Estimate of the extreme wave height in the South China Sea using GPD method

H J Cao[1], **F Yi**[2] **and W B Feng**[2]

[1] College of Oceanography, Hohai University, Nanjing 210098, China

[2] College of Harbour, Coastal and offshore Engineering, Hohai University, Nanjing 210098, China

hohaijin@hhu.edu.cn

Abstract. The South China Sea (SCS) experiences severe typhoon impact every year. That also cause extreme wind waves. The estimate of extreme wave height is of great importance for human activities in the SCS. Since the generation of typhoon is still unpredictable, it is very difficult to estimate the storm waves. In this paper, we did a statistics analysis of the tropical cyclones that can affect the SCS. Based on a hindcast wave simulation, we use the generalized Pareto distribution (GPD) method to estimate the extreme wave height, and discuss the determination of threshold parameters. It is found that GPD method is a more favourable method for the estimation of extreme wave height of storm waves, which is of great significance for ocean engineering design.

1. Introduction

The South China Sea (SCS) is a typhoon-prone sea area, along with extremely large waves. Then the extreme hydrological conditions here are quite of importance for ocean engineering. The critical thing to obtain reasonable extreme wave parameters is to find an appropriate extreme probability distribution function. Currently, many researcher have studied the extreme wave parameters of the South China Sea (e.g. Ge et al. [1], Zheng et al. [2], Zhao et al. [3]) However, previous studies normally used a generalized extreme value distribution which easily cause an overestimation with limited annual extreme data. GPD (Generalized Pareto Distribution) was firstly introduced and applied in the economic field. These years, this method has been widely applied to the field of meteorological and hydrological extremum estimation. Ribereau et al. [4], Keiven et al. [5] Philip et al. [6] have made some improvements for a better application. Few studies are reported on its application in extreme wave estimate of the SCS. This paper presents an example of GPD in the SCS and compares the results to other methods.

2. Wave data

The third-generation ocean wave numerical model SWAN was used to obtain the hindcast 20-year (1995-2014) wave field in the SCS. The wind forcing is from a WRF simulation. The model adopts a horizontal resolution of 0.125°, and vertically divided from the surface to a height of 50 hPa into 36 layers with unequal distances. The modelling range is (10°S~30°N, 100°E~135°E), the north-south direction is divided into 320 grids, and the east-west direction is divided into 280 grids, covering the northwest Pacific including the SCS. The background wind fields come from the NCEP reanalysis data (http://dss.ucar.edu/dsszone/ds083.2) with a horizontal resolution of 1°×1°. The basic data of the typhoon fusion come from the tropics of the Japan Meteorological Agency's Regional Meteorological

Center, cyclone best path data set (JMA data). The wave simulation ranges from 1°N to 25°N in latitude and 105°E to 125°E in longitude, SCS as shown in Figure 1. The model uses an unstructured triangular mesh with a maximum resolution of 0.125°. In-situ and satellite altimeter data are employed to verify the simulated wave height, with the correlation coefficient above 0.9 and the root mean square error below 0.5.

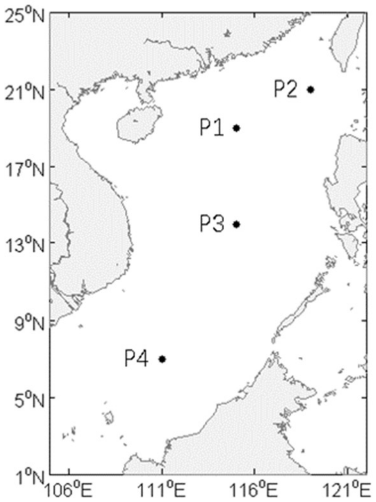

Figure 1. The simulated sea area and Pacific TC selected

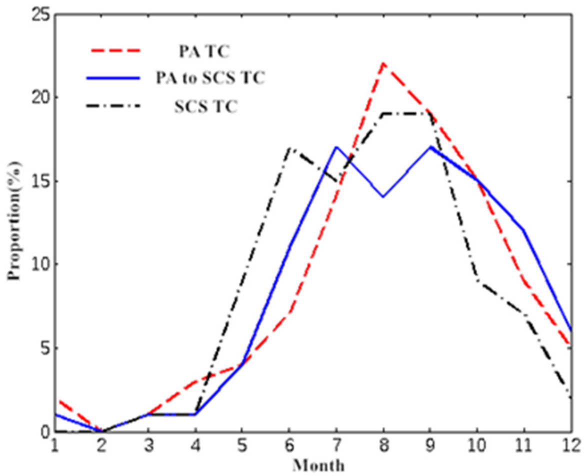

Figure 2. Monthly proportion of the SCS and study positions

3. Tropical Cyclones in the SCS

3.1. statistics of TC in SCS and Pacific Ocean

Tropical Cyclone (TC) is a huge mediation system formed on the sea surface of tropical and subtropical regions. The Tropical Cyclone Best Path Data Set (JMA data) of the Japan Meteorological Agency's Regional Meteorological Center is employed to conduct a preliminary statistical analysis of tropical cyclones that affects the Pacific Ocean and the SCS. The data include the path and maximum wind speed of tropical cyclones with wind speeds exceeding 17m/s from 1977 to 2016. A total of 1035 tropical cyclones occurred in the entire Pacific Ocean in the past 50 years, affecting 405 times on the SCS, in which 129 were generated in the SCS. The results show that most months experience the impact of TC a lot besides the period between January and April.

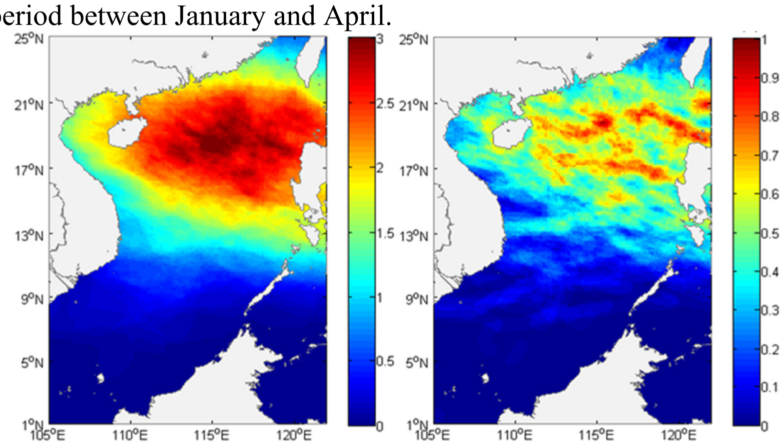

Figure 3. The spatial distribution of the number of TCs affected by the SCS, R17 (left) and RMW (right)

3.2. Spatial and temporal distribution of the SCS TC

According to different definitions, the size of TC can range from tens of kilometres to several hundred km. In order to determine the scope of influence of the TC, different definitions are used to count the number of times that all grid points are affected by the TC. The first method is R17, which is a 17 m/s wind speed radius. The second is RMW, which is the maximum wind speed radius, and only the areas with the strongest influence of TC are counted. As shown in Figure 3, (a) and (b) are the results with R17 and RMW as the influence radius of TC. Overall, the SCS is affected by TCs very frequently, and the annual average number shows a spatial distribution that decreases from north to south. Both the quantity and scope of the TC impacts present a diving line at 13°N, indicating the intensity of TC activities.

4. Estimate of Extreme wave height

Two commonly-used methods of estimating extreme wave height are introduced here. The first is called the generalized extreme value distribution (GEV). It has a distribution function as follows

$$F(x)=\begin{cases}\exp\left\{-\left[1+\gamma(\frac{x-u}{\sigma})^{-\frac{1}{\gamma}}\right]\right\} & \gamma\neq 0\\ \exp\left[-\exp(-\frac{x-u}{\sigma})\right] & \gamma=0\end{cases}\tag{1}$$

in which γ、σ、u are shape, size and position parameters respectively. If $\gamma=0$, GEV is simplified as Gumbel distribution; if $\gamma<0$, it is simplified as Tippett type II distribution; if $\gamma>0$, it is Weibull distribution. The generalized extreme value distribution (GEV) uses the maximum statistics method when selecting the sample sequence. Only one sample is taken each year, resulting in fewer samples and uneven distribution, which is not suitable in the requirements of typhoon conditions.

Another is the generalized Pareto distribution (GPD) distribution that uses part of the time series when sampling. Picking out all the data larger than the threshold value to compose the sample sequence solves the problem of the maximum statistical method to some extent. Its distribution function [7] is written as

$$F(x)=\begin{cases}1-(1+\gamma\frac{x-u}{\sigma})^{-\frac{1}{\gamma}} & \gamma\neq 0,\ x>u\\ 1-\exp(-\frac{x-u}{\sigma}) & \gamma=0,\ x>u\end{cases}\tag{2}$$

in which: γ is shape parameter, σ is size parameter, u is threshold. For a given return period year, Cheng et al. [8] introduced the extremum x_T for the return period.

$$x_T=\begin{cases}u-\frac{\sigma}{\gamma}\left[1-(\lambda T)^{\gamma}\right] & \gamma\neq 0\\ u+\sigma\ln(\lambda T) & \gamma=0\end{cases}\tag{3}$$

where λ is the annual crossover rate, which is the mean of data that are larger than the threshold per year, equal to the ratio of the total number of samples exceeding the given threshold to the total number of data. Three methods are usually used to estimate the parameters of the probability distribution function, the moment method, the least squares method, and the maximum likelihood estimation method. Compared with the other two methods, the least squares method has many advantages such as simplicity and clarity [9, 10]. Therefore in this paper, the least squares method will be used to estimate the parameters of the distribution function.

4.1. Threshold for GPD

The threshold in the GPD is critical and determines the number of sample sequences. If u is too small, the sample size is too large and the parameter estimation is too cumbersome and the estimator will become a biased estimate; if u is too large, the sample size is too small to highlight the advantages of the partial time series statistic method. At the same time, it is also the premise of correctly estimating γ and σ ensuring the reliability of the results. The threshold u selection is based on the average overrun and the stability of the shape and size parameters. In order to explain the process of threshold determination, we select P1 (115°E, 19°N) in the northern SCS as the test point to the show the steps of determining the threshold. The analysis found that when $u < 5.5$ the shape parameters and dimensional parameters are basically monotonically increasing or decreasing, but when $u > 5.5$ parameters are basically stable afterwards, it is reasonable to set the threshold value to 5.5 in terms of comprehensive average overshoot.

4.2. Comparison between GEV and GPD

In order to compare the applicability of GEV and GPD in the SCS, point P1 (115°E, 19°N), P2 (119°E, 21°N), and P3 (115°E, 14°N) and P4 (111°E, 7°N) were selected in the north, northeast, central and southern parts of the SCS respectively, the specific location of the distribution shown in Figure 1. The GEV and GPD were used to fit the cumulative frequency curves (Figure 4), and the Kolmokov Test (KS) statistics, correlation coefficient, and error were used to test the curve fitting. Finally the wave height of the test sites of 10, 20, 50 and 100-year return period by two different methods is calculated (shown in Tables 1).

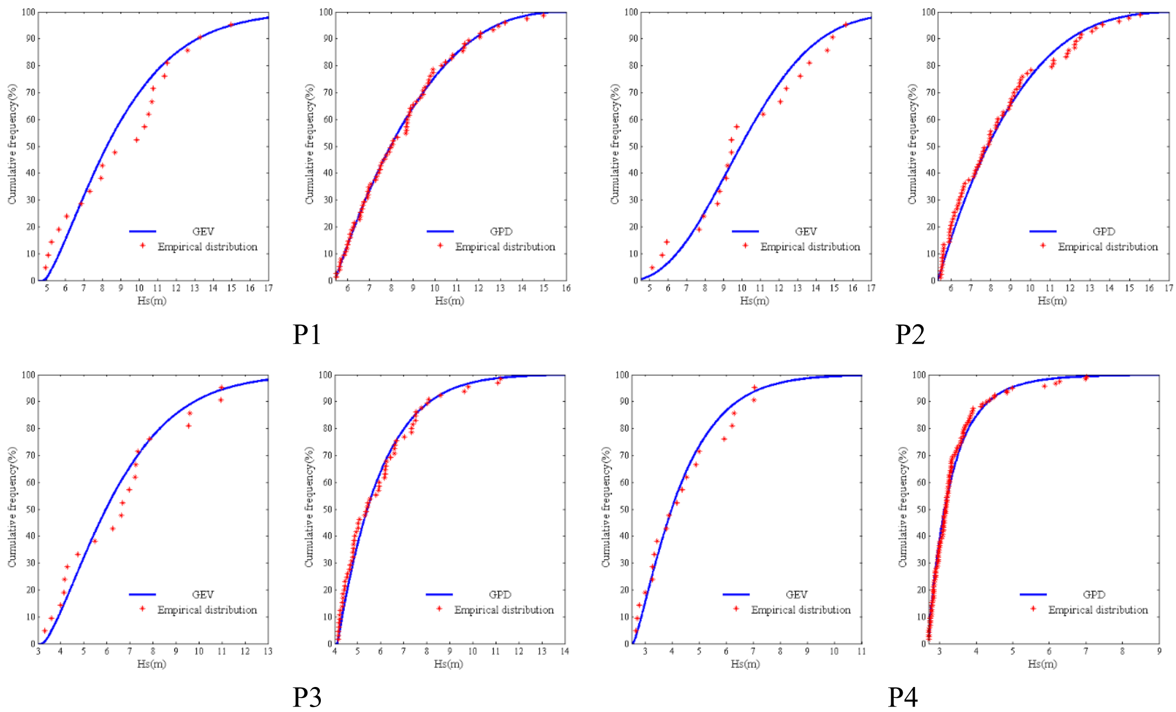

Figure 4. GEV and GPD fitting cumulative frequency curves at different positions

From the fitting test indicators, the fitting results of GPD are superior to GEV in all aspects. Especially P1 and P2, the advantages of GPD are more obvious. From the wave height of the return periods calculated from the two methods, the results are similar in the projections of P3 and P4, but at the two points P1 and P2, the projections of GEV are significantly larger than those of GPD. According to the statistics in the first section of this chapter, this is because P1 and P2 in the north and northeastern sea areas may be affected by TCs many times in one year, but some years are not even affected by TCs. GEV uses annual maximum sampling and in the method, the difference in the sampled wave height is large when the numble of samples is small, so that the wave height of the calculated return period is too large. In this case, GPD uses partial duration serial statistical method sampling to obtain more abundant sample data, which have more stability and better fitting curve. In summary, GPD has a higher degree of fitness and stability when compared with GEV when extreme wave height is calculated in the SCS affected by TC.

Table 1. Comparison of GEV and GPD projections

postion	method	10-year	20-year	50-year	100-year
P1	GPD	11.12	12.41	14.04	15.20
	GEV	13.02	14.72	16.63	17.90
P2	GPD	12.39	13.99	15.23	16.63
	GEV	14.32	15.64	17.16	18.53
P3	GPD	8.87	10.22	11.89	13.10
	GEV	9.72	11.02	12.61	13.73
P4	GPD	5.85	6.68	7.73	8.50
	GEV	6.03	6.82	7.84	8.52

4.3. Calculation of return-period wave height

Four positions in the north (P1), northeast (P2), middle (P3) and south (P4) of SCS as the studied region. And we also estimate the extreme wave height in different wave directions (Table 2). The 100-year return-period wave height in the northeastern sea area most affected by tropical cyclones can reach 16 m or more, and the return-period wave height is highest in the E direction; the 100-year return-period wave height in the northern and central seas reaches 15 m and above 13 m, the maximum 100-year return-period wave height appear in the NE direction, while in the southern waters, it is only 8 to 9 m and the N direction is the dominant direction.

Table 2. Estimate of extreme wave height in different wave directions

	direction	10-year	20-year	50-year	100-year
P1 North	all directions	11.12	12.41	14.04	15.20
	N	9.05	10.09	11.46	12.87
	NE	10.43	11.56	13.22	14.56
	E	8.95	9.86	11.02	12.77
	S	9.36	10.23	11.73	13.05
	SW	8.22	9.06	10.38	11.21
P2 Northeast	all directions	12.39	13.99	15.23	16.63
	N	10.11	11.21	12.56	13.83
	NE	11.19	12.77	14.14	15.32
	E	11.42	13.03	14.32	15.64
	SE	10.36	12.03	13.75	14.64
	S	8.96	9.89	11.07	12.65
	SW	8.46	9.23	10.59	11.42
P3 Middle	all directions	8.87	10.22	11.89	13.10
	N	8.18	8.98	9.95	11.04
	NE	8.79	9.74	11.36	12.49
	SW	7.09	7.62	8.41	9.16
	W	6.43	6.87	7.32	8.07
P4 South	all directions	5.85	6.68	7.73	8.50
	N	5.69	6.31	7.12	7.85
	NE	4.02	4.68	5.22	5.96

E	2.96	3.32	3.86	4.42
SW	3.02	3.34	3.98	4.56
W	3.75	4.17	4.89	5.36
NW	3.55	4.06	4.62	5.12

5. Conclusions
This paper first makes a preliminary statistical analysis of the tropical cyclones in the SCS and then compares the GEV and GPD methods in estimating the extreme wave height. Four positions were selected in the north, northeast, central and southern parts of the SCS. GEV and GPD were used to fit the cumulative frequency curves, and three indicators of the Kolmokov test (KS) statistics, correlation coefficient and error was used to test the fitting curve. The results show that the fitting results of GPD are better than GEV in all aspects, especially in the north and northeast where there may be more than one year of mega-value, the advantage of GPD is more obvious. The main reason is that when there are few years of GPD data, sampling using part-time serial statistics method make that sample data are more abundant and the fitting curve is better. In addition, using the above method to calculate the extreme wave height in the sub-region and sub-direction of the SCS, it is concluded that the worst-perceived tropical cyclone in the SCS has a 100-year return-period wave height of more than , and the E-direction return-period wave height is the largest; In the northern and central seas, the 100-year return-period wave height also reach 15m and 13m; while in the southern waters, it was only 8~9 m and the N direction was the dominant direction.

Acknowledgments
Authors wish to acknowledge the founding provided by the National Natural Science Foundation of China (51709092), "the Fundmental Research Funds for the Central Universities" grant (2017B03414), and the Natural Science Foundation of Jiangsu Province (BK20170865).

References
[1] Ge L, Qu Y, Zhang Z, Yan Y and Qi Y 2009 *China Offshore Oil and Gas* 3(4) 207-210 (in Chinese)
[2] Zheng C and Zhang X 2011 *Meteorology and Disaster Reduction Research* 34(1) 48-55 (in Chinese)
[3] Zhao Z, Kuang G and Zheng W 2018 *J Hydro.* A1 (in Chinese)
[4] Ribereau P, Naveau P and Guillou A 2011 *Adv. Water Resour.* 2011 34(10) 1215-1221
[5] Jonathan P and Ewans K 2007 *Ocean Eng.* 34(14–15) 1977-1994
[6] Guan F and Xie Q 1984 *Marine Science Bulletin* (4) 21-29
[7] Cheng B, Ding Y, Zhang J 2008 *Plateau Meteorology* 27(5) 1004-1009 (in Chinese)
[8] You Z J and Yin B S 2006 *China Ocean Eng.* 20(2) 225-241
[9] Chen X 2008 *Meteorology and Disaster Reduction Research* 31(4) 52-54
[10] Martins E S and Stedinger J R 2000 *Water Resour. Res.* 36(3) 737-744