



This discussion paper is/has been under review for the journal Ocean Science (OS).
Please refer to the corresponding final paper in OS if available.

A Monte Carlo simulation of multivariate general Pareto distribution and its application

L. Yao¹, W. Dongxiao², Z. Zhenwei³, H. Weihong¹, and S. Hui⁴

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

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

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

⁴China Water Resources Pearl River Planning Surveying & Designing Co., Ltd., Guangzhou, China

Received: 20 August 2014 – Accepted: 21 October 2014 – Published: 8 December 2014

Correspondence to: W. Dongxiao (dxwang@scsio.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

2733

Abstract

This paper presents a multivariate general Pareto distribution (MGPD) method and builds a method for solving MGPD through the use of a Monte Carlo simulation for marine environmental extreme-value parameters. The simulation method has proven to be feasible in the analysis of the joint probability of wave height and its concomitant wind from a hydrological station in the South China Sea (SCS). The MGPD is the natural distribution of the multivariate peaks-over-threshold (MPOT) sampling method, and is based on the extreme-value theory. The existing dependence functions 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 wave and the concomitant wind from a period of 23 years (1960–1982), the number of the wave and wind selected is averaged to 19 per year. For the joint conditional probability of the MGPD, the relative error is rather small in the Monte Carlo simulation method.

1 Introduction

Statistical modeling of extreme values (EV) plays a crucial role in design and risk evaluation in ocean engineering, and multivariate extreme-value distributions have been extensively developed over the last decades. This has been shown by several studies (Morton and Bowers, 1996; Sheng, 2001; Yang and Zhang, 2013). Problems concerning ocean environmental extremes are often multivariate in character. An example of this is that ocean environments (including waves, wind and currents) all contribute to the forces experienced by offshore systems during typhoons. Thus the severity of such a typhoon event may be described by a function of both wind speed peak and concomitant wave height. When the force of a system is dominated by both wind and concomitant wave, it may be sufficient to employ a 50-year return wave and 50-year return wind as a design criterion. However, the 50-year return wind and 50-year return

2734

wave do not occur 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). Therefore, analyzing the encounter probability among the ocean environments by means of the multivariate distribution can offer useful reference in evaluating a project's safety and cost.

In multivariate EV theory, two sampling methods – the block maxima method and the peaks-over-thresholds (POT) method – have been developed. These methods respectively correspond to two natural distributions: multivariate EV distribution (MEVD) and multivariate general Pareto distribution (MGPD). MEVD is the natural distribution of the block maxima of all components. A typical example is that a block is a year and the block maxima are the annual maxima. MGPD is the theoretical distribution of the multivariate peaks-over-threshold (MPOT) method, in which the sample includes all values which are larger than a suitable threshold. Rootzén and Tajvidi (2006), based on the research of Tajvidi (1996), suggest that MGPD should be characterized by the following few properties: (i) exceedances (of suitably coordinated levels) asymptotically have a multivariate GP distribution if and only if componentwise maxima asymptotically are EV distributed, (ii) the multivariate GP distribution is the only one which is preserved under (a suitably coordinated) change in exceedance levels. The MPOT method has a high utilization rate of raw data and a more stable calculation result; as a result it has recently become widely used. The study of Morton and Bowers (1996) is based on the response function with wave and wind speed of anchoring semi-submersible platforms, enabling analysis of extreme anchorage force and corresponding wave height and wind speed through the use of logical extreme-value distribution. In that study, the authors do not use the natural distribution of the MPOT method, MGPD, to fit samples but instead bivariate extreme-value distribution to fit the POT samples. Coles and Tawn (1994) also follow the same idea. 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.

2735

MGPD has been the focus of certain studies, and further detail about it can be found in Rootzén and Tajvidi (2005), Tajvidi (1996), Beirlant et al. (2005) and Falk et al. (2004).

However, due to difficulties regarding MGPD in the solving procedure, (in general, with the dimension increased, the calculated quantity and complexity also rapidly increase), the application of MGPD in ocean engineering has been restricted. The use of Monte Carlo simulation is feasible to solve these problems because it only changes inner product operation and the complexity of the algorithm does not increase as 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) presents a new parameter estimation method of bivariate extreme-value distribution that uses 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 of the Monte Carlo simulation methods of multidimensional extreme distribution.

We have developed a procedure to handle the application of MGPD in marine engineering design. This paper uses the Monte Carlo simulation to solve the MGPD equation, and is structured as follows. The Monte Carlo method is introduced in Sect. 2. Fundamental to the application of MGPD is the choice of the optimal joint threshold and the estimation of the joint density: these aspects, including a case study, are discussed in Sect. 3. Finally, in Sect. 4, the advantages of MGPD and its Monte Carlo simulation are outlined.

2 Monte Carlo simulation of MGPD

2.1 MGPD theory

The MGPD method is based on the extreme-value theory, and has been widely used around the world. Generalized extreme value distribution (GEVD) is the theoretical

2736

If we assume that $\mu = \int_{R_{d-1}} \varphi(z) dz > 0$ and constant $c_0 < 0$ exist in a neighborhood of zero, then the simulation method of MGPD is as follows: (1) generate uniform random numbers on unit simplex $\overline{R_{d-1}}$; (2) generate random vector (z_1, \dots, z_d) based on the density function $f(z) = \frac{\varphi(z)}{\mu}$ of $Z = (z_1, \dots, z_{d-1})$ in the Pickands polar coordinate combined with the acceptance–rejection method; (3) generate uniform random numbers on $(c_0, 0)$; and (4) calculate vector $(cz_1, \dots, cz_{d-1}, c - c \sum_{i=1}^{d-1} z_i)$, which is a random vector for satisfying the multivariate over-threshold distribution.

The c_0 above is the joint threshold in the MGPD method. This paper determines the threshold following the principle of 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. API (American Petroleum Institute), DNV (DET NORSE VERITAS) and so on were not proposed an explicit method as design criteria for marine structures although they made some relevant rules. API (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”. The joint return period of two variables needs to be considered for the probability of encounter between variables. Conditional probability can represent the probability of encounter between the extreme value of main marine environmental elements and the extreme value of its simultaneous marine environmental elements. For example, the probability of a 50-year return wave and a 50-year return wind speed occurring simultaneously at the same place is very small. Therefore, it is critical to use conditional probability to describe the probability of their joining together and analyze the effect of all kinds of marine environmental elements with regard to engineering.

2739

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

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

$W_x(x)$ and $W_y(y)$ are marginal distributions of x and y , respectively. Conditional extreme-value distribution can be as follows:

5 Conditional probability 1:

$$\Pr(X \geq x | Y \geq y) = \frac{\Pr(X \geq x, Y \geq y)}{\Pr(Y \geq y)} = \frac{1 - W_x(x) - W_y(y) + W(x, y)}{1 - W_y(y)}; \quad (8)$$

conditional probability 2:

$$\Pr(X \leq x | Y \geq y) = \frac{\Pr(X \leq x, Y \geq y)}{\Pr(Y \geq y)} = \frac{W_x(x) - W(x, y)}{1 - W_y(y)}; \quad (9)$$

conditional probability 3:

$$10 \Pr(X \geq x | Y \leq y) = \frac{\Pr(X \geq x, Y \leq y)}{\Pr(Y \leq y)} = \frac{W_y(y) - W(x, y)}{W_y(y)}; \quad (10)$$

conditional probability 4:

$$\Pr(X \leq x | Y \leq y) = \frac{\Pr(X \leq x, Y \leq y)}{\Pr(Y \leq y)} = \frac{W(x, y)}{W_y(y)}. \quad (11)$$

Another four conditional probability distributions can be deduced by swapping two variables.

3 Case study

3.1 Sample selection of over-threshold values and marginal distribution

The raw data of the paper are wave height and synchronous wind speed observed four times a day over 23 years from an ocean hydrological station in the South China Sea

(SCS). In the sample, the maximum winds reached 40 ms^{-1} and the maximum wave height was 8.50 m. The extreme wind speed and its corresponding wave height are selected as research samples. The sample of the over-threshold method is from the extreme value of blocks, and the principle behind declustering is to maintain sample independence. In the SCS, typhoons occur frequently and are the cause of almost all extreme wind speeds and wave heights. Generally, a typhoon may last several days or 1 week in the SCS, and so this paper declusters by 5-day intervals, taking the maximum value of a block. If the interval between two extremes is less than 2 days, then we need to delete smaller values from the samples in order to keep independence. Except for some individual processes of the storm which last a long time, most of the data meet the requirements of independence.

After the sample has been fixed and completed according to the requirements of independence, 1436 groups of extreme wind speed and corresponding wave height are selected. Their marginal distributions can be described by means of GEVD. GEVD includes three types of extreme-value distribution, and both marginal distributions use three variables for GEVD in this paper.

$$F(x) = P(X < x) = \exp \left\{ - \left[1 - \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{1/\xi} \right\}, \xi \neq 0, \quad (12)$$

where ξ , σ and μ are the three variables of GEVD; these are estimated by means of a maximum-likelihood estimate. Figure 1 shows the probability plot of marginal distribution. For the annual maximum of wind speed and wave height, a Pearson type III distribution is used to obtain return period values of wind speed and wave height in one dimension (see Fig. 2). The Pearson type III distribution is

$$F(x) = P(X < x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{-\infty}^x (x - \mu)^{\alpha-1} \exp[-\beta(x - \mu)] dx. \quad (13)$$

2741

3.2 The joint probability distribution

The bivariate logistic generalized Pareto distribution was selected, and the data were converted to negative exponential distribution in the interval $(-1, 0)$ due to the active interval of the method being $(-\infty, 0)$. The MGPD model of the paper is based on multivariate extreme-value distribution; the joint threshold can be calculated using the method developed by Coles and Tawn (1994). The joint threshold is $c_0 = -0.7$, and there are 450 groups of the combination of wind speed and wave height over c_0 . Figure 3a shows the samples of over-threshold values. In the left-hand panel of Fig. 3a, $c_0 = -0.7$ is a curve, and the right side of the curve shows values over the threshold. In the right-hand panel of the Fig. 3a, $c_0 = -0.7$ is a line, and the area to the top right of the line represents over-threshold values the converted data in polar coordinates. The joint distribution is shown in Fig. 3b.

3.3 Comparison of stochastic simulation results

Figure 4 shows the over-threshold values and the data of stochastic simulation by $N = 50\,000$ and $N = 100\,000$, respectively.

The simulation results are in agreement with the actual situation, showing that the MGPD method was successful. The scatter diagrams show the results directly, but they require further quantitative analysis in order to show the differences of them objectively. A couple of the conditional probabilities mentioned above are used in this paper: (1) $P(H > h | V > v)$ and (4) $P(H < h | V < v)$, which mean (1) the probability of the wave height over h under the wind speed over v and (4) the probability of the wave height less than h under the wind speed less than v , respectively. Both of these actually respond to the probability of extreme-value wave height and its corresponding wind speed joint occurrence.

Figure 5 shows the calculation of the conditional probability $P(H > h | V > v)$ by group $h = 7.99 \text{ m}$ and $v = 37 \text{ ms}^{-1}$. The Monte Carlo method is used to calculate its conditional probability through the result of the simulation based on the definition of con-

ditional distribution. As is shown in Fig. 5, the difference values of the simulation and the model are related to the simulation times N . A relative difference value has been reduced with the increase in simulation times. When the simulation times are up to 2×10^6 , the relative error value of simulation results and calculation results is 0.1 %, which shows that the error of simulation results is acceptable.

Based on the results with simulation times 2×10^6 , Tables 2 and 3 represent the calculation results of two different conditional probabilities.

Tables 2 and 3 show, for five groups, the calculation and stochastic simulation results of conditional probability 1 and 4 for different combinations of wave height and wind speed. In the two tables, probability 20, 10, 5, 2 and 1 % represent 5, 10, 20, 50 and 100-year return wave heights and wind speeds, respectively, which are obtained by the annual maximum value and Pearson type III distribution. The results for the calculation and the stochastic simulation are similar. For instance, the calculated result of the probability of greater than the 10-year return value (7.08 m) for wave height when wind speed is greater than the 10-year return value (37 m s^{-1}) is 94.67 %, whereas the stochastic simulation result of the same conditional probability is 94.44 %. The relative error is 0.24 %. For conditional probability 4, the calculated result of the probability of less than the 5-year return value (4.65 m) for wave height when wind speed is less than the 50-year return value (43.41 m s^{-1}) is 98.78 %, whereas the stochastic simulation result of the same conditional probability is 95.73 %. The relative error is 3.09 %. Synthesizing all conditional probabilities among the different extreme sea environments is beneficial for finding a balance between investment and risk with regard to engineering, and can provide a scientific basis for pre-estimates of risk.

4 Conclusions

1. This paper presents the theoretical method of MGPD, which is based on existing multivariate extreme-value distribution and can describe various dependence relationships among different extreme-value variables. The model is based on the

2743

theory of extreme values, which is well founded, and the intrinsic properties of all extreme variables are taken into consideration.

2. Through analysis of conditional probability, the Monte Carlo method of MPOT has only small errors, and provides a solution for the analysis of multivariate and complex cases, and thus the technique shows promise for future use.
3. Conditional probability includes the probability of extreme events being encountered, and provides the theoretical basis for finding the best balance point between engineering cost and risk.
4. The model of MGPD has the ability to describe the probability of multivariate extreme-event occurrence at the same time. A larger sample size than traditional annual extreme-value methods allows for the extreme features of the raw data to be maintained as best as possible.

References

- API RP 2A-LRFD: Planning, Designing and Constructing for Fixed Offshore Platforms Load and Resistance Factor Design, ISO 13819-2, 127, 1995.
- Beirlant, J., Segers, J., and Teugels, J.: *Statistics of Extremes, Theory and Applications*, Wiley, Chichester, 2005.
- Coles, S. G. and Tawn, J. A.: Modelling extreme multivariate events, *J. Roy. Stat. Soc. B*, 53, 377–392, 1991.
- Coles, S. G. and Tawn, J. A.: Statistical methods for multivariate extremes: an application to structural design, *Appl/ Statist.*, 43, 1–48, 1994.
- Falk, M., Hüsler, J., and Reiss, R. D.: *Laws of Small Numbers: Extremes and Rare Events*, 2nd edn., Birkhäuser, Basel, 2004.
- Joe, H., Smith, R. L., and Weissman, I.: Bivariate threshold methods for extremes, *J. Roy. Stat. Soc. B*, 54, 171–183, 1992.
- Liu, D., Shi, J., and Zhou, Z.: The application of Monte-Carlo method in the marine engineering design, *Acta Oceanol. Sin.*, 12, 236–243, 1990 (in Chinese).

2744

Morton, I. D. and Mould, G.: Extreme value analysis in a multivariate offshore environment, *Appl. Ocean Res.*, 18, 303–317, 1996.

Caperaa, P.: Estimation of a bivariate extreme value distribution, *Extreme*, 3, 311–329, 2000.

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

5 Rootzen, H. and Tajvidi, N.: Multivariate Generalized Pareto Distributions, Preprint, 2005:38, Chalmers Tekniska Högskola Göteborg, available at: <http://www.math.chalmers.se/Math/Research/Preprints/2005/38.pdf> (last access: 25 November 2014), 2005.

Rootzén, H. and Tajvidi, N.: Multivariate generalized Pareto distributions, *Bernoulli*, 12, 917–930, 2006.

10 Shi, D.: Moment estimation for multivariate extreme value distribution in a nested logistic model, *Ann. I. Stat. Math.*, 51, 253–264, 1999.

Smith, R. L.: Multivariate threshold methods, in: *Extreme Value Theory and Applications*, edited by: Galambos, J., Lechner, J., and Simiu, E., 225–248, Kluwer, Dordrecht, 1994.

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

Tajvidi N.: Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalised Pareto Distributions, Department of Mathematics, Chalmers Tekniska Hogskola, Goteborg, 1996.

20 Yang, X. and Zhang, Q.: Joint probability distribution of winds and waves from wave simulation of 20 years (1989–2008) in Bohai Bay, *Water Science and Engineering*, 6, 296–307, 2013.

Yue, S.: The Gumbel logistic model for representing a multivariate storm event, *Adv. Water Resour.*, 24, 179–185, 2001.

Table 1. Parameters of marginal distribution.

	ξ	σ	μ
Wind speed	0.108	2.807	13.488
Wave height	0.008	0.861	1.763

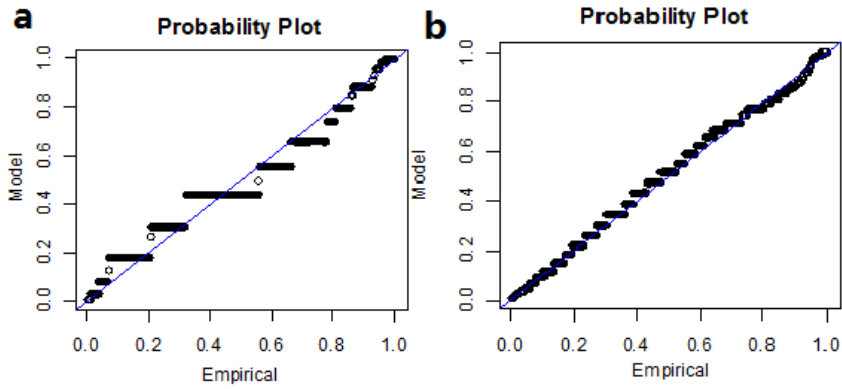


Figure 1. Fitting testing of marginal distribution: (a) wind speed and (b) wave height.

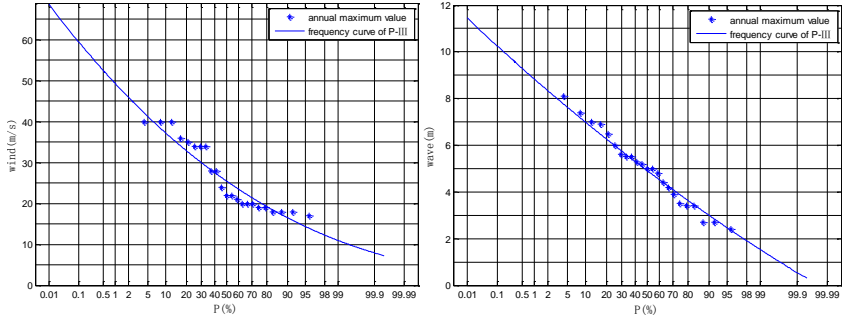


Figure 2. Calculation of wind speed samples and wave height return values.

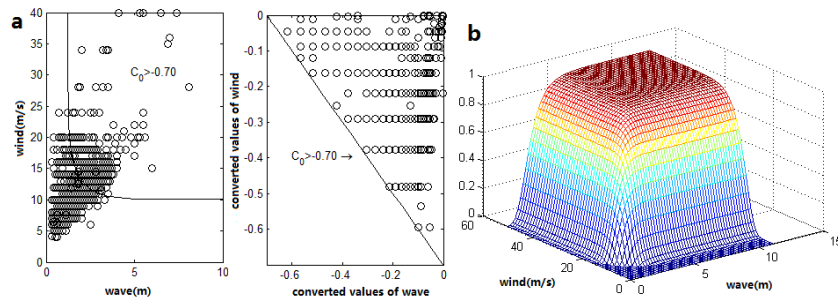


Figure 3. (a) Over-threshold values of wave height and wind speed. (b) Joint distribution of extreme values of wave height and wind speed.

2751

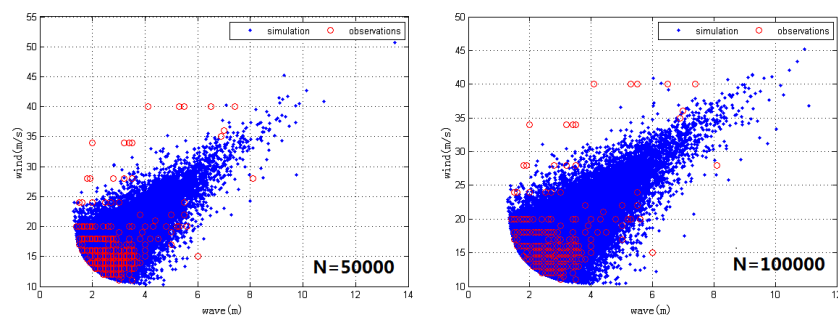


Figure 4. Over-threshold values and data of stochastic simulation.

2752

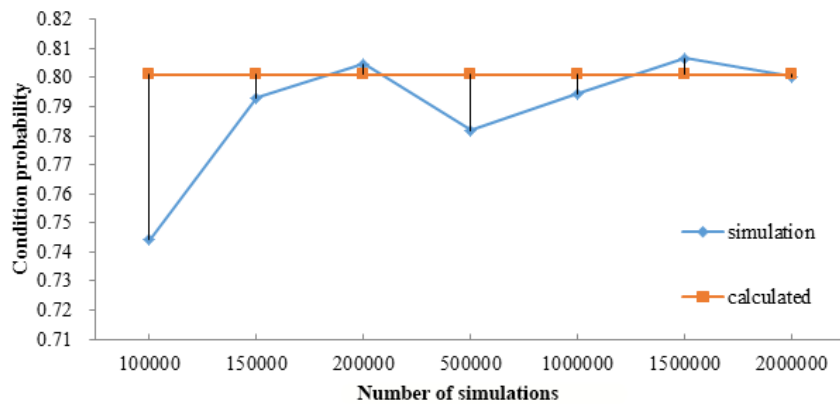


Figure 5. The change in simulation accuracy.