A Monte Carlo Simulation of Multivariate General Pareto Distribution and its Application

Luo Yao[1], Sui Dandan[2], Wang Dongxiao[2,*], Zhou Zhenwei[3], He Weihong[1], Shi Hui[4]

1 South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

2 State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

3 School of Civil Engineering and Transportation, South China University of Technology, China

4 China Water Resources Pearl River Planning Surveying & Designing CO., LTD, China

Correspondence to: yaoluo@scsio.ac.cn (Luo Yao),

*Corresponding author: dxwang@scsio.ac.cn (Wang Dongxiao)

Abstract：

The paper presents a MGPD (Multivariate General Pareto Distribution) method and builds the solving method of MGPD by a Monte Carlo simulation for the marine environmental extreme value parameters. The simulation method is proved to be feasible by analyzing the joint probability of wave weight and its concomitant wind from a hydrological station in SCS (South China Sea). The MGPD is the natural distribution of the MPOT (Multivariate Peaks Over Threshold) sampling method, and has the extreme value theory background. The existing dependence function can be used in the MGPD, so it may describe more variables which have different dependence relationships. The MGPD method improves the efficiency of the extremes in raw data. For the wind and the concomitant wave in 23 years (1960-1982), the number of the wind and wave selected is averaged 19 each year. Finally, by using the CP (Conditional Probability), a Monte Carlo Simulation method based on the MGPD is adopted in case of the return period of base shear.

Key Words： MPOT, extreme wave, extreme value theory, Monte Carlo

1. Introduction

Statistical modeling of extreme values (EV) plays a crucial role in the design and risk evaluation of ocean engineering. Problems concerning ocean environmental extremes are often multivariate in character. An example is the ocean environments (including waves, wind and currents) are all contributing to the forces experienced by offshore systems during the typhoon. So the severity of such a typhoon event may be described by a function of wind speed peak and concomitant wave height, currents, etc. (or switch their order). When the force of a system is dominated by both wind and concomitant wave it may be sufficient to employ the 50-year return wave and 50-year return wind as a design criterion. However, the 50-year return wind and 50-year return wave are frequently not occurred at the same time. Therefore, any simple analysis assuming a perfect correlation between the wind and waves is likely to overestimate the design value (Morton and Bowers, 1996). So, analyzing the encounter probability among the ocean environments by the multivariate distribution can offer useful reference to evaluate project′s safety and cost.

In Multivariate EV theory, two sampling methods, the Block Maxima method and the POT (Peaks Over Thresholds) method have been developed. They correspond respectively to two natural distributions, MGEVD (Multivariate Generalized Extreme Value Distribution) and MGPD. MGEVD is the natural distribution of the block maxima of all components. A typical example is that block is a year and the block maxima are the Annual Maxima Series (AMS). And MGEVD have developed extensively during the last decades too. This is witnessed by several literatures (Morton and Bowers, 1996; Sheng, 2001; YANG and ZHANG, 2013); MGPD should be the theoretical distribution of MPOT method, in which the sample includes all extreme values which are larger than a suitable threshold. MPOT improves the efficiency of the extremes in raw data, and is superior to other methods (such as the annual maximum, Luo and Zhu, 2014). Rootzén and Tajvidi (2006) suggest, based on the research by Tajvidi (1996), that MGPD should be characterized by the following couple properties: (i) exceedances (of suitably coordinated levels) asymptotically have a MGPD if and only if componentwise maxima asymptotically are EVD, (ii) the MGPD is the only one which is preserved under (a suitably coordinated) change of exceedance levels. MPOT method has a high utilization rate of raw data. Besides, Luo et al. (2012) selected 20 or 30-year 6 samples from about 60-year wave raw data arbitrarily. The 6 samples were analyzed by POT and AMS respectively, and the result shows that the return levels with the POT method are closer to the return levels from 60-year wave raw data and have smaller fluctuations than AMS. Because the POT method can get as much extreme information as possible from raw data, its result may be more stable. MGPD is widely used recently, Morton and Bowers (1996) are based on the response function with wave and wind speed of anchoring semi-submersible platforms enabling to analyze extreme anchorage force and corresponding wave height and wind speed by using logical extreme value distribution. They didn’t use the natural distribution of the MPOT method - MGPD, to fitting samples but bivariate extreme value distribution to fitting the MPOT samples. Coles and Tawn (1994) using the same mind. MGPD theory has improved greatly in recent years, but the definition of MGPD still needs further research. Bivariate threshold methods were developed by Joe et al. (1992) and Smith (1994) based on point process theory. MGPD has been the focus of research and the detail about MGPD can be found in Rootzén and Tajvidi (2005), Tajvidi (1996), Beirlant et al. (2005) and Falk et al. (2004).

However, due to difficulties of MGPD in solving procedure, (in general, with the dimension increased, the calculated quantity and complexity rapidly do), the application of MGPD in ocean engineering has been restricted. Monte Carlo simulation is feasible to solve these problems because it only changes inner product operation and the complexity of algorithm doesn't increase with dimension decreases. Liu et al. (1990) use the Monte Carlo simulation for the design of offshore platforms, and practical examples prove its fast calculation speed and high precision in Compound Extreme Value Distribution. Philippe (2000) presented a new parameter estimation method of bivariate extreme value distribution by using Monte Carlo simulation. Shi (1999) presents a Monte Carlo method from a simple trivariate nested logistic model.

Stephenson (2003) gives methods for simulating from symmetric and asymmetric versions of the multivariate logistic distribution, and compares many the Monte Carlo simulation methods of multi-dimensional extreme distribution.

We develop a procedure to handle the application of MGPD in marine engineering design. The paper uses the Monte Carlo simulation to solve the MGPD equation. The Monte Carlo method is introduced in section 2. Fundamental to the application of MGPD is the choice of the optimal joint threshold and the estimation of the joint density. These aspects included in an example are discussed in section 3. Finally, the advantage of MGPD and its Monte Carlo simulation are outlined.

2. MONTE CARLO SIMULATION OF MGPD

2.1 MGPD THEORY

It is well known that MGEVD (Multivariate Generalized Extreme Value Distributions) arise, like in the univariate case, as the limiting distributions of suitably scaled componentwise maxima of independent and identically distributed random vectors. If for independent $X_1,...,X_n$ following F , there exist vectors , a_n , $b_n \in \mathbf{R}^d, a_n > 0$, such that

$$P\left(\frac{\max_{i=1,...,n}(X_i - b_n)}{a_n}\right) = F^n(a_n x + b_n) \xrightarrow{n\to\infty} G(x) \quad (1)$$

where $G(x)$ is a MGEVD, and F is in the domain of attraction of G . We note this by $F \in D(G)$.

The MGPD is based on the extreme value theory and has been widely used in many fields. In one dimension, Generalized Extreme Value distribution (GEVD) is the theoretical distribution of all variation block maxima. GPD describes the properties of extreme of all variation over threshold after declustering, so called POT distribution. Based on the relationship of GPD and GEVD: $H(x) = 1 + \log(G(x)), \log(G(x)) > -1$, the distribution function of MGPD can be deduced:

$$W(X) = 1 + \log(G(x_1,...,x_d))$$

$$= 1 + \left(\sum_{i=1}^{d} x_i\right) D\left(\frac{x_1}{\sum_{i=1}^{d} x_i}, ..., \frac{x_{d-1}}{\sum_{i=1}^{d} x_i}\right), \quad \log\left(G\left(x_1,...,x_d\right)\right) > -1 \tag{2}$$

where $(x_1,...,x_d) = x \in U$, U is a neighborhood of zero in the negative quadrant $(-\infty,0)^d$, D is the Pickands dependence function in the unit simplex $\overline{R_{d-1}}$ on the domain of definition, $\overline{R_d} = \{x \in [0,\infty)^d \mid \sum_{i=1}^{d} x_i = 1\}$; $G(x_1,\cdots,x_n)$ is a MGEVD function which marginal distribution is negative exponential distribution (detail in René Michel, 2007). The MGPD can simply use existing multivariate extreme dependence function due to it is deduced from MGEV, greatly enriched the expression of correlation of MGPD. MGPD has a variety of different types of distribution functions (Coles et al., 1991) varying from Pickands dependence function eq. (3). Logistic dependence function is ease to use and has the favorable statistical properties, and widely used to hydrology, financial and other fields.

$$D_r\left(t_1,...,t_{d-1}\right) = \left(\sum_{i=1}^{d-1} t_i^r + \left(1 - \sum_{i=1}^{d-1} t_i\right)^r\right)^{1/r} \tag{3}$$

$$W_r(x) = 1 - \left(\sum_{i=1}^{d} (-x_i)^r\right)^{1/r} = 1 - \|x\|_r , \tag{4}$$

where *r* is the correlation parameter of dependence function and *r*>1. x_i in the interval (-1, 0), are variables of standardization. The Bivariate Logistic GPD density function is

$$w(x,y) = \frac{\partial W}{\partial x \partial y} = (r-1)(xy)^{r-1}[(-x)^r + (-y)^r]^{1/r-2} \qquad x<0, y<0 \tag{5}$$

The correlation parameter *r* can be evaluated by step by step method: evaluate by using bivariate extreme value distribution firstly then introduce to distribution function; or be evaluated by global method, estimate the parameter by using the maximum likelihood for the density function w . The global method evaluated results more reliable due to the final function form are to be concerned, but the processes of evaluate are more complex. The maximum-likelihood function is

$$L(r)=\sum_{i=1}^{n}\ln(w_r(x_i,y_i)) \tag{6}$$

2.2 SIMULATION METHOD

The Monte Carlo Simulation method of multivariate distribution is relatively complex, because of generating multivariate random and relevant vectors involved. By a transformation method, the variables become independence. And then, every variable is generated a random vector. Finally by the inverse transformation, the random vectors of the multivariate distribution are obtained. The simulation method was suggested by René(2007).

Using polar coordinate to demonstrate the simulated method of MGPD better:

$$T_p(x_1,...,x_d)=(\frac{x_1}{x_1+...+x_d},...,\frac{x_{d-1}}{x_1+...+x_d},x_1+...+x_d)=(z_1,...,z_{d-1},c), \tag{7}$$

where Tp change vector $(x_1,...,x_d)$ into polar coordinate. $\mathbf{C}=x_1+\cdots x_d$ and $\mathbf{Z}=(x_1/\mathbf{C},\cdots,x_{d-1}/\mathbf{C})$ are radial component and angular component, respectively. They called Pickands polar coordinate.

In the Pickands polar coordinate, $W(X)$ presents different properties. Presume that $(X_1,...,X_d)$ follow multivariate generalized Pareto distribution $W(X)$ and its Pickands dependence function D exists d order differential, define the Pickands density of H(X)

$$\phi(z,c)=|c|^{d-1}\left(\frac{\partial^d}{\partial x_1,\cdots,\partial x_d}H\right)T_p^{-1}(z,c) \tag{8}$$

Presume $\mu=\int_{R_{d-1}}\phi(z)dz>0$ and constant $c_0<0$ existing in a neighborhood of zero, then the simulation method of MGPD is: 1) generate uniform random numbers on unit simplex $\overline{R_{d-1}}$; 2) generate random vector $(z_1,\cdots z_d)$ based on the density function $f(z)=\frac{\phi(z)}{\mu}$ of $\mathbf{Z}=(z_1,\cdots,z_{d-1})$ in the Pickands polar coordinate combined with Acceptance-Rejection Method; 3) generate uniform random numbers on $(c_0,0)$; and 4) calculate vector $\left(cz_1,\cdots,cz_{d-1},c-c\sum_{i=1}^{d-1}z_i\right)$ which is random vector of satisfy the multivariate over threshold distribution.

The c_0 above is the joint threshold in MGPD method. This paper determines the threshold by using the principle on Coles and Tawn (1994).

2.3 JOINT PROBABILITY DISTRIBUTION

With the development of offshore engineering, joint probability study for extreme sea environments such as wind, waves, tides and streams is beginning to receive much more attention. But for the selection of design criteria for marine structures, a clear standard about the joint probability has not been found. API (American Petroleum Institute), DNV (Det Norske Veritas) and so on didn’t propose an explicit method for joint variables although they made some relevant rules. API RP2A-LRFD (1995) suggests three options, one of which is “Any ‘reasonable’ combination of wind speed, wave height, and current speed that results in the 100-year return period combined platform load, e.g. base shear or base overturning moment”, but does not provide details of how to determine the appropriate extreme conditions in a multivariate environment.

The joint the return period of two variables need to consider the probability of encounter between variables. In other word, 50-years return wave may not counter 50-year return wind speed. CP can represent the probability of encounter between extreme value of main marine environmental elements and extreme value of its accompanied marine environmental elements. So, it is critical to use CP to describe the probability of their joint together and analyze the effect of all kinds of marine environmental elements to the engineering.

The joint distribution of bivariate Pareto distribution function W(x, y) is

$$W(x,y)=Pr(X<x,Y<y)\text{.}$$

And $W_x(x)$ and $W_y(y)$ are marginal distribution of x and y respectively. Conditional extreme value distribution can be

CP 1: $$Pr(X \geq x \mid Y \geq y) = \frac{Pr(X \geq x, Y \geq y)}{Pr(Y \geq y)} = \frac{1 - W_X(x) - W_Y(y) + W(x,y)}{1 - W_Y(y)} \quad (9)$$

CP 2: $$Pr(X \leq x \mid Y \geq y) = \frac{Pr(X \leq x, Y \geq y)}{Pr(Y \geq y)} = \frac{W_X(x) - W(x,y)}{1 - W_Y(y)} \quad (10)$$

CP 3: $$Pr(X \geq x \mid Y \leq y) = \frac{Pr(X \geq x, Y \leq y)}{Pr(Y \leq y)} = \frac{W_Y(y) - W(x,y)}{W_Y(y)} \quad (11)$$

CP 4: $$Pr(X \leq x \mid Y \leq y) = \frac{Pr(X \leq x, Y \leq y)}{Pr(Y \leq y)} = \frac{W(x,y)}{W_Y(y)} \quad (12)$$

Other four CP distributions can be deduced by swapping two variables.

3. CASE

The ocean environment is multivariate with waves, wind and currents all contributing to the forces experienced by marine structures. Any simple assumes that there is either perfect dependence or independence among them are unreasonable for a design criterion. The section combined with CP for analysis of the design criteria (taking the 50-year return value as an example) about base shear of a plat. Assume the function of base shear and wind velocity and wave height for a certain type of structure is

$$Z(x,y) = 0.44x^2 + 20.18y^2 \quad (13)$$

where x and y are wind speed and wave height, with their units m/s and m respectively. Z represents base shear and the unit is KN .

3.1 THE DATA

The data are a sequence of 23 (1960-1982) years of wave height records and synchronous wind speed records, which was recorded four times a day from an ocean hydrological station (ZL) near 22.6^{o}N, 115.5^{o}E (fig. 1) in SCS, which is a typical typhoon sea area. In the raw data, the maximum winds reached 40m/s and the maximum wave height is 8.50m. Due to restrictions with observation technologies at the time, the wind speed and the wave height were kept only the integer and one decimal place respectively. This will influence the level of precision of extreme value, so in fig. 2 (a), the all the wind values in the range 0.4 and 0.6 (standard unit) show the same probability.

3.2 DECLUSTERING

The sample of MPOT method is from the extreme value of blocks, so the first stage in a multivariate extreme value analysis is to identify a set of declustered events. For satisfying independent random distribution assumption, the principle of declustering is keeping samples independence. In SCS, typhoon occurs frequently and causes almost all extreme wind speed and wave height. Generally the influence of typhoon for one point may last several days or one week. Exploratory analysis suggested that a window corresponding to five days is appropriate. But, if the interval between the adjacent extremes is less than 2 days, then we need to delete smaller values from the adjacent extremes in order to keep independence.

After declustering, the block maxima should be extracted as extreme values. Generally, in a bivariate analysis of wave heights and wind speeds, there are four definitions of "maximum" data set in a certain block: (1) maximum wave height and maximum wind speed in a block; (2) maximum wave height and its "concomitant" wind speed in a block; (3) maximum wind velocity and its "concomitant" wave height in a block; Liu et al. (2002) suggests that definition (1) is the simplest one, but it may lead to over-conservative results for they are not observed at the same time; for definitions (2) and (3), only one of the extreme observations is selected, so we have to consider which has a major influence for offshore structures. For example, if wave load has a major influence on a certain structure, definition (2) would be a good choice. In case that we do not know which load has a dominating influence on a structure, we usually make extreme value analysis on definitions (2) and (3) respectively, and finally choose the more dangerous one.

In the present study, the case is used solely for testing the validity of the simulation method of MGPD and giving an example of its application. So definition (3) is an arbitrary choice. That is wind speed is the "dominating" variable. After declustering according to the requirements of independence, 1436 groups of extreme wind speed and corresponding wave height are selected.

3.3 MARGINAL TRANSFORMATION AND JOINT THRESHOLD

In section 2.1, the variables of MGPD must be in a neighborhood of zero in the negative quadrant. By a suitable marginal transformation, we can transfer a margin into a uniform margin in a neighborhood of zero.

After many experiments, it is found that marginal distributions of extreme wind speeds and wave heights in 1436 data sets can be described by GEVD:

$$F(x)=P(X<x)=exp\left\{-[1-\xi(\frac{x-\mu}{\sigma})]^{1/\xi}\right\},\xi\neq 0 \quad (14),$$

where ξ,σ and μ are three variables of GEVD. They are estimated by using maximum likelihood estimate. This is an approach of estimation suggested in Section 5.1 of Beirlant et al. (2005) and Section 3.3 of Coles (2001). Fig. 2 shows the probability plot and probability plots (including the 95% confidence intervals) of marginal distribution before fitting MGPD.

To standardize the margins, the marginal distribution of MGPD is negative exponential distribution. According to Taylor expansion, we get

$$logF(x_i)=log(1+F(x_i)-1)\approx F(x_i)-1 \quad (15),$$

where $F(x_i)$ is GEVD of variable i $(i=1,2)$,which represent wind and wave respectively (detail in René Michel, 2007).

The dependence models between extreme variables have been suggested: Logistic, Bilogistic and Dirichlet. However, it appears that the choice of dependence model is not usually critical to the accuracy of the final model (Morton and Bowers, 1996). So the simple bivariate Logistic GPD was selected. The MGPD model of the paper is based on multivariate extreme value distribution, the joint threshold can be calculated by the method in section 2.2. The joint threshold is $c_0=-0.7$, and there are 450 groups of combination of wind speed and wave height over c_0. Fig. 3 (a) shows that the samples of over threshold value. In the left subfigure, $c_0 = -0.7$ is a curve, and the right side of the curve are over threshold value. In the right subfigure $c_0 = -0.7$ is a line, the top-right side of the line are over threshold value of the data converted.

The correlation parameter r of dependence function is estimated by the maximum-likelihood method, and the joint distribution is showed in Fig. 3 (b).

3.4 COMPARISON OF STOCHASTIC SIMULATION RESULTS

For the annual maximum of wind speed and wave height, Pearson type Ⅲ distribution is used to obtain return period values of wind speed and wave height in one-dimensional (Fig. 4). The Pearson type Ⅲ distribution is

$$F(x) = P(X < x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_{-\infty}^{x} (x-\mu)^{\alpha-1} exp[-\beta(x-\mu)]dx \tag{16}$$

Fig. 5 shows the data of stochastic simulation by N=50000 and N=100000. Due to the accuracy of the raw data, the 40m/s wind event is so much that they can't be fitted by MGPD well. So the simulation results and the measurement points are not overlap completely for the 40m/s wind event in Fig. 5. But the simulation results are in basic conformity with the actual situation, and this represents that the MGPD simulation method is worked. The scatter diagrams show the results directly, they need further quantitative analysis to show the differences of them objectively. A couple of CP are to be used to the paper: 1) $P(H > h | V > v)$ and 4) $P(H < h | V < v)$ which means 1) the probability of the wave height over a stander h under the wind speed over the stander v and 4) the probability of the wave height less than a stander h under the wind speed less than the stander v, respectively. Both of them are actually responding the probability of extreme value wave height and its corresponding wind speed joint occur.

Fig. 6 represent calculating the CP $P(H > h | V > v)$ by group $h = 7.99m$ and $v = 37m/s$. Using the Monte Carlo method calculate its CP by the result of simulation based on the definition of conditional distribution. As it's showed in Fig. 6, the difference value of the simulation and the model are related to the simulation times N.

Relative difference value has been reduced with the increase of simulation times. When the simulation times up to 2×10^{6}, the relative error value of simulation results and calculation results is 0.1%, shows that the simulation results error is acceptable.

Based on the results by simulation times 2e6, Tab. 2 and Tab. 3 show the calculation results of two different CP. The tables represent 5 groups' calculation and stochastic simulation results of CP 1 and 4 on different combination of wave height and wind speed. The two results are closely. For instance, the calculated result of the probability of wave height of the design frequency more than 10% encounter wind speed of the design frequency more than 10% is 94.67% where stochastic simulation result is 94.44%, the relate error only 0.24%. The calculated result of the probability of wave height of the design frequency more than 2% encounter wind speed of the design frequency more than 10% is 40.05% where stochastic simulation result is 38.25%, the relate error only 4.49%. Synthesizing the appearing probability of extreme sea environment is to the benefit of find a balance between the engineering investment and risk, and can provide the scientific basis for the risk pre-estimates.

3.5 ANALYSIS OF RETURN VALUE

For 'reasonable' combination of wind speed and wave height, the method of analyzing the multivariate extreme environment is offered by CP. Because wind speed is the "dominating" variable in the paper, we use CP 1: $P(H>h\,|\,V>v)$ as a standard. The return value of the base shear follows the couple principles: (1) if $P(H>h_{50}\,|\,V>v_{50})$ >98%, the 50-year base shear is the combination of 50-year wind speed and 50-year wave height; (2) if $P(H>h_{50}\,|\,V>v_{50})$ <98%, the 50-year base shear is the combination of 50-year wind speed and the wave height corresponding to $P(H>h_{50}\,|\,V>v_{50})$ =98%. The value 98% was selected only to show how to determine design criterion by the CP, and the paper do not analyze whether the value is appropriate for the marine structure. In the method, it can be avoided that the encounter probability of wind speed and wave height is smaller.

The result of CP 1: $P(H > h \mid V > v)$ is in tab. 2. We can know that the estimates of 50-year wind speed and wave height are 43.41m/s and 9.13m respectively, and their CP1 is 91.58. According to the above principles, this is meeting the principle (2). Using the Monte Carlo method, we can get $P(H > 9.01 \mid V > v_{50})$ =98%. So the 50-year base shear is $Z(43.41, 9.01)$=2467.4KN.

4. Discussion and Conclusions

4.1 NEW POSSIBILITIES BASED ON MONTE CARLO SIMULATION

The design sea states for a particular location are often used in design and assessment of coastal and ocean engineering. These design sea states might be jointly decided by multivariate ocean environment factors, such as combinations of wave conditions and water levels with given joint return periods. A potentially better approach is possible based on Monte Carlo simulation of a wide range of ocean environment factors, from which the structure variables in different return periods can be determined directly. Once the long-term (such as thousands of years) sea state data has been simulated, several structure variables or ocean environment factors can be assessed quite quickly by the law of large numbers. If the raw data is enough, the MGPD can be used in extreme analysis based on 3 or more-variables. In this case, Monte Carlo simulation becomes a must-have tool, because it is difficult to solve the high-dimension MGPD and get the structure variables in different return periods.

4.2 CONCLUSIONS

The MGPD is the nature distribution of MPOT method, which can dig up more extreme information from the raw data. The model based on the extreme value theory which is well-founded and the intrinsic properties of all extreme variables are into consideration. The MGPD is a useful basis for extreme value analysis of the offshore environment. The Monte Carlo simulation of MGPD provides estimates of the 50-year base shear, taking into account the conditional probability, which represents the encounter probability between variables such as wind speed and wave height. CP1 includes the encounter probability of extreme events and provides the theory basis for finding the best balance point between engineering cost and risk. The analysis of the return mooring forces at SCS illustrates the practicalities of Monte Carlo simulation of MGPD.

Acknowledgments

This work was supported by the 'Strategic Priority Research Program' of the Chinese Academy of Sciences (XDA11010302);

Referencesd

API RP 2A-LRFD-Planning, Designing and Constructing for Fixed Offshore Platforms Load and Resistance Factor Design, ISO 13819-2: 1995(E), 127.

Beirlant J, Segers J, Teugels J. Statistics of extremes, theory and applications. Wiley, Chichester 2005.

Coles S G, Tawn J A. Modelling extreme multivariate events. Journal of the Royal Statistical Society 1991; B53, 377-392.

Coles S G, Tawn J A. Statistical methods for multivariate extremes: an application to structural design. Applied Statistics 1994; 43, 1-48.

Falk M, Hüsler J, Reiss, R D. Laws of Small Numbers: Extremes and Rare events. 2nd edn. Birkhäuser, Basel 2004.

Joe H, Smith R L, Weissman I. Bivariate threshold methods for extremes. Journal of the Royal Statistical Society 1992; B54, 171-183.

Liu Defu, Shi Jiangang, Zhou Zhigang. The application of Monte-carlo method in the Marine engineering design, Acta Oceanologica Sinica 1990; 12(2) 236-243. (in chinese)

Liu, D. F., Wen, S. Q. and Wang, L. P.. Poisson-Gumbel mixed compound distribution and its application, Chin. Sci. Bull. 2002; 47(22) 1901–1906.

Luo Yao, Zhu Liangsheng. A New Model of Peaks over Threshold for Multivariate Extremes [J], 2014. China Ocean Eng., 28(6):761~776

Luo Yao, Zhu Liangsheng, Hu Jinpeng. A new method for determining threshold in using PGCEVD to calculate return values of typhoon wave height [J]. China Ocean Eng., 2012, 26(2): 251−260.

Morton I D, Mould G. Extreme value analysis in a multivariate offshore environment. Applied Ocean Research 1996; 18, 303-317.

Philippe Caperaa. Estimation of a Bivariate Extreme Value Distribution. Extreme 2000; 3(4), 311-329

René M. Simulation of certain multivariate generalized Pareto distributions. Extremes 2007; 10, 83–107.

Rootzen, Tajvidi, Multivariate Generalized Pareto Distributions, Preprint 2005:38, Chalmers Tekniska Häogskola Gäoteborg, http://www.math.chalmers.se/Math/Research/Preprints/2005/38.pdf 2005.

Rootzén H, Tajvidi N. Multivariate generalized Pareto distributions. Bernoulli 2006; 12, 917-930.

Shi D.. Moment estimation for multivariate extreme value distribution in a nested logistic model. Ann. Inst. Statist. Math. 1999; 51, 253-64.

Sheng Yue. The Gumbel logistic model for representing a multivariate storm event. Advances in Water Resources 2001; 24. 179-185.

Smith R L. Multivariate threshold methods. In Galambos, J., Lechner, J., Simiu, E., editors, Extreme Value Theory and Applications 1994; pages 225-248. Kluwer, Dordrecht.

Stephenson A. Simulating multivariate extreme value distributions of logistic type. Extremes 2003; 6, 49–59.

Tajvidi N, Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions, Department of Mathematics, Chalmers Tekniska Hogskola Goteborg, http://www.math.chalmers.se/ »nader/thesis.ps 1996

Xiao-chen YANG, Qing-he ZHANG, Joint probability distribution of winds and waves from wave simulation of 20 years (1989-2008) in Bohai Bay, Water Science and Engineering 2013; 6(3): 296-307.

Tab.1 parameters of marginal distribution

	ξ	σ	μ
Wind speed	0.362	2.868	15.403
Wave height	0.008	0.861	1.857

Tab. 2 comparision of the results of CP 1

RP (year)		5		10		20		50		100	
RP	V(m/s) / H(m)	26.38		37		40.01		43.41		45.67	
		c	s	c	s	c	s	c	s	c	s
5	4.65	85.71	85.59	99.89	100.0	99.97	100.0	99.99	100.0	100.0	100.0
10	7.08	11.38	11.41	94.67	94.44	98.72	99.03	99.75	98.95	99.92	97.62
20	7.99	4.05	4.14	80.10	80.02	94.93	95.79	99.02	98.95	99.68	97.62
50	9.13	1.13	1.11	40.05	38.25	75.42	73.14	94.75	91.58	98.26	97.62
100	9.98	0.45	0.49	17.83	18.91	45.36	48.22	83.08	80.00	94.15	92.86

RP: return periods, a: calculation results by analytic solution, s: simulation results

Tab. 3 comparision of the results of CP 4

RP (year)		5		10		20		50		100	
RP	V(m/s) / H(m)	26.38		37		40.01		43.41		45.67	
		c	s	c	s	c	s	c	s	c	s
5	4.65	99.24	97.29	98.79	95.77	98.78	95.74	98.78	95.73	98.78	95.73
10	7.08	100	99.99	99.95	99.82	99.94	99.79	99.94	99.78	99.94	99.78
20	7.99	100	100	99.99	99.96	99.98	99.94	99.98	99.93	99.98	99.92
50	9.13	100	100	100	100	100	99.99	100	99.98	99.99	99.98
100	9.98	100	100	100	100	100	100	100	99.99	100	99.99

RP: return periods, a: calculation results by analytic solution, s: simulation results

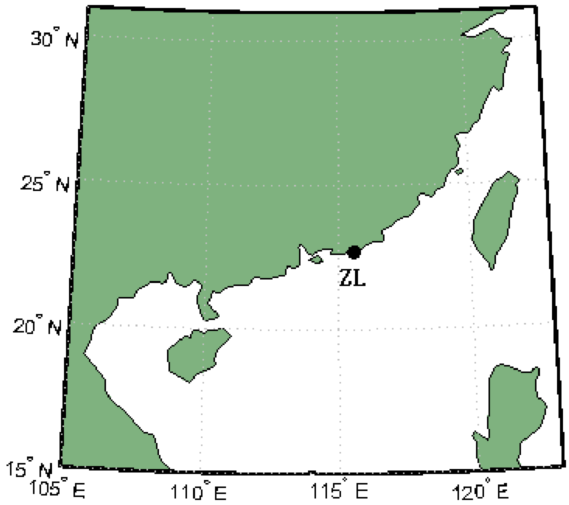


Fig. 1 Location of ZL ocean hydrological station

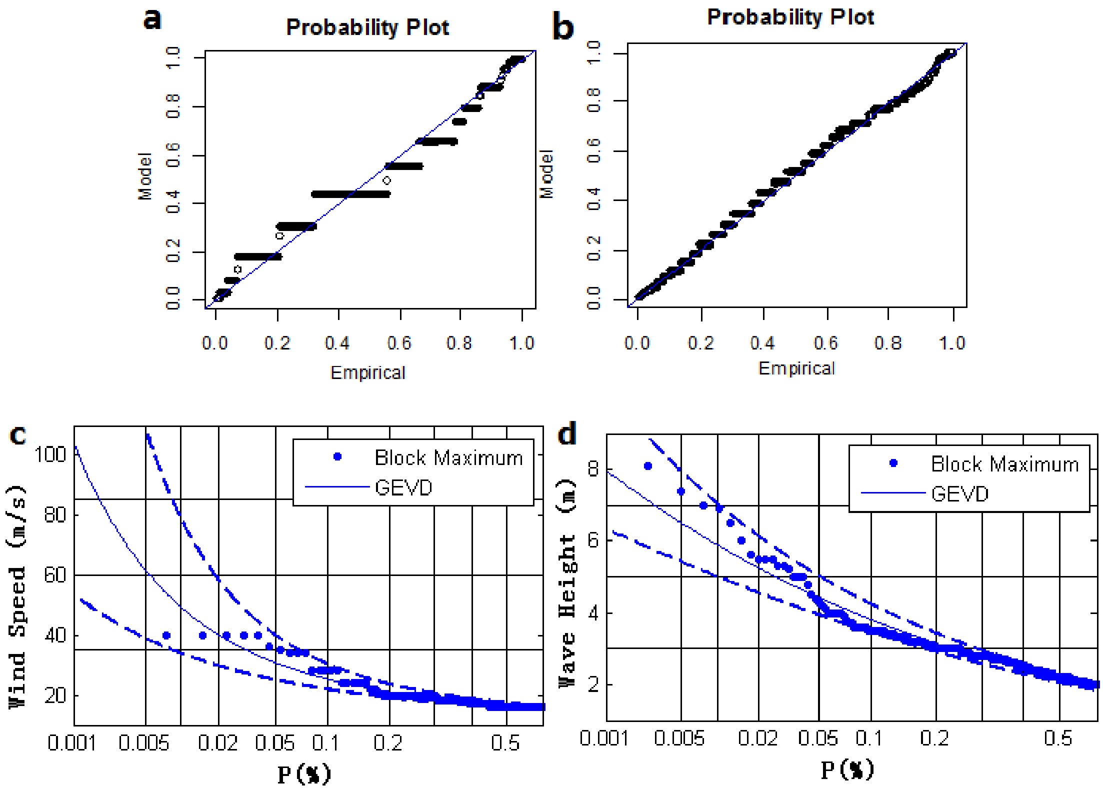


Fig. 2 fitting testing of marginal distribution, (a) wind speed (b) wave height

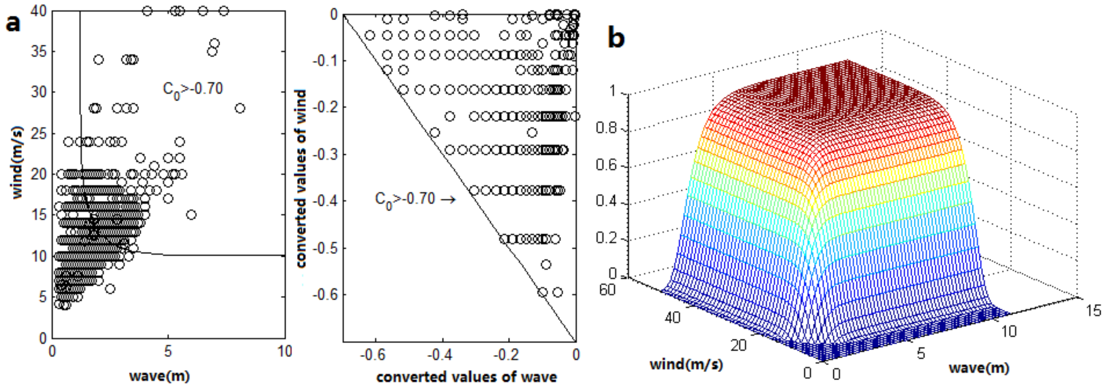


Fig. 3 (a) over threshold value of wave height and wind speed

(b) joint distribution of extreme value of wave height and wind speed

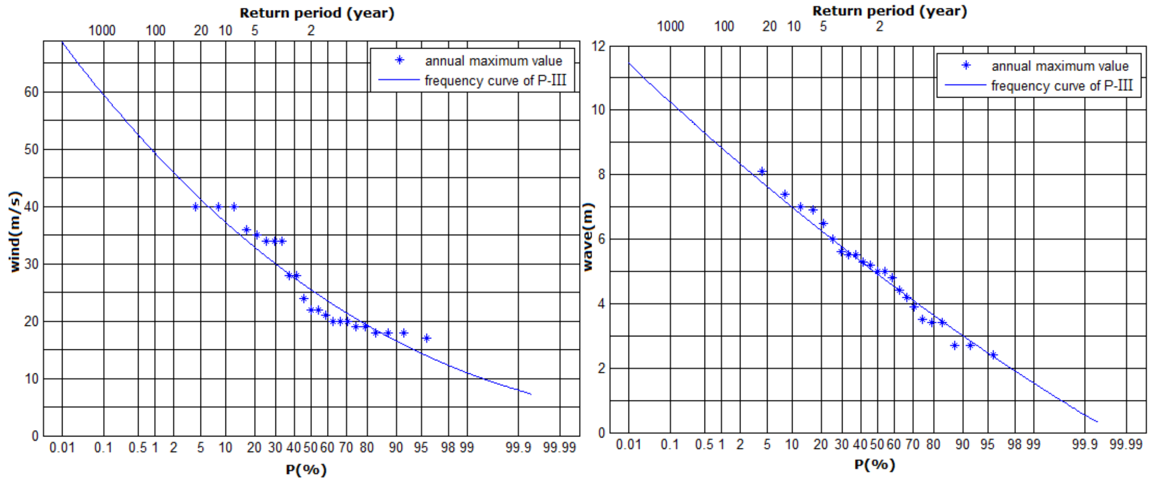
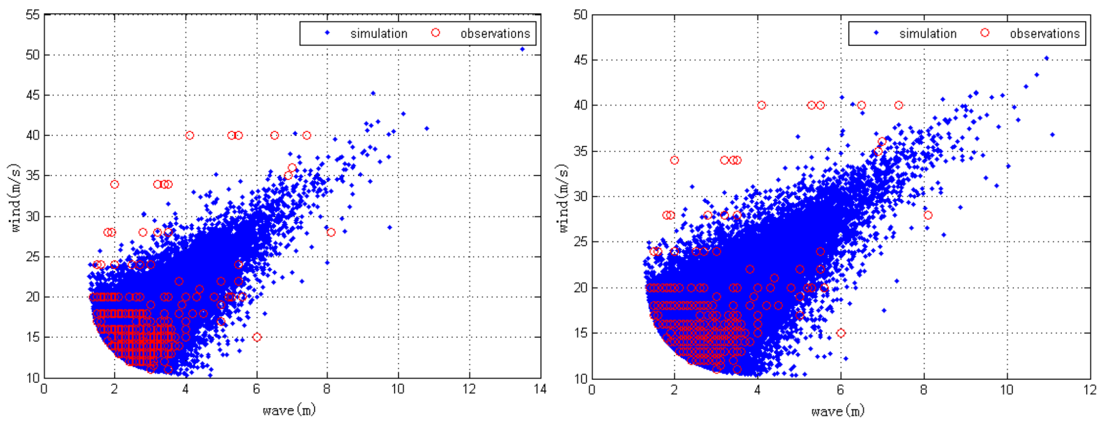


Fig. 4 calculation of wind speed samples and wave height return value



N=50000 N=100000

Fig. 5 value of over threshold and data of stochastic simulation

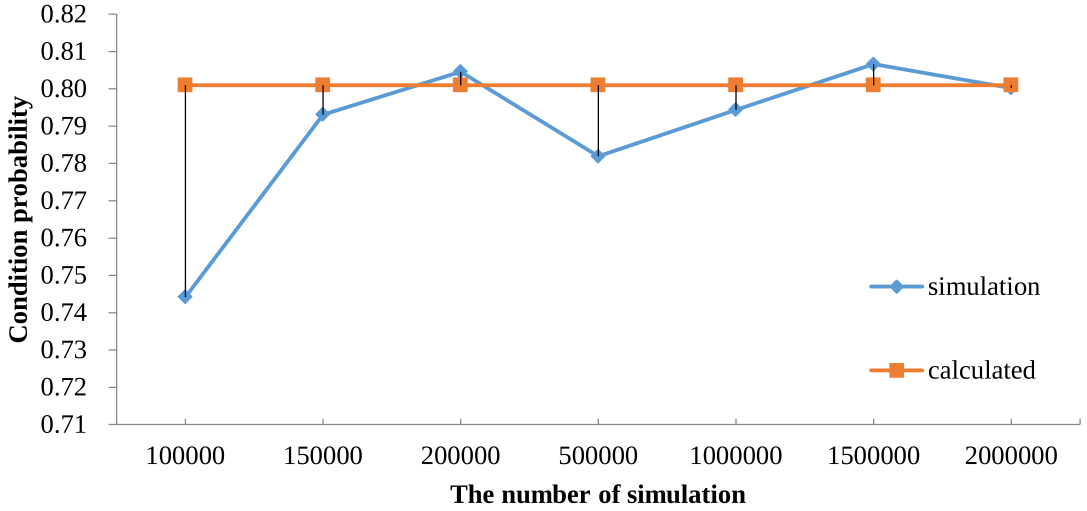


Fig. 6 The change of simulation accuracy