Journal of Marine Science and Application (2023) 22: 128-136
https://doi.org/10.1007/s11804-023-00314-0

RESEARCH ARTICLE

Binomial-Bivariate Log-Normal Compound Model and its Application on Probability Estimation of Extreme Sea State

Jinghua Ding[1], Weichen Ding[2,3], Botao Xie[4] and Liang Pang[1]

Received: 01 June 2022 / Accepted: 30 December 2022

Abstract

Extreme value analysis is an indispensable method to predict the probability of marine disasters and calculate the design conditions of marine engineering. The rationality of extreme value analysis can be easily affected by the lack of sample data. The peaks over threshold (POT) method and compound extreme value distribution (CEVD) theory are effective methods to expand samples, but they still rely on long-term sea state data. To construct a probabilistic model using short-term sea state data instead of the traditional annual maximum series (AMS), the binomial-bivariate log-normal CEVD (BBLCED) model is established in this thesis. The model not only considers the frequency of the extreme sea state, but it also reflects the correlation between different sea state elements (wave height and wave period) and reduces the requirement for the length of the data series. The model is applied to the calculation of design wave elements in a certain area of the Yellow Sea. The results indicate that the BBLCED model has good stability and fitting effect, which is close to the probability prediction results obtained from the long-term data, and reasonably reflects the probability distribution characteristics of the extreme sea state. The model can provide a reliable basis for coastal engineering design under the condition of a lack of marine data. Hence, it is suitable for extreme value prediction and calculation in the field of disaster prevention and reduction.

Keywords Bivariate compound extreme value distribution; Double-threshold sampling; Extreme sea state; Short-term data; Probabilistic prediction

Article Highlights

- Considering short-term data and the correlation of different environmental elements, the binomial-bivariate log-normal compound extreme value distribution model is proposed to predict the extreme sea state.
- The reliability of binomial-bivariate log-normal compound extreme value distribution is verified by sample data of different years, which can provide the reference for engineering design.
- The combination of the peaks over threshold method and compound extreme value theory makes the model more suitable for short-term data.

✉ Liang Pang
pang@ouc.edu.cn

[1] College of Engineering, Ocean University of China, Qingdao 26610, China

[2] State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

[4] CNOOC Research Institute, Beijing 100025, China

1 Introduction

China has a total coastline of 32 000 km, which has good prospects for marine development. As the frontier and base of marine development, coastal areas are also vulnerable to marine disasters, such as surges and waves induced by storms. The above extreme sea states can usually be described by extreme value theory. Through this theory, using the marine data of the marine observation station or verified reanalysis data and filtering the extreme value sequence according to certain standards, a corresponding fitted model can be developed. Then, the design values of various marine elements in different return periods can be calculated.

In such studies, the common sampling method is to select annual maximum data to fit an extreme value model, such as Gumbel distribution, Weibull distribution, and log-normal distribution, to estimate the design value of marine elements in different return periods. However, the long time-series span and lack of some domestic observation station data result in fitted results that often cannot meet engineering

requirements. Past long-term data may not be applicable to current sea states. The current sea state is in a long-term change, and future ocean statistics will not be adequately represented by past sea states. Liu et al. (2010) demonstrated that long-term sea-level changes increase the amount of wave runup, and thus it is complex to estimate the design of the sea level in long-term sea conditions. In addition, in actual engineering conditions, more attention is paid to high quantiles. Therefore, for the accuracy of the extreme value calculation of a long return period, the sample size is significant. The generalized Pareto distribution model based on the peaks over threshold (POT) and compound extreme value distribution (CEVD) models based on process sampling can effectively expand the sample sequence, but it still relies on long-term (usually no less than 20 years) observation or reanalysis data. To solve this problem.

The ocean engineering structure is faced with a complex environment in which multiple elements work together, and there is a complex correlation between marine environmental elements, and the multivariate extreme value model has become a topic of extreme theory. Gumbel and Mustafi (1967) first proposed the multivariate extreme value theory symmetric logistic model. With the acceleration of research progress on the multivariate extreme value theory, Galambos (1977) and Leadbetter et al. (1983) summarized the theory into a volume from the aspects of probability statistics and random sequence processes. In the 1990s, Tawn et al. studied the multivariate extreme value theory in detail and proposed asymmetric logistic model (Tawn 1990), negative asymmetric logistic model (Joe 1989) and Dirichlet model (Coles and Tawn 1991), which provided a theoretical basis for the engineering community to solve the above problems. With these foundations, many scholars began to explore the application of the multivariate extreme value theory in complex sea states. Gupta and Manohar (2005) develop the extreme values associated with a vector of mutually correlated, stationary, and Gaussian random processes. Li and Song (2006) proposed that a joint event with 100-year return period could be approximated by either including a 100-year return period wave height and a 10-year return period surge, or a 10-year return period wave height and a 100-year return period surge, or the consisting of 50-year return period wave height and a 50-year return period surge. Pei et al. (2012) utilized a stochastic hurricane simulation program along with the ADCIRC model to simulate 5000 years of hurricanes and the corresponding storm surge heights for the City of Charleston, SC. After that, Park et al. (2013) coupled the Gaussian discriminative analysisand Gaussian mixture models and investigated variations in wind field characteristics by comparing the joint probability distribution functions of several wind field features. Jia and Sasani (2021) presented a methodology to estimate the joint exceedance probability for wind and flood hazards using a copula-based joint probability mode, which could evaluate the compounding threats of coastal storms, designcoastal structures, and estimate building performance under coastal storms. At the same time, the research on the correlation between different marine elements has made progress. Chen et al. (2019) found that the structural response (sampling method shows the best performance in describing correlations between extreme wave heights and surges, particularly in the typhoon-affected areas, in comparison with wave dominated and surge- dominated sampling methods. Afterward, Yang et al. (2020) analyzed the joint distributions of the destructive factors using copulas, and discussed the combination design method of the destructive factors. Through this method, they optimized the combination of rains and tides for different situation. Xi et al. (2021) applied the JPM method to tropical cyclone rainfall hazard estimation, they found it is important to include all three important variables (maximum intensity when the storm is near the point of interest (POI), duration of the storm, and the minimal distance) into the probability assignment process. Simão et al. (2022) presented an approach for obtaining an analytical probabilistic model of environmental parameters, including linear and directional variables. The model can well represent wind, sea, and swell waves and wind and current parameters at the studied location. The related structural changes between different combinations of engineering environmental loads are complex, and the expressions of multivariate extreme value models are mostly implicit. Only through complex iterative solutions can they be applied in engineering applications. Shi and Sun (2001) established the ternary nested logistic model and its explicit expression and calculated its parameters through the moment method and maximum likelihood method, which provided a solution for different combinations of engineering environmental loads with complex changes. At the same time, the CEVD theory has rapidly developed. Ma and Liu (1979) proposed CEVD, composed of a discrete distribution and continuous distribution, which has been widely recognized in the engineering field (Liu and Li 2001; Liu et al. 2002; Liu et al. 2007; Pang et al. 2015). Then, Liu et al. extended the theory from one dimension to multiple dimensions, such as the Poisson–Gumbel mixed compound distribution and Poisson-nested logistic compound extreme value model (Liu and Dong 2004; Liu et al. 2006; Pang et al. 2013).

As explained above, clearly, in the field of probability prediction of extreme sea states, the key problems are the reasonable sampling and combination of different environmental factors (Cheng et al. 2018; Yan et al. 2020). This paper proposes the binomial-bivariate log-normal CEVD (BBLCED) based on short-term observation samples of the POT. It considers not only the frequency of the extreme sea state process but also the correlation between two environmental elements. The model can be used to solve the design values of the wave height and period in different return periods with short-term (5 years) data.

2 Model construction methodology

2.1 Binomial log-normal CEVD

When the number of annual measured extreme sea state data (k) conforms to binomial distribution, we have:

$$p_k = \begin{bmatrix} m \\ k \end{bmatrix} \dot{p}^k (1 - \dot{p})^{m-k} = \begin{bmatrix} 365 \\ k \end{bmatrix} \dot{p}^k (1 - \dot{p})^{365-k} \quad (1)$$

where the observation data of m days are statistically independent and identically, $\bar{p}$ is the mean value of p_k, The expression of binomial log-normal compound extreme value Distribution is:

$$F_0(x) = \sum_{k=0}^{\infty} \begin{bmatrix} m \\ k \end{bmatrix} \dot{p}^k (1 - \dot{p})^{m-k} \left[G(x) \right]^k = \left[\dot{p} \cdot G(x) + 1 - \dot{p} \right]^m \quad (2)$$

$$G(x) = 1 - \frac{1}{\dot{p}} \left(1 - R^{1/365} \right) \quad (3)$$

If $G(x)$ conforms normal distribution, then

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X_R} e^{-t^2/2} \, \mathrm{d}t = 1 - \frac{365}{\dot{n}} \left(1 - R^{1/365} \right) \quad (4)$$

where $\bar{n}$ is the average value of daily maximum wave height taken every year, and R is the cumulative probability value.

When the sequence conforms to the log-normal distribution, the following conversion can be performed:

$$X_R = \frac{\ln \left(H_\tau - H_0 \right) - a}{\sigma} \quad (5)$$

where $a = \frac{1}{N} \sum_{i=1}^{N} \ln (H_i - H_0)$, $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[\ln (H_i - H_0) \right]^2 - a^2}$. n is the total number of measured daily maximum values; H_i is the measured daily maximum ($i = 1, 2, ..., n$); a is the mean value of $\ln H_i$; σ is the variance of $\ln H_i$; H_T is the design value of T-Year return period; H_0 is the threshold.

2.2 BBLCED

According to the multivariate compound extreme value distribution theory proposed by Wang (2005), a binomial bivariate Log-normal model is established.

Definition with a univariate discrete probability distribution of

$$\begin{pmatrix} 0 & 1 & 2 & \cdots & k & \cdots \\ p_0 & p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

And a bivariate continuous probability distribution of $G(x, y)$, let

$$F_0(x, y) = \sum_{k=1}^{\infty} P_k \cdot k \cdot \int_{-\infty}^{y} \int_{-\infty}^{x} G_x^{k-1}(u) \, \mathrm{d}u \mathrm{d}v \quad (6)$$

Let n represent the annual frequency of extreme sea states, and its distribution is p_k; The maximum of a marine element (wave height) in each extreme sea states and its "concomitant" another marine element (wave period) are noted as (ξ, η), its probability density function is $g(x, y)$, and corresponding joint cumulative distribution function is $G(x, y)$, $G_x(x)$ is Marginal distribution of $G(x, y)$, (X, Y) is the annual maximum of (ξ, η).

In practical application, the main problem is to give an $R(0 < R < 1, R = 1 - P)$, and solve the equation:

$$F(x, y) = R \quad (7)$$

Let

$$T = \frac{1}{P} = \frac{1}{1 - R} \quad (8)$$

If (x_R, y_R) satisfies the Eq. (8), then we called (x_R, y_R) return value of T years.

Usually, we calculate the design value with a return period of more than ten years, i.e., $0.9 < R < 1$, so there is a low limit of R. Therefore, in solving Eq. (8) we usually define

$$R_0 < R < 1$$

According to inference of Wang (2005), when solving Eq. (7), It can be changed to solve $F_0(x, y) = R$ instead. If there is no extreme sea states in this year, there is no need to calculate the distribution function, so as to simplify the problem.

Ignoring the situation of no extreme sea states, the bivariate compound extreme value (BCEV) distribution is

$$F_0(x, y) = P_0 + F(x, y) \quad (9)$$

i.e.,

$$F_0(x, y) = P_0 + \sum_{k=1}^{\infty} P_k \cdot k \cdot \int_{-\infty}^{y} \int_{-\infty}^{x} G_x^{k-1}(u) \, g(u, v) \, \mathrm{d}u \mathrm{d}v \quad (10)$$

The discrete distribution adopts binomial distribution. Let the observation data of m days are statistically independent and identically distributed random variables (X_i, Y_i), $i = 1, 2, \cdots, n$, and its distribution function is $F(x, y)$. For a sufficiently large threshold H_0, if $X_j > H_0$, $j = 1, 2, \cdots, n$, then we call X_j is the data over threshold, X_j and corresponding Y_j obeys bivariate Log-normal distribution. The length of the sample sequence is k. N obeys the binomial

distribution of parameter (m, p), i. e., $\Pr(N = k) = C_m^k \bar{p}^k (1 - \bar{p})^{m-k}$, and the expression of its distribution function is:

$$Y(x) = \sum_{k=0}^{m} \binom{m}{k} \bar{p}^k (1 - \bar{p})^{m-k} \tag{11}$$

where $\bar{p}$ is the mean value of p_k, N is the number of samples exceeding the threshold.

When n conforms to binomial distribution, the formula is converted to the following form:

$$F_0(x, y) = (1 - p)^n + \sum_{k=1}^{\infty} \binom{n}{k} \bar{p}^k (1 - p)^{n-k} \cdot k \cdot \int_{-\infty}^{y} \int_{-\infty}^{x} G_x^{k-1}(u)\, g(u, v)\, du dv \tag{12}$$

After derivation, we obtain the probability density function

$$f(x, y) = \sum_{k=1}^{\infty} \binom{n}{k} \bar{p}^k (-p)^{n-k} \cdot k \cdot G_x^{k-1}(u) \cdot g(x, y) \tag{13}$$

Bivariate Normal distribution is as follows:

$$U, V \sim \mathrm{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}) \tag{14}$$

where μ_x, μ_y are respectively the mean value of variable X and Y; σ_x^2, σ_y^2 are respectively the variance of variable X and Y; ρ_{xy} is correlation coefficient of variables X and Y.

Let $X = \exp(U)$, $Y = \exp(V)$ and take the logarithm of them, the bivariate Log-normal distribution is obtained:

$$f(x, y) = \frac{1}{2\pi \sqrt{1 - \rho_{xy}^2}\, \sigma_x \sigma_y xy} \cdot \exp\left[\frac{-1}{2(1 - \rho_{xy}^2)} \cdot \left(\frac{(\ln x - \mu_x)}{\sigma_x^2} - \frac{2\rho_{xy}(\ln x - \mu_x)(\ln y - \mu_y)}{\sigma_x \sigma_y} + \frac{(\ln y - \mu_y)}{\sigma_y^2} \right) \right] x > 0,\ y > 0 \tag{15}$$

where ρ_{xy} is the correlation coefficient of U, V:

$$\rho_{xy} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{E(XY) - E(X)E(Y)}{\sigma_x \sigma_y} \tag{16}$$

where ρ is the correlation coefficient of X, Y:

$$\rho = \frac{\exp(\rho_{xy} \sigma_x \sigma_y) - 1}{\sqrt{(\exp(\sigma_x^2) - 1)(\exp(\sigma_y^2) - 1)}} \tag{17}$$

And the marginal distribution:

$$G_x(u) = \int_0^x \frac{1}{\sqrt{2\pi}\, \sigma_x x} \cdot e^{-\frac{(\ln x - \mu_x)^2}{2\sigma_x^2}} du \tag{18}$$

When combining Eqs. (12), (15) with Eq. (18), we obtained BBLCED model.

3 Application of BBLCED model to extreme sea states prediction

3.1 Sampling method

3.1.1 Threshold selection

The reasonable threshold selection is the key to the successful fitting of POT model. If the threshold is too small, the difference between the sample sequence and the extreme value model may be distinct, and the estimated value will produce biased estimation (Hua and Zhang 2009; Liu 2014); If the threshold is too large, and the number of samples exceeding the threshold decreases, the fitting effect of the model will be affected,which can cause the variance of the parameter estimation to be too large (Sun 2014; Cheng et al. 2019). Therefore, we should get the reasonable threshold of wave height data and filter the wave height data over threshold, and find out the corresponding period data, i. e., the sample sequences of BBLCED model (X, Y).

The wave height and period data (every six hours) of a marine observation station in Yellow Sea, China, from 1992 to 1996 were used as the original sample. this thesis uses the hill diagram method to select the threshold. The selection principle is to find the relatively stable line segment at the tail index (α) of the figure as the starting point, and the data corresponding to the abscissa of the point is the threshold H_0.

As shown in Figure 1, the trend of the lines become stable after the threshold is less than 3.4. In order to test whether the threshold is reasonable, this paper further judges it in

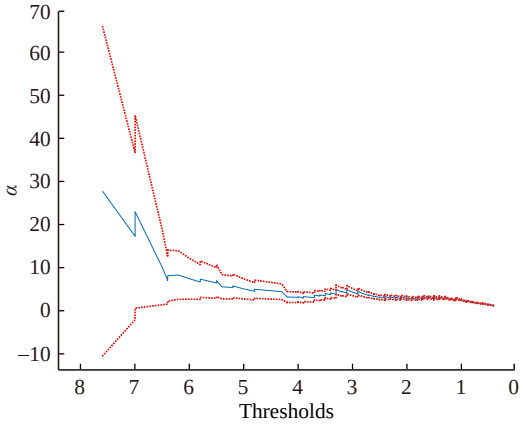


Figure 1 Hillplot

combination with the parameter stability estimation diagram. The judgment method is to take H_0=3.4 as the initial threshold, and evenly select 81 values between H_0= 3.2 and u = 3.6, then observe whether the calculated maximum likelihood estimation value remains relatively stable. The variation range of parameters is shown in Figure 2.

Figure 2 shows that the parameters remain relatively stable within (3.2, 3.395). To ensure the accuracy of the POT model fitting, this study selected the larger value in the relatively stable interval as the final threshold, i.e., H_0 = 3.395.

Figure 3 shows that when the wave height threshold is 3.395, the maximum likelihood estimation value of each parameter of the corresponding period is stable within (4.7, 5.1), so the threshold is relatively accurate.

3.1.2 Double threshold sampling

The double threshold method is a widely applied method. In addition to the data exceeding threshold, the interval between these data should also exceed a certain time to eliminate the impact of the same extreme sea states. Li et al. (2012) proposed a method to filter the wave data of Rottnest:

1) At least one recording must exceed the storm peak threshold, H_0; storm duration is measured as the time recordings exceed the duration threshold, Hs_{dur};

2) The interval between two consecutive storms (storm peak to storm peak) is not less than 30 hours. Otherwise, they are regarded as the continuation of a single storm;

3) The storm break is not shorter than three hours; otherwise, they are regarded as the continuation of one storm.

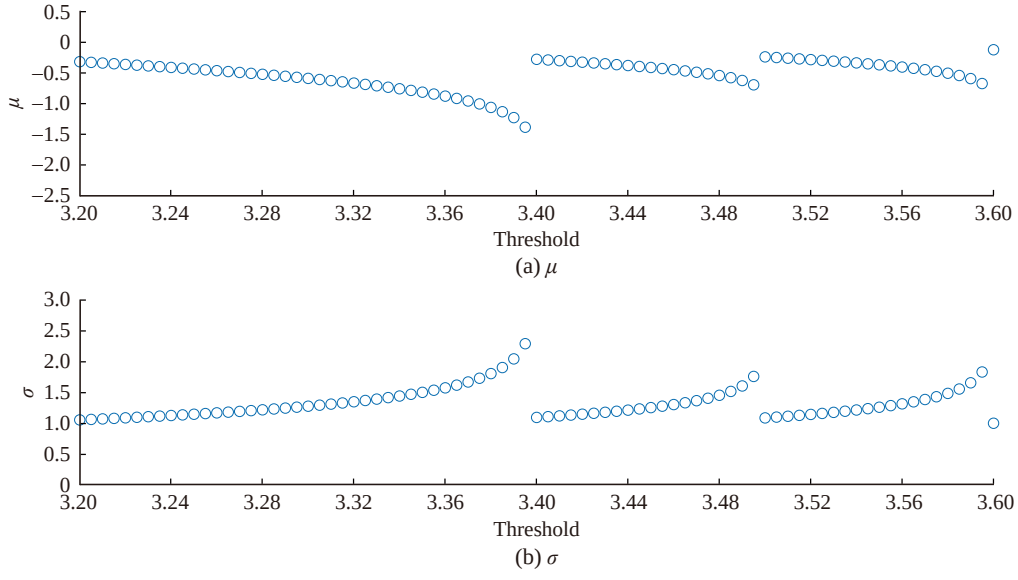


Figure 2 Maximum likelihood estimation for each parameter at different thresholds of wave height

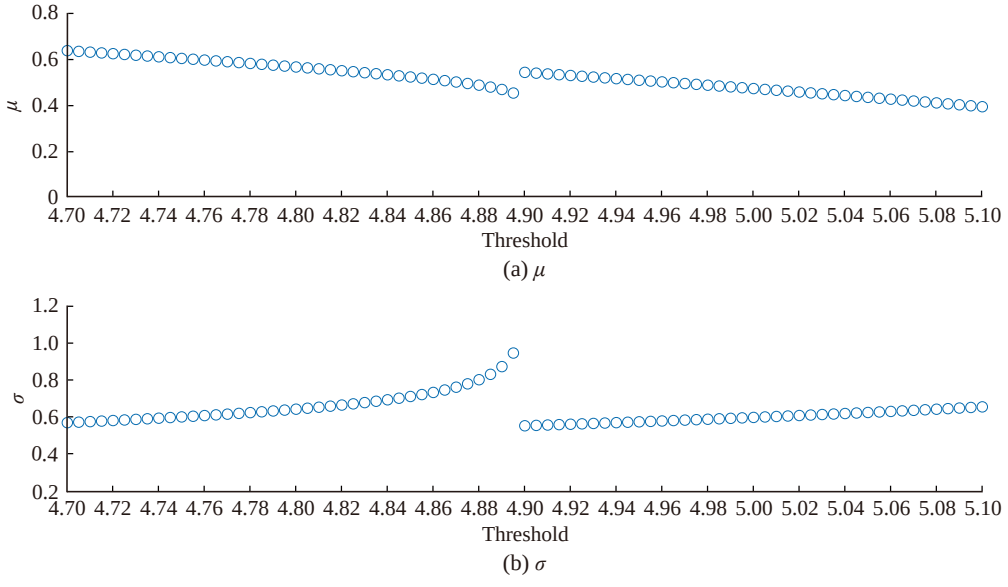


Figure 3 Maximum likelihood estimation for each parameter at different wave period thresholds

H_0=3.395, Hs_{dur} needs artificial judgment, let Hs_{dur}= 24 hours. Finally, 57 groups of data exceeding the threshold were extracted from 7304 groups of data.

3.2 Parameter estimation

After obtained the optimal threshold, the maximum likelihood estimation method is used to estimate the statistical parameters in the BBLCED model function. The value of parameters are shown in Table 1.

Table 1 Statistical parameters of BBLCED model

μ_x	μ_y	σ_x	σ_y	ρ_{xy}	p
1.46	1.92	0.27	0.15	0.78	0.004

It can be seen from Table 1 the value of parameters. By substituting the above parameters into Eq. (14), the joint distribution function can be obtained.

3.3 Parameter test

Based on the above parameter estimation results and distribution function, it is necessary to test the fitting effect of BBLCED model .

Figures 4 and 5 are the diagnostic check of Log-normal distribution model fitting the wave height data and period data (every six hours) from 1992 to 1996 respectively. It can be seen from the Figures P–P (Figures 4(a) and 5(a)) and Q–Q (Figures 4(b) and 5(b)) show that all points are distributed on or near the two sides of the line, which means that the model has a good fit effect. The cumulative distribution function (CDF) and empirical accumulation function (Figures 4(c) and 5(c)) basically coincide, and the probability density function (PDF) and frequency histogram (Figures 4(d) and 5(d)) also coincides. Therefore, the four diagnostic checks support the fitted Log-normal distribution model, i.e., the model has a great goodness of fit.

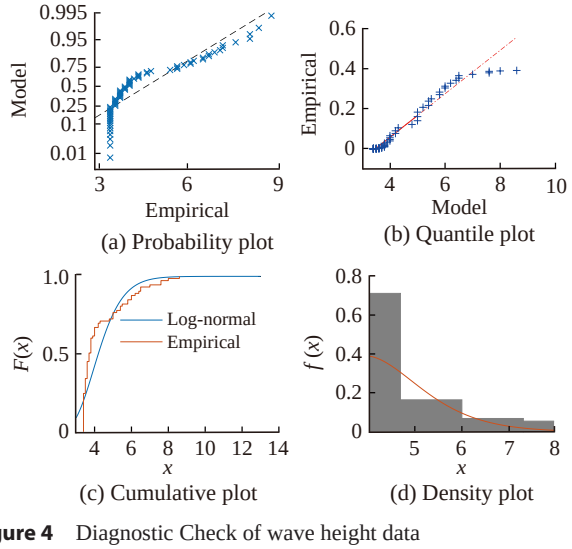


Figure 4 Diagnostic Check of wave height data

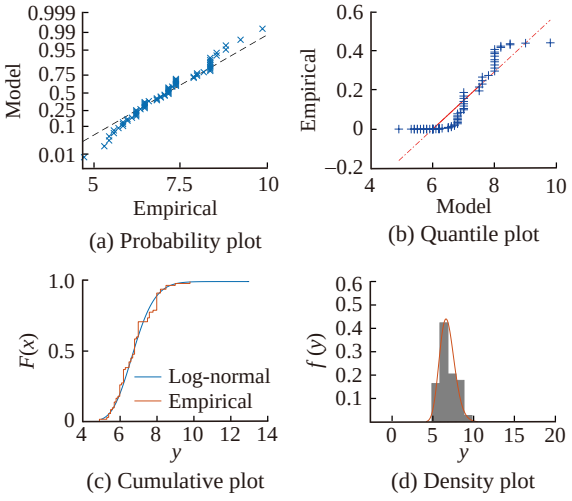


Figure 5 Diagnostic Check of wave period data

3.4 Comparison of multi-year return wave height and period with fourPrediction methods

Based on the short-term data of Yellow Sea for five years (1992–1996), the BBLCED model was selected to calculate the design values of multi-year return wave height and period. In addition, the 20-year annual maximum of the wave height and period data of the Yellow Sea was used in the other 3 models. the results calculated by Poisson-Gumbel mixed model are used as the comparison of CEVD, and the multi-year return wave height and period design values calculated by Gumbel mixed model and log-normal model are used as the comparison of traditional methods. The results obtained by these four methods are shown in Fig. 6–Fig. 10. Fig. 6 shows the cumulative distribution function (CDF) diagram and probability density function (PDF) diagram of BBLCED model, and Fig. 7–Fig. 10 shows the wave height and period design values in different return periods found of the four models. The calculation results are summarized into Table 2 and Table 3.

As shown in Tables 2 and 3, there is little difference between the wave height and wave period predictions of different return periods with the BBLCED model and the other three models. In particular, the results of the BBLCED model are closer to those obtained from the bivariate log-normal model sampled by the annual maximum. Compared with the other three models, the wave period predictions of the 10-year and 20-year return periods under the BBLCED model are smaller, but the prediction is close in the other three situations, i.e., 50-year, 100-year, and 200-year periods. The most striking result to emerge from the data is that the BBLCED model established in this study selects the data of the 5-year period, which can achieve similar results with the other three models with 20-year data, which is the biggest advantage of this model.

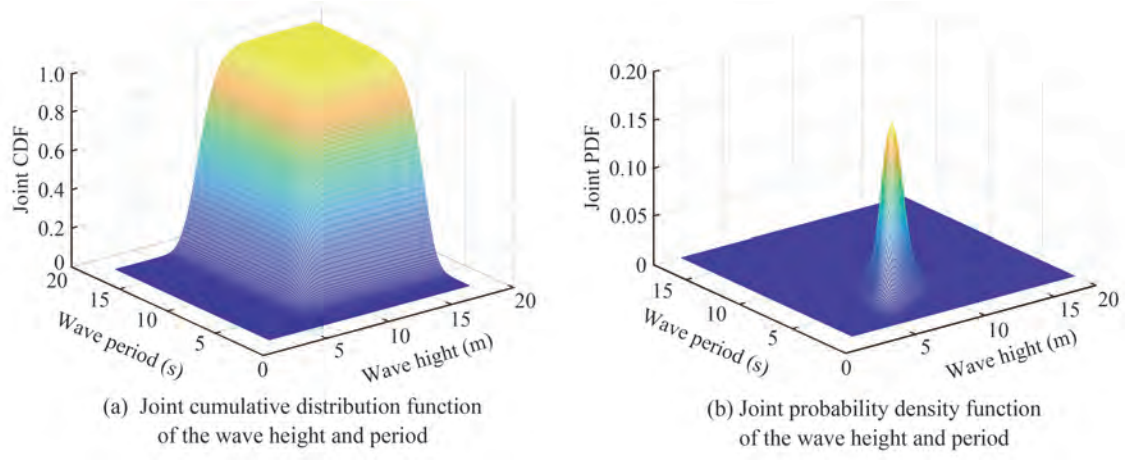


Figure 6 Joint CDF and PDF of the wave height and period

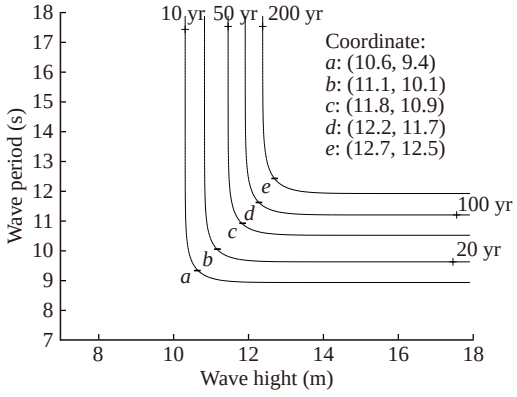


Figure 7 Design values of wave height and period of BBLCED model in different return periods

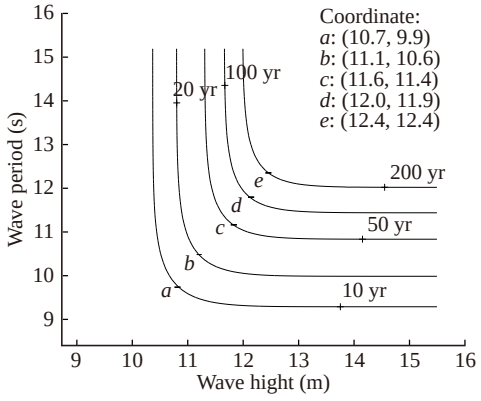


Figure 9 Design values of wave height and period of bivariate log-normal model in different return periods

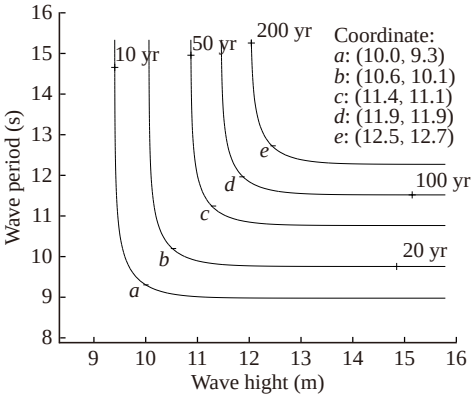


Figure 8 Design values of wave height and period of Poisson-Gumbel mixed compound model in different return periods

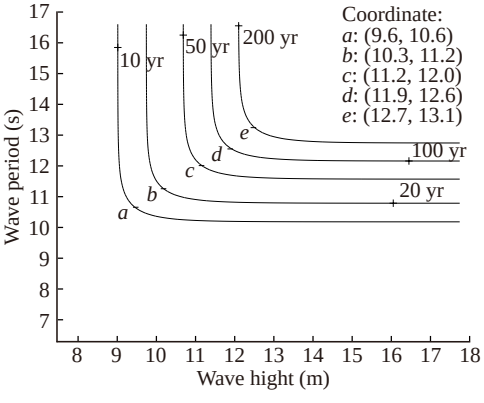


Figure 10 Design values of wave height and period of Gumbel mixed model in different return periods

Table 2 Wave Height in different return periods of the four models (m)

Computational model	10-year	20-year	50-year	100-year	200-year
BBLCED	10.6	11.1	11.8	12.2	12.7
PGMCED	10.0	10.6	11.4	11.9	12.5
Bivariate log-normal	10.7	11.1	11.6	12.0	12.4
Gumbel mixed	9.6	10.3	11.2	11.9	12.7

Table 3 Wave Period in different return periods of the four models (s)

Computational model	10-year	20-year	50-year	100-year	200-year
BBLCED	9.4	10.1	10.9	11.7	12.5
PGMCED	9.3	10.1	11.1	11.9	12.7
Bivariate log-normal	9.9	10.6	11.4	11.9	12.4
Gumbel mixed	10.6	11.2	12.0	12.6	13.1

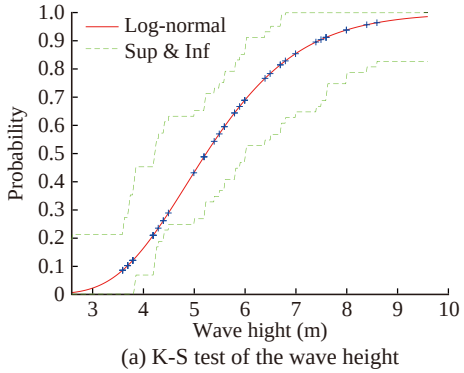
3.5 Reliability of BBLCED model

The return period of wave height and wave period has been calculated with the sea states data of 1992–1996. The time range of the data is extended to every five consecutive years of 1990–1997, i.e., 1990–1994, 1991–1995, 1992–1996, 1993–1997. Kolmogorov-Smirnov (K-S) test is used to check the stability of the results of BBLCED model. As a contrast, annual maxima sequence is calculated as well.

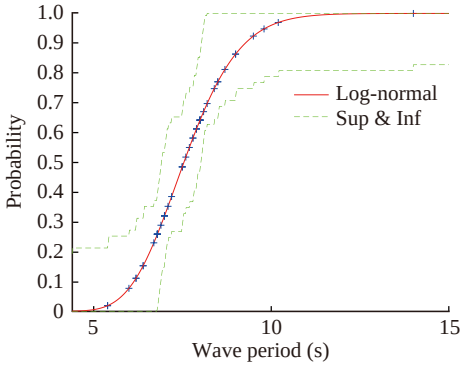
As shown in Table 4, $D_N \leqslant D_{N,1-\alpha}$. In Figure11, the CDF of the wave height and period sequences is between the supremum and infimum, which means it fits well. Due to the length of sequence in different years is not the same, the statistics, i.e., $D_N, D_{N,1-\alpha}$ changed accordingly. On the whole, the results of BBLCED model are relatively stable. As for the annual maximum series, it is clear that the wave height sequence fit well too.

Table 4 K-S test statistic of the BBLCED and AMS model

Model		D_N of wave height sequence	D_N of wave period sequence	$D_{N,1-\alpha}$
BBLCED	1990-1994	0.139	0.118	0.192
	1991-1995	0.160	0.145	0.212
	1992-1996	0.173	0.143	0.248
	1993-1997	0.196	0.127	0.294
AMS		0.188	0.224	0.301



(a) K-S test of the wave height



(b) K-S test of the wave period

Figure 11 K-S test of wave height and wave period

4 Conclusions

Based on the POT model, the sample sequences over the threshold were filtered, and the frequency of the sample sequences conformed to the binomial distribution. On the premise of occurring extreme sea conditions, the wave height data and their "accompanying" period data conform to the log-normal distribution. Therefore, combining the binomial distribution with the bivariate log-normal model, we obtained a new CEVD, i.e., the BBLCED model. The following main conclusions are drawn:

1) Through the POT method, wave data were fully utilized, which makes up for the shortage of short-term data. The model based on five years (1992–1996) of wave height and period data of the Yellow Sea area has good fitness. The extreme wave height and period in the 10-year, 20-year, 50-year, 100-year, and 200-year return periods were predicted, and the stability of the calculation results was tested using the K-S test, as shown in Section 3.4. The results indicate that the calculation results of the BBLCED model fit well, and they are close to the results of the traditional model. The method has good stability, and the probability distribution characteristics of extreme sea states can be reasonably reflected. Hence, the BBLCED model using 5-year data can replace the traditional extreme value method using 20-year data.

2) At the same time, the existing data are difficult to achieve complete accuracy, and various complex statistical models will also produce errors, so no method can be applied in all cases.

3) The BBLCED model based on the POT method makes the best use of the effective data, and the requirement for data sequence length is reduced. On this basis, the model not only considers the occurrence frequency of extreme sea states but also integrates the correlation of the wave height and period, which undoubtedly proposes a more reasonable and reliable method for the design standard of practical engineering. In addition, it has a wide application prospect

in the prediction of various engineering design and disaster prevention fields.

References

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

Cheng YJ, Pang L, Dong S (2019) Study on the estimation of very long return-period significant wave height during hurricane in the region of South China Sea. Journal of Ocean University of China, 49(S2): 125-132. DOI: 10.16441/j.cnki.hdxb.20180015

Cheng YJ, Yan ZD, Pang L, Liu WW (2018) Probability analysis on the typhoon induced sea states of the South China Sea. 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, Japan, 1-10. DOI: 10.1109/OCEANSKOBE.2018.8559105

Coles SG, Tawn JA (1991) Modelling extreme multivariate events. Journal of the Royal Statistical Society, Series B, Methodological 53(2): 377-392. DOI: 10.1111/j.2517-6161.1991.tb01830.x

Galambos J (1977) The theory and applications of reliability with Bayesian and nonparametric methods. Academic Press, New York, 151-164. https://doi.org/10.1016/B978-0-127-02101-0.X5001-X

Gumbel EJ, Mustafi CK (1967) Some analytical properties of bivariate extremal distributions. Publications of the American Statistical Association 62(318): 569-588. DOI: 10.2307/2283984

Gupta S, Manohar CS (2005) Multivariate extreme value distributions for random vibration applications. Journal of Engineering Mechanics 131(7): 712-720. DOI: 10. 1061/(ASCE)0733-9399(2005)131:7(712)

Hua YJ, Zhang ZY (2009) Comparative research on extreme risk of stock market based on BMM and POT model. Journal of Industrial Engineering/Engineering Management 23(4): 104-115. (in Chinese)

Jia Y, Sasani M (2021) Modeling joint probability of wind and flood hazards in Boston. Natural Hazards Review 22(4): 04021047. DOI: 10.1061/(ASCE)NH.1527-6996.0000508

Joe H (1989) Families of min-stable multivariate exponential and multivariate extreme value distributions. Statistics Probability Letters 9(1): 75-82. DOI: 10.1016/0167-7152(90)90098-R

Leadbetter MR, Lindgren G, Rootzen H (1983) Extreme and related properties of random sequences and series. Springer-Verlag, New York, 1-141. DOI: 10.2307/2283984

Li CW, Song Y (2006) Correlation of extreme waves and water levels using a third-generation wave model and a 3D flow model. Ocean Engineering 33(5-6): 635-653. DOI: 10. 1016/j.oceaneng.2005.06.003

Li FJ, Bicknell C, Lowry R, Li Y (2012) A comparison of extreme wave analysis methods with 1994–2010 offshore Perth dataset. Coastal Engineering 69: 1-11. DOI: 10.1016/j.coastaleng.2012.05.006

Liu DF, Li HJ (2001) Prediction of extreme significant wave height from daily maxima. China Ocean Engineering 15(1): 97-106. (in Chinese)

Liu DF, Wen SQ, Wang LP (2002) Poisson-Gumbel mixed compound extreme value distribution and its application. Chinese Science Bulletin 47(17): 1356-1360. (in Chinese)

Liu DF, Dong S (2004) Stochastic engineering oceanography. China Ocean University Press, Qingdao, 107-117. (in Chinese)

Liu DF, Wang LP, Pang L (2006) Theory of multivariate compound extreme value distribution and its application to extreme sea state prediction. Chinese Science Bulletin 23(51): 2926-2930. DOI: 10.1007/s11434-006-2186-x

Liu DF, Pang L, Xie BT, Wu YK (2007) Study on typhoon disaster zoning and Fortification Criteria in China–double nested multi-objective joint probability model and its application. Science in China Series E Technological Sciences 51(7): 1038-1048. DOI: 10.1007/s11431-008-0053-5

Liu SS (2014) The selection and application of the threshold of POT model. Master thesis, Jilin University, Changchun, 12, 28. (in Chinese)

Liu JC, Lence BJ, Isaacson M (2010) Direct joint probability method for estimating extreme sea levels. Journal of Waterway Port Coastal and Ocean Engineering 136(1): 66-76

Ma FS, Liu DF (1979) Compound extreme value distribution theory and its applications. Acta Mathematicae Applacatae Sinica 2(4): 366-375. (in Chinese)

Pang L, Chen X, Li YL (2013) Long-term probability prediction on the extreme sea states induced by typhoon of the South China sea. 2013 Advanced Materials Research, Guilin, China, 726-731, 833-841. DOI: 10.4028/www.scientific.net/AMR.726-731.833

Pang L, Xu F, Gong X, Zhan YF (2015) Study of the influence of tropical cyclone on offshore wind turbine egenerator system. Periodical of Ocean University of China 45(10): 109-113. DOI: 10.16441/j.cnki.hdxb.20130319

Park J, Smarsly K, Law KH, Hartmann D (2013) Multivariate analysis and prediction of wind turbine response to varying wind field characteristics based on machine learning. ASCE International Workshop on Computing in Civil Engineering, Los Angeles, 113-120. DOI: 10.1061/9780784413029.015

Pei B, Pang WC, Testik F, Ravichandran N (2012) Joint distributions of hurricane wind and storm surge for the city of charleston in South Carolina. ATC & SEI Conference on Advances in Hurricane Engineering 2012, Miami, 703-714. DOI:10.1061/9780784412626.062

Shi DJ, Sun BK (2001) Moment estimation in a nested logistic model. Systems Engineering-Theory & Practice 21(1): 53-60. (in Chinese)

Simão ML, Sagrilo LVS, Videiro PM (2022) A multi-dimensional long-term joint probability model for environmental parameters. Ocean Engineering 255: 111470. https://doi.org/10.1016/j.oceaneng.2022.111470

Sun LL (2014) Measurements and application of the extreme risk of financial data. Master thesis, Chongqing University, Chongqing, 19-20. (in Chinese)

Tawn JA (1990) Modelling multivariate extreme value distributions. Biometrika 77(2): 245-253. DOI: 10.2307/2336802

Wang LP (2005) Multivariate compound extreme value distribution theory and its engineering applications. PhD thesis, Ocean University of China, Qingdao, 68-72. (in Chinese)

Xi D Z, Lin N, Nadal-Caraballo NC (2021) A joint-probability model for tropical cyclone rainfall hazard assessment. GEO-EXTREME 2021: Climatic Extremes and Earthquake Modeling, 329: 1-10. DOI: 10.1061/9780784483695.001

Yan ZD, Pang L, Dong S (2020) Analysis of extreme wind speed estimates in the Northern South China Sea. Journal of Applied Meteorology and Climatology 59(10): 1625-1635. DOI: 10.1175/JAMC-D-20-0046.1

Yang X, Wang J, Weng SG (2020) Joint probability study of destructive factors related to the "Triad" phenomenon during typhoon events in the coastal regions: Taking Jiangsu Province as an example. Journal of Hydrologic Engineering 25(11): 05020038. DOI: 10.1061/(ASCE)HE.1943-5584.0002007