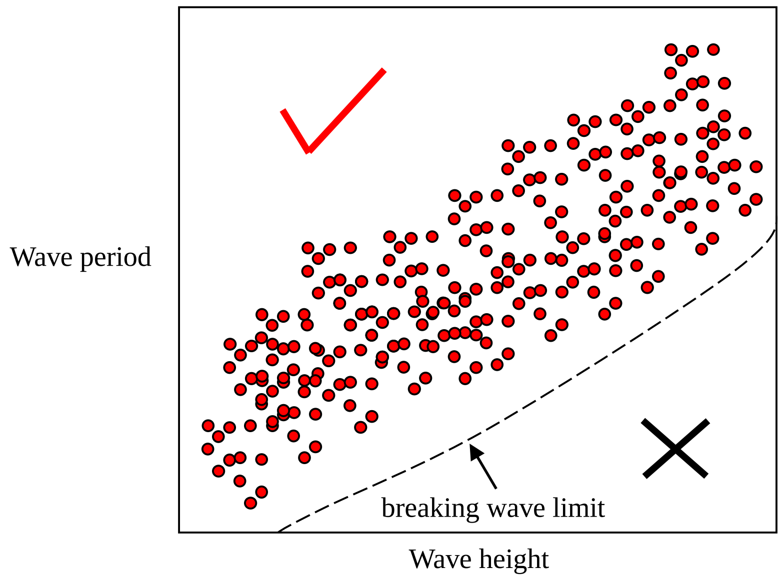


Figure **2** Asymmetric domain of wave data caused by physical phenomenon.

These effects can be commonly observed in most collected ocean datasets. The ignorance of such asymmetric dependencies would lead to unreliable multivariate models and associated estimates, for example, of long term return values. However, the traditional copula approach is not capable of handling these effects efficiently. A more advanced statistical approach is therefore required.

4. Asymmetric copulas

In general, as discussed in the above, the most commonly used copulas cannot meet the current needs for modeling ocean data. An accurate modeling of the asymmetric dependencies is highly demanded. To remedy this problem, we introduce several groups of asymmetric copulas as well as the basic concepts in measuring the asymmetry of a copula.

4.1 Measure of asymmetry and tail dependency

First of all, it is necessary to introduce the basic definition of symmetry for the copula model. For a given copula $C(u_1, \dots, u_n)$, if

$$C\left(u_1,\ldots,u_i,\ldots,u_j,\ldots,u_n\right)=C(u_1,\ldots,u_j,\ldots,u_i,\ldots,u_n) \text{ is true for any pair } u_i,u_j \in \mathbf{I},$$

then u_i and u_j are said to be exchangeable and the copula $C(u_1,\ldots,u_n)$ is said to be symmetric (Genest and Nešlehová, 2013). On the other hand, if a copula function does not satisfy the above condition, it is considered to be asymmetric. Based on this idea, a measure of asymmetry in a copula was developed by Klement and Mesiar (2006).

For instance, a natural measure of asymmetry for a 2-dimensional copula can be calculated by the following equation Klement and Mesiar (2006)

$$\eta_p\left(C\right)=\left\{\int_0^1\int_0^1\left|C\left(u_1,u_2\right)-C\left(u_2,u_1\right)\right|^p du_1du_2\right\}^{1/p} \tag{4}$$

where p can be set at any value that is greater than or equal to 1, $p \geq 1$. In other words, the measure of asymmetry is represented as the distance between C and its transpose C^T, like the norm. Moreover, it is usually more convenient to compute the value when p approaches infinity, which leads to

$$\eta_\infty\left(C\right)=\sup_{(u_1,u_2)\in[0,1]^2}\left|C\left(u_1,u_2\right)-C\left(u_2,u_1\right)\right|. \tag{5}$$

Thus, when this measure goes up to a certain value, the copula is considered as non-exchangeable and is understood to be asymmetric. In the observation of bivariate ocean data, the measure of asymmetry as calculated by Eq. (5) could serve as a measure of exchangeability for the data. An extension of Eq. (4) and Eq. (5) for measuring the asymmetry of two variables in high dimensional copula can also be derived as following

$$\eta_\infty\left(C\right)=\sup_{(\boldsymbol{u},\boldsymbol{u}_\pi)\in[0,1]^n}\left|C\left(\boldsymbol{u}\right)-C\left(\boldsymbol{u}_\pi\right)\right|, \tag{6}$$

where $\boldsymbol{u}=(u_1,\ldots,u_n)$ and $\boldsymbol{u}_\pi=(u_{\pi(1)},\ldots,u_{\pi(n)})$, $\pi \in S_n$ is a permutation of $\{1,\ldots,n\}$. Generally, it calculates the maximal distance between a copula and a version of it where the arguments are permuted (Harder and Stadtmüller, 2013).

Besides the measure of asymmetry, the asymmetric characteristics can also be observed from the differences in the tail dependencies. According to the coefficient of tail dependence proposed by Joe (2014), there are four coefficients that can be used to describe the tail dependence for bivariate data, namely, *lower-lower*, *lower-upper*, *upper-lower*, *upper-upper* tail dependence coefficients. For example, if an *n*-dimensional copula $C(u_1, \dots, u_n)$ is given, these tail coefficients can be calculated as follows:

$$\lambda_{i|j}^{l,l}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u), \dots, x_i \le F_i^{-1}(u), \dots, x_n \le F_n^{-1}(u) \mid x_j \le F_j^{-1}(u)\right), \quad (7)$$

$$\lambda_{i|j}^{l,u}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u), \dots, x_i \ge F_i^{-1}(1-u), \dots, x_n \le F_n^{-1}(u) \mid x_j \le F_j^{-1}(u)\right), \quad (8)$$

$$\lambda_{i|j}^{u,l}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u), \dots, x_i \le F_i^{-1}(u), \dots, x_n \le F_n^{-1}(u) \mid x_j \ge F_j^{-1}(1-u)\right), \quad (9)$$

$$\lambda_{i|j}^{u,u}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u), \dots, x_i \ge F_i^{-1}(1-u), \dots, x_n \le F_n^{-1}(u) \mid x_j \ge F_j^{-1}(1-u)\right), (10)$$

where $F_1^{-1}(.), \dots, F_n^{-1}(.)$ are inverse marginal distribution functions for the variables $x_1, \dots, x_n$. Therefore, these four equations provide measures of the dependence in four different tails of two variables in a multivariate setting. More commonly, if a bivariate copula $C(u_1, u_2)$ is analyzed, the tail dependence can be derived as (Nelsen (2006))

$$\lambda_{1|2}^{l,l}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u) \mid x_2 \le F_2^{-1}(u)\right) = \lim_{u \to 0+} \frac{C(u,u)}{u}, \quad (11)$$

$$\lambda_{1|2}^{l,u}(C) = \lim_{u \to 0+} P\left(x_1 \ge F_1^{-1}(1-u) \mid x_2 \le F_2^{-1}(u)\right) = 1 - \lim_{u \to 0+} \frac{C(u,1-u)}{u}, \quad (12)$$

$$\lambda_{1|2}^{u,l}(C) = \lim_{u \to 0+} P\left(x_1 \le F_1^{-1}(u) \mid x_2 \ge F_2^{-1}(1-u)\right) = 1 - \lim_{u \to 0+} \frac{C(1-u,u)}{u}, \quad (13)$$

$$\lambda_{1|2}^{u,u}(C) = \lim_{u \to 0+} P\left(x_1 \ge F_1^{-1}(1-u) \mid x_2 \ge F_2^{-1}(1-u)\right) = 2 - \lim_{u \to 0+} \frac{1 - C(1-u,1-u)}{u}. \quad (14)$$

The value range of these four coefficients is [0, 1], i.e. $\lambda^{l,l}, \lambda^{l,u}, \lambda^{u,l}, \lambda^{u,u} \in [0,1]$ where a value of 0 indicates asymptotical independence. Equations (11) and (14) are also known as coefficients of upper and lower tail dependence. For multivariate case, if the copula function is known, Eqs. (7) to (10) can be further expressed as follows

$$\lambda_{i|j}^{l,l}(C) = \lim_{u \to 0+} \frac{C(u, \dots, u)}{u}, \quad (15)$$

$$\lambda_{i|j}^{l,u}(C) = \lim_{u \to 0+} \frac{\breve{C}_i(u,\ldots,u)}{u}, \quad (16)$$

$$\lambda_{i|j}^{u,l}(C) = \lim_{u \to 0+} \frac{\breve{C}_j(u,\ldots,u)}{u}, \quad (17)$$

$$\lambda_{i|j}^{u,u}(C) = \lim_{u \to 0+} \frac{\breve{C}_{j,i}(u,\ldots,u)}{u}, \quad (18)$$

where $\breve{C}_i(.)$ and $\breve{C}_j(.)$ are copulas modified from the base copula $C(.)$ which can be shown by the following relationship

$$\breve{C}_k(u_1,\ldots,u_n) = C(u_1,\ldots,u_{k-1},1,u_{k+1},\ldots,u_n) - C(u_1,\ldots,u_{k-1},1-u_k,u_{k+1},\ldots,u_n), \quad (19)$$

and $\breve{C}_{j,i}(.)$ can be expressed by

$$\breve{C}_{j,i}(u_1,\ldots,u_n) = \breve{C}_j(u_1,\ldots,u_{i-1},1,u_{i+1},\ldots,u_n) - \breve{C}_j(u_1,\ldots,u_{i-1},1-u_i,u_{i+1},\ldots,u_n). \quad (20)$$

Tail dependencies can help to understand differences in the dependence structure for different tails. This provides useful information about the properties of extreme values from the intrinsic dependencies. In other words, tail dependencies provide a measure for relating one margin exceeding a certain quantile threshold while the other has already exceeded that quantile threshold. When assessing the asymmetry of a copula, the *lower-upper* and *upper-lower* tail coefficients can be utilized. The special case of a symmetric copula is included in this model. A symmetric copula can have its variables exchanged, the copula function values $C(u, 1-u)$ in Eq. (12) and $C(1-u, u)$ in Eq. (13) are identical. Further, for the symmetric case, the value of the *lower-upper* tail coefficient equals the *upper-lower* tail coefficient. If these coefficients are different, the copula is asymmetric. However, it should be noticed that the number of coefficients will grow exponentially as the dimension increases, and the interpretation of each coefficient becomes more difficult and indeterminate/vague. More detailed explanations are required when tail dependences are calculated for high dimensional data.

4.2 Asymmetric copulas constructed by products

The construction of asymmetric copulas can be pursued in various ways. In recent years, many methods have been developed in this direction (Grimaldi and Serinaldi, 2006; Mesiar and Najjari, 2014; Mazo et al., 2015). These include plenty of techniques that are utilized to capture the asymmetric dependencies in the multivariate data (Patton, 2006). However, not all of the asymmetric copulas can be practically applied. The application of some asymmetric copulas may need extra functions to characterize the complex dependencies. For instance, the well known Archimax copula, which is proposed by Capéraà et al. (2014), needs to have the Pickands dependence function for its construction. The construction of Pickands dependence is quite difficult and sometimes required complex statistical derivations (Pickands, 1981). Therefore, from a practical point of view, the most popular and practical alternatives among these asymmetric copulas are reviewed in this study. We choose to focus on the asymmetric copula families that can be easily constructed from various base copulas, e.g. Archimedean copulas. Asymmetric copulas with a very complicated way of construction are not explored in the present study.

One popular construction principle for asymmetric copulas is to formulate a product of copulas (see Liebscher, (2008)). A general form to obtain an asymmetric copula is

$$C_{product}\left(u_1,...,u_n\right)=\prod_{i=1}^{k} C_i\left(f_{i1}\left(u_1\right),..., f_{in}\left(u_n\right)\right) , \tag{21}$$

where $C_1, \dots, C_k$: $[0,1]^d \rightarrow [0,1]$ are all *n*-dimensional copulas, f_{ij}: $[0,1] \rightarrow [0,1]$ for *i*=1,…,*k*, *j*=1,…,*n* are functions that are strictly increasing or identically equal to 1. To ensure this product of copulas is also a copula, the functions f_{ij} have to satisfy the following additional properties:

1. $f_{ij}(1) = 1$ and $f_{ij}(0) = 0$,
2. f_{ij} is continuous on $(0,1]$,
3. If there are at least two functions $f_{i_1 j}, f_{i_2 j}$ with $1 \leq i_1, i_2 \leq k$ which are not identically equal to 1, then $f_{ij}(x) > x$ holds for $x \in (0,1)$, *i*=1,…,*k*.

From the above it is easy to see that the constructed copula is generally an asymmetric copula. The properties of these asymmetric copulas are derived from the fundamental properties of copula model. All the functions f_{ij} play a role in the asymmetric dependence modeling. This technique is also known as an extension of Khoudraji's device (1995). For example, by utilizing type I individual function (see Table 1) and set k, n=2, Eq. (21) becomes exactly the Khoudraji copula. The n-dimensional copulas $C_1, \ldots, C_k$ can be selected from various groups of parametric copulas, e.g. the Gumbel, the Clayton, the Frank, the Gaussian and etc. It is also possible to use independent or Fréchet-Hoeffding bounds for the individual copulas. As for the individual functions f_{ij}, Liebscher (2008) has provided a list of candidate functions which are suitable for the application. The most applicable individual functions are presented in Table 1. This flexibility can allow this asymmetric copula to be extended to much more complex multivariate models. However, even with these individual functions provided, the number and type of individual copulas are still unknown and need to be decided. Moreover, more advanced numerical methods are required for the simulation and use of this type of copula. Fortunately, certain simulation techniques have already been developed and utilized in the statistical analysis software. For example, the simulation of Khoudraji copula can be easily done by using a package named "copula" in R (Hofert et al., 2016). The use of Khoudraji-Liebscher copulas has already been very popular in the hydrology community, for example, both in terrestrial hydrology (e.g., Durante and Salvadori (2010); De Michele et al. (2013)) and in maritime hydrology (e.g., Salvadori et al. (2013, 2014, 2015)).

Table 1 Examples of individual functions

Individual function	Parameters	Value range
I. $f_{ij}(u) = u^{\theta_{ij}}$	$\sum_{i=1}^{k}\theta_{ij} = 1$	$\theta_{ij} \in [0,1]$
II. $f_{ij}(u) = u^{\theta_{ij}} e^{(u-1)\alpha_{ij}}$	$\sum_{i=1}^{k}\theta_{ij} = 1$, $\sum_{i=1}^{k}\alpha_{ij} = 0$	$\theta_{ij} \in (0,1)$, $\alpha_{ij} \in (-\infty,1)$, $\theta_{ij} + \alpha_{ij} \geq 0$
III. * $f_{1j}(u) = \exp\left(\theta_j - \sqrt{\lvert\ln u\rvert + \theta_j^2}\right)$, $f_{2j}(u) = u\exp(-\theta_j + \sqrt{\lvert\ln u\rvert + \theta_j^2})$	θ_j for $j \in \{1,\ldots,n\}$	$\theta_j \geq \frac{1}{2}$

*Note: type III individual functions can only be used for the asymmetric copula having two individual copulas (e.g. k=2).

4.3 Asymmetric copulas constructed by linear convex combinations

Another way of an algebraic construction of an asymmetric copula is by linear convex combinations of copulas. However, the direct linear convex combination of copulas is not suitable to create asymmetric copulas. Since most fundamental copulas are symmetric, the linear convex combination of these copulas would also only produce symmetric copulas. Wu (2014) has proposed a way to modify the fundamental copulas in order to account for asymmetric properties. In his theorem, a new kind of copula is proposed as

$$\breve{C}_h\left(u_1,\ldots,u_n\right)=C\left(u_1,\ldots,u_{h-1},1,u_{h+1},\ldots,u_n\right)-C\left(u_1,\ldots,u_{h-1},1-u_h,u_{h+1},\ldots,u_n\right) \qquad (22)$$

where $C(.)$ is the original n-dimensional base copula. It can be seen that the variable u_h is not exchangeable with other variables in the developed copula. This type of copula is also known as flipped copula as specified by Salvadori et al. (2007). Such copulas which have flipped dependence structures are already available in the R package "copula" (e.g. *rotCopula*()).

Therefore, with such amendment, $\breve{C}_h(.)$ can be used to fit data exhibiting unequal tail dependencies along the hth variable. Furthermore, in order to model asymmetric properties in multiple variables, one may use the following equation to construct the copulas:

$$C_{addition}\left(u_1,\ldots,u_n\right)=\sum\nolimits_{h=0}^{n} p_h\breve{C}_h\left(u_1,\ldots,u_n\right) \qquad (23)$$

where p_h is a weighting factor satisfying the conditions $0 \leq p_h \leq 1$ and $\sum_{h=0}^{n} p_h = 1$. When h=0, $\breve{C}_0(u_1,\ldots,u_n) = C(u_1,\ldots,u_n)$. That is, an asymmetric copula is obtained by linear convex combinations of copulas. The compound copula is now a combination of various base copulas with different individual tail dependencies. As in the approach in Section 4.2, a large group of copula families can be selected for the base copula $C(u_1,\ldots,u_n)$. For the case of a bivariate copula $C(u_1,u_2)$, Eq. (22) can be further expressed as

$$\breve{C}_1\left(u_1,u_2\right)=u_2-C\left(1-u_1,u_2\right), \qquad (24)$$

$$\breve{C}_2\left(u_1,u_2\right)=u_1-C\left(u_1,1-u_2\right), \qquad (25)$$

which can also be called the horizontal- and vertical-flipped copulas (Salvadori et al. 2007). Therefore, the constructed asymmetric copula can be generally written as

$$C_{addition}\left(u_1,u_2\right)=p_0C\left(u_1,u_2\right)+p_1\breve{C}_1\left(u_1,u_2\right)+p_2\breve{C}_2\left(u_1,u_2\right) \qquad (26)$$

where $p_0,p_1,p_2 \geq 0$ and $p_0+p_1+p_2=1$. Using this formula, we can easily adjust the values of weight factors assigned to each base copula in order to characterize the asymmetry properties of bivariate data along different variables. In other words, the individual copula $\breve{C}_1(u_1,u_2)$ or $\breve{C}_2(u_1,u_2)$ can only capture the asymmetry in one variable. As such, we can also point out the differences between the current construction method and Liebscher's method. That is, the current method constructs asymmetric copulas that present the asymmetric property in one variable each at a time, whereas Liebscher's method constructs the copulas for variables having asymmetric properties all at a time.

4.4 Skewed copula

Besides the algebraic construction methods, another way of modeling asymmetrically dependent data is utilizing skewed copulas. This approach originated from skewed multivariate distributions and generalizes the original distribution to allow non-zero skewness. The idea is to transform a multivariate distribution to an asymmetric one by introducing a parameter, which can regulate the skewness (Koll et al., 2013). However, there are only few skewed copulas available in the literature. The most popular one is the *skew Gaussian copula*.

Before introducing the skew Gaussian copula, we recall some basics about the Gaussian copula. An n-dimensional Gaussian copula is defined by

$$C_{Gaussian}\left(u_1,...,u_n\right)=\Phi_n\left(\Phi^{-1}\left(u_1\right),...,\Phi^{-1}\left(u_n\right);\mathbf{\Sigma}\right) \quad (27)$$

where $\Phi^{-1}(.)$ denotes the inverse of the standard normal distribution function, $\Phi_n(.)$ represents the n-dimensional normal distribution function, and $\sum$ stands for the covariance matrix. The Gaussian copula is a member of the elliptical copula family. The function is very like a multivariate normal distribution function and therefore can only be used to model variables having symmetric dependencies. To overcome this limitation, the basic formula is modified to account for asymmetries based on skew Gaussian distribution functions. A general n-dimensional skew Gaussian copula is given by

$$C_{skew-Gaussian}\left(u_1,...,u_n;\mu,\Sigma,\beta\right)=F_{n,skew}\left(F_{1,skew}^{-1}\left(u_1;\mu_1,1,\beta_1\right),...,F_{1,skew}^{-1}\left(u_n;\mu_n,1,\beta_n\right);\mu,\Sigma,\boldsymbol{\beta}\right) \quad (28)$$

where $F_{1,skew}^{-1}(.)$ is the inverse of the univariare skew normal distribution $SN(\mu_i, 1, \beta_i)$, $F_{n,skew}(.)$ is the n-dimensional skew normal distribution with mean parameter μ**,** shape parameters $\boldsymbol{\beta}$ and covariance matrix $\sum$. The density function of a multivariate skew normal distribution for n-dimensional random variables

$\boldsymbol{X}(x_1,\ldots,x_n)$ is given by

$$f_n\left(X;\mu,\Sigma,\beta\right)=2\phi_n\left(X;\mu,\Sigma\right)\Phi_n\left(\boldsymbol{\beta}^T X;\mu,\Sigma\right) \tag{29}$$

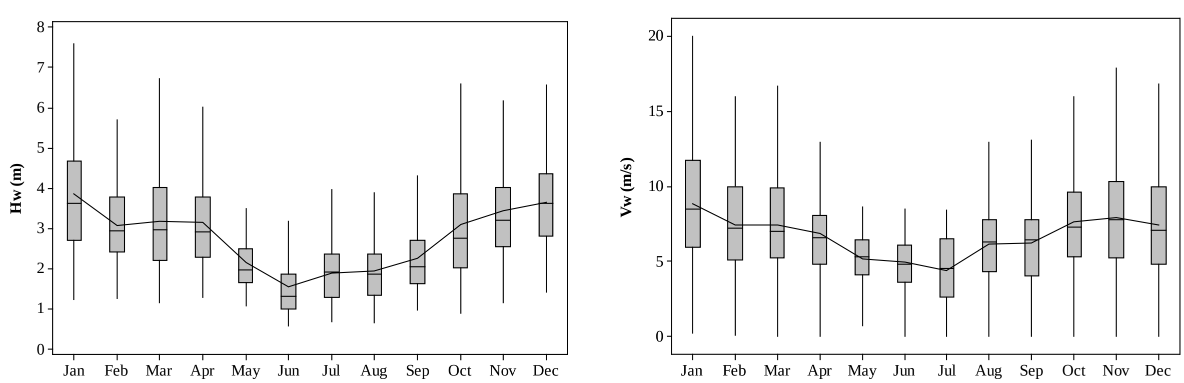
where $\phi_n(\cdot;\mu,\sum)$ and $\Phi_n(\cdot;\mu,\sum)$ represent the probability density function and cumulative distribution function of n-dimensional normal distribution with mean μ and covariance $\sum$ (Azzalini and Dalla Valle, 1996). Usually, for the ease of modeling, the mean values are all set at zero. The asymmetric property thus only results from the shape parameters. When $\boldsymbol{\beta}$=0, the copula becomes the standard Gaussian copula with no skewness. If $\boldsymbol{\beta}$ increases, the skewness of the distribution increases. Once $\boldsymbol{\beta}$ changes its sign, the skewness is reflected in the opposite side of the axis. The asymmetric properties can be characterized by the shape parameters either for the marginals or for the multivariate distribution.

From the comparison between Eq. (28) and Eq. (21), it can be seen that the skew copula is in fact a special case of the constructed copulas as given in Section 4.2. Compared to the general form, the skew copula has only one individual copula (k=1) and this individual copula (C_i) and the individual function (f_{ij}) are coming from the same family (skew Gaussian distribution). However, it is still interesting to discuss the use of skew copulas since no previous work has been done on its application in the modeling of ocean data.

5. Data analysis

To demonstrate the advantages of asymmetric copulas over the other models in a real case application, a comparative study is performed based on ocean data from the National Data Buoy Center, US (NDBC, 2016). The data were collected at a site in the Aleutian Trench, off the south coast of Alaska (52.785˚N 155.047˚W Buoy No. 46066) which has a water depth of 4545 m. The hourly recorded ocean data from the years 2014 and 2015 are extracted for the investigation (2014/1/1 01:00-2015/12/31 23:00). We choose to study three ocean parameters: significant wave height (H_w), average wave period (T_a) and wind speed (V_w). The unit of the measured wave height is in meters, while the units of wave periods and wind speed are second and meter per second, respectively. Here, only non-braked waves are recorded. The record of the ocean data shows a clear seasonal variation as depicted in Fig. 3. The data indicate more severe conditions in winter compared to summer.

As the time varying feature of the data record is very obvious, it is not reasonable to consider a simple statistical analysis to all the multivariate data at one time. A data partitioning is necessary to separate different groups of data for the analysis. To simplify the problem, we analyze a specific period of data which are critical and short. On this basis we can assume that all ocean data are quasi stationary for the statistical analysis. In this study, the ocean data covering the most severe period from November to February are chosen for the investigation. Figure 3 shows the variations of the mean and standard deviation of H_w, T_a and V_w over this period. However, the statistical testes are required to check whether the data are significantly different from each other in different months. Here, *t*-tests are applied to test whether the data of H_w, T_a and V_w are significantly different in different months. The highest p-values for H_w, T_a and V_w are 0.324, 0.187 and 0.443 which imply that the hypothesis of data from different months show statistically different means is rejected. Therefore, it is believed that the data within the period from November to February may approximately be represented by the same statistical model. Although the data within this period might not be perfectly homogeneous, the time varying effects associated with H_w, V_w and T_a can be neglected. Therefore, the data set (H_w, V_w, T_a) from the four months winter season in 2014 and 2015 are used for the subsequent statistical analysis. However, it should be pointed out the ocean data at this ocean site for the whole period is not completely collected. It has an amount of missing data in the hourly record. Only 3910 out of 5952 observations are collected. The detailed information and a general statistical summary of H_w, V_w and T_a is provided in Table 2. The differences in statistical properties between different ocean data are large. Individual characteristics of the ocean parameters H_w, V_w and T_a have to be investigated separately.



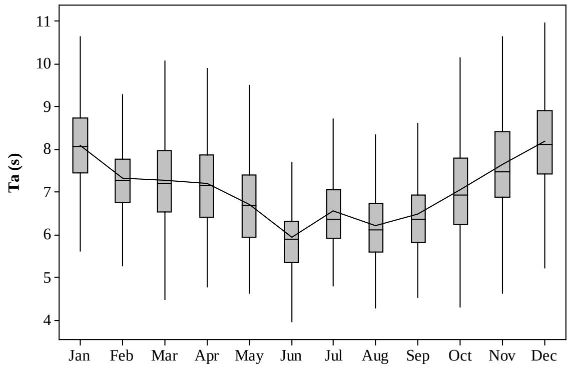


Figure **3** Monthly box plot of H_w, V_w and T_a over one year

Table **2** Basic statistics of H_w, V_w and T_a.

Dataset	Number of data	Mean	Std. Deviation	Skewness	Kurtosis
H_w	3910	3.509	1.276	1.041	1.802
V_w	3910	7.687	3.726	0.324	-0.318
T_a	3910	7.814	1.073	0.277	0.248

Another issue that needs to be considered before the statistical modeling is the serial correlation. To observe the serial dependence, the autocorrelation functions for H_w, V_w and T_a are plotted in Fig. 4. The figure shows the autocorrelation function values of the time series are not very strong when time lag is about 3 hours. In fact, the dependence between the observation and the one 3 hours later can be negligible (e.g. autocorrelation function value is within the rejection region for test of individual autocorrelations, see Fig. 4). Therefore, in this study, we generally assume the serial dependence is weak and the collected data can be directly used for statistical analysis. However, it should be realized the time series data need to be pre-processed when serial correlation is very strong. The current analysis only adopts a relaxed assumption in this data pre-processing. Typical steps about how to remove the serial correlations contained in time series data can be found in Vanem (2016).

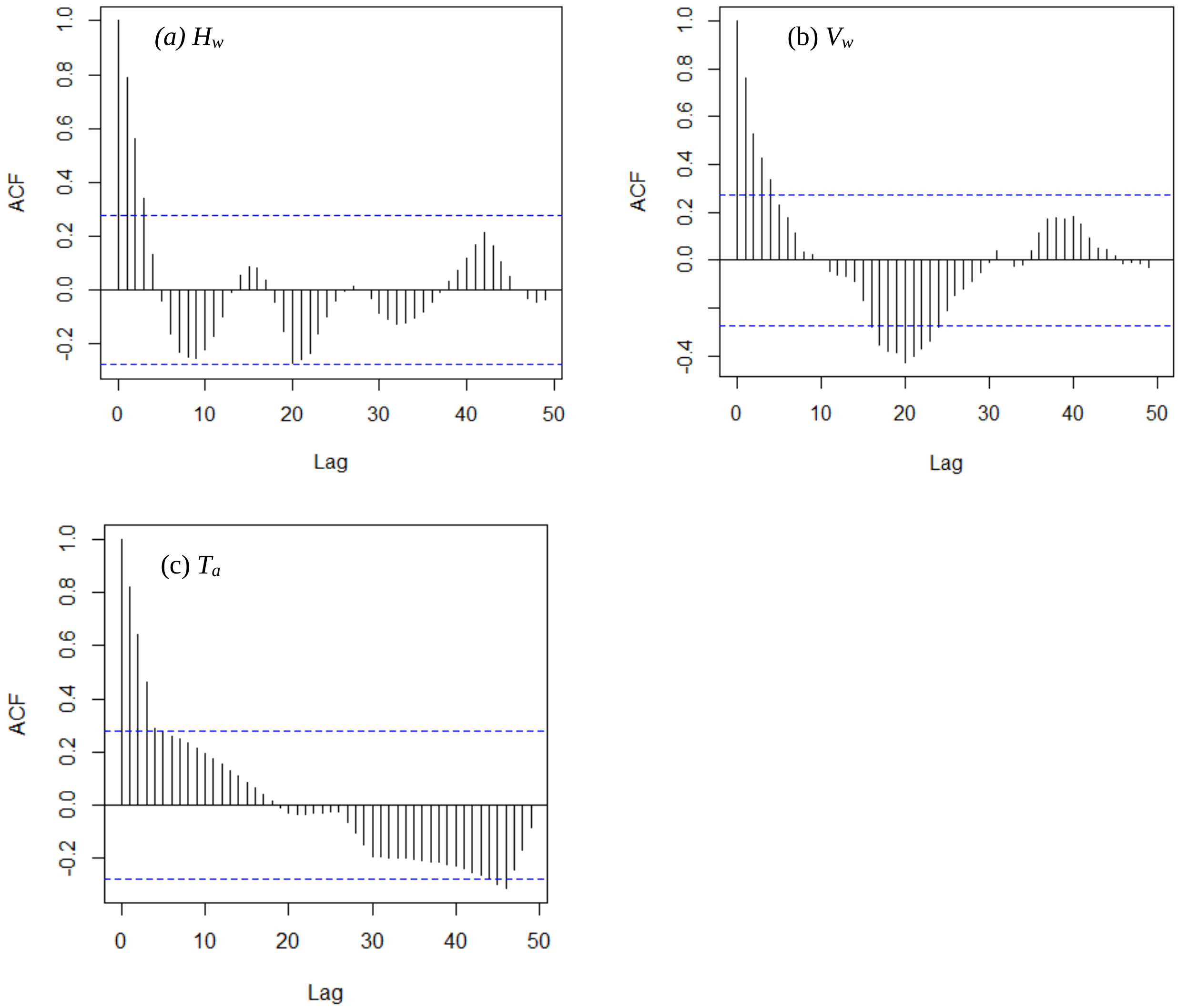


Figure 4 Autocorrelation function of H_w, V_w and T_a for the selected period

As a first step, like in all the copula approaches, marginal distributions are determined for the variables to put up an asymmetric copula. In order to make a fair judgment, a group of distributions are utilized to fit the individual ocean data. These include Weibull, Normal, Lognormal, Rayleigh, Extreme value, Exponential and Gamma distributions. Maximum likelihood method is used to estimate the model parameters for each variable. The Akaike Information Criterion (AIC) is used to select the best models. Table 3 summarizes the calculated statistics for each model. It indicates the best models are Lognormal distribution for H_w, Weibull distribution for V_w and Gamma distribution for T_a. The goodness of fitting of these models to the variables can be seen from Fig. 5. The 95% bounds of the empirical cumulative distribution function are also included. The p-value in the Kolmogorov-Smirnov test shows that the fitted model is a valid option at a significant level of 1% for each of the ocean variables.

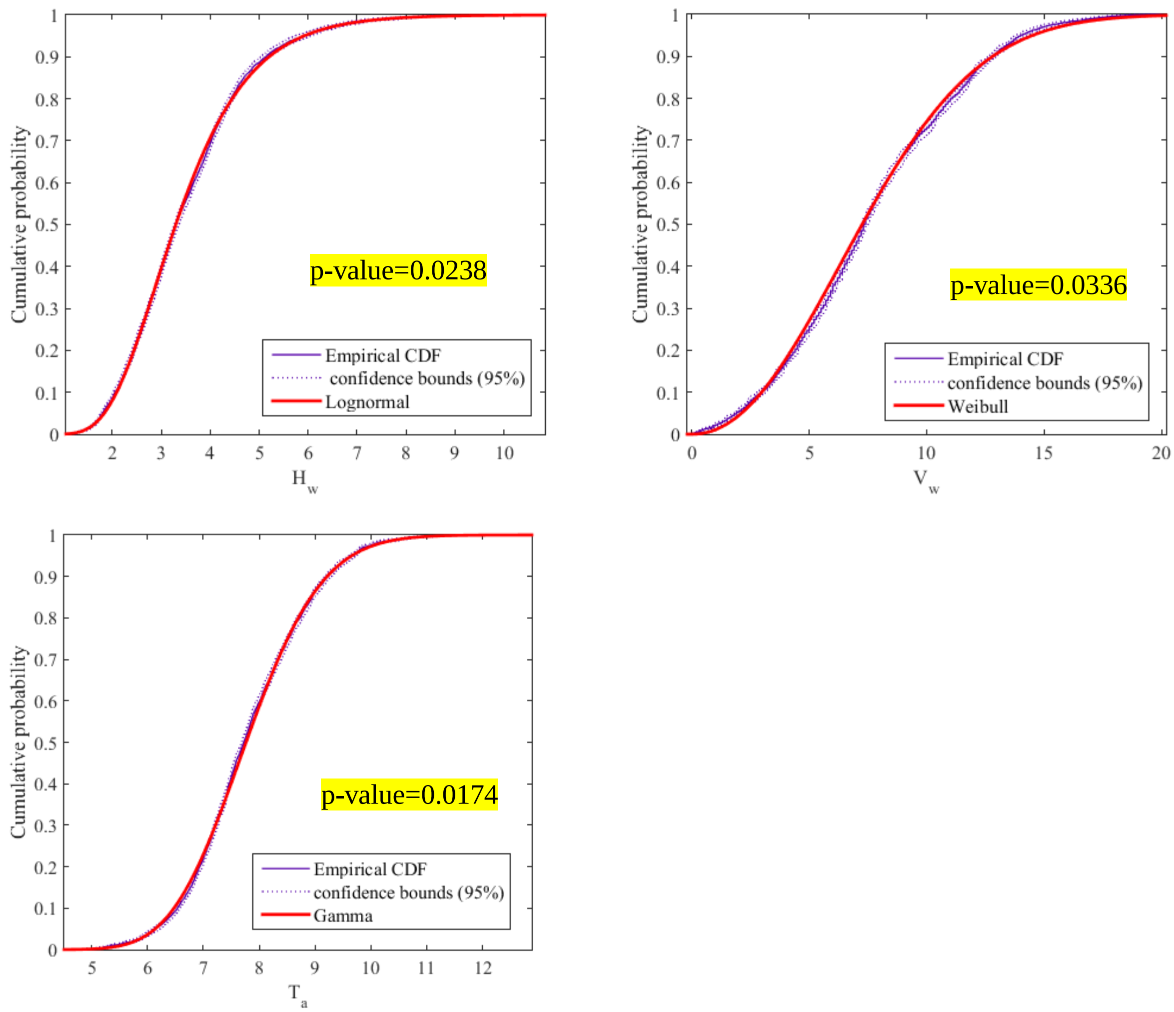


Figure **5** Distribution fitting to the collected ocean data

Table **3** Calculated AIC statistics for the marginal distribution model fitting

	Weibull	Normal	Lognormal	Rayleigh	Extreme value	Exponential	Gamma
H_w	125814	128922	119792*	127976	149104	156506	121004
V_w	139494*	151950	151300	187970	162430	228980	151000
T_a	186032	187578	188190	189672	198086	218644	186006*

*The lowest AIC value indicates the best model.

Despite the differences in the individual characteristics, the multivariate statistical properties in between the ocean variables are studied. The dependence measure concepts including Kendall's tau, Spearman's rho and correlation coefficient are calculated for each pair of the dataset as summarized in Table

4. Furthermore, the measure of the asymmetry is computed for the paired data and presented in the table. Here, the calculation of the measure of asymmetry adopts an infinity value for p as following Eq. (5). As can be seen from this table, the dependence measure concepts show similar values. The dependence in (H_w, V_w) and (H_w, T_a) is relatively stronger than that in (V_w, T_a) (as indicated by the value and independence tests). On the other hand, the results of asymmetric measures for (H_w, T_a) and (V_w, T_a) are slightly higher than (H_w, V_w). A test of exchangeability is also conducted for the bivariate data based on pseudo observations. The calculated p-values for (H_w, T_a), (V_w, T_a) and (H_w, V_w) are 0.001, 0.003 and 0.008 according to the method given by Genest et al. (2012). This indicates all the bivariate data show obvious asymmetric dependency. A general feeling of these complicated dependencies can be developed using the scatter plot in Fig. 6. It can be seen that all the dependencies between the ocean variables are not perfectly linear. For example, (H_w, T_a) data points are only available in the left upper domain in the scatter plot. Moreover, it should be noticed that the tail dependency of these bivariate data is quite different. (H_w, T_a) data has a very strong tail dependency in the maxima extremes whereas (H_w, V_w) data has a weak tail dependency in the maxima extremes (can be seen as the linearity). The characteristics of tail dependency must also be accounted for in the multivariate data modeling.

Table **4** Summary of the ocean bivariate data (p-values of the independence tests are provided in the bracket)

Data	Number of data	Kendall's tau	Spearman's rho	Correlation coefficient	Measure of asymmetry η_∞
H_w, V_w	3910	0.342 ($< 2.2 \cdot 10^{-16}$)	0.492 ($< 2.2 \cdot 10^{-16}$)	0.545 ($< 2.2 \cdot 10^{-16}$)	0.005
H_w, T_a	3910	0.483 ($< 2.2 \cdot 10^{-16}$)	0.666 ($< 2.2 \cdot 10^{-16}$)	0.652 ($< 2.2 \cdot 10^{-16}$)	0.028
V_w, T_a	3910	-0.070 (0.7029)	-0.105 (0.6732)	-0.095 (0.7112)	0.023

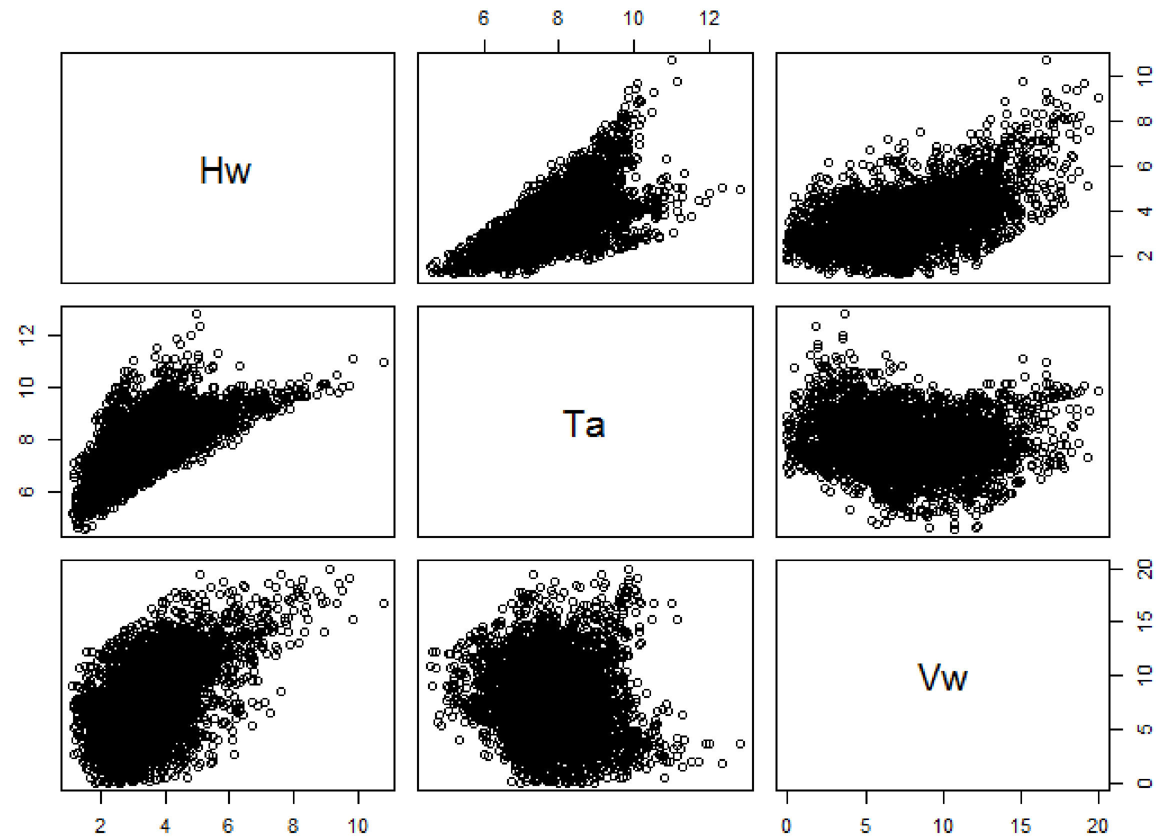


Figure 6 Scatter plot of H_w, T_a and V_w

For the dataset (V_w, T_a), all the dependence measure values are close to zero implying a nearly independent relationship between the variables. From a statistical point of view, for such a weak dependency, a multivariate modeling is not necessary. Therefore, the subsequent study is limited to the datasets (H_w, T_a) and (H_w, V_w) for the copula modeling. Based on the selected marginal distribution models, the multivariate ocean data are transformed to the copula domain. A scatter plot of (H_w, T_a) and (H_w, V_w) in the copula domain is presented in Fig. 7. The pseudo-observations show clear asymmetric dependence structures. Compared to (H_w, T_a), the bivariate data (H_w, V_w) are distributed over a broader region in the copula domain. From the scatter plot it can be observed that the data (H_w, T_a) centralize at both the minimum and the maximum extremes, while the data (H_w, V_w) only centralize at maximum extremes. This can be seen even clearer from the probability density plot. As shown in Fig. 8, the peak density values appear at both extremes in (H_w, T_a) but only at the maximum extremes for (H_w, V_w).

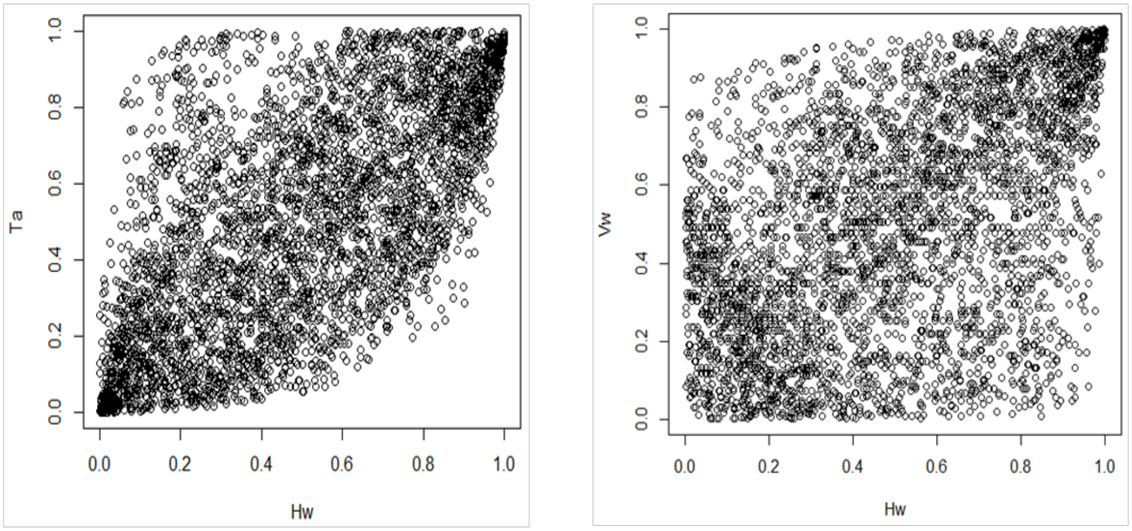


Figure 7 Scatter plot of (H_w, T_a) and (H_w, V_w) in the copula domain

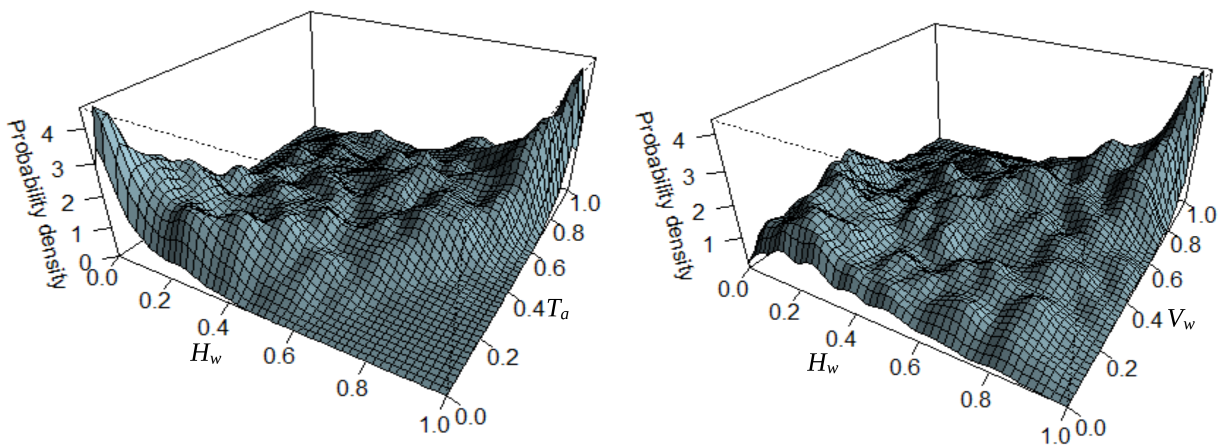


Figure 8 Empirical probability density of (H_w, T_a) and (H_w, V_w) in the copula domain

To better understand tail dependencies of (H_w, T_a) and (H_w, V_w), the four tail dependence measure concepts as discussed in Section 4.1 are applied to the two data sets. For varying quantile values, the tail dependence coefficient values are plotted in Figs. 9 and 10. It can be seen that the *upper-upper* ($\lambda^{u,u}$) and the *lower-lower* ($\lambda^{l,l}$) tail dependence coefficients show similar values with respect to each other for both (H_w, T_a) and (H_w, V_w). However, the value of the *upper-lower* ($\lambda^{u,l}$) and the *lower-upper* tail ($\lambda^{l,u}$) dependence coefficients show large differences when the quantile values change. This is especially obvious for the high quantile tail extremes (e.g. $u \to 0$). Compared to (H_w, T_a), the data of (H_w, V_w) do not show too much difference in the value of the *upper-lower* and the *lower-upper* tail dependence coefficients. Generally, the results show that the dependence between H_w and T_a is asymmetric and nonlinear whereas the dependence between H_w and V_w is almost symmetric and linear. These tail dependence characteristics are considered in the subsequent evaluation of the copula approaches.

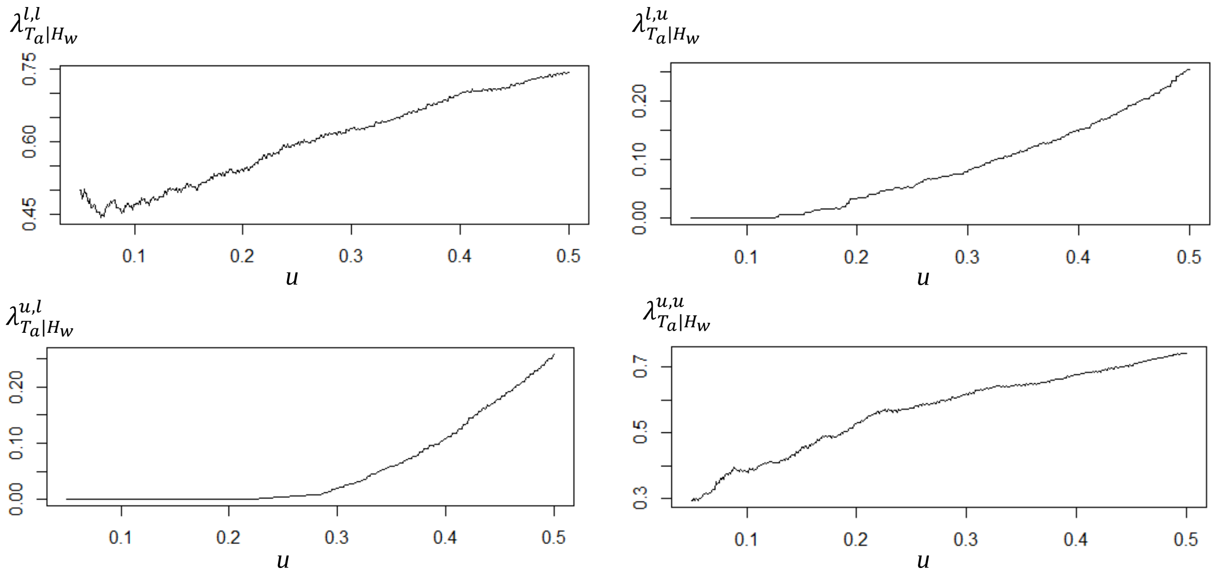


Figure 9 Estimated empirical tail dependences for the data (H_w, T_a)

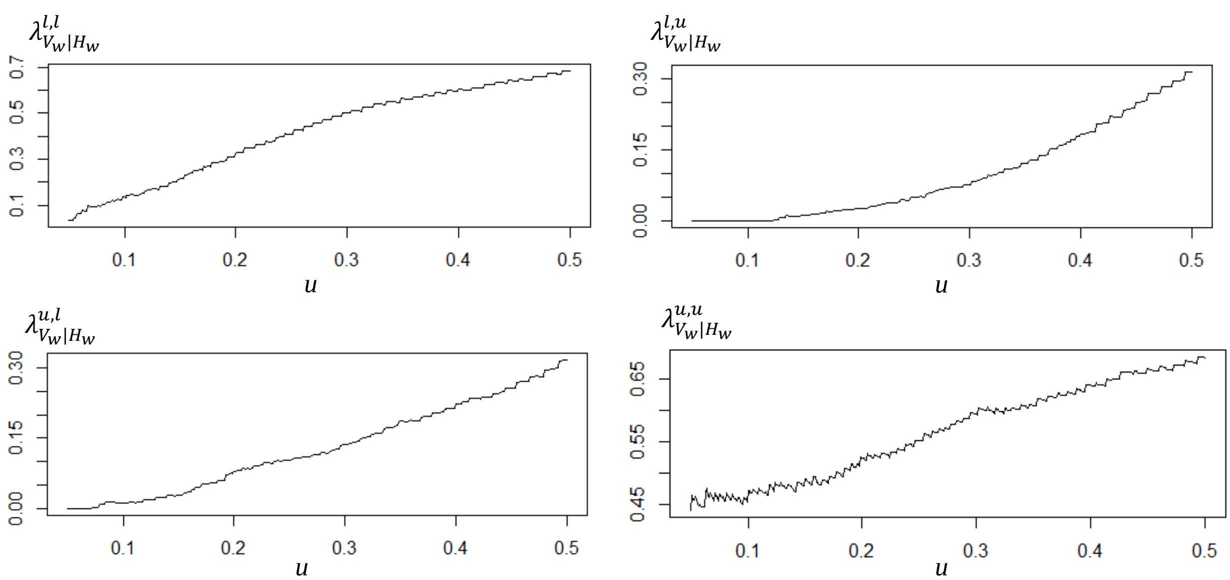


Figure 10 Estimated empirical tail dependences for the data (H_w, V_w)

Before proceeding with the multivariate analysis, it is important to address the problem of repeated observations in the data sample (Genest et al. 2011; Bücher and Kojadinovic, 2016). In particular, when a copula is chosen and fitted on a set of collected data, it is believed that the considered variables originate from a continuous joint distribution. However, due to the limitations of measurement equipment, the collected data of H_w, T_a and V_w are rounded to a numeric number having two decimal places in the current case. A summary of the quantity of tied data among the collected ocean data is provided in Table 5. It can be seen the amount of repeated observations are quite large. Such discretized version of the actual measurement may cause some errors in the statistical analysis. In fact, if the marginals are discontinuous, the data might contain ties (repeated observations) and perhaps the probabilistic model associated with the underlying random vectors cannot be uniquely determined, see the exact survey by Genest and Neslehova (2007) and also the simulation study by Pappadà et al. (2016). Therefore, herein, a reasonable procedure to test how much the randomization of the data may affect the statistical analysis is carried out in advance to the multivariate modeling.

Table 5 Summary of the quantity of tied data.

	H_w	V_w	T_a	(H_w, V_w)	(H_w, T_a)
Percentage of repeated observations	84.78%	86.32%	91.12%	81.46%	76.11%

An easy and practical way to tackle this issue is to add random components to each of the observations (De Michele et al. 2013; Salvadori et al. 2014). Here, the multivariate ocean data are modified to contain a random term by the following formulas

$$\tilde{V}_{w_i} = V_{w_i} + \Delta_V \alpha_i, \qquad \tilde{H}_{w_i} = H_{w_i} + \Delta_H \beta_i \qquad \text{and} \qquad \tilde{T}_{a_i} = T_{a_i} + \Delta_T \gamma_i, \; i = 1,\dots,n, \qquad (30)$$

where n is the sample size, Δ_V, Δ_H and Δ_T are the data resolutions for H_w, T_a and V_w and α, β and γ are random samples drawn from the standard uniform distribution between 0 and 1. For the collected data, the resolutions are Δ_V =0.01m/s, Δ_H =0.01m and Δ_T =0.01s. Therefore, through such procedures, the original data becomes randomized. $\widetilde{V_w}$, $\widetilde{H_w}$ and $\widetilde{T_a}$ are now continuous and could represent the continuous original variables wrongly recorded as discrete ones.

As discussed in the above, it is necessary to show how significant the tie effects are. In other words, we need to test whether the randomization procedure affects the statistical analysis. Generally, this can be seen from the comparison between the original data and the randomized data. In this work, we test both the univariate marginals and the joint behavior of the bivariate data. Figure 11 shows the CDF plot of original and randomized data for H_w, V_w and T_a respectively. The adopted parametric models as given in Table 3 are also plotted in the figure. The differences between CDFs of the original data and of the randomized data are quite small. The adopted models fit well in both cases. The model parameters of the marginal distribution are estimated for the randomized data and compared with the original data as shown in Table 6. The calculated p-values show that the randomized data can be well fitted by the adopted marginal distributions. Furthermore, the scatter plots of (H_w, V_w) and (H_w, T_a) are also presented in Fig. 12. It is observed that the randomized data almost overlaps the original data when they are plotted in the same figure. Finally, the KS test is performed in all these comparisons. All the statistics results indicate the randomized data and the original data show nearly the same characteristics. The fitted parametric models are also accepted since the calculated p-values are all larger than 5%. As a result, in all cases, the randomized data do not show any significant differences with respect to the original data. In other words, the randomization procedure does not significantly spoil the statistical analysis of the data. As a consequence, from a practical point of view, we believe that the collected data could represent the continuous random variables and thus can be used for performing the further analysis.

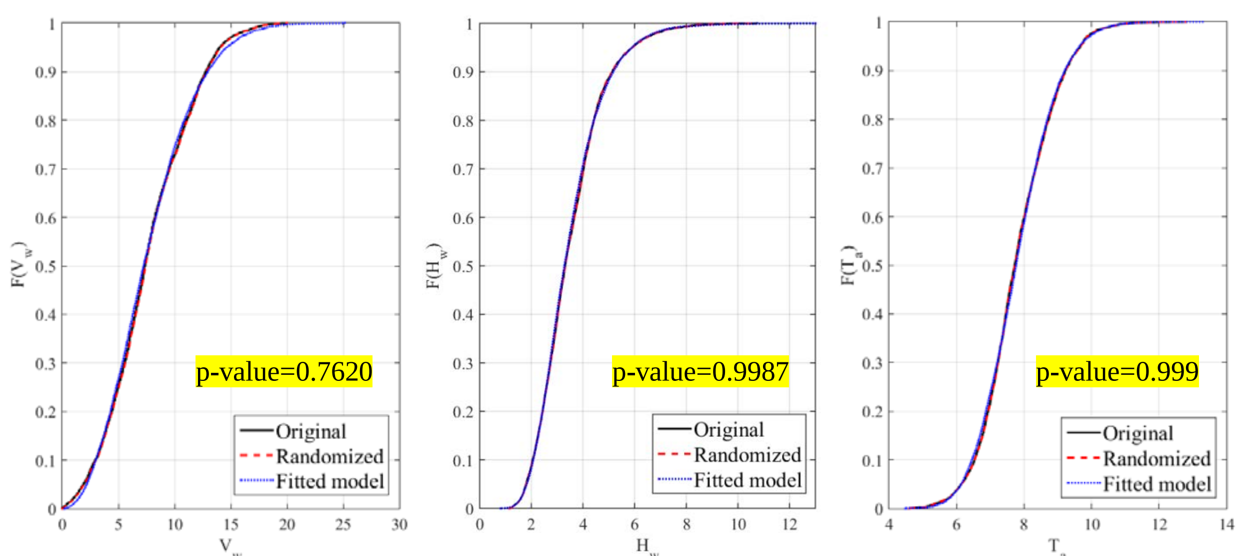


Figure 11 CDF plots of original and randomized data for H_w, V_w and T_a

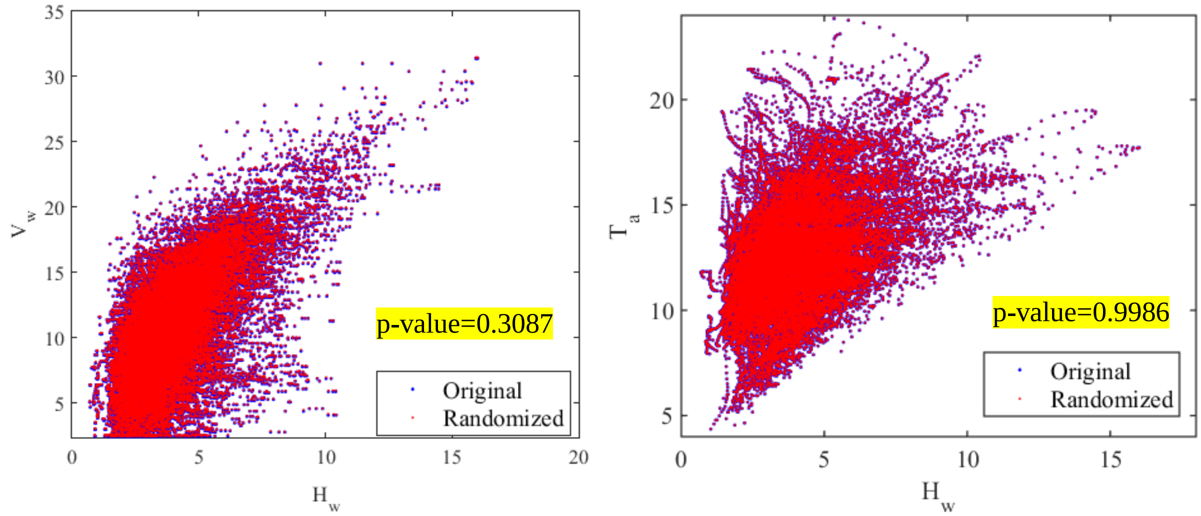


Figure 12 Scatter plots of original and randomized data for (H_w, V_w) and (H_w, T_a)

Table 6 Comparison of the estimated model parameters between original and randomized data (p-values of the KS tests between the simulated data and the fitted model for the original data are provided in the bracket)

	H_w	V_w	T_a
Original data	μ=1.1924, σ=0.3561	a=8.6267, b=2.1107	k=53.0181, θ=0.1474
Simulated data	μ=1.1940, σ=0.3555 (p-value=0.271)	a=8.6379, b=2.1197 (p-value=0.167)	k=53.0824, θ=0.1473 (p-value=0.165)

The asymmetric copulas discussed in Section 4 are used to model the bivariate data. In order to make a comparison between asymmetric copulas and the traditional ones, the commonly applied symmetric copulas are also considered. Nevertheless, it is still difficult to choose the candidate copula models as many parametric copula families and combination rules are available in the literature. To simplify the problem, this work is limited to the use of Archimedean copulas. That is, the Gumbel, Clayton and Frank copulas, which feature different tail dependencies, are used to construct the asymmetric copulas. The asymmetric copulas are formed according to the options discussed for combining the three selected copula models. Specifically, based on different rules, the following four categories of copula models are considered:

I. **One parameter copula**: The first group considers the traditional symmetric copulas from the Archimedean family. These are Gumbel, Clayton and Frank copulas.

II. **Asymmetric copulas constructed by products**: The second group of copulas is constructed by the product of copulas which are following the rules given in Section 4.2. The model combines two base copulas from the selected Archimedean copulas. These include three combinations, namely, Gumbel-Clayton, Gumbel-Frank and Clayton-Frank. To see the influence of individual functions in Eq. (21), all the three types of individual functions in Table 1 are considered in the construction.

III. **Asymmetric copulas constructed by linear convex combination**s: The third group of copulas is constructed based on linear convex combination of copulas as presented in Section 4.3. The selected Gumbel, Clayton and Frank copulas are considered as the base copulas in Eq. (23).

IV. **Skewed copula**: The fourth group of copula is the skewed Gaussian copula, which was introduced in Section 4.4. Since this copula does not need any base copulas, there is only one model in this category.

To make a fair comparison amongst all the candidate models, the corrected Akaike Information Criterion (AICc) is used. This is given by

$$\mathrm{AICc} = -2l(p) + 2p + \frac{2(p+1)(p+2)}{n-p-2} \tag{31}$$

where n is the sample size, p is the number of parameters used in each model, and $l(p)$ is the maximum log-likelihood for the model. The AIC measures the relative quality of statistical models for a given set of data whereas AICc gives a correction for finite sample sizes. A smaller AIC value indicates a better model. Note there are several other ways to judge the quality of a multivariate model. For example, the Bayesian information criterion and the cross-validation criterion can be found in (Grønneberg and Hjort, 2014). A discussion on the performances of these criterions is also provided by Jordanger and Tjøstheim (2014).

6. Comparison of symmetric and asymmetric copulas

The parameter estimates, log-likelihood and AICc statistics values for each of the models are presented in Table 7 and Table 8. The values of the model parameters are estimated based on the minimization of Cramer-von Mises statistic, which is explained in Appendix A. The best models in each of the four categories including different types of individual functions are highlighted in the tables.

Table 7 Comparison of parameter estimates and AICc values for the data of (H_w, T_a)

Copula type		Parameter estimate	Total log-likelihood	No. of parameters	AICc
I. One parameter copula	Gumbel	γ=1.744	-10973	5	21956
	Clayton	γ=1.253	-10950	5	21910
	Frank	γ=5.291	-10856	5	21722*
II. Asymmetric copulas constructed by products	Gumbel-Clayton Type1	γ_1=3.766, γ_2=3.917, θ_{11}=0.616, θ_{12}=0.287, θ_{21}=0.384, θ_{22}=0.713	-10700	10	21420*
	Gumbel-Frank Type1	γ_1=4.417, γ_2=6.019, θ_{11}=0.517, θ_{12}=0.238, θ_{21}=0.483, θ_{22}=0.762	-10735	10	21490
	Frank-Clayton Type1	γ_1=6.053, γ_2=5.021, θ_{11}=0.548, θ_{12}=0.818, θ_{21}=0.452, θ_{22}=0.182	-10873	10	21766
	Gumbel-Clayton Type2	γ_1=2.876, γ_2=8.743, θ_{11}=0.785, θ_{12}=0.739, θ_{21}=0.215, θ_{22}=0.261, α_{11}=-0.061, α_{12}=-0.538, α_{21}=0.061, α_{22}=0.538	-10636	14	21300
	Gumbel-Frank Type2	γ_1=3.025, γ_2=10.468, θ_{11}=0.722, θ_{12}=0.663, θ_{21}=0.278, θ_{22}=0.337, α_{11}=-0.014, α_{12}=-0.467, α_{21}=0.014, α_{22}=0.467	-10627	14	21282*
	Frank-Clayton Type2	γ_1=6.302, γ_2=16.588, θ_{11}=0.741, θ_{12}=0.831, θ_{21}=0.259, θ_{22}=0.169, α_{11}=-0.655, α_{12}=-0.275, α_{21}=0.655, α_{22}=0.275	-10684	14	21396
	Gumbel-Clayton Type3	γ_1=2.017, γ_2=2.454, θ_1=0.899, θ_2=0.835	-10868	8	21752
	Gumbel-Frank Type3	γ_1=1.879, γ_2=6.091, θ_1=1.024, θ_2=0.987	-10837	8	21690*
	Frank-Clayton Type3	γ_1=-0.976, γ_2=4.296, θ_1=1.912, θ_2=1.479	-11098	8	22212
III. Asymmetric copulas constructed by linear convex combinations (LCC)	Gumbel-LCC	γ=2.122, p_0=0.968, p_1=0.031, p_2=0.001	-11336	8	22688
	Clayton-LCC	γ=3.460, p_0=0.891, p_1=3.79·10^{-9}, p_2=0.108	-11398	8	22812
	Frank-LCC	γ=5.961, p_0=0.964, p_1=0.015, p_2=0.020	-11132	8	22280*
IV. Skewed copula	Skew-Gaussian	β_1=-0.418, β_2=0.964, $\boldsymbol{\beta}$=[-3.002, 3.884]	-9661	8	19338*

*Minimum AICc value indicates the best model in each type.

Table **8** Comparison of parameter estimates and AICc values for the data of (H_w, V_w)

Copula type		Parameter estimate	Total log-likelihood	No. of parameters	AICc
I. One parameter copula	Gumbel	γ=1.507	-16108	5	32226*
	Clayton	γ=0.461	-16571	5	33152

	Frank	γ=3.397	-16240	5	32490
II. Asymmetric copulas constructed by products	Gumbel-Clayton Type1	γ_1=1.898, γ_2=10, θ_{11}=0.590, θ_{12}=0.958, θ_{21}=0.410, θ_{22}=0.042	-16155	10	32330
	Gumbel-Frank Type1	γ_1=1.718, γ_2=5.373, θ_{11}=0.820, θ_{12}=0.999, θ_{21}=0.180, θ_{22}=0.001	-16128	10	32276*
	Frank-Clayton Type1	γ_1=3.574, γ_2=5.202, θ_{11}=0.999, θ_{12}=0.999, θ_{21}=0.001, θ_{22}=0.001	-16241	10	32502
	Gumbel-Clayton Type2	γ_1=1.606, γ_2=16.797, θ_{11}=0.999, θ_{12}=0.001, θ_{21}=0.001, θ_{22}=0.999, α_{11}=-0.237, α_{12}=0.723, α_{21}=0.237, α_{22}=-0.723	-16060	14	32148*
	Gumbel-Frank Type2	γ_1=1.001, γ_2=6.160, θ_{11}=0.999, θ_{12}=0.001, θ_{21}=0.001, θ_{22}=0.999, α_{11}=-0.001, α_{12}=1.001, α_{21}=0.001, α_{22}=-1.001	-16136	14	32300
	Frank-Clayton Type2	γ_1=5.111, γ_2=4.987, θ_{11}=0.999, θ_{12}=0.001, θ_{21}=0.001, θ_{22}=0.999, α_{11}=0.001, α_{12}=0.999, α_{21}=-0.001, α_{22}=-0.999	-16163	14	32354
	Gumbel-Clayton Type3	γ_1=1.676, γ_2=1.001, θ_1=0.537, θ_2=0.500	-16177	8	32370*
	Gumbel-Frank Type3	γ_1=3.130, γ_2=0.435 θ_1=0.957, θ_2=0.841	-16262	8	32540
	Frank-Clayton Type3	γ_1=-9.589, γ_2=6.129, θ_1=1.372, θ_2=1.052	-16359	8	32734
III. Asymmetric copulas constructed by linear convex combinations (LCC)	Gumbel-LCC	γ=1.586, p_0=0.998, p_1=$1.225 \cdot 10^{-9}$, p_2=0.001	-16116	8	32248*
	Clayton-LCC	γ=0.501, p_0=0.981, p_1=0.003, p_2=0.016	-16723	8	33462
	Frank-LCC	γ=3.767, p_0=0.979, p_1=0.020, p_2=0.001	-16420	8	32856
IV. Skewed copula	Skew-Gaussian	β_1=-0.877, β_2=0.081, $\boldsymbol{\beta}$=[-1.416, 0.593]	-16722	8	33460*

*Minimum AICc value indicates the best model in each type.

Again, in order to see the influence of “ties” to the multivariate data, a randomization study is conducted. Following the same procedures provided in Section 5, the randomized multivariate data are generated by adding the random component. The dependency measures including the Kendall’s τ_k, Spearman’s ρ_s and correlation coefficient for the randomized data are estimated and compared with the original data in Table 9. It can be seen the differences in the values of dependency measures are really small (around 1%) indicating the influence of ties is minimal in the degree of dependencies for multivariate data.

A further comparison is also done for the copula parameter estimates. The parameter estimates for the best copula model for both the original data and randomized data are also provided in Table 9. The estimates are based on the average value over 1000 randomized data set. The uncertainties of the associated parameters are quite small. This can be seen from the box plot for showing the uncertainties of parameter estimates from the randomized data samples, see Fig. 13. The differences between the parameter estimates are very small. The calculated p-value also implies the randomized data can be well fitted by the adopted copula model. Therefore, it is believed that the ties of data will not largely affect the statistical properties of the multivariate data and thus the data can be analyzed by the copula model.

Table 9 Comparison of the estimated model parameters between original and randomized data (p-values of the KS tests between the randomized data and the fitted model for the original data are provided in the bracket)

	Kendall's τ_k	Spearman's ρ_s	Correlation coefficient ρ	Best copula model parameter
(H_w, T_a)				
Original data	0.483	0.666	0.652	β_1=-0.418, β_2=0.964,
Randomized data	0.478	0.661	0.644	β_1=-0.416, β_2=0.965, (p-value=0.106)
(H_w, V_w)				
Original data	0.342	0.492	0.545	γ_1=1.606, γ_2=16.797, θ_{11}=0.999, θ_{12}=0.001, θ_{21}=0.001, θ_{22}=0.999, α_{11}=-0.237, α_{12}=0.723, α_{21}=0.237, α_{22}=-0.723
Randomized data	0.347	0.498	0.552	γ_1=1.611, γ_2=16.820, θ_{11}=0.999, θ_{12}=0.001, θ_{21}=0.001, θ_{22}=0.999, α_{11}=-0.242, α_{12}=0.758, α_{21}=0.249, α_{22}=-0.751 (p-value=0.213)

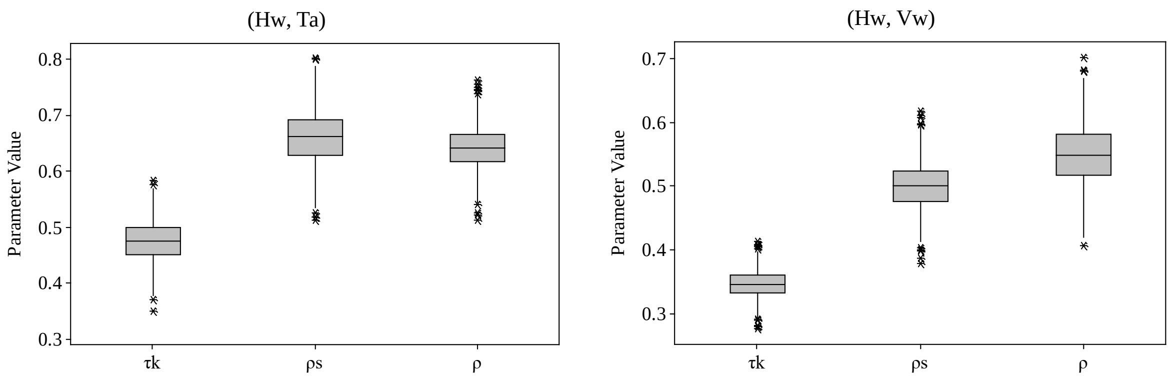


Figure 13 Estimates of the parameters from randomized data sets

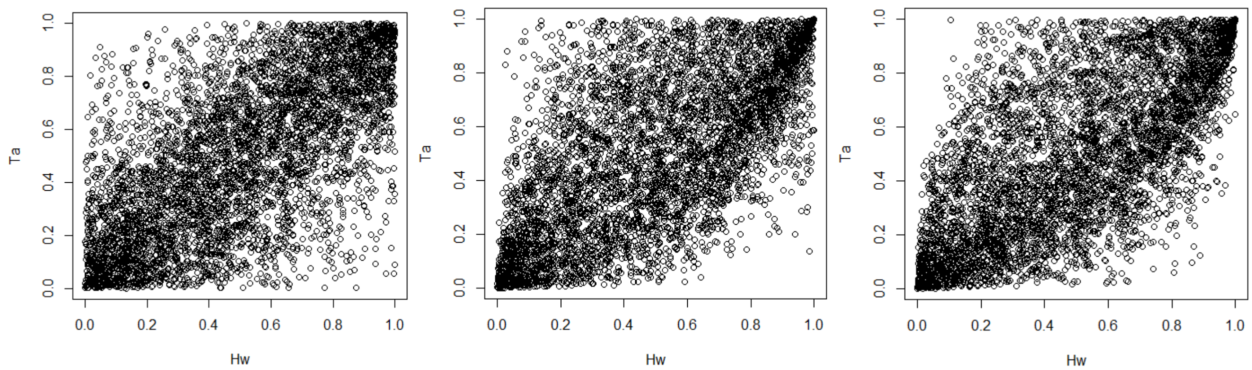
For the data (H_w, T_a), the result shows that the best model is the skewed Gaussian copula. Compared to the symmetric copula, nearly all the best asymmetric copula models show an AICc value lower than the symmetric copula. However, this does not imply the all the asymmetric copulas are better choices. The quality of an asymmetric copula model can be significantly affected by the selection of base copulas. For example, when the asymmetric copula is constructed by products of copulas, the combination of Clayton and Frank copulas is not a good choice. If the base copulas are Clayton and Frank, the AICc value becomes quite large compared to other combinations no matter which individual functions are applied, as can be seen in the tables. This indicates that the combination of Clayton and Frank cannot characterize the dependency among the data very well. Besides the selection of the base copulas, the construction rules are also a dominant factor in the construction of asymmetric copulas. From the comparison, it can be seen that a skewed copula and an asymmetric copula constructed by products of copulas show a better performance than the other two categories of copulas. This results from the fact that these two kinds of copulas are not only good for modeling asymmetric dependencies but also capable of for capturing the nonlinear dependencies. Apparently, the samples investigated in this work suggest that copulas constructed by linear convex combinations may suffer from a potential weakness.

The situation changes quite a lot for the other data set. For the data (H_w, V_w), the Gumbel-Clayton product copula with Type 2 individual function outperforms all the other copula models. Compared to the data (H_w, T_a), the asymmetric copula does not show very significant advantages over the symmetric copula in modeling (H_w, V_w). In fact, the Gumbel copula is even a better choice than most constructed asymmetric copulas. For the asymmetric copulas constructed by products, all the candidate models yield similar results. This indicates that the selection of base copulas and individual functions is not critical in this case. The asymmetric copulas constructed by linear convex combinations are still showing poor performance for the bivariate data. Additionally, the skewed copula does not show a good performance in this modeling. It actually produces the largest AICc value amongst all choices. The major reason for such a big difference is the only very weak asymmetry of the bivariate data (H_w, V_w) compared to (H_w, T_a). Clearly, quasi symmetric data do not require an asymmetric modeling. In such cases, the use of asymmetric copulas is even disadvantageous.. Since most asymmetric copulas are utilizing quite a number of parameters, the model can somehow tend to be over parameterized. Some of the degrees of freedom created by the parameters are therefore not necessarily to be added. Therefore, the symmetric model would be the better choice.

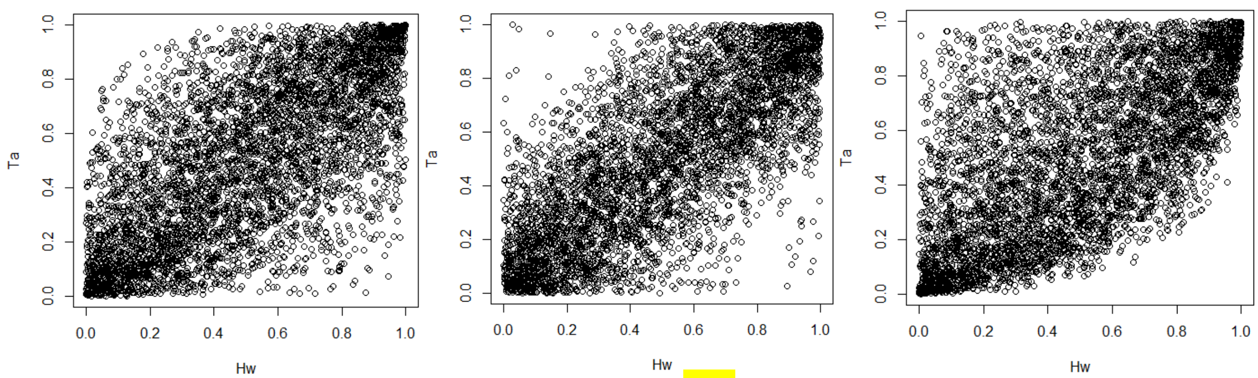
Another reason why symmetric and asymmetric copulas perform so differently in modeling bivariate data sets is the link to physical phenomena. As for (H_w, T_a), the breaking wave limit is causing asymmetry in the bivariate data. This physical feature can be implemented by the asymmetric copula while the traditional copula cannot capture this feature. However, as for (H_w, V_w), there is no such limit. The asymmetric copula would, thus, lose its advantage in the modeling.

In the final part of this study, a short reference is made to the comparison between symmetric and asymmetric copula models for the prediction of extreme values. This can be achieved through a comparison of simulated data from the established models with the empirical data. The simulated data for (H_w, T_a) and (H_w, V_w) are plotted in Figs. 14 and 15, respectively. It can be seen that the nonlinear dependencies are well displayed in the simulated data for (H_w, T_a). The asymmetric dependency is also obvious in the scatter plot, especially for the skewed copula model. This agrees well with the AICc result, which indicates that the skewed copula is the best choice for modeling (H_w, T_a). Moreover, compared to the symmetric model (Frank copula), the data simulated from the skewed copula are closer to reality. For example, there are no data beyond the breaking wave limit in the data simulated from the skewed copula.

This physical phenomenon is well captured by the skewed copula. In fact, all the asymmetric copulas show a better performance in capturing the breaking wave limit while the symmetric copula cannot capture this phenomenon, see Fig. 14. On the other hand, the simulated data for (H_w, V_w) do not have this problem. As can be seen in Fig. 15, there are no clear nonlinear dependencies in the simulated data. The simulated data for (H_w, V_w) are scattered over the whole copula domain without any limits. Some even clearer view can be obtained from the contour plots of the simulated data in the original domain of ocean parameters as shown in Figs. 16 and 17. The contour lines are obtained based on the procedures provided in Montes-Iturrizaga and Heredia-Zavoni (2015). The inverting Rosenblatt transformation is used to derive the contour points corresponding to a certain exceeding probability (sometimes is referring to a reliability index in the first order reliability method) whereas the copula model provides the conditional joint distribution. Five levels of the probability density function values are plotted in the figures. The plots indicate that the quality of model fitting to the empirical data is decreasing with the decrease of contour level values. In fact, the low value contour lines are representing extremes in different directional tails. Therefore, a good quality of fitting to the low value contour lines indicates a good characterization of tail values. From the comparison of all these plots it can be observed that the asymmetric copulas either constructed by products or by linear convex combinations do not lead to a good prediction of lower-upper tail extremes in the domain of (H_w, T_a). And symmetric copula cannot predict the upper-lower tail extremes very well for (H_w, T_a). As for the data (H_w, V_w), the skewed copula shows a bad fitting at both the upper-lower and the lower-upper tails. These models cannot predict all the extremes with sufficient quality. The selection of the most appropriate model for predicting the extreme values has to be made based on the tail fitting capabilities. Nevertheless, the comparative studies do tell that the asymmetric copulas are very applicable to ocean data where asymmetric dependencies are obvious.

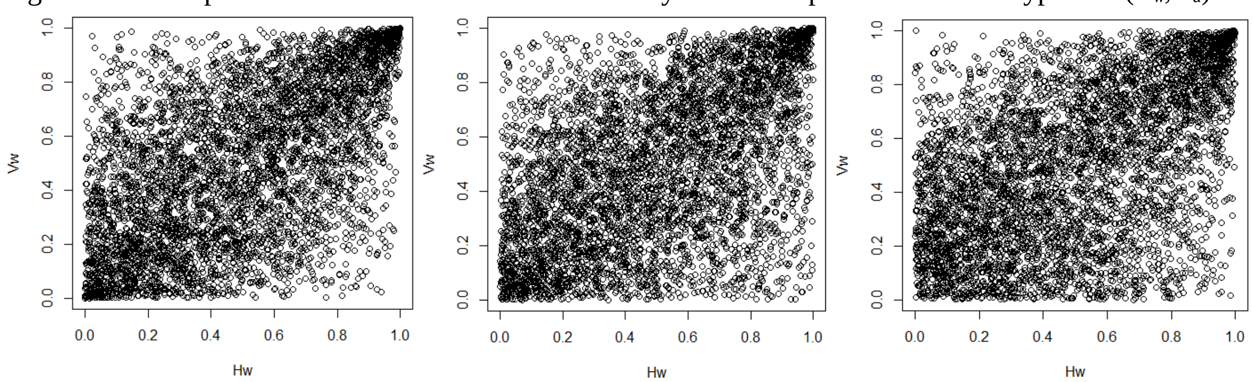


(a) Frank (b) Gumbel-Clayton Type1 (c) Gumbel-Frank Type2

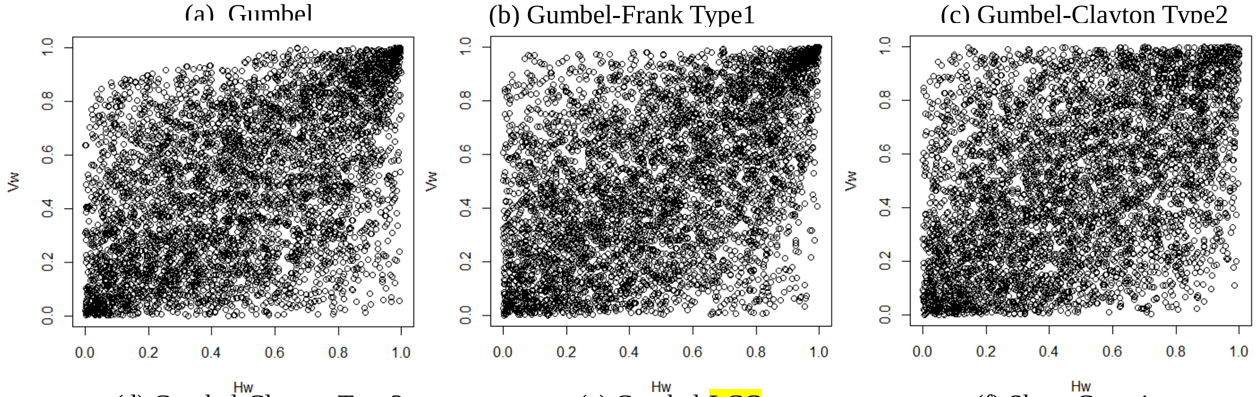


(d) Gumbel-Frank Type3 (e) Frank-LCC (f) Skew-Gaussian

Figure 14 Scatterplots of simulated data from the best asymmetric copulas in different types for (H_w, T_a).

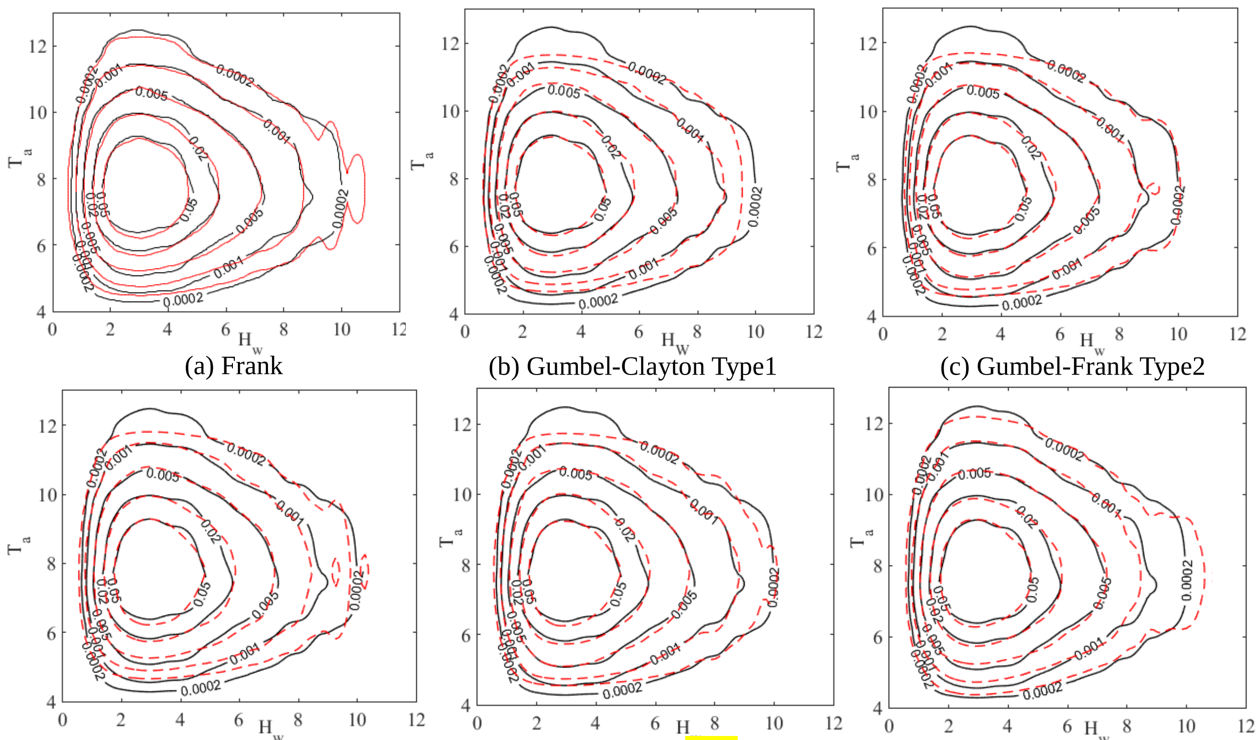


(a) Gumbel (b) Gumbel-Frank Type1 (c) Gumbel-Clayton Type2



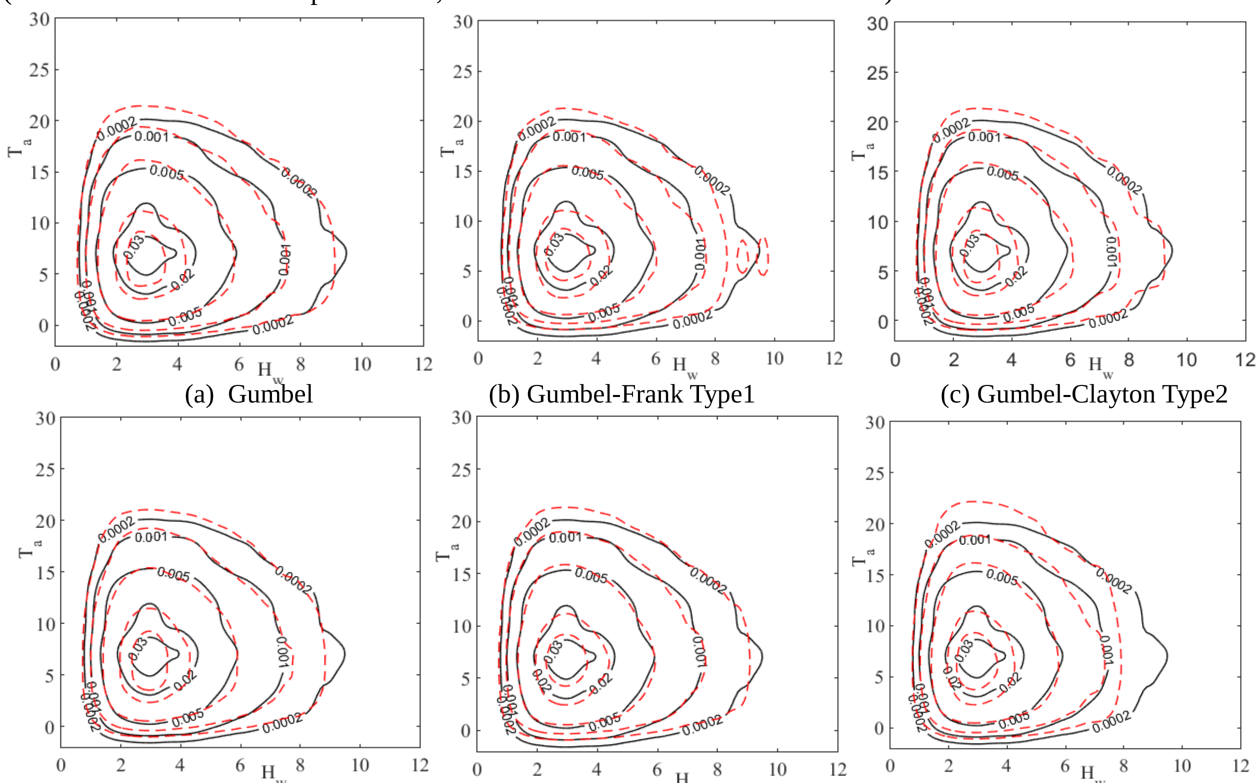
(d) Gumbel-Clayton Type3 (e) Gumbel-LCC (f) Skew-Gaussian

Figure 15 Scatterplots of simulated data from the best asymmetric copulas in different types for (H_w, V_w).



(a) Frank (b) Gumbel-Clayton Type1 (c) Gumbel-Frank Type2
(d) Gumbel-Frank Type3 (e) Frank-LCC (f) Skew-Gaussian

Figure 16 Comparison of contour plot between original data and best fitted copula models for (H_w, T_a) (black line indicates the empirical data; red dash line indicates the fitted model).



(a) Gumbel (b) Gumbel-Frank Type1 (c) Gumbel-Clayton Type2
(d) Gumbel-Clayton Type3 (e) Gumbel-LCC (f) Skew-Gaussian

Figure 17 Comparison of contour plot between original data and best fitted copula models for (H_w, V_w) (black line indicates the empirical data; red dash line indicates the fitted model).

It is noted that the results derived in the current study need to be interpreted separately for different kinds of ocean multivariate data. Asymmetric copulas are suitable for modeling asymmetrically dependent variables. In this context, the bivariate data (H_w, T_a) can be well modeled by the asymmetric copula. The results can only be explained specifically for the studied data set for the filtered time period. This analysis of ocean data is only valid for the investigated specific ocean location. Another important limitation of present study is the simplified data pre-processing. It should be realized the pre-processing of the data is very arbitrary. The analyzed ocean data in the current analysis is influenced by many environmental factors which may need to be removed before the statistical analysis. Also, the serial correlation is not removed in the present work. If the collected time series data show very strong serial correlation, this simplification is not allowed. A more accurate data pre-processing is required. However, the results can be used to explain significant features of using asymmetric copulas for modeling ocean data in general. The modeling of asymmetric dependencies can produce more reliable and realistic statistical joint models. Physical phenomena can be well handled in this manner. Additionally, it should also be noted that asymmetric copula models are much more flexible compared to the traditional copula models. A large number of base copulas and combination rules can be implemented for constructing an asymmetric copula. This provides great flexibility and needs to be considered in the application of asymmetric copulas and in the interpretation of the results.

Finally, the findings of the present study may help design engineers, marine exploiters or ship owners who are working in the open sea. Offshore engineering practitioners may be empowered to make better decisions based on more insight into the dependencies between the ocean variables. This can support the design of more robust offshore structures to resist the environmental loading when considering the ocean dependencies. It can also help to identify which environmental factors are the most dominant ones that affect other factors in the ocean. Moreover, the empirical results from this study may help researchers to exploit features of asymmetric copulas in a targeted manner.

7. Conclusions

In this research, asymmetric copulas were utilized to model ocean parameters in a multivariate setting. The fundamentals associated with asymmetric copulas including the measures of asymmetric dependencies and tail dependencies have been reviewed in detail in the context of ocean data modeling. Three different ways of constructing an asymmetric copula were introduced. The asymmetric copulas were compared with the traditional copula, specifically on the modeling of bivariate data (H_w, T_a) and (H_w, V_w) collected from an ocean site in the south coast of Alaska in an example study. The copula models were constructed for the ocean data after preconditioning to obtain quasi stationary data records. It was found that the asymmetric copula models can be more realistic in describing ocean data with asymmetric dependencies. The results suggest that one may utilize the asymmetric copula models when the observed data show an asymmetric dependence structure as inferred by the scatter plot of the pseudo-observations. However, the construction of a robust asymmetric copula model largely depends on the selection of the base copulas and combination rules. The study also shows that asymmetric copula models are more powerful than the symmetric copulas in capturing extreme contours for the present analyzed ocean data. Thus, it is expected that asymmetric copulas have great advantages in risk assessment of marine and coastal structures due to extreme ocean environment. Future work seems necessary to investigate the significance of this approach, including uncertainties, on the offshore engineering practical designs and operation of coastal structures.

Acknowledgments

This study is partly sponsored by Japanese Society for the Promotion of Science (JSPS) for the Grant-in-Aid for Scientific Research (B) under the project No. 16H04398. The first author, Yi Zhang, is sponsored by “The JSPS Postdoctoral Fellowship for Foreign Researchers” Program. Such financial aids are gratefully acknowledged.

Appendix A Fundamental knowledge of asymmetric copulas

We briefly introduce the fundamentals of parameter estimation and simulation for using asymmetric copulas.

All the related theoretical proofs and mathematical justifications can be found in Nelson (2006) and Joe (2014). Some discussions of these methods are simplified to a bivariate problem. However, the same concept can be generalized to higher dimensional problems.

The estimation of the parameters of asymmetric copulas can be realized in various ways. The most widely applied approach is the maximum likelihood estimation. In general, the maximum likelihood estimation method selects that set of values of the model parameters, which maximizes the likelihood function. This concepts is very convenient for estimating copula parameters especially for one parameter copulas. However, the major difficulty in applying this approach for asymmetric copula models is the maximization of the likelihood function. Since most asymmetric copulas contain multiple parameters, such maximization requires a very robust and efficient optimization algorithm in searching the optimum solution from the parameter domain. The computational effort can become tremendous when the number of estimated parameters increases.

Alternatively, the parameters can be determined through a distance based estimation method. For seeking the most appropriate model parameters $\boldsymbol{\Theta} = \{\theta_1, \ldots, \theta_n\}$ of the copula, the Cramer-von Mises statistic $\boldsymbol{S}$ can be employed. The Cramer-von Mises statistic $\boldsymbol{S}$ calculates the distances between the theoretical copula distribution function and the empirical copula distribution function. The estimates for the copula parameters are determined in such a way that they minimize this statistic. For a bivariate copula, this parameter estimation method can be formulated as

$$\Theta = \underset{\theta_1,\ldots,\theta_n}{\arg\min} S = \underset{\theta_1,\ldots,\theta_n}{\arg\min} \sum_{i=1}^{N} \left\{ C_{empirical}\left(u_i, v_i\right) - C_{\Theta}\left(u_i, v_i\right) \right\}^2 \qquad \text{(A.1)}$$

where $C_{empirical}$ stands or the empirical copula function, C_{Θ} represents the candidate parametric copula and Θ is the set of copula parameters which need to be determined. Therefore, the distance between the theoretical copula and empirical copula is evaluated for each of the observed data points (u_i, v_i). One should note this "distance based method" is same as the Least Square estimation. However, the dimension of the parameters' space does not change, and, in general, there is no guarantee that this approach performs better than the traditional maximum likelihood method. Meanwhile, it should be pointed out this method is just a numerical interpolation method, not a "statistical" one.

The simulation from asymmetric copulas can be performed using the same algorithm as for traditional copulas. The most general approach is based on a sequence of conditional distributions derived from the copula function. For example, the conditional approach developed in Devroye (1986) has utilized the Rosenblatt transform to simulate the random vector. Other similar concepts regarding the simulation of random vectors from asymmetric copulas can be found in Mai and Scherer (2012). However, it should be pointed out that the random generation requires a root finding procedure. A robust and efficient method is needed for this purpose to achieve reasonable computational efficiency.

Appendix B Examples of copulas

Many copula families and classes have been developed in the literature. Each one has its own advantages which could characterize certain types of data.

Archimedean copulas

In practice, the Archimedean copulas are most frequently applied. The family of Archimedean copulas consists of a wide range of parametric copula groups. They have been applied widely in data modeling utilizing their feature to be constructed easily. Generally, an *n*-dimensional Archimedean copula can be constructed based on an algebraic method using a generating function $\varphi(.)$:

$$C_{Archimedean}\left(u_1,...,u_n;\theta\right)=\varphi^{[-1]}\left(\varphi\left(u_1;\theta\right)+...+\varphi\left(u_n;\theta\right);\theta\right) \quad \text{(B.1)}$$

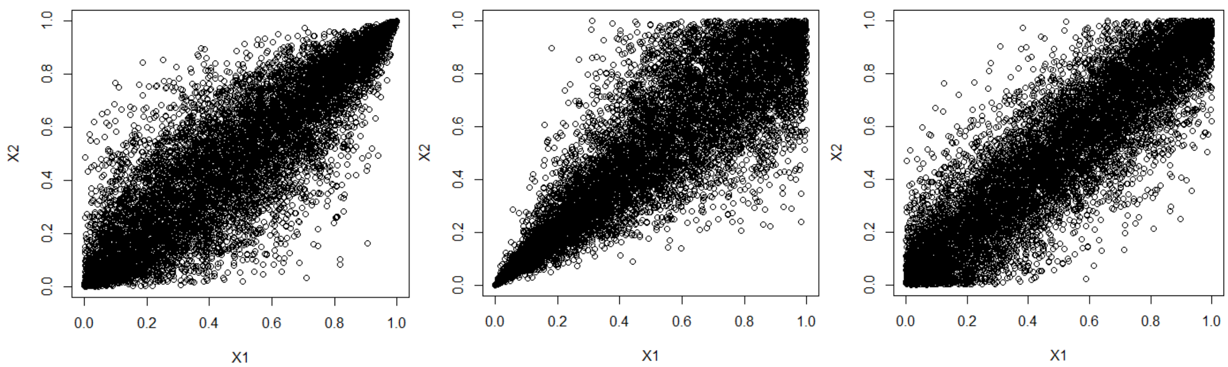
where φ: $[0, 1]\times\mathbf{\Theta}\rightarrow[0,\infty)$ is a completely monotone function with $\varphi(1)=0$. θ is a parameter with the domain $\mathbf{\Theta}$. $\varphi^{[-1]}$ is the pseudo-inverse of φ defined by

$$\varphi^{[-1]}\left(t;\theta\right)=\begin{cases}\varphi^{-1}\left(t;\theta\right) & if\ 0\le t\le\varphi\left(0;\theta\right)\\ 0 & if\ \varphi\left(0;\theta\right)\le t\le\infty\end{cases}. \quad \text{(B.2)}$$

Therefore, the construction of a multivariate copula model depends on the generating functions used in Eq. (B.2). Many well known copula families are Archimedean, as widely studied in the literature (Embrechts, 2009). The generating functions of the most popular single parameter Archimedean copulas such as Gumbel, Frank and Clayton are presented in Table B.1.

Table B.1 Examples of Archimedean copulas

Copula	Bivariate Formula $C_\gamma(u, v)$	Generating Function $\varphi_\gamma(t)$	$\gamma \in$
Gumbel	$\exp\left\{-\left[(-\ln u)^{\gamma}+(-\ln v)^{\gamma}\right]^{\frac{1}{\gamma}}\right\}$	$(-\ln t)^{\gamma}$	[1,+∞)
Frank	$\frac{1}{\gamma}\ln\left(1+\frac{(e^{u\gamma}-1)(e^{v\gamma}-1)}{e^{\gamma}-1}\right)$	$\ln\frac{e^{\gamma t}-1}{e^{\gamma}-1}$	(-∞, +∞)
Clayton	$(u^{-\gamma}+v^{-\gamma}-1)^{-\frac{1}{\gamma}}$	$t^{-\gamma}-1$	(1, +∞)



(a) Gumbel (b) Clayton (c) Frank

Figure B.1 Comparison of different bivariate copulas for correlation coefficient equals to 0.7.

An illustration of these Archimedean copulas is provided by means of scatter plots shown in Fig. B.1. Each generating function in an Archimedean copula characterizes a special tail dependency. For example, the Gumbel copula is an appropriate candidate model for data having stronger dependencies at high values (less spread) compared to low values, whereas the Clayton copula can characterize data exhibiting strong low value dependencies. On the other hand, the Frank copula is deemed quite appropriate for data having relatively weak dependencies at both tails. For the coastal engineering practice, in some ocean data sets, stronger dependencies are observed at the extremes. For instance, storms usually induce large wave heights associated with high wind speed, and thus one can expect the data to show stronger dependencies between wave height and wind speed for large values than for smaller values of wave height and wind speed. Hence, to be suitable, a copula model for this case should capture these characteristics observed in the data.

References

Andrew, J., Patton, A., 2012. Review of copula models for economic time series. Journal of Multivariate Analysis 110, 4-18.

Antão, E.M., Guedes Soares, C., 2014. Approximation of bivariate probability density of individual wave steepness and height with copulas. Coastal Engineering 89, 45-52.

Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715–726.

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

Bitner-Gregersen, E.M., Guedes Soares, C., Machado, U., Cavaco, P., 1998. Comparison of different approaches to joint environmental modeling. Proceedings of the 17th International Conference on Offshore Mechanics and Arctic Engineering, ASME, New York, Vol. II.

Bitner-Gregersen, E.M., Haver, S., 1989. Joint long term description of environmental parameters for structural response calculation. Proceedings of the 2nd International Workshop on Wave Hindcasting and Forecasting.

Bücher, A., Kojadinovic, I., 2016. An overview of nonparametric tests of extreme-value dependence and of some related statistical procedures. In: Dey, D., Yan, J. (Eds.), Extreme Value Modeling and Risk Analysis: Methods and Applications. CRC Press, pp. 377–398.

Burton, T., Sharpe, D., Jenkins, N., Bossanyi, E., 2001. Wind Energy Handbook. John Wiley, Chichester, England.

Charpentier, A., Fougères, A.L., Genest, C., Nešlehová, J.G., 2014. Multivariate Archimax copulas. Journal of Multivariate Analysis 126, 118-136.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. The Wiley Finance Series. Wiley.

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

De Michele, C., Salvadori, G., Vezzoli, R., Pecora, S., 2013. Multivariate assessment of droughts: frequency analysis and Dynamic Return Period. Water Resour. Res. 49 (10), 6985–6994.

Devroye, L., 1986. Non-uniform Random Variate Generation. Springer-Verlag, New York.

deWaal, D.J., van Gelder, P.H.A.J.M., 2005. Modelling of extreme wave heights and periods through copulas. Extremes 8(4), 345–356.

DNV RP-C205, 2010. Environmental conditions and environmental loads. Høvik, Norway.

DNV RP-C205, 2014. Environmental conditions and environmental loads. Høvik, Norway.

Dong, S., Tao, S., Li, X., Guedes Soares, C., 2015. Trivariate maximum entropy distribution of significant wave height, wind speed and relative direction. Renewable Energy 78, 538-549.

Dong, S., Wang, N., Liu, W., 2013. Bivariate maximum entropy distribution of significant wave height and peak period, Ocean Engineering 59, 86-99.

Durante, F., Salvadori, G., 2010. On the construction of multivariate extreme value models via copulas. Environmetrics 21, 143–161.

Durante, F., Sempi, C., 2015. Principles of copula theory. CRC/Chapman & Hall, Boca Raton, FL.

Embrechts, E.,2009. Copulas: a personal view. Journal of Risk and Insurance 76(3), 639–650.

Ernst, B., Seume, J.R., 2012. Investigation of site-specific wind field parameters and their effect on loads of offshore wind turbines. Energies 5(10), 3835-3855.

Ewans, K., Jonathan, P., 2014. Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. Journal of Marine Systems 130, 124-130.

Ferreira, J.A., Guedes Soares, C., 2002. Modelling bivariate distributions of significant wave height and mean wave period. Applied Ocean Research 24(1), 31-45.

Genest, C., Favre, A., 2007. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. Journal of Hydraulic Engineering 12(4), 347-368.

Genest, C., Neslehova J., 2007. A primer on copulas for count data. Astin Bull. 37 (2), 475–515.

Genest, C., Neslehova, J., Ruppert, M., 2011. Discussion: Statistical models and methods for dependence in insurance data. J. Korean Statist. Soc. 40 (2), 141–148.

Genest, C., Nešlehová, J.G., 2013.Assessing and modeling asymmetry in bivariate continuous data. P. Jaworski, F. Durante, W.K. Härdle (Eds.), Copulae in Mathematical and Quantitative Finance, Proceedings of the Workshop Held in Cracow, 10–11 July 2012, Springer, Berlin Heidelberg, Germany, 91–114.

Grimaldi, S., Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. Advances in Water Resources 29(8), 1155-1167.

Grimaldi, S., Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. Advances in Water Resources 29, 1155–1167.

Grønneberg, S., Hjort, N. L., 2014. The copula information criteria. Scandinavian Journal of Statistics 41 (2), 436–459.

Guedes Soares, C., Lopes, L.C., Costa, M.D.S., 1988. Wave Climate Modelling for Engineering Purposes, Computer Modelling in Ocean Engineering. Rotterdam:A.A. Balkema Pub.

Harder, M., Stadtmüller, U., 2014. Maximal non-exchangeability in dimension. Journal of Multivariate Analysis 124, 31-41.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J. 2016. Copula: multivariate dependence with copulas. R package version 0.999-15 URL http://CRAN.R-project.org/package=copula

Huseby, A.B., Vanem, E., Natvig, B., 2013. A new approach to environmental contours for ocean engineering applications based on direct Monte Carlo simulations. Ocean Eng 60,124-135.

Hutchinson, T., Lai, C., 1990. Continuous bivariate distributions, emphasising applications. Rumsby Scientific Publishing.

Jane, R., Dalla Valle, L., Simmonds, D., Raby, A., 2016. A copula-based approach for the estimation of wave height records through spatial correlation. Coastal Engineering 117, 1-18.

Joe, H., 2014. Dependence Modeling with Copulas. Chapman & Hall/CRC.

Jonathan, P., Ewans, K., 2011. Modeling the seasonality of extreme waves in the Gulf of Mexico. Journal of Offshore Mechanics and Arctic Engineering 133, 0211041-49.

Jonathan, P., Flynn, J., Ewans, K., 2010. Joint modeling of wave spectral parameters for extreme sea states. Ocean Engineering 37(11-12), 1070-1080.

Jondeau, E., 2016. Asymmetry in tail dependence in equity portfolios. Computational Statistics & Data Analysis 100, 351-368.

Jones, M., Randell, D., Ewans, K., Jonathan, P., 2016. Statistics of extreme ocean environments: Non-stationary inference for directionality and other covariate effects. Ocean Engineering 119, 30-46.

Khoudraji, A., 1995. Contributions à l'étude des copules et àla modélisation des valeurs extrêmes bivariées, Ph.D. Thesis, University Laval Quebec, Canada.

Klement, E.P., Mesiar, R., 2006. How non-symmetric can a copula be? Comment. Math. Univ. Carolin. 47, 141–148.

Kollo, T., Selart, A., Visk, H., 2013. From multivariate skewed distributions to copulas. In: Combinatorial matrix theory and generalized inverses of matrices. Springer, New York.

Liebscher, E., 2008. Construction of asymmetric multivariate copulas. Journal of Multivariate Analysis 99(10), 2234-2250.

Lucas, C., Guedes Soares, C., 2015. Bivariate distributions of significant wave height and mean wave period of combined sea states. Ocean Engineering 106(15), 341-353.

Mai, J.-F., Scherer, M., 2012. Simulating Copulas. Vol. 4 of Series in Quantitative Finance. Imperial College Press, London (UK).

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

Mazo, G., Girard, S., Forbes, F., 2015. A class of multivariate copulas based on products of bivariate copulas. Journal of Multivariate Analysis 140, 363-376.

McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative risk management. Concepts, techniques and tools. Princeton Series in Finance. Princeton University Press, Princeton, NJ.

Mesiar, R., Najjari, V., 2014. New families of symmetric/asymmetric copulas. Fuzzy Sets and Systems 252, 99-110.

Montes-Iturrizaga, R., Heredia-Zavoni, E., 2015. Environmental contours using copulas. Applied Ocean Research 52, 125-139.

Morton, I.D., Bowers, J., 1996. Extreme value analysis in a multivariate offshore environment. Applied Ocean Research 18(6), 303-317.

Muraleedharan, G., Lucas, C., Martins, D., Guedes Soares, C., Kurup, P.G., 2015. On the distribution of significant wave height and associated peak periods. Coastal Engineering 103, 42-51.

Nataf, A., 1962. Détermination des distributions de probabilities dont les marges sont données Comptes Rendus de l'Academie des Sciences 225, 42–43.

National Data Buoy Center (NDBC), 2016. NDBC station locations and information. National Oceanic and

Atmospheric Administration.

Nelson, R.B., 2006. An introduction to Copulas. Springer, New York.

Ochi, K., 1992. New approach for estimating the severest sea state from statistical data. Coastal Engineering Conference, 512–525.

Pappadà, R., Durante, F., Salvadori, G., 2016. Quantification of the environmental structural risk with spoiling ties: is randomization worthwhile? Stochastic Environmental Research and Risk Assessment-Doi: 10.1007/s00477-016-1357-9.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. International Economic Review 47(2), 527-556.

Pickands, J., 1981. Multivariate extreme value distributions (with a discussion). Proceedings of the 43rd Session of the International Statistical Institute. Bull. Inst. Internat. Statist. 49, 859–878, 894–902.

Prince-Wright, R., 1995. Maximum likelihood models of joint environmental data for TLP design, Proceedings of the 14th International Conference on Offshore Mechanics and Arctic Engineering.

Repko, A., Van Gelder, P.H.A.J.M., Voortman, H.G., Vrijling, J.K., 2004. Bivariate description of offshore wave conditions with physics-based extreme value statistics. Applied Ocean Research 26(3-4), 162-170.

Sagrilo, L.V.S., Naess, A., Doria, A.S., 2011. On the long-term response of marine structures. Applied Ocean Research 33(3), 208-214.

Salvadori, G., De Michele, C., 2007. On the use of copulas in hydrology: theory and practice. J. Hydrol. Eng. 12 (4), 369–380, (Special Issue: Copulas in Hydrology; doi: 10.1061/(ASCE)1084-0699(2007)12:4(369)).

Salvadori, G., De Michele, C., Kottegoda, N., Rosso, R., 2007. Extremes in Nature. An approach using Copulas. Vol. 56 of Water Science and Technology Library Series. Springer, Dordrecht, ISBN: 978-1- 4020-4415-1.

Salvadori, G., De Michele, C., 2010. Multivariate multiparameter extreme value models and return periods: A copula approach. Water Resour. Res. 46, W10501.

Salvadori, G., De Michele, C., Kottegoda, N., Rosso, R., 2007. Extremes in Nature. An approach using Copulas. Vol. 56 of Water Science and Technology Library Series. Springer, Dordrecht, ISBN: 978-1-4020-4415-1.

Salvadori, G., Tomasicchio, G.R., D'Alessandro, F. 2013. Multivariate approach to design coastal and offshore structures. Journal of Coastal Research, SI(65), pp. 386-391.

Salvadori, G., Durante, F., Tomasicchio, G. R., D'Alessandro, F., 2015. Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering. Coastal Engineering 95, 77–83.

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

Silva-González, F., Heredia-Zavoni, E., Montes-Iturrizaga, R., 2013. Development of environmental contours using Nataf distribution model. Ocean Engineering 58, 27-34.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8: 229–231.

Vanem, E., 2016. Joint statistical models for significant wave height and wave period in a changing climate. Marine Structures 49, 180-205.

Winterstein, S.R., Ude, T.C., Cornell, C.A., Bjerager, P., Haver, S., 1993. Environmental parameters for extreme response: inverse FORM with omission factors. In: Proceedings ICOSSAR-93.

Wist, H.T., Myrhaug, D., Rue, H., 2004. Statistical properties of successive wave heights and successive wave periods. Applied Ocean Research 26(3-4), 114-136.

Wu, S., 2014. Construction of asymmetric copulas and its application in two-dimensional reliability modeling. European Journal of Operational Research 238(2), 476-485.

Zhang, Y., Beer, M., Quek, S.T., 2015. Long-term performance assessment and design of offshore structures. Computers & Structures 154, 101-115.

Zhang, Y., Cao, Y., 2015. A fuzzy quantification approach of uncertainties in an extreme wave height modeling. Acta Oceanologica Sinica 34(3), 90—98.

Zhang, Y., Cheng, T., 2016 Characterizing the seasonal and directional varying properties in a marine environment. China Ocean Engineering 30(4), 549-564.

Zhang, Y., Kim, C.W., Tee, K.F., 2017a. Maintenance management of offshore structures using Markov process model with random transition probabilities. Structure and Infrastructure Engineering 13(8), 1068-1080.

Zhang, Y., Kim, C.W., Tee, K.F., Lam, J.S.L. 2017b. Optimal sustainable life cycle maintenance strategies for port infrastructures. Journal of Cleaner Production 142, 1693-1709.

Zhang, Y., Lam, J.S.L., 2015. Reliability analysis of offshore structures within a time varying environment. Stochastic Environmental Research and Risk Assessment 29(6), 1615—1636.

Zhang, Y., Lam, J.S.L., 2014. Non-conventional modeling of extreme significant wave height through random sets. Acta Oceanologica Sinica 33(7), 125-130.