The performance of the copulas in estimating the joint probability of extreme waves and surges along east coasts of the mainland China

Jiangxia Li[a,b], Shunqi Pan[c], Yongping Chen[d,e,*], Min Gan[e]

a School of Hydraulic Engineering, Changsha University of Science and Technology, Changsha, 410114, China

b Key Laboratory of Water-Sediment Sciences and Water Disaster Prevention of Hunan Province, Changsha, 410114, China

c Hydro-environmental Research Centre, School of Engineering, Cardiff University, Cardiff, CF24 3AA, United Kingdom

d State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China

e College of Harbour, Coastal and Offshore Engineering, Hohai University, Nanjing, 210098, China

Abstract

In designing coastal and nearshore structures, the joint probability of the wave heights and storm surges is essential in determining the possible highest total water level. The key elements to accurately estimate the joint probability are the appropriate sampling of the extreme values and selection of probability functions for the analysis. This study is to provide a full assessment of the performance of the different methods employed in the joint probability analysis. The bivariate extreme wave height and surge samples are analysed using 2 different probability distributions and the performance of 4 copulas, namely: Gumbel–Hougaard copula, Clayton copula, Frank copula and Galambos copula, is assessed. The possible highest total water levels for 100-year return period along the coastline of the mainland China are estimated by the joint probability method with the Gumbel–Hougaard copula. The results show that the wave heights and surges are highly correlated in the areas of dense typhoon paths. The distributions of the possible highest total water levels show a higher value in the southeast coast and lower value in the north. The results also indicate that at the locations where the sea states are energetic, the joint probability approach can improve the accuracy of design.

Key words: Coast of the mainland China; Joint probability; Copula; Extreme wave height; Extreme surge level

1. Introduction

Designing the coastal and offshore structures requires the consideration of a broad range of ocean factors due to complexity of the environment surrounding them. Amongst them, extreme waves and storm surges are two main factors. Under severe meteorological conditions, such as those during typhoons or cold storms, the extreme waves and storm surges can be closely correlated due to their driving forces. Joint probability analysis commonly becomes essential in estimating the extreme water levels to ensure the effective and sustainable designs of coastal engineering structures as demonstrated in the studies of Serafin and Ruggiero (2014) and Wahl et al (2015).

In joint probability analysis, a wide range of probability distributions of simultaneous environmental variables are obtained with the bivariate methods, as used by Ferreira and Guedes Soares (2002), Galiatsatou and Prinos (2007). Furthermore, Bruun and Tawn (1998) compared the properties of two extreme value methods: the univariate structure variable method and multivariate joint probability method, and found that the latter provided more useful and accurate design information when applied to several sites along Dutch coastlines. Based on a marginal distribution function fitted to the water level and wave height and their dependence, Hawkes et al. (2002) conducted a joint probability analysis, which was seen to perform better than the commonly used structure variable approach and joint exceedance approach.

However, during the last few decades, the copula theory, which has been initially used in finance, insurance and other economic sectors, has been widely adopted for joint probability analysis in the fields of hydrology (Mikosch, 2006) and coastal engineering (Salvadori et al., 2015). A copula function can connect different environmental variables without any hypothesis about their marginal distributions, and provides a powerful tool for the joint analysis of multivariate data. Recent examples of adopting the copula theory in hydrology fields include the study of extreme rainfalls (Salvadori and De Michele, 2004; Zhang and Singh, 2007), flood frequency for rivers (Chen et al., 2012; Sraj et al., 2015) and droughts (De Michele et al., 2013).

In coastal engineering applications, the copula theory has been found to be useful in providing increased flexibility in modelling the joint probabilities of ocean hydrodynamic variables. As stated in Coles et al (1999), quantifying dependence plays an importance role in the joint probability analysis. In dealing with the dependence between two variables, Wist and Myrhaug (2004) modelled two successive wave heights exceeding a certain threshold by a Gaussian copula and compared the results with field measurements and laboratory data. De Waal and van Gelder (2005) analysed the joint probability of extreme wave height and wave period using the Burr–Patero–Logistic copula. Similar studies were also conducted by Montes–Iturrizaga and Heredia–Zavoni (2015), as well as Vanem and Erik (2016). Wahl et al. (2010) carried out a study between two storm surge parameters using the Gumbel–Hougaard (GH) copula. Chini and Stansby (2010) used an integrated modelling system to investigate the joint probability of the extreme wave heights and water levels at Walcott, on the eastern coast of UK for determining the changes in the overtopping rates. Gruhn et al. (2012) used the Frank copula to estimate the joint probability of the water level residuals and significant wave heights along the coast of the Baltic Sea. Wahl et al. (2012) applied Archimedean copula functions in the German Bight to determine the exceedance probabilities of storm surges and wind waves. Masina et al. (2015) used a copula-based approach to estimate the joint probability of the water levels and waves at the Ravenna coast in Italy. The probability of failure/inundation was estimated by the direct integration method, and the coastal flooding risks were calculated. Galiatsatou and Prinos (2016) applied the copula method to investigate the changes in the joint probabilities of extreme wave heights and corresponding storm surges with time in the Aegean Sea. Ward et al. (2018) used the copula models to analyse the dependence between sea level and river discharge as well as the probability of flooding events in global deltas and estuaries. Bevacqua et al. (2019) discovered a higher probability of compound flooding from precipitation and storm surge in Europe under climate change using a copula-based multivariate probability model.

For extreme events, Gudendorf and Segers (2010) proposed the extreme value copulas for extreme multivariate analysis due to their capability of describing the upper tail dependence well. Mazas and Hamm (2017) used an event-based approach for determining extreme joint probabilities of waves and sea levels by focusing on the sampling of extreme events. In their study, three extreme value copulas (GH copula, Galambos copula, Husler–Reiss copula) were compared, and their results showed that different extreme value copulas would yield similar results, but the sampling methods could cause a large difference in the joint probability results. The samples could be selected by different ways. For example, in the sampling of extreme wave heights and surges, some researchers sample the extreme wave heights (or surges) and the simultaneous surges (or wave heights) by the block maxima method (Li and Song, 2006). Others consider the "impact" of the events and select the samples according to a defined response function, i.e. total water levels, overtopping and run-up (Gouldby et al., 2014; Serafin et al., 2014; Rueda et al., 2016). Also, the extreme pairs of samples by defining the storm events using certain thresholds of variables are used (Li et al., 2014; Wahl et al., 2016).

For multivariate cases, the dependence among a large range of extreme ocean elements like wave height, water level, wave period, storm duration, etc. was assessed. Corbella and Stretch (2012, 2013) investigated the dependence between storm parameters: significant wave height, peak wave period, duration, inter-arrival time, and water level, by applying a copula-based statistical model under varying climatic conditions. Li et al. (2014) analysed the variates of extreme storm events (wave height, wave period, sea level, wave direction, and storm duration) under deep-water wave conditions, where the Monte Carlo method and four other methods to construct the dependency structures based on the copula functions, physical relationship, and extreme value theory were adopted. It was found that the Gaussian copula model was the most suitable wave climate simulation method for the Dutch coast. Rueda et al. (2016) used the generalized extreme value (GEV) distributions and Gaussian copula to model the dependence between multivariate extremes related to coastal floods for different weather patterns. Lin-Ye et al. (2016) applied a hierarchical Archimedean copula to characterize storm intensity based on the storm energy, unitary energy, peak wave period, and duration on the Catalan coast. Montes–Iturrizaga and Heredia–Zavoni (2016) developed a multivariate model for the joint distributions of environmental variables using vine copulas, which was applied to build trivariate environmental contours of the wave height, period, and wind velocity at the Gulf of Mexico. Zhang et al. (2018) modelled multivariate ocean data using asymmetric copulas and compared the results with those obtained by traditional copulas.

The applications to the coastal waters of China are also seen rapidly emerging in recent years. Tao et al. (2013) developed a criterion to classify the intensity grade of a storm surge by the joint return period of the extreme water levels and wave heights in Qingdao. Yang and Zhang (2013) applied the GH copula to analyse the joint probability of extreme winds and wave heights at the Bohai Bay. Dong et al. (2015) used the Clayton copula to clarify the relations between the group height and length of ocean waves based on laboratory data and field wave data near the coast of Zhejiang province. Dong et al. (2017) studied the joint return probability of the wind speed and rainfall intensity in a typhoon-affected sea area close to Shanghai using the Weibull distribution and GH copula. More recently, Yin et al. (2018) estimated the extreme sea levels in the Yangtze estuary using the quadrature joint probability optimal sampling method (JPM-OS) with consideration of the typhoon field, wave height, and sea level in the studied region. Yang and Qian (2019) analysed the joint probability of typhoon-induced surges and rainstorms at Shenzhen and derived trivariate joint distributions and conditional distributions of these variables based on the copula method.

To estimate the desired design combination of wave height and surge accurately under extreme conditions can be rather challenging. Many studies have outlined that a univariate frequency analysis may not be capable of assessing the occurrence probability of extremes if the events are characterized by interrelated random variables (Chebana and Ouarda, 2011; Masina et al., 2015). According to Marcos et al. (2019), the return periods of extreme sea levels are underestimated in 30% of the coasts around the world if dependence is neglected. In particular, along the coasts of China, Li and Song (2006) analysed the correlations between the extreme wave heights and extreme water levels in the coastal waters of Hong Kong using the Gumbel–logistic model. The result proved that applying the commonly used empirical method to estimate the total water level (by directly adding the univariate extreme values) may not be sufficiently accurate to derive the coastal design criteria. On the other hand, because of the lack of long-term matched oceanic data, most of the previous studies only focused on a limited area or specific observation station. Therefore, it is necessary to carry out further research to clarify the relationships between extreme wave heights and storm surges and devise a realistic and safe design in coastal and offshore engineering.

Built on the model results from the previous work of Chen et al. (2019), which used the GH copula in analysing the joint probability of the wave height and surge along the coast of the mainland China, this study is to fully examine the performance of four different types of copulas using the existing model results from Chen et al. (2019) in estimating the joint probability. This study uses the annual N-largest sampling method with a detailed analysis of the predominance of joint extreme samples, in an attempt to effectively increase the available sample size compared to previous studies. Then a comprehensive analysis of the dependence between wave height and surge is conducted on the extreme samples obtained. The established joint probability model is subsequently applied to 87 selected locations representing the entire mainland China coast, to estimate the extreme combined water levels (CWLs) for flood risk assessment.

2. Study area and data

In this study, the model results of the significant wave heights (Hs) and surge levels (S) over a 35-year (1979–2013) period as detailed in Li et al. (2018) are used. For the sake of completeness, the model setup and applications are briefly presented here. The computational domain covered an area from 105 °E to 140 °E and from 15° N to 41 °N, as shown in Fig. 1. A coupled wave (FVCOM-SWAVE) and hydrodynamic (FVCOM) model (Qi et al., 2009), which was well calibrated and validated in the nearshore and offshore area by Li et al. (2018), was used. The model used an unstructured mesh with a spatial resolution of 1 degree at the open boundaries and finer than 0.1 degree in the coastal areas. Along the open boundaries, the model was driven by the tide conditions obtained from TPXO database. The modified ECMWF re-analysis wind data with a parametric typhoon model in order to account for the effects of 862 typhoons during the simulation period was used as the sea surface forcing. Hourly wave height and surge data from the model at nine nearshore locations which are identical to those in Chen et al. (2019), as shown in Fig. 1, are extracted from the model results and used for the joint probability analysis in this study. The selection of those nearshore locations is mainly due to the availability of field measurements for validating the hydrodynamic model.

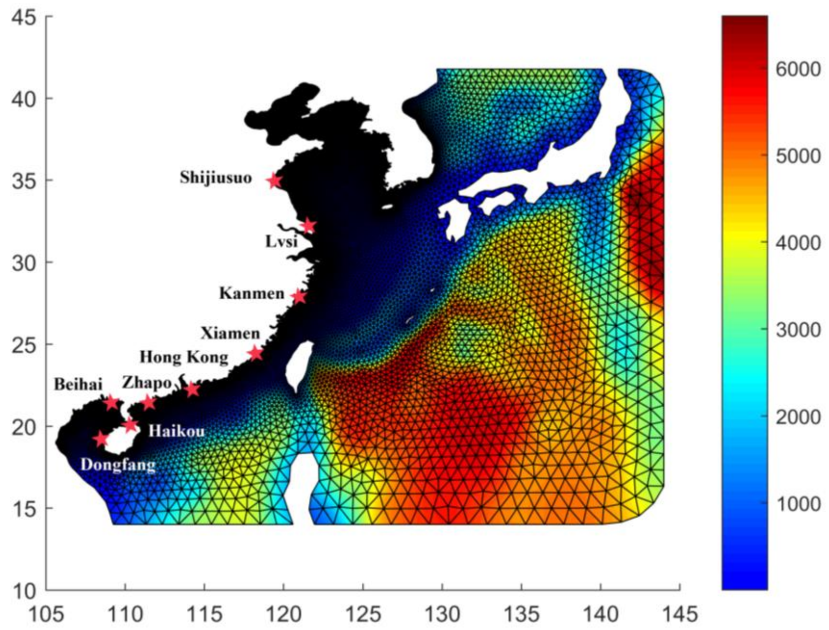

Fig. 1 Model area/mesh and the locations of the nine nearshore stations along the mainland China coast for applications (colour represents the water depth, m)

3. Methodology

3.1 Sampling method

For extreme analysis, sampling of the extreme values from the time series is a key step. When the data length is sufficiently long, the annual maximum (AM) method is commonly used to select the joint extreme samples to ensure the independence of extreme samples (Sraj et al., 2015; Yang and Zhang, 2013). However, according to the studies of Bernardara et al. (2014) and Mazas and Hamm (2017), for effective bivariate analysis, the sample size should be normally more than 300. Therefore, in most cases, the AM method may only generate a small sample size of extreme events, insufficient to effectively capture the information of the dependence between the variables. To overcome this, the peak over threshold (POT) method can be the effective one for selecting multivariate samples (Li et al., 2014; Mazas et al, 2014). Compared to the block maxima approach, POT method is advantageous when selected peaks result from different storm events. However, the POT-based joint sampling methods can present with the major difficulty in determining the values of the thresholds, particularly in the cases of highly variable hydrodynamic conditions temporally and spatially over a large study area such as this study.

Based on the block maximum sampling method for univariate analysis (Galiatsatou, 2011), in this study, an annual N-largest (ANL) joint extreme sampling method is proposed. This method selects the top N samples in each year such that it can capture more information than the AM method. Unlike the POT methods, the number of samples selected per year can be pre-determined in the ANL method, so that the extreme conditions can be fairly represented over the study area. In addition, to ensure the independence of the extreme events selected, a standard storm length covering both sides of each peak is considered. The standard storm length generally ranges from 24 to 72 hours in coastal storm analysis, following several previous studies (Basco and Walker, 2010; Martzikos at al., 2021; Marcos at al., 2019). It is set to be 48 hours in this study after conducting a sensitivity test suggested by Tawn (1988): provided the storm length is approximately correct the estimates of quantiles should not change too much by making small changes to this length. The simultaneous S is selected within the standard storm length along with the N-largest Hs to account for the possible time lag between extreme Hs and S. The number of samples per year (N) can be set accordingly to meet the required sampling size. Thus, in this study, by considering data length available over the 35-year period and the required sample size for joint probability analysis suggested by Mazas et al. (2014), N = 10 is used.

3.2 Univariate probabilistic distributions

Before establishing the dependence between wave height (Hs) and surge level (S), a frequency analysis would be required for each variable to define its marginal distribution. The two probabilistic distributions as shown in Table 1 are tested in this study for searching the best fit of the samples:

Table 1 The cumulative distribution function (CDF) of two probabilistic distributions

Distribution	CDF
Pearson-III (P3)	$F_p(x)=\frac{[\frac{2}{\bar{x}C_vC_s}]^{\frac{4}{C_s^2}}}{\Gamma(\frac{4}{C_s^2})}\int_{a_0}^{x}(x-\bar{x}+\frac{2C_v}{C_s}\bar{x})^{\frac{4}{C_s^2}-1}\cdot\exp(-\frac{2}{\bar{x}C_vC_s}(x-\bar{x}+\frac{2C_v}{C_s}\bar{x}))dx$
	where, Γ is the gamma function; $\bar{x}$ is the mean value of the samples;

	C_v and C_s are the coefficients of variation and skewness.
Generalized Extreme Value (GEV)	$F_{gev}(x) = \exp(-(1+k\frac{x-\mu}{\sigma})^{-1/k})$ where, μ, σ and k are the location, scale and shape parameters respectively.

3.3 Copulas

According to the theory of Sklar (1959), there exists a copula, C, that can connect the marginal distributions, $u_1 = F_X(x)$ and $u_2 = F_Y(y)$, to form the CDF (Genest and Favre, 2003) expressed as:

$$F(x,y) = C(F_X(x), F_Y(y)) \quad (1)$$

The commonly used copula families include Gaussian copula, t-copula, extreme value copula (EV-copula) and Archimedean copula. Among them, the Archimedean copula family has been frequently applied to the hydrologic fields. Meanwhile, Gudendorf and Segers (2010) suggested that EV-copula could also well describe the upper tail dependence for an extreme multivariate analysis. Thus, in this study, three commonly used copulas under the Archimedean family: Gumbel–Hougaard (GH) copula, Frank copula, and Clayton copula, together with an EV-copula, Galambos copula, are examined. The EV-copula is a type of copula which not only satisfies all the definitions and properties of copulas, but also meets the max-stable property for fixed integer n, i.e.

$$\lim_{n\to\infty} C_F(u_1^{1/n},\ldots,u_d^{1/n})^n = C(u_1,\ldots,u_d), \quad (u_1,\ldots,u_d)\in[0,1]^d \quad (2)$$

In fact, GH copula fits the properties of both Archimedean copula and EV-copula groups.

The generator function, CDF and probability density function (PDF) of these copulas are listed in Table 2, where u_1 and u_2 are the marginal distributions and θ is the parameter of copula which describes the dependencies. The Galambos copula which belongs to EV-copulas does not have a generator function.

Table 2 The generator function, CDF and PDF of four copulas

Copula	Function	Functions

names	names	
Gumbel–Hougaard copula	generator function	$\varphi(t)=(-\ln t)^{\theta}$
	CDF	$C(u_1,u_2,\theta)=e^{-[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{1/\theta}}$
	PDF	$c(u_1,u_2,\theta)=\frac{\{-1+\theta+[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{1/\theta}\}}{u_1u_2e^{[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{1/\theta}}}$ $\times(-\ln u_1)^{-1+\theta}\times(-\ln u_2)^{-1+\theta}[(-\ln u_1)^{\theta}+(-\ln u_2)^{\theta}]^{-2+1/\theta}$
Frank copula	generator function	$\varphi(t)=-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$
	CDF	$C(u_1,u_2,\theta)=-\frac{1}{\theta}\ln[1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}]$
	PDF	$c(u_1,u_2,\theta)=\frac{\theta\cdot e^{\theta(1+u_1+u_2)}(1+e^{\theta})}{(e^{\theta}-e^{\theta+\theta u_1}+e^{\theta u_1+\theta u_2}-e^{\theta+\theta u_2})^2}$
Clayton copula	generator function	$\varphi(t)=\frac{1}{\theta}(t^{-\theta}-1)$
	CDF	$C(u_1,u_2,\theta)=(u_1^{-\theta}+u_2^{-\theta}-1)^{-1/\theta}$
	PDF	$c(u_1,u_2,\theta)=(1+\theta)(u_1u_2)^{-\theta-1}(u_1^{-\theta}+u_2^{-\theta}-1)^{-2-1/\theta}$
Galambos copula	CDF	$C(u_1,u_2,\theta)=u_1\times u_2\times\exp\{((-\ln u_1)^{-\theta}+(-\ln u_2)^{-\theta})^{-\frac{1}{\theta}}\}$
	PDF	$c(u_1,u_2,\theta)=e^{[(-\ln u_1)^{-\theta}+(-\ln u_2)^{-\theta}]^{-1/\theta}}$ $\times\{1+(-\ln u_1)^{-\theta-1}\times(-\ln u_2)^{-\theta-1}\times[(-\ln u_1)^{-\theta}+(-\ln u_2)^{-\theta}]^{-2/\theta-2}$ $\times[1+(1+\theta)[(-\ln u_1)^{-\theta}+(-\ln u_2)^{-\theta}]^{1/\theta}]-[(-\ln u_1)^{-\theta}+(-\ln u_2)^{-\theta}]^{-1/\theta-1}$ $\times[(-\ln u_1)^{-\theta-1}+(-\ln u_2)^{-\theta-1}]\}$

3.4 Dependence

Several methods are available to determine the dependence structure between two random variables X and Y. They are commonly used to calculate the correlation coefficients, for example, Pearson's r correlation coefficient, Spearman's ρ coefficient, or Kendall's τ coefficient. In this study, Kendall’s τ coefficient is chosen to quantify the dependence between the Hs and S samples. It describes the dependence between the samples by ranking the variables with the following expression:

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ disconcordant\ pairs)}{n(n-1)/2} \quad (3)$$

where n is the total number of pairs. Any pair of observations, (x_i, y_i) and (x_j, y_j), where $i \neq j$, is reckoned to be concordant if the ranks for both the elements agree, i.e., both $x_i < x_j$ and $y_i < y_j$ holds or both $x_i > x_j$ and $y_i > y_j$ holds, and otherwise is regarded as the discordant pair. Therefore, $\tau = 0$ indicates the perfectly independent cases and $\tau = 1$ indicates perfectly dependent cases.

Generally, in the extreme analysis, the dependency is determined for the extreme values. However, the correlation coefficients for the extreme values can be less capable of fully capturing the asymptotic dependency (Mazas et al., 2014). Thus, in this study, the chi-plots are used as graphical tools to assess the dependence between the extreme Hs and S. It supplements an ordinary scatterplot of the data by providing a graph that has characteristic patterns depending on whether the variates are independent, with some degree of monotone relationship or more complex dependence structure. Two variables $(\Lambda_i, \mathrm{X}_i)$ as suggested by Fisher and Switzer (1985, 2001) are used in the scatterplots as:

$$\Lambda_i = 4S_i \max\left\{\left(F_i - \frac{1}{2}\right)^2, \left(G_i - \frac{1}{2}\right)^2\right\} \quad (4)$$

$$\mathrm{X}_i = \left(H_i - F_iG_i\right) / \{F_i(1-F_i)G_i(1-G_i)\}^{\frac{1}{2}} \quad (5)$$

where,

$$S_i = sign\left\{\left(F_i - \frac{1}{2}\right)\left(G_i - \frac{1}{2}\right)\right\} \quad (6)$$

$$F_i = \sum_{j \neq i} I(x_j \leq x_i) / (n-1) \quad (7)$$

$$G_i = \sum_{j \neq i} I(y_j \leq y_i) / (n-1) \quad (8)$$

$$H_i = \sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i) / (n-1) \quad (9)$$

and I is the indicator function.

The relationships between Kendall's coefficient τ and the correlation index, θ, for copulas introduced in Section 3.3 are listed in Table 3.

Table 3 Relationships between Kendall's coefficient τ and parameter θ for different copulas

Copula	Relationship
Gumbel–Hougaard copula	$\tau = 1 - 1/\theta$
Clayton copula	$\tau = \theta / (\theta + 2)$
Frank copula	$\tau = 1 + \frac{4}{\theta}[\frac{1}{\theta}\int_0^{\theta} \frac{t}{e^t - 1} dt - 1]$
Galambos copula	$\tau = \frac{\theta + 1}{\theta}\int_0^1 (\frac{1}{s^{1/\theta}} + \frac{1}{(1-s)^{1/\theta}} - 1)^{-1} ds$

3.5 Return period

In joint probability analysis, the bivariate return period can be defined. The OR return period (T_o) indicates that at least one of the variable exceeds a certain value, and the AND return period (T_a) indicates that both the variables exceed a certain value. They can be calculated using the following expressions:

$$T_o(x, y) = \frac{1}{1 - F(x, y)} \tag{10}$$

and

$$T_a(x, y) = \frac{1}{1 + F(x, y) - F_X(x) - F_Y(y)} \tag{11}$$

where $F_X(x)$ and $F_Y(y)$ are the marginal distributions and $F(x, y)$ is calculated by Eq. (1) by combining the CDF of the copula and corresponding marginal distributions.

4. Results

4.1 Dependence of extreme samples

For the joint probability analysis, it is necessary first to examine the dependency between the extreme Hs and S. As an example, the extreme wave height (Hs) and surge (S) sampled at nine nearshore stations are shown in Fig. 2. It indicates that those two variables are partly related as the data points present a clear linear relation at all stations, but with a high degree of scattering. Relatively stronger dependencies between the extreme Hs and S are found at Haikou, Zhapo, Hong Kong, Xiamen and Kanmen stations because the scatters show a more obvious linear trend, but at other stations, such dependency appears relatively weaker. It is also noticed that the stations with stronger dependencies are located in the coastal areas facing the open sea and are easily affected by typhoon events. Stations Beihai and Dongfang are to some extent sheltered by the land. Stations Lvsi and Shijiusuo are located in the mid-north coast where fewer typhoon events occur. This result indicates that the dependencies between the extreme Hs and S at certain locations can be influenced by typhoon events.

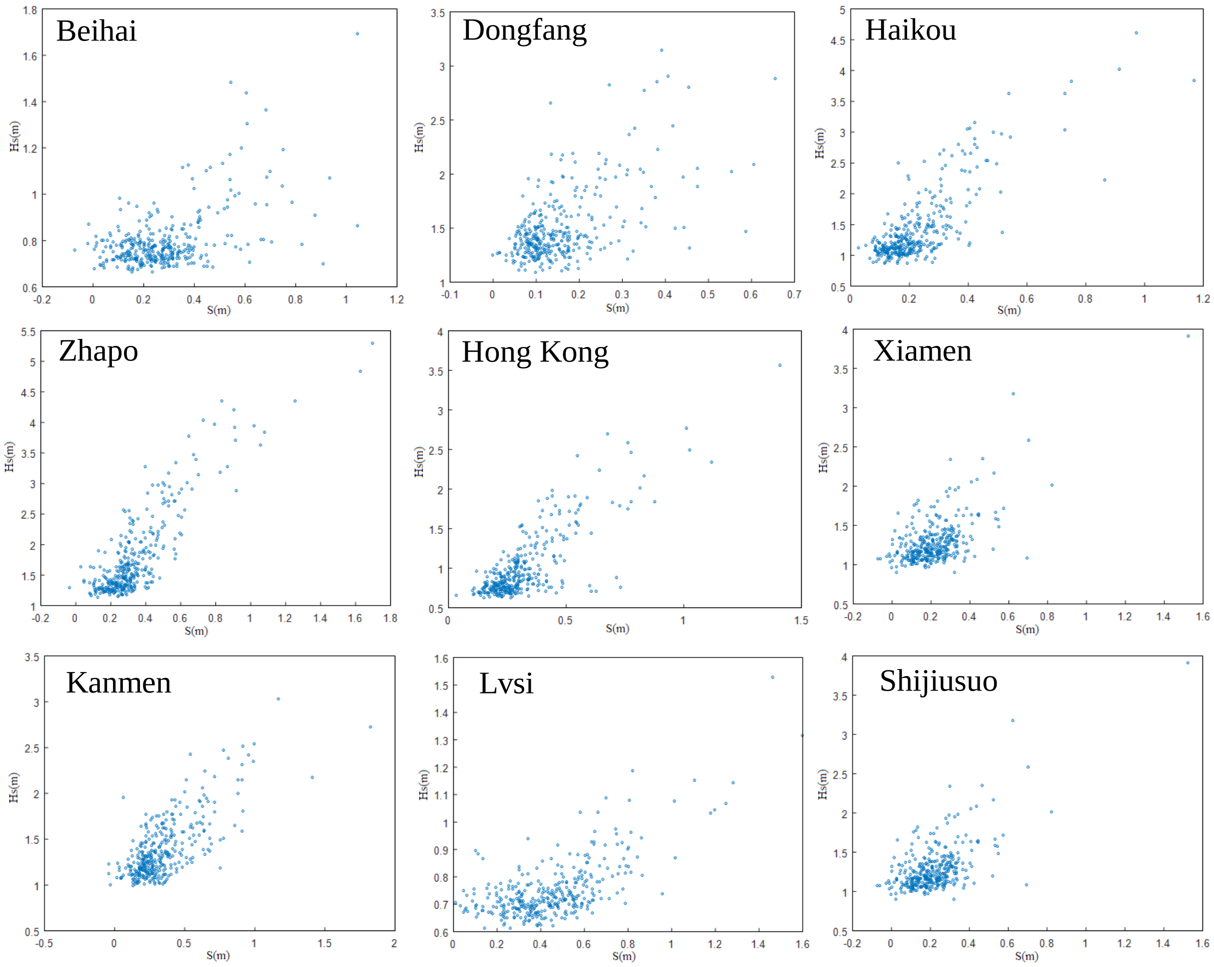

Fig. 2 Scatterplot of the N-Largest joint samples

The chi-plots for all nine stations are shown in Fig. 3. In the chi-plot, Λ_i measures the distance of a pair of variables from their medians: a positive (negative) value implies that both variables are on the same (opposite) side of their respective medians and a value close to 1 (0) implies they are larger or smaller relative to (close to) their respective medians, and X_i measures the dependence: a positive (negative) value describes a positive (negative) dependence, while a value close to zero suggests independence (Mazas et al., 2014). From Fig. 3, it can be seen that there is a clear positive dependence between the extreme Hs and S at all the stations. However, for the events where Λ_i is negative, there is only one population at all stations, whilst for positive Λ_i, two different populations are observed, namely, the upper and lower “lobes” as suggested by Fisher and Switzer (2001). The upper “lobe” corresponds to pairs where both the Hs and S are larger than their median, exhibiting a relatively strong dependence. This is because higher Hs and S are generally caused by the same extreme atmospheric event. In contrast, the lower “lobe” corresponds to a pair where both the Hs and S are smaller than their medians, exhibiting weak dependence. At most stations, there are two distinct upper and lower lobes, which indicates the bimodal dependence of wave height and surge due to relatively large events (such as typhoons) or weaker events. In other words, this bimodal dependence could be caused by two extreme situations: typhoon related extremes and non-typhoon related extremes. At the Shijiusuo station, however, the boundaries of the two “lobes” are obscure, which may be attributed to the low frequency of typhoon events at this location.

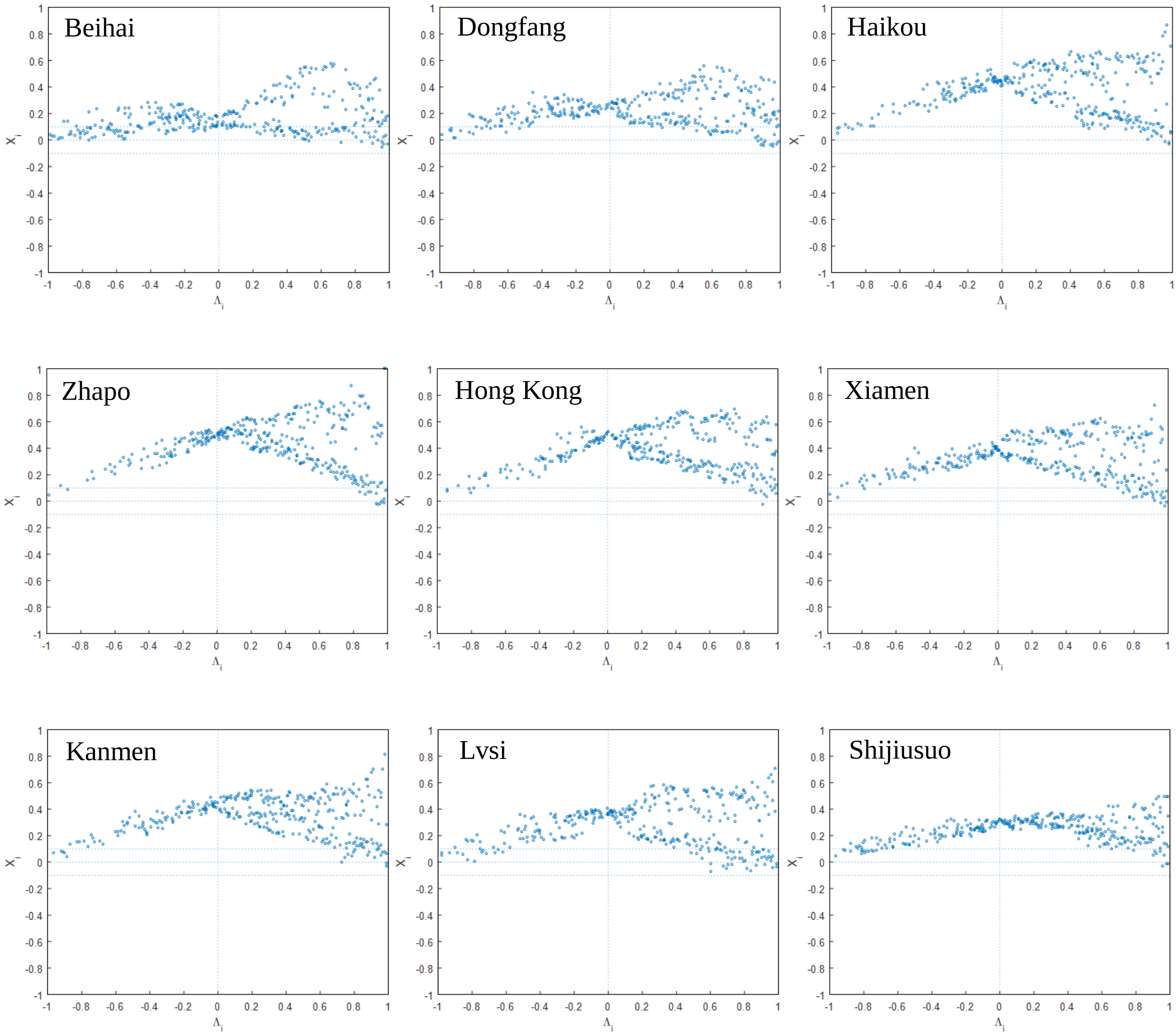

Fig. 3 Chi-plots of the N-Largest joint samples

Furthermore, the distribution of Kendall's coefficient τ over the computational domain is also calculated, as shown in Fig. 4a. The results clearly show that the coefficients in the southeast area of the computational domain are remarkably larger than those at other locations, which coincides well with the areas along the paths of frequent typhoons during the 35-year (1979–2013) period (Fig. 4b). Specifically, the dependence between the extreme Hs and S increases in the areas where the sea states are more energetic, which was also reported in Hawkes et al. (2002). It is found that this character could not be fully revealed with the AM sampling method as used in Chen et al. (2019), which also serves as an indication of the improvement when the ANL sampling method is used.

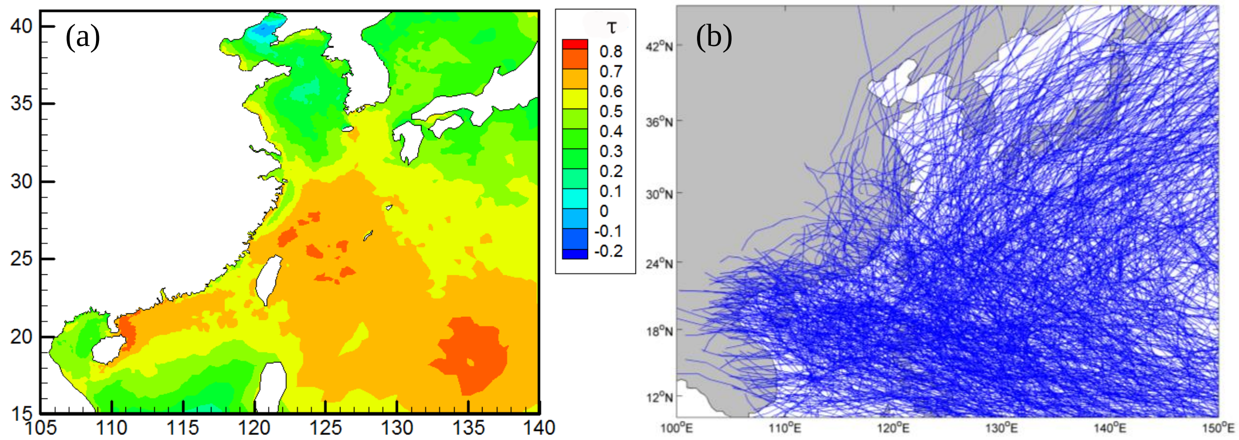

Fig. 4 Distribution of (a) Kendall's coefficient τ and (b) the typhoon tracks from 1979 to 2013 in the study area

More specifically, the Kendall's coefficient τ at the nine nearshore locations is shown in Fig. 5. The results indicate that at Beihai, Dongfang, Lvsi, and Shijiusuo stations, values of τ are generally smaller compared to other stations, just below 0.35, suggesting relatively weak dependence between the extreme Hs and S at these stations. At other stations, particularly Zhapo and Hong Kong, the dependence between the extreme Hs and S is strong. According to Fig. 4 and Fig. 5, it seems that the stations in the areas frequently affected by typhoons tend to have large τ coefficients, as expected since typhoon events can be a significant cause for the extreme Hs and S.

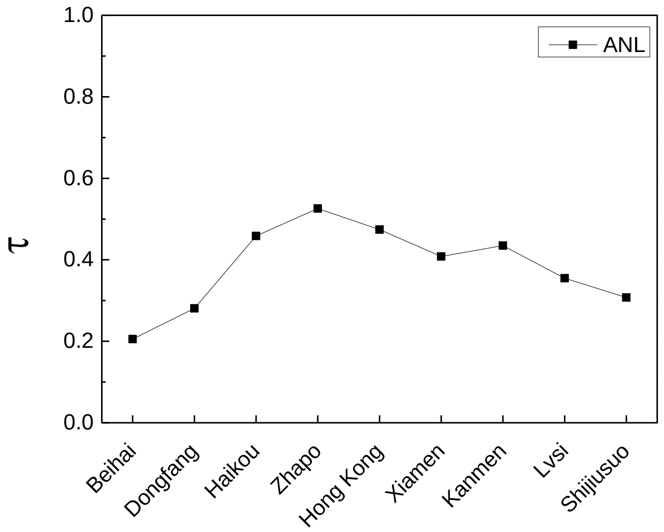

Fig. 5 Kendall's coefficient τ between Hs and S at the nine nearshore stations

4.2 Marginal distributions

An advantage of applying the copula theory to bivariate or multivariate probability analysis is that copulas allow different types of marginal distributions to be used for different variables. To examine the performance and fitness of copulas, in this study both Hs and S in the joint extreme samples at the nine nearshore locations are fitted with two univariate distributions as introduced previously. Since the study area is observed to be wave-predominated (discussed further in Section 6), extreme Hs data is sampled by the ANL method, and extreme S data is sampled based on the sampled Hs. By subsampling the N-largest data to annual maxima, two probability distributions introduced in Section 3.2 are used to fit the samples. The parameters in GEV are estimated by the maximum likelihood estimation method. In Fig. 6, the fittings of the Hs and S in the joint extreme samples with different distributions at the Kanmen station are plotted. Fig. 6 (a) shows that the P3 distribution fits better the extreme Hs samples at Kanmen than GEV distribution, whereas in Fig. 6 (b) the GEV distribution can comparatively better fit the extreme S samples.

To quantify the fitting results, Pearson's coefficient *r* (Pearson, 1895) between the samples (dot in Fig. 6) and theoretical values (line in Fig. 6) are calculated at Kanmen station and listed in Table 4. The Pearson's coefficient *r* could be calculated by,

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} . \tag{12}$$

where, X_i and Y_i are the sample values and the theoretical values; $\bar{X}$ and $\bar{Y}$ are the averaged values of X_i and Y_i. The largest correlation coefficients coincide with the best fit distribution chosen by Fig. 6.

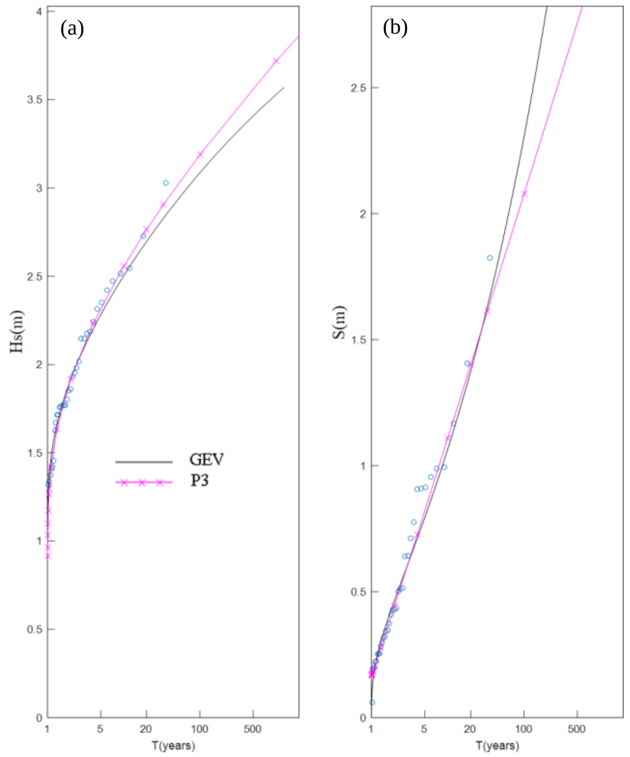

Fig. 6 Fitting of the samples with different distributions at Kanmen station: (a) wave height; (b) surge level

Table 4 Correlation coefficients between the samples and different distributions at Kanmen station (the best fit distributions are indicated in bold)

	GEV	P3
Wave height	0.9876	**0.9877**
Surge level	**0.9921**	0.9833

By combining the results in Fig. 6 and Table 4, the best fit distributions for the nine nearshore stations are summarized in Table 5. It can be seen that GEV distributions fit the extreme Hs samples better than P3 at 6 out of 9 stations, and the GEV distribution fits the extreme S samples better at all stations in the study area. Although not shown here, the 95% confidence intervals of the selected marginal are also examined to ensure a proper fit. It is reasonable to see that the confidence intervals increase from the lower tail to the upper tail. Therefore, the distributions of the Hs and S are determined by the selected probability distributions in this study.

Table 5 Chosen distributions for the Hs and S in the joint samples at the nine nearshore stations

Station	Beihai	Dongfang	Haikou	Zhapo	Hong Kong	Xiamen	Kanmen	Lvsi	Shijiusuo
Hs	GEV	GEV	GEV	P3	P3	GEV	P3	GEV	GEV

S	GEV	GEV	GEV	GEV	GEV	GEV	GEV	GEV	GEV

4.3 Selection of copulas

To determine the best fit copulas for the data sets in this study with the chosen marginal distributions of the extreme Hs and S as described previously, it is essential to examine the characteristics of each copula. Fig. 7 shows the probability density distributions of the GH copula, Clayton copula, Frank copula, and Galambos copula. It is clear that both GH and Galambos copulas have a pronounced upper tail density, suggesting that they are capable of describing the dependence in the upper tail of the distribution, i.e. upper tail dependence. However, the density distribution of the Clayton copula has a thick lower tail density, suggesting that it can better describe the dependence in the lower tail of the distribution, i.e. lower tail dependence. The Frank copula has a symmetric tail, i.e. no tail dependence, which can only be suitable for the symmetrical distributed samples.

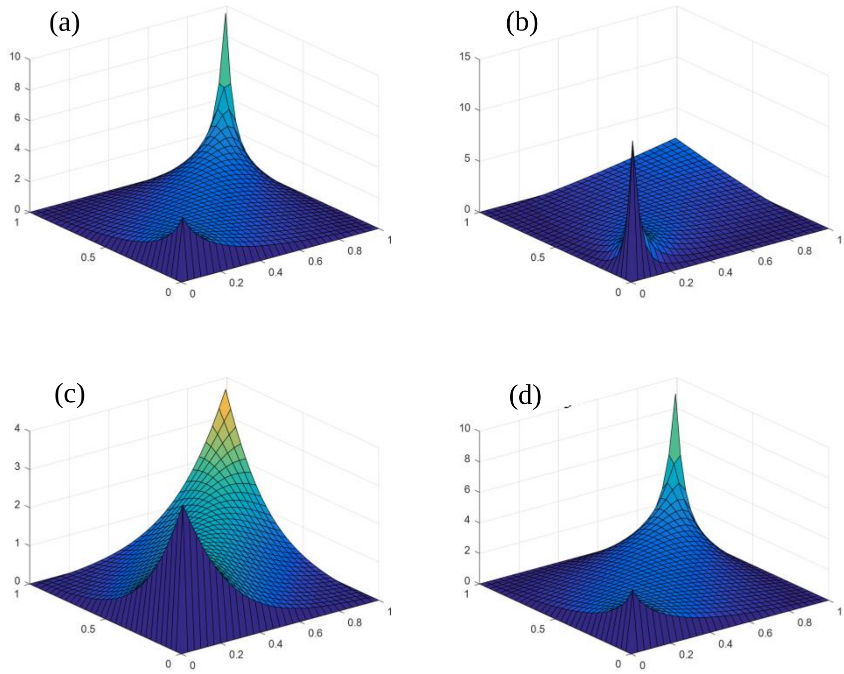

Fig. 7 Probability density distributions of (a) GH copula, (b) Clayton copula, (c) Frank copula, and (d) Galambos copula

To achieve the best match of the characteristics of the copulas shown in Fig. 7 with the samples in this study, the extreme samples at all nine stations are examined with the binary frequency histograms of the Hs and S. As shown in Fig. 8, at all stations, a thick upper tail density can be observed, although the frequency distributions are slightly different at different stations. In general, there is a clear suggestion that the GH copula and Galambos copula can be chosen in the probability analysis as they match well with all density distributions at those stations.

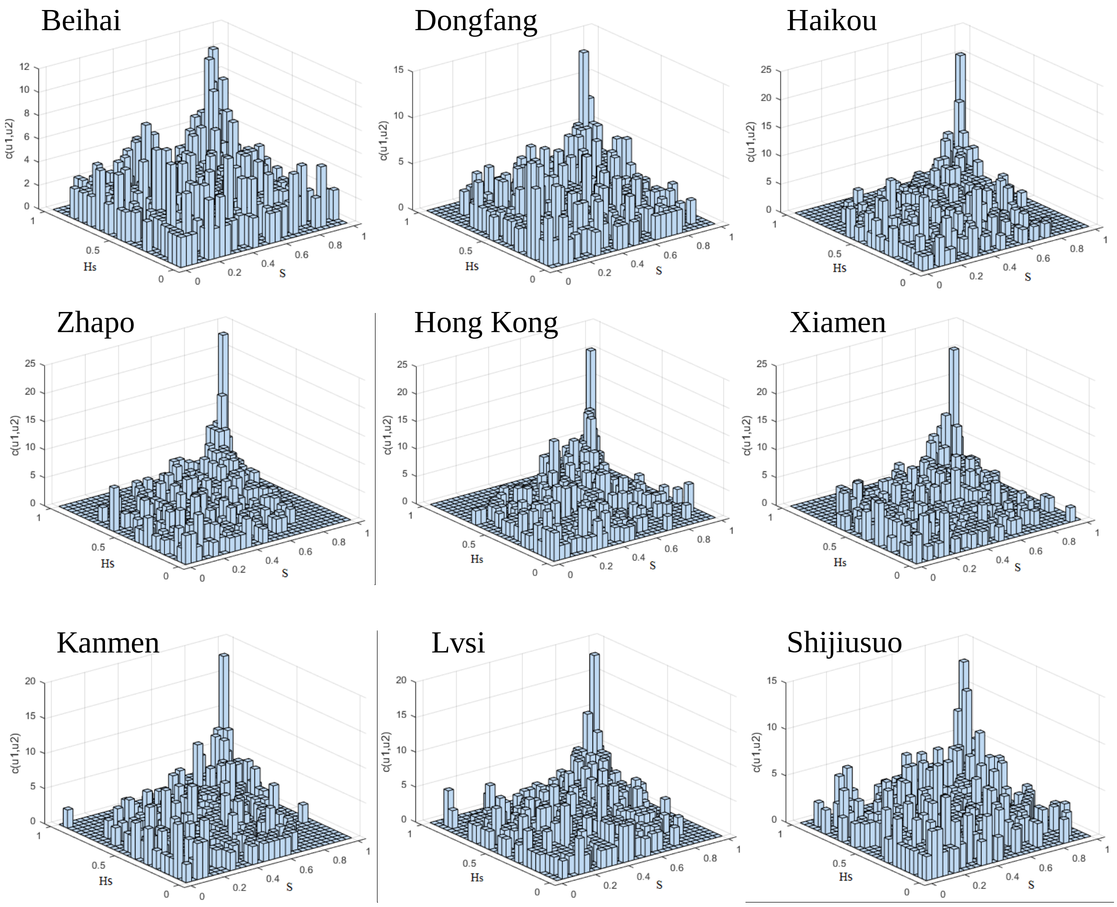

Fig. 8 Binary frequency histograms of Hs and S in the joint samples

However, for the completeness of analysis, all four copulas are also used to fit the joint extreme samples using Kendall's coefficient as introduced in Section 3.5. Fig. 9 shows their joint cumulative probabilities in comparison with those of the empirical copula at all nine stations. As the probability of the empirical copula is directly calculated based on the samples, any copula in the test that has the best fit with the empirical copula will be regarded as the optimal copula for the samples. It can be seen from the comparisons that the contours of four copulas provide a very similar fit in the mid-range of probabilities. However, Clayton and Frank copula perform poorly with tendency of underestimating the probability in the upper tail region while overestimate the probability in the lower tail region. This is related to the density distribution of those tested copulas. The results clearly show a general trend of good match of the GH and Galambos copulas with the empirical copula, better than the other two copulas, while Frank copula has the worst fit.

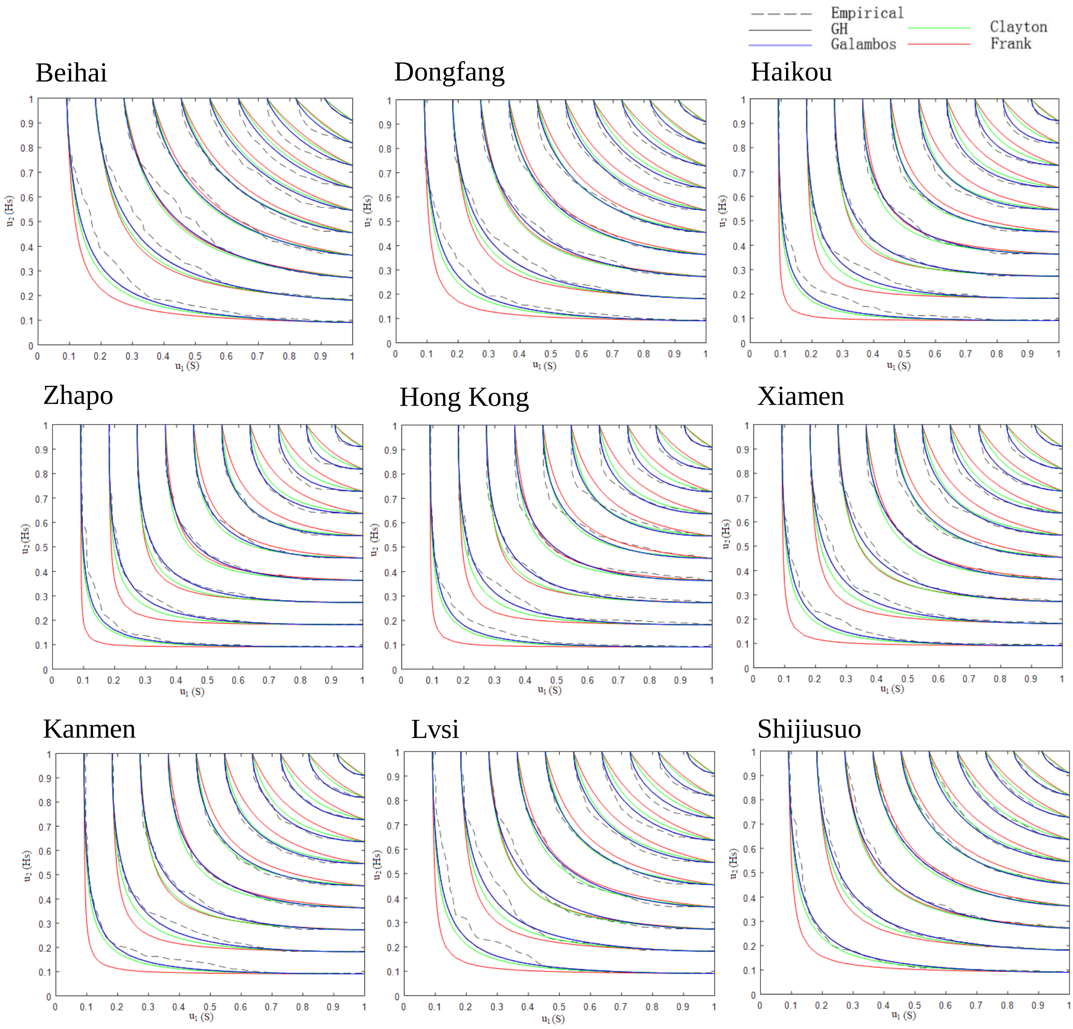

Fig. 9 Comparison of joint probability of four copulas with that of empirical copula

In addition, the Cramér-von Mises (CVM) test is carried out to compare the performance of the four copulas with that of the empirical copula quantitatively, using the following equation (Mazas and Hamm, 2017; Genest and Rivest, 1993):

$$S_n = \sum_{i=1}^{N}[C_n(U_i,V_i) - C_{\hat{\theta}}(U_i,V_i)]^2 , \qquad (13)$$

where, N is the sample size, (U_i,V_i) is the sample of the normalized ranks, C_n is the copula in test, and $C_{\hat{\theta}}$ is the empirical copula. The CVM statistics at all stations are shown in Fig. 10. It is clear that CVM values for the Galambos and GH copulas are the lowest amongst all 4 copulas, while GH copula preforms slightly better than Galambos copula. The results again confirm the outcomes of the probability density analysis of these copulas as shown in Fig. 7 and Fig. 8.

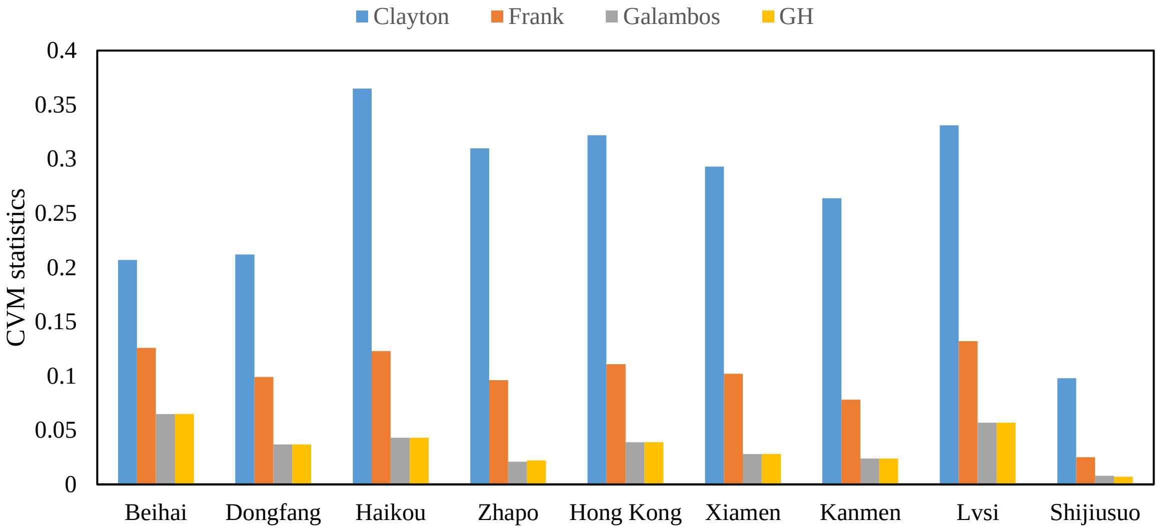

Fig. 10 CVM statistics at the nine nearshore stations

From the results presented in Fig. 10, it can be concluded that both GH and Galambos copulas, which have the lowest CVM values amongst all, are deemed to be the optimal ones for studying the joint probability of the extreme Hs and S along the east coast of the mainland China. It also highlights the necessity of using an EV-copula to conduct the joint probability analysis of extreme values. Considering that the GH copula has a simpler function than the Galambos copula, therefore it is decided that the GH copula is adopted in this study.

4.4 Joint probability

For the joint probability, both AND and OR return periods are assessed at all station. As an example, the isolines of the joint events with both return periods at the Kanmen station are shown in Fig. 11. In general, for the same joint event, the AND return period is found to be larger than the OR return period. Specifically, when calculating the joint probability of the variables, the selection of the different types of return period should be according to the aim of the study. In the following analysis in this study, the AND return period is applied. Concurrently, according to a previous study (Chen et al., 2019), the shapes of the isolines are diverse at different locations because the joint probability is location-specific, particularly in the nearshore areas. Because the distributions of the joint events at different locations are discussed in detail in a previous study (Chen et al., 2019), the isolines of the joint events at other stations are not provided here.

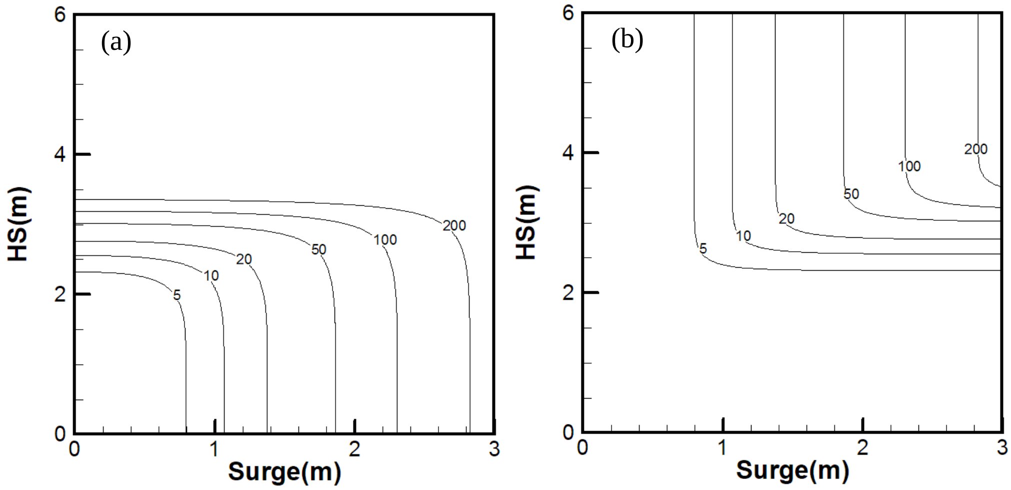

Fig. 11 Isolines of (a) the AND return period and (b) the OR return period at the Kanmen station

From Fig. 11, it can be seen that different combinations of the Hs and S can have the same return period along an isoline when calculating the joint events by using a cumulative probability. Thus, to search for the most probable joint event for a certain return period, the joint probability density is the best function to be used. The combined water level (CWL) which is the sum of the Hs and S is analysed in this study for engineering application. To determine the most probable CWL, the joint probability density is calculated to obtain the failure probability by integration over the failure region (Masina et al., 2015; Chen et al., 2019). Along the isoline of the failure probability, the point corresponding to the highest probability density is the most probable extreme event, which is the tangential point between the isoline of the failure probability (indicated by straight lines in Fig. 12) and a particular isoline of the probability density (indicated by curves in Fig. 12). Then the extreme CWL is calculated by adding the Hs and S of the most probable extreme event. Fig. 12 shows the isolines of the joint probability density and failure probability at the nine representative nearshore stations. The most probable joint events with a 50-year and 100-year return period at the nine nearshore stations are then determined according to Fig. 12, as shown in Table 6.

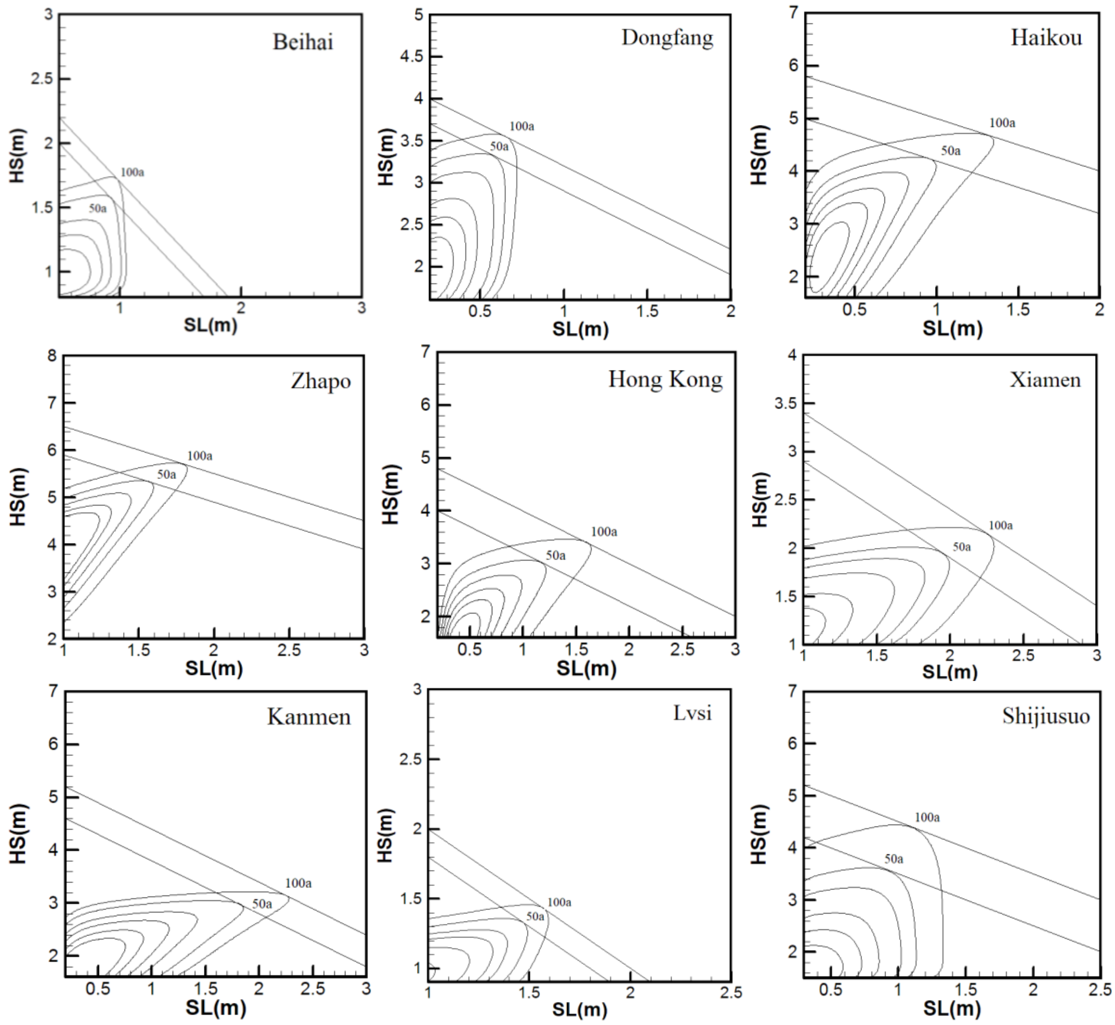

Fig. 12 Isolines of the joint probability density and failure probability

Table 6 Most probable 50-year and 100-year return level joint events

Station	50-year			100-year		
	Hs(m)	S(m)	CWL(m)	Hs(m)	S(m)	CWL(m)
Beihai	1.55	0.95	2.50	1.70	1.00	2.70
Dongfang	3.30	0.60	3.90	3.55	0.65	4.20
Haikou	4.25	0.95	5.20	4.65	1.35	6.00
Zhapo	5.25	1.65	6.90	5.65	1.85	7.50
Hong Kong	3.00	1.20	4.20	3.40	1.60	5.00
Xiamen	1.95	1.95	3.90	2.20	2.20	4.40
Kanmen	2.90	1.90	4.80	3.05	2.35	5.40
Lvsi	1.35	1.45	2.80	1.40	1.60	3.00
Shijiusuo	3.55	0.95	4.50	4.30	1.20	5.50

In engineering practice, when lacking the analysis of joint probability, the joint event for a certain period is typically estimated by an addition of the single event with specified return period. For example, a 100-year return level joint event is sometimes approximated by the sum of a 100-year

Hs (100Hs) and 10-year S (10S), the sum of a 10-year Hs (10Hs) and 100-year S (100S), or the sum of a 50-year Hs (50Hs) and 50-year S (50S) (Li and Song, 2006), or an addition of 100-year Hs and S (Code of Hydrology for Harhour and Waterway, JTS 145-2015, China). To compare the outcome of these combinations and the joint probability results, the Hs and S sampled by the univariate method without considering their dependence are used to calculate the Hs and S with 100-year, 50-year and 10-year return periods at all nine locations. The CWLs calculated by four empirical combinations described above are compared with those calculated by the joint probability method with the 50-year and 100-year return periods at the nine nearshore stations, as shown in Fig. 13, where the ranges of the CWLs from the 50-year to 100-year return levels calculated by joint probability method are presented for the sake of clarity. It can be seen from the figure that the 100-year CWLs calculated by joint probability method are larger than the “100Hs+10S,” “50Hs+50S,” and “10Hs+100S” combinations but are smaller than the “100Hs+100S” combination. In general, the “100Hs+10S” and “50Hs+50S” combinations are close to the 50-year return level CWLs calculated by joint probability which could be recommended to estimate the 50-year return level situation when the joint probability data is unavailable. Meanwhile, “50Hs+50S” combination is always within the range of the 50-year and the 100-year return levels calculated by joint probability method which could be a meaningful indicator for joint events between 50-year and 100-year return levels. The “10Hs+100S” combination is relatively smaller than other combinations, especially at Beihai, Dongfang, Haikou, Zhapo, and Shijiusuo stations, which indicates a strong wave predominant property.

The result suggests that three of the empirical combinations may lead to an unsafe design, with only the “100Hs+100S” combination being safe for an engineering design at all nine locations. However, the method of using the “100Hs+100S” combination to estimate a 100-year joint event is proposed based on the assumption that the Hs and S are independent random variables. This assumption may be unrealistic because it has been proved that the Hs and S are partly dependent in this study, as described in Section 4. Thus, it is necessary to conduct a joint probability analysis when designing engineering structures.

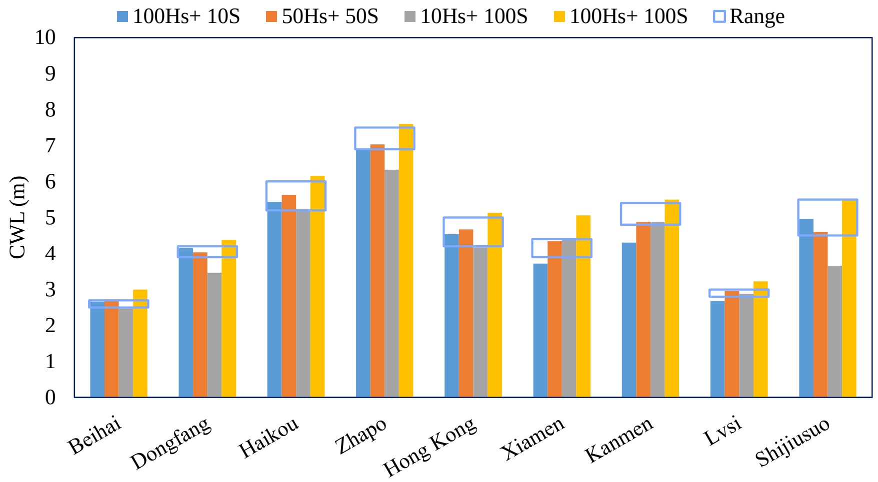

Fig. 13 The ranges of the CWL calculated by the joint probability method with GH copula for 50-year and 100-year return periods (shown as a box) in comparison with the CWLs calculated by the empirical combinations

5. Discussion

From the detailed comparison of the CWLs calculated by the empirical combination and joint probability method along the east coast of the mainland China, it is clear that the "100Hs+100S" combination is the only method which can lead to a safe design among the four empirical combinations. If this approach is adopted for the entire coastline, the extreme CWLs can be estimated for wide engineering applications. For the purposes of inter-comparison, Fig. 14 shows the distributions of the 100-year return level Hs and S at the studied coastline with 87 uniformly distributed locations. It can be seen from Fig. 14 (a) that the 100-year Hs along the southeast coast of the mainland China are remarkably larger than those at the other sites. However, in Fig. 14 (b), the distributions of the extreme S are rather uniform along the entire coast, with S being generally larger than 2 m from the mid-east to the south coast.

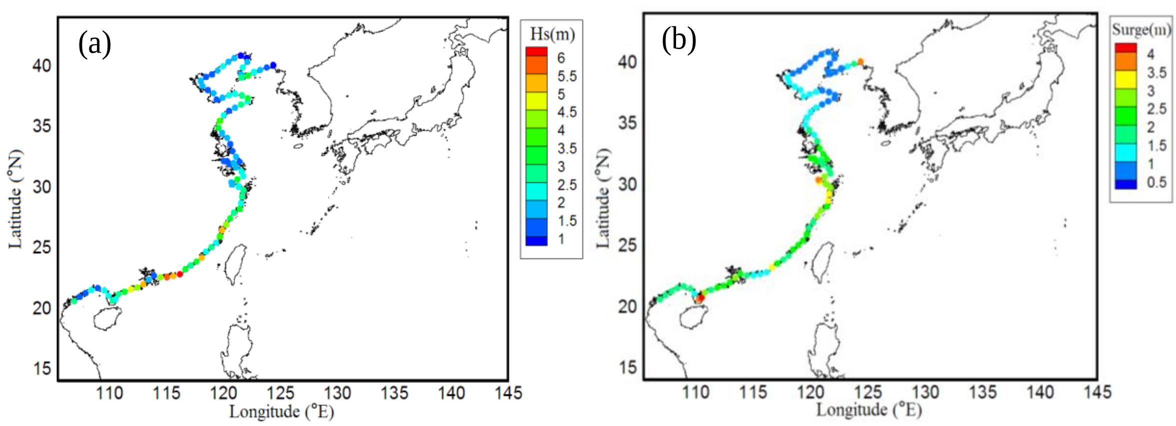

Fig. 14 Distributions of the 100-year return level: (a) wave height and (b) surge level along the coasts of the mainland China without considering their dependence

To investigate the hydrodynamic conditions for different areas in detail, a response coefficient (D) is defined as:

$$D = (Hs + S) / Hs \quad (14)$$

where Hs is the 100-year return level wave height and S is the 100-year return level surge.

Although partly related, wave and surge are characterized by different dynamic, and have different magnitudes and spatial scales. As a coastal environment is usually defined as wave-predominated or surge-predominated based on the relative contributions of the wave and surge on coastal processes studied, as well as on coastal morphodynamics, the coefficient D could give a first impression on the relative significance of these two variables.

With the response coefficient (D) presenting the relative contributions of the Hs and S for the same return period at different locations, the hydrodynamic conditions there can then be described as wave-predominated or surge-predominated. D is generally larger than 1. If D is between 1 and 2, the location could be described as wave-predominated since the Hs has larger impact than S; otherwise, if over 2, it is surge-predominated. The value of D could reflect the relative value of Hs and S. Higher values indicate a larger impact of S. Fig. 15 shows the distribution of coefficient D along the mainland China coast. It can be seen that coefficient D at most of the sites along the mainland China coast is between 1 and 2, which suggests that most of the areas along the mainland China coast are wave-predominated. The extreme wave height is obviously larger than surge level, which indicates a larger wave impact at these locations. This justifies the way that the joint extreme samples were selected in a wave-predominated manner in Section 3.1. For the southeast coast, the coefficient D is a little bit larger than 1, as these areas are facing open seas and are found in deep waters, which enhance the wave energy and mitigate the surge. However, in a few sites, D coefficients are far larger than 2, for example, the points in the Yangtze River estuary and Hangzhou bay. The water depths are small at these locations, and the shape of the estuary coastline may have caused surge to concentrate, resulting in those sites becoming surge-predominated.

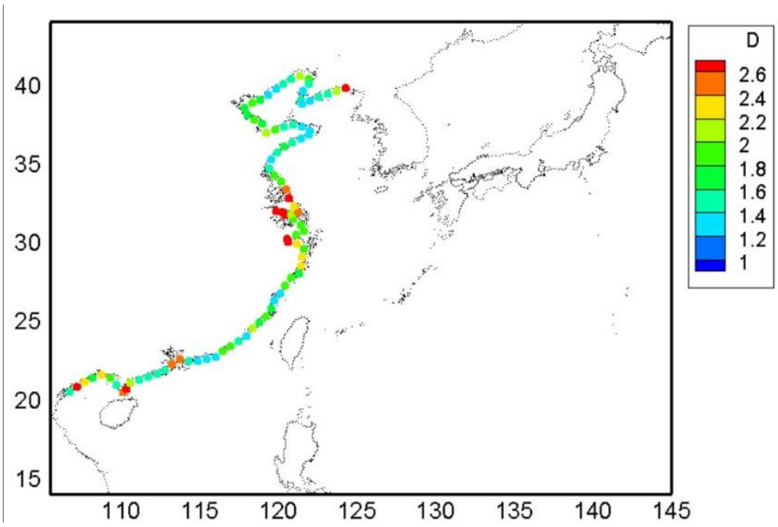

Fig. 15 Distribution of the response coefficient (D) along the coasts of the mainland China

Furthermore, Fig. 16 shows the distributions of the 100-year return level CWLs, calculated by the empirical method (100Hs+100S) and joint probability method. The distributions of the extreme CWLs calculated by the joint probability method show a relatively higher value in the southeast coast and lower value in the north. Although with the similar distribution patter, it is clear that using the empirical method by assuming Hs and S being independent random variables can yield a higher water level, but using the joint probability method can yield relatively more economical design conditions.

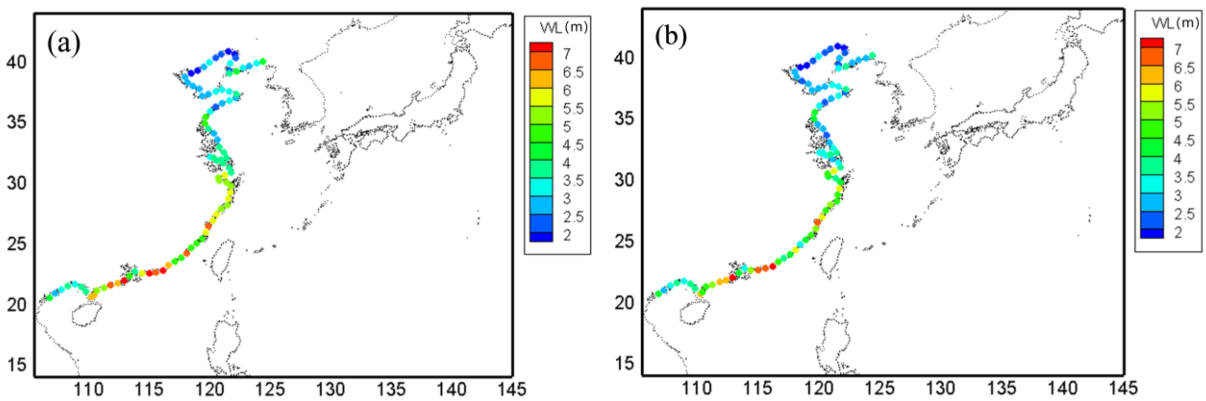

Fig. 16 Distribution of 100-year return level CWL calculated by the (a) empirical method (100Hs+100S) and (b) joint probability method along the coasts of the mainland China

As an indication of the improvement made between the 100-year CWLs calculated by the empirical and joint probability methods, a parameter, Q, is introduced and defined as:

$$Q = \frac{CWL_e - CWL_j}{CWL_e} \times 100\% , \tag{15}$$

where, CWL_e is the water level calculated by the empirical method, and CWL_j is the water level calculated by the joint probability method.

The distribution of the Q parameter shown in Fig. 17 indicates that the improvements in the joint probability method compared with the empirical method are not notable in the north coast (Bohai Sea coast) and south coast with a Q of under 6%, which means that the use of empirical combinations at these locations is relatively reasonable. However, for the mid-east mainland coast and southeast mainland coast, Q is relative larger, over 25% at its maximum. Thus, using the design water level calculated by the empirical method at these locations may be inaccurate. In other words, the joint probability method can yield better results at the sites where the hydrodynamic conditions are generally complex or energetic. For example, two major estuaries (Yangtze and Hangzhou Bay) are located the mid-east coast. The southeast coast is frequently affected by typhoon events, particularly near the Taiwan Strait, which can incur stronger hydrodynamic processes and cause larger diversity between water levels calculated by empirical and joint probability methods.

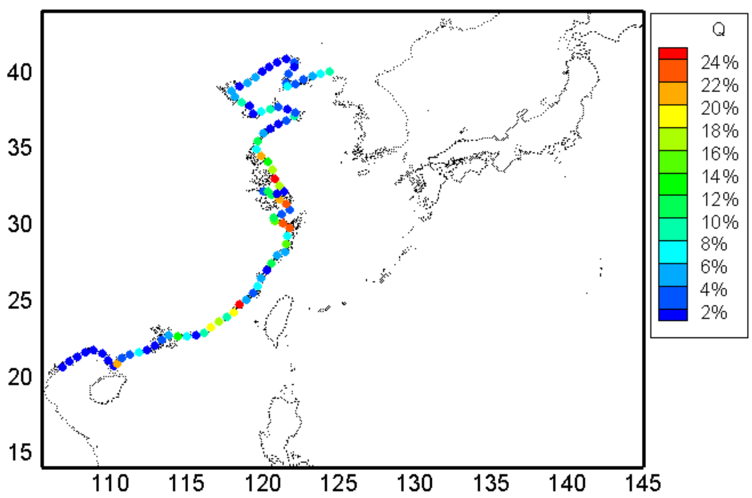

Fig. 17 Distribution of the improvement coefficient along the mainland China coast

6. Conclusions

This study uses long-term (35 years) model results to examine the suitability and performance of 4 copulas in the joint probability analysis of the extreme wave height (Hs) and surge (S) along the coasts of the mainland China. The extreme data is extracted with the annual N-largest sampling method and the dependencies between the Hs and S in the joint extreme samples at the nine selected nearshore stations are fully analysed. The performance of the four commonly used copulas, i.e. Gumbel-Hougaard, Clayton, Frank and Galambos copulas, in estimating the joint probability of extreme samples are assessed. The optimal copula identified is used for predicting combined water levels (CWLs, sum of Hs and S) in the study area with 50- and 100-year return periods and the accuracy is quantified.

Two theoretical univariate probabilistic distributions, i.e. GEV and P3, are used to fit the marginal of Hs and S samples. The results show that either GEV or P3 distributions could appropriately fit the extreme wave samples which depends on their location, while the GEV distribution provides the best fit to the extreme surge samples for all the selected locations along the mainland China coast. After assessing the performance of the copulas, the extreme value copula group is found to be the optimal copula group to describe the joint probability of extreme Hs and S. The Gumbel–Hougaard copula that belongs to the extreme value copula group is finally chosen to conduct the joint probability analysis of the Hs and S along the mainland China coast owing to its precision and conciseness.

By adopting the GEV/P3 distribution and applying the copula theory, the joint exceedance probabilities and joint probability densities at the nine representative nearshore stations are calculated. The results at these locations show that there are no uniform distribution patterns of joint distributions at different locations. The failure probability analysis is applied to calculate the most probable CWLs. The analysis is also extended to the entire coastline of the study site at 87 uniformly distributed locations, where the coastline is clearly identified with the predominance of the waves and surges. The empirical value of "100Hs+10S" and "50Hs+50S" combinations is recommended to estimate the 50-year return level situation when the joint probability data is unavailable and the "50Hs+50S" combination could be a meaningful indicator for events between 50-year and 100-year return levels. In comparison with the commonly used empirical design approaches, the improvement coefficient (Q) is introduced and calculated, which suggests that applying the joint probability approaches to the mid-east coast and southeast coast can improve the accuracy in predicting extreme combined water levels with the given return period. The results from this study provide reliable and realistic design guidelines for coastal engineering applications.

Acknowledgements

The work was partly supported by the National Key R&D Program of China (2017YFC0405401); the National Natural Science Foundation of China (51909013, 51979076) and the Fundamental Research Funds for the Central Universities, China (B200204017, 2019B04814)

References

Basco, D.R., Walker, R.A., 2010. Application of the Coastal Storm Impulse (COSI) parameter to predict coastal erosion. In: Proceedings of the Coastal Engineering Conference.

Bernardara, P., Mazas, F., Kergadallan, X., Hamm, L., 2014. A two-step framework for over-threshold modelling of environmental extremes. Nat. Hazard. Earth Syst. Sci. 14, 635-647. https://doi.org/10.5194/nhess-14-635-2014.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., Widmann, M., 2019. Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. Sci. Adv. 5. https://doi.org/10.1126/sciadv. aaw5531

Bruun, J.T., Tawn, J.A., 1998. Comparison of approaches for estimating the probability of coastal flooding. J. R. Stat. Soc. 47, 405-423. https://doi.org/10.1111/1467-9876.00118.

Chebana, F., Ouarda, T.B.M.J., 2011. Multivariate quantiles in hydrological frequency analysis. Environ. 22, 63-78. https://doi.org/ 10.1002/env.1027.

Chen, L., Singh, V.P., Guo, S., Hao, Z., Li, T., 2012. Flood coincidence risk analysis using multivariate copula functions. J. Hydrol. Eng. 17, 742–755. https://doi.org/10.1061/ (ASCE) HE.1943-5584.0000504.

Chen, Y., Li, J., Pan, S., Gan, M., Pan, Y., Xie, D., Clee, S., 2019. Joint probability analysis of extreme wave heights and surges along China's coasts. Ocean. Eng. 177, 97-107. https://doi.org/10.1016/j.oceaneng.2018.12.010.

Chini, N., Stansby, P., Leake, J., Wolf, J., Roberts-Jones, J., Lowe, J., 2010. The impact of sea level rise and climate change on inshore wave climate: A case study for East Anglia (UK). Coast.

Eng. 57, 973–984. https://doi.org/10.1016/j.coastaleng. 2010.05.009.

Code of Hydrology for Harhour and Waterway (JTS 145-2015). China. 2015.

Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extreme value analyses. Extrem. 2, 339-365. https://doi.org/10.1023/A:1009963131610

Corbella, S., Stretch, D.D., 2012. Predicting coastal erosion trends using non-stationary statistics and process-based models. Coast. Eng. 70, 40–49. https://doi.org/10.1016/j.coastaleng. 2012.06.004.

Corbella, S., Stretch, D.D., 2013. Simulating a multivariate sea storm using Archimedean copulas. Coast. Eng. 76, 68–78. https://doi.org/10.1016/j.coastaleng.2013.01.011.

De Michele, C., Salvadori, G., Vezzoli, R., Pecora, S., 2013. Multivariate assessment of droughts: frequency analysis and dynamic return period. Water Resour. Res. 49, 6985–6994. https://doi.org/10.1002/wrcr.20551.

De Waal, D.J., van Gelder, P.H.A.J.M., 2005. Modelling of extreme wave heights and periods through copulas. Extrem. 8, 345–356. https://doi.org/10.1007/s10687-006-0006-y.

Dong, S., Jiao, C.S., Tao, S.S., 2017. Joint return probability analysis of wind speed and rainfall intensity in typhoon-affected sea area. Nat. Hazard. 86, 1193-1205. https://doi.org/10.1007/ s11069-016-2736-8.

Dong, S., Wang, N., Lu, H., Tang, L., 2015. Bivariate distributions of group height and length for ocean waves using copula methods. Coast. Eng. 96, 49-61. https://doi.org/10.1016/ j.coastaleng.2014.11.005.

Ferreira, J.A., Guedes Soares, C., 2002. Modelling bivariate distributions of significant wave height and mean wave period. Appl. Ocean Res. 24, 31–45. https://doi.org/10.1016/S0141-1187(02)00006-8.

Fisher, N.I., Switzer, P., 2001. Graphical assessment of dependence: is a picture worth 100 tests?. Am. Stat. 55, 233-239. https://doi.org/10.1198/000313001317098248.

Fisher, N. I., and Switzer, P., 1985. Chi-plots for assessing dependence. Biom. 72, 253-265. https://doi.org/10.2307/2336078.

Galiatsatou, P., 2011. Bivariate analysis of extreme wave and storm surge events. determining the failure area of structures. Open Ocean Eng. J. 4, 3-14. https://doi.org/10.2174/ 1874835X 01104010003.

Galiatsatou, P., Prinos, P., 2007. Bivariate models for extremes of significant wave height and period. An application to the Dutch Coast. In: Proceedings of 2nd IMA International Conference on Flood Risk Assessment. University of Plymouth, UK.

Galiatsatou, P., Prinos, P., 2016. Joint probability analysis of extreme wave heights and storm surges in the aegean sea in a changing climate. In: FLOODrisk 2016 - 3rd European Conference on Flood Risk Management.

Genest, C., Favre, A.C., 2003. Everything You Always Wanted to Know about copula Modeling but Were Afraid to Ask. Symmetries In Nuclear Structure: An Occasion to Celebrate the 60th Birthday of Francesco Iachello.

Genest, C., Rivest, L.P., 1993. Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Assoc. 88, 1034–1043. https://doi.org/10.1080/01621459.1993.10476372.

Gouldby, B., Méndez, F.J., Guanche, Y., Rueda, A. and Mínguez, R. 2014. A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. Coast. Eng. 88, 15-26. https://doi.org/10.1016/j.coastaleng.2014.01.012

Gruhn, A., Salecker, D., Fröhle, P., Schüttrumpf, H., Thorenz, F., 2012. Flood protection dunes – An approach for reliability assessment by means of fragility curves as part of a risk and damage analysis. In: Proceedings of 33rd International Conference on Coastal Engineering. Santander, Spain.

Gudendorf, G., Segers, J., 2010. Extreme-value copulas, in: Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (Eds.), copula Theory and Its Applications, Lecture Notes in Statistics. Springer, Berlin Heidelberg.

Hawkes, P.J., Gouldby, B.P., Tawn, J.A., Owen, M.W., 2002. The joint probability of waves and water levels in coastal engineering design. J. Hydraul. Res. 40, 241-251. https://doi.org/10. 1080/00221680209499940.

Ji, X., Jing, D., Shen, H.W., Salas, J.D., 1984. Plotting positions for pearson type-III distribution. J. Hydrol. 74, 1-29. https://doi.org/10.1016/0022-1694(84)90137-9.

Li, C., Song, Y., 2006. Correlation of extreme waves and water levels using a third-generation wave model and a 3D flow model. Ocean Eng. 33, 635-653. https://doi.org/ 10.1016/ j.oceaneng. 2005.06.003.

Li, F., van Gelder, P.H.A.J.M, Ranasinghe, R., Callaghan, D.P., Jongejan, R.B., 2014. Probabilistic modelling of extreme storms along the dutch coast. Coast. Eng. 86, 1-13. https://doi. org/10.1016/j.coastaleng.2013.12.009.

Li, J., Pan, S., Chen, Y., Fan, Y., Pan, Y., 2018. Numerical estimation of extreme waves and surges over the northwest Pacific Ocean. Ocean. Eng. 153, 225-241. https://doi.org/10.1016/ j.oceaneng.2018.01.076.

Lin-Ye, J., Garcia-Leon, M., Gracia, V., Sanchez-Arcilla, A., 2016. A multivariate statistical model of extreme events: an application to the Catalan coast. Coast. Eng. 117, 138-156. https://doi. org/10.1016/j.coastaleng.2016.08.002.

Martzikos, N.T., Prinos, P.E., Memos, C.D., Tsoukala, V.K., 2021. Key research issues of coastal storm analysis. Ocean. Coast. Manag. 199, 105389. https://doi.org/10.1016/ j.ocecoaman.2020. 105389

Marcos, M., Rohmer, J., Vousdoukas, M. I., Mentaschi, L., Le Cozannet, G., Amores, A., 2019. Increased extreme coastal water levels due to the combined action of storm surges and wind waves. Geophys. Res. Lett. 46, 4356-4364. https://doi.org/10.1029/2019GL082599

Masina, M., Lamberti, A., Archetti, R., 2015. Coastal flooding: a copula based approach for estimating the joint probability of water levels and waves. Coast. Eng. 97, 37-52. https://doi.org/ 10.1016/j.coastaleng.2014.12.010.

Mazas, F., Hamm, L., 2017. An event-based approach for extreme joint probabilities of waves and sea levels. Coast. Eng. 122, 44-59. https://doi. org/10.1016/j.coastaleng.2017.02.003.

Mazas, F., Kergadallan, X., Garat, P., Hamm, L., 2014. Applying POT methods to the revised joint probability method for determining extreme sea levels. Coast. Eng. 91, 140-150. https://doi. org/10.1016/j.coastaleng.2014.05.006.

Mikosch, T., 2006. Copulas: Tales and facts. Extrem. 9, 3–20. https:// doi.org/10.1007/s10687-006 -0015-x.

Montes-Iturrizaga, R., Heredia-Zavoni, E., 2015. Environmental contours using copulas. Appl. Ocean. Res. 52, 125-139. https://doi.org/10.1016/j.apor.2015.05.007.

Montes-Iturrizaga, R., Heredia-Zavoni, E., 2016, Multivariate environmental contours using c-vine copulas. Ocean. Eng. 118, 68-82. https://doi.org/10.1016/j.oceaneng.2016.03.011.

Pearson, K., 1895. Mathematical contributions to the theory of evolution. iii. regression, heredity, and panmixia. Philosophical Transactions of the Royal Society of London A, 186(4), 343-414.

https://doi.org/10.1098/rsta.1895.0010

Qi, J., Chen, C., Beardsley, R.C., Perrie, W., Cowles, G.W., Lai, Z., 2009. An unstructured grid finite-volume surface wave model (FVCOM-SWAVE): implementation, validations and applications. Ocean Model. 28, 153–166. https://doi.org/10.1016/j.ocemod.2009.01.007

Rueda, A., Camus, P., Tomás, A., Vitousek, S., Méndez, F. J., 2016. A multivariate extreme wave and storm surge climate emulator based on weather patterns. Ocean. Model. 104, 242-251. https://doi.org/10.1016/j.ocemod.2016.06.008.

Serafin, K.A. and Ruggiero, P., 2014. Simulating extreme total water levels using a time-dependent extreme value approach. J. Geophy. Res. Ocean. 119, 6305-6329. https://doi.org/10.1002/2014jc010093

Salvadori, G., De Michele, C., 2004. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. Water Resour. Res. 40, W12511. https://doi.org/10.1029/2004WR003133.

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. Coast. Eng. 95, 77-83. https://doi.org/10.1016/j.coastaleng.2014.09.007.

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

Sraj, M., Bezak, N., Brilly, M., 2015. Bivariate flood frequency analysis using the copula function: a case study of the litija station on the sava river. Hydrol. Processes. 29, 225-238. https://doi.org/10.1002/hyp.10145.

Tao, S., Dong, S., Wang, N., Soares, C.G., 2013. Estimating storm surge intensity with Poisson bivariate maximum entropy distributions based on copulas. Nat. Hazard. 68, 791-807. https://doi.org/10.1007/s11069-013-0654-6.

Tawn, J.A., 1998. An extreme-value theory model for dependent observations. J. Hydrol. 10, 227-250. https://doi.org/10.1016/0022-1694(88)90037-6.

Vanem, Erik, 2016. Joint statistical models for significant wave height and wave period in a changing climate. Mar. Struct. 49, 180-205. https://doi.org/10.1016/j.marstruc.2016.06. 001.

Wahl, T., Plant, N. G., Long, J.W., 2016. Probabilistic assessment of erosion and flooding risk in the northern gulf of mexico. J. Geophys. Res. Ocean. 121, 3029-3043. https://doi.

org/10.1002/2015JC011482.

Wahl, T., Jensen, J., Mudersbach, C., 2010. A multivariate statistical model for advanced storm surge analyses in the North Sea. In: 32rd International Conference on Coastal Engineering. Shanghai, China.

Wahl, T., Mudersbach, C., Jensen, J., 2012. Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on copula functions. Nat. Hazard. Earth Syst. Sci. 12, 495–510. https://doi. org/10.5194/nhess-11-2925-2011.

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I.D., Hendry, A., Muis, S., Veldkamp, T.I.E., Winsemius, H.C., and Wahl, T., 2018. Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. Environ. Res. Lett.13, 084012 https://doi. org/10.1088/1748-9326/aad400

Wist, H.T., Myrhaug, D., 2004. Statistical properties of successive wave heights and successive wave periods. Appl. Ocean. Res. 26, 114–136. https://doi.org/10.1115/1.2829561.

Yang, X., Qian, J., 2019. Joint occurrence probability analysis of typhoon-induced storm surges and rainstorms using trivariate Archimedean copulas. Coast. Eng. 171, 533-539. https://doi. org/ 10.1016/j.oceaneng.2018.11.039.

Yang, X.C., Zhang, Q.H., 2013. Joint probability distribution of winds and waves from wave simulation of 20 years (1989-2008) in bohai bay. Water. Sci. Eng. 6, 296-307. https://doi.org/10.3882/j.issn.1674-2370.2013.03.006.

Yin, K., Xu, S., Huang, W., 2018. Estimating extreme sea levels in Yangtze Estuary by quadrature Joint Probability Optimal Sampling Method. Coast. Eng. 140, 331-341. https://doi.org/ 10. 1016/j.coastaleng.2018.08.007.

Zhang, L., Singh, V.P., 2007. Bivariate rainfall frequency distributions using Archimedean copulas. J. Hydrol. 332, 93–109. https://doi.org/10.1016/j.jhydrol.2006.06.033.

Zhang, Y., Kim, C.W., Beer, M., Dai, H., Soares, C.G., 2018. Modeling multivariate ocean data using asymmetric copulas. Coast. Eng. 135, 91-111. https://doi.org/10.1016/j.coastaleng. 2018.01.008.