Indian Journal of Geo Marine Sciences
Vol. 50 (09), September 2021, pp. 685-692

Estimation of extreme wave heights and wind speeds in the South China Sea

J C Wang*, S D Zang, Y J Ji & Y H Zhang
College of Science, China University of Petroleum, Qingdao – 266580, China
*[Email: wangjc@upc.edu.cn]

Received 29 September 2019; revised 30 July 2021

Extreme events have significant impacts on the coastal and offshore regions. In the present paper, the return values of significant wave height and wind speed are estimated based on three different models namely generalized extreme value distribution, generalized Pareto distribution and polynomial approximation to acquire a universal and trustworthy model estimate. Here, six sites in the South China Sea with diverse geographical characteristics are considered to perform the extreme value analysis and the datasets used in the experiments were derived from the ERA-Interim reanalysis. Then the advantages and shortcomings of three different methods are discussed and analyzed in detail. The polynomial approximation method is analyzed and compared with the other methods, and it is found that this new method predominantly resolves the drawbacks encountered by the other typical extreme value estimation methods and it is suitable for estimating the return values in the South China Sea.

[**Keywords**: Extreme events, Polynomial approximation, Return value, Significant wave height, Wind speed]

Introduction

Estimates of extreme events have been recognized as a significant concern for the security and economic development in both coastal and offshore regions, as changes in wave height and wind speed will impact the coastal infrastructure and also leads to beach erosion. China has a long coastline of about 18000 km with many international ports along the coast and a large majority of the population lives in the eastern coastal areas. Therefore, extreme events could have significant impacts on the safety operability of shipping and building structures. The designs of solid constructions which take the extreme wave heights and wind speeds into sufficient account are the indispensable guarantee for the normal progress of these activities. Calculating return values for wind speed and significant wave height is one of the most fundamental approaches to analyze extreme marine events. Therefore, to this end, much attention has been paid to return values for these two wave parameters. Meanwhile, the training data with high quality and sufficient volume as well as appropriate estimation method are the core for the successful estimation of return values[1].

The challenge in determining extreme values is provoked by the differences in statistical characteristics of extreme and non-extreme events, which requires being analyzed separately[2]. So far, researchers have put forward many different methods to analyze the extreme events. It has become a trend to study the extreme value prediction by using extreme value distribution since the theories of which have been put forward[3,4]. Although many different methods have been elaborated, there is no best way to fit all the datasets. Caires & Sterl[5] stated that there exists no unified probability distribution having the ability to present the statistical information of some geophysical quantity on all variable scales. The most widely used approaches for the return values of assessment are the Initial Distribution Method (IDM), the Annual Maximum Method (AMM) and the Peaks Over Threshold method (POT)[2].

Related information about IDM can be found in the recent articles for details[6,7]. In general, the IDM employs the whole of recorded data and then fits a Cumulative Distribution Function (CDF) to the available data material. For the selection of CDF functions, there are usually three typical functions to choose from:

The Gumbel distribution:

$$F(x) = \exp\left[-\exp\left(-\frac{x-A}{B}\right)\right] \quad \ldots (1)$$

Where, $F(x)$ is the CDF of x, parameters A and B are identified in the process of fitting.

The Weibull two-parameter distribution:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{B}\right)^k\right] \quad \ldots (2)$$

Where, k represents a fitting parameter.

The Weibull three-parameter distribution:

$$F(x) = 1 - \exp\left[-\left(\frac{x-A}{B}\right)^k\right] \quad \ldots (3)$$

All the parameters mentioned in above distributions are calculated by fitting the model CDF to the empirical cumulative distribution on the basis of maximum likelihood or least squares method[8]. However, IDM has an obvious drawback when estimating the return values. As available observations are generally produced by either 4 or 6 hours intervals, using all the available data resources to fit the distribution function will certainly destroy the conditions of identity and independence in distribution.

In order to overcome the dependence of the available data, the AMM method only using the maximum value in per year is proposed. A sequence of annual maximum for a certain period such as 30 years or 100 years will obey a generalized extreme value (GEV) distribution. However, the AMM-GEV method requires an adequately long dataset of 30 years or more to provide the expected estimates. This is unrealistic in some areas where the expected data are missing or incomplete. As a method to prevent too little or too long data, the POT method which is a compromise between these two above approaches has become popular. When using the POT method, all data exceeding the threshold defined in advance will be recorded for extreme value analysis. According to the extreme valuetheory[9], the distribution of data exceeding a suitable threshold defined in advance follows the Generalized Pareto Distribution (GPD). Vinoth & Young[7] pointed out that although the POT-GPD method is theoretically convincing, the primary barrier lies in the selection of threshold. Inaccurate thresholds can affect the accuracy of the estimates. The specific methods for how to choose a satisfactory threshold will be mentioned in the next part. Note that both the AMM-GEV and POT-GPD method are based on the IDM method.

Although these methods both have inevitable shortcomings, the AMM-GEV method and POT-GPD method are currently the standard and commonly-accepted methods in mainstream extreme analysis. For example, GEV and GPD are adopted in the estimation of wave height based on a 31-year wave hindcast data series in the eastern Arabian Sea[10]. Salcedo-Castro *et al.*[11] explored the spatial features of wave extremes in the South Atlantic Ocean by employing the POT-GPD method and multiple altimeter platform data resources during 1993 – 2015.

At the same time, Polnikov & Gomorev[12] proposed a new method that uses the extrapolation $F_{ap}(H)$ of a polynomial approximation designed for the shorter part of the tail of probability function to generate estimation of the return values. Developing this new polynomial approximation method is dedicated to form a novel way of state characteristics evaluation for wind speed and wave height along the Indian coast. When this newly proposed method is utilized to estimate the return values along the Indian coast, the underestimation of extreme values faced by the GEV and GPD methods can be avoided to a certain degree.

This study sets out to assess the practical feasibility and credibility of various mainstream methods (AMM-GEV method, POT-GPD method and polynomial approximation method) and models in the South China Sea domain. The significant advantages of the polynomial approximation method compared with other classical methods are presented in this article. Because the performance of extreme value prediction relies heavily on the quality and length of input data, a dataset covering 30 years ERA-interim reanalysis data was considered, which is reasonable for extreme wave height and wind speed analysis[5].

Materials and Methods

Data description

In order to conduct the extreme analysis and compare the differences between these different methods, the datasets of wind speed and significant wave height (Hs) spanning from 1988 to 2017, *i.e.* of duration of 30 years data collected from ERA-interim reanalysis were considered. ERA-Interim reanalysis data is produced by the European Center for Medium-Range Weather Forecasts (ECMWF), which generates atmospheric reanalysis from 1979 on a global scale, constantly refreshed near real time. The spatial resolution of data used in the present paper is 0.5×0.5 degrees, and the time resolution is 6 h. Six important locations in the South China Sea are selected for specific extreme analysis. The specific locations of the several sites chosen for use in the present paper are displayed in the Figure 1.

The specific latitude and longitude of these different locations are shown in Table 1. The reasons for choosing the six locations are that they are of pivotal importance in the South China Sea in terms of economic, military and political aspects, therefore the

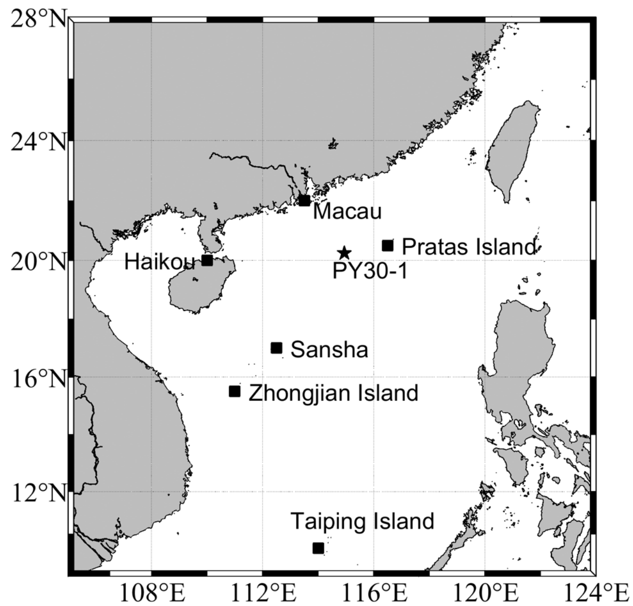

Fig. 1 — The six different locations and the wave observation radar (WOR) on board the oil platform PY30-1 in the South China Sea

Table 1 — Coordinates of the locations used in extreme analysis along the South China Sea

No.	Location	Lat./N	Lon. /E
1	Haikou city	20.0°	110.0°
2	Macau city	22.0°	113°40’
3	Taiping Island	10.0°	114°30’
4	Zhongjian Island	15°30’	111.0°
5	Sanshacity	17.0°	112°30’
6	Pratas Island	20°30’	116°30’

stability and safety of them has decisive significance for the development of the South China Sea. Therefore, it is necessary to make a more accurate estimate of the return values for wave height and wind speed along the regional water of selected locations.

Facility which can obtain measurements directly in the South China Sea, especially under extreme conditions, is desperately lacking. To compare with ERA-Interim, observations on Hs derived from the wave observation radar (WOR) on board the oil platform PY30-1 is applied, which is located at 114°56’28’’ N, 20°14’42’’ E and spans the period from January 2011 to May 2012 (See the black pentagram in the Fig. 1). The C-band WOR serves as a direct sensor that can calculate the wave height by measuring the velocity of water particles, and the measuring precision can reach 0.2 m. The WOR is operated at 5.8 GHz on the basis of range-gated pulse Doppler radar technology. The radar not only can operate in a low sweep angle mode, but also can be employed as a directional wave sensor by using multiple antennas. To objectively assess the performance of ERA-interim, three statistical measures are employed, including the Bias, Root Mean Squared Error (RMSE) and Coefficient of Correlation (R). Bias is the discrepancy between the ERA-interim and WOR measurements. RMSE is the most commonly used statistical parameter to reflect the average deviation degree between the reanalysis and the actual value. R can reflect the degree of linear association between ERA-interim and the observations. The Hs results of the ERA-interim series and WOR observation are very consistent in the Figure 2.

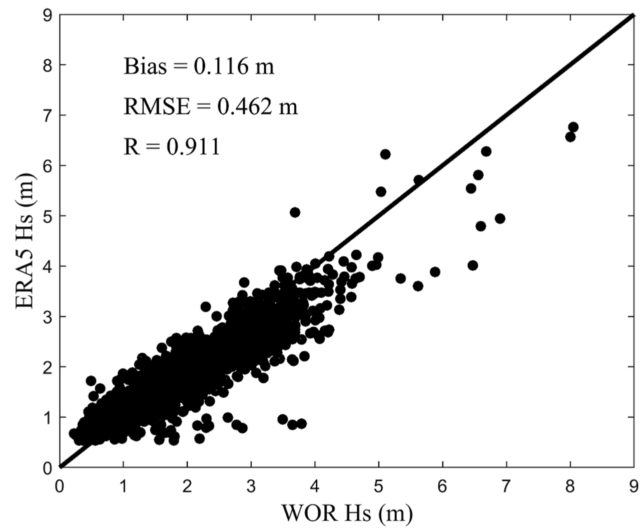

Fig. 2 — A scatter plot of the data collected from ERA-interim and WOR

Methods

In this section, two mainstream extreme value analysis methods and a polynomial approximation method are introduced in detail. There are several numerical methods available for the parameters of the estimation models (*i.e.* GEV model and GPD model) applied in the present paper. Such as Probability-Weighted Moments (PWM)[13], L-moments[14] and Maximum Likelihood Estimate (MLE)[15]. These mature methods have their own advantages and disadvantages in practical work. However, each distribution is suitable for fitting the available data using the MLE method in the following applications as the MLE method has been proved more flexible and convenient with the increasing quantity of parameters[16].

Generalized extreme value distribution model

In the light of extreme value theory, the sampled observations should be independent so as to generate a reasonable distribution, which means that successive observations should not have associations with one another[16]. The IDM method is not a good approach to conduct extreme value analysis, because all the recorded data are used and they violate the condition of independence in the distribution. Therefore, observations used in the extreme analysis invalidate the performance of the routine statistical methods. The AMM-GEV method is generally used to mitigate this problem occurred in the IDM method where the sample selected by means of AMM method could meet the requirement of independency. The maximum value of each year which follows a GEV model is considered in AMM method. However, if only the maximum value of the observed data is used for estimation each year, this will result in less available data and a longer number of years to compensate, which is a serious limitation of the AMM approach.

Here, the GEV distribution for a given random variable H has the cumulative distribution function as:

$$GEV(H;\mu,\sigma,\zeta) = \begin{cases} \exp\left(-\left(1-\zeta\left(\frac{H-\mu}{\sigma}\right)\right)^{\frac{1}{\zeta}}\right), & \text{for } \zeta \neq 0 \\ \exp\left(-\exp\left(-\frac{(H-\mu)}{\sigma}\right)\right), & \text{for } \zeta = 0 \end{cases} \quad \ldots (4)$$

Where, the μ, σ and ζ denote the location, scale and shape parameters of the GEV distribution, respectively. Note that the range of $-\infty < \mu, \zeta < \infty$ and $\sigma > 0$. Then once the number of years forecasted N_R (such as 30 years or 100 years) is given and the cumulative distribution function is obtained, the return value H_R can be calculated through the inversion of cumulative distribution function as follows:

$$H_R = F^{-1}\left(1-\frac{1}{N_R}\right) \quad \ldots (5)$$

Where, F represents the cumulative distribution function. That is,

$$H_R = \begin{cases} \mu - \frac{\sigma}{\mu}\left(1-\left(-\log\left(1-\frac{1}{N_R}\right)\right)^{\zeta}\right), & \text{for } \zeta \neq 0 \\ \mu - \sigma \ln\left(-\log\left(1-\frac{1}{N_R}\right)\right), & \text{for } \zeta = 0 \end{cases} \quad \ldots (6)$$

Generalized Pareto distribution model

The shortcomings of AMM method which have been stated is that, it needs a long-term dataset in practice and wastes a number of information which could impact the estimation of return value. An alternative method that extracts and chooses the recorded data, known as POT method, consists of the peak excesses over a certain high threshold μ of a time series. Then the peaks over threshold sampled data also satisfy the obligatory of independency and overcome the shortcomings of less available data in AMM-GEV method. The peaks over threshold sampled data can be fitted to the GPD distribution. The cumulative distribution function of GPD is computed as:

$$GPD(H;\mu,\sigma,\zeta) = \begin{cases} 1-\left(1-\zeta\left(\frac{H-\mu}{\sigma}\right)\right)^{\frac{1}{\zeta}}, & \text{for } \zeta \neq 0 \\ 1-\exp\left(-\frac{(H-\mu)}{\sigma}\right), & \text{for } \zeta = 0 \end{cases} \quad \ldots (7)$$

Where, the μ, σ and ζ are also the threshold, scale and shape parameters, respectively. Note that the range of $-\infty < \zeta < \infty, 0 < H < \infty$ and $\sigma > 0$.

The threshold μ for the GPD distribution is a particular parameter as must often it is notestimated as the other ones (*i.e.* shape and scale parameters). Selecting sufficient events to decrease the variance is the primary objective of threshold setting. Therefore, the choice of threshold μ should place emphases on the threshold stability of the GPD distribution[5]. There are several useful tools to define a reasonable threshold. Such as mean residual life plot[17], L-Moments plot and dispersion index plot. For different methods, the threshold selection may be different in practical work. The mean residual life plot would be used to select a suitable threshold in the present paper, which is a widely used method.

Hence, when the parameters of GPD distribution are determined, the 1/T yr return value of the POT-GPD method can be given as:

$$H_T = \begin{cases} \mu + \frac{\sigma}{\zeta}\left(1-(\lambda T)^{-\zeta}\right), & \text{for } \zeta \neq 0 \\ \mu + \ln(\lambda T), & \text{for } \zeta = 0 \end{cases} \quad \ldots (8)$$

Where, $\lambda = \frac{N_\mu}{N}$, with N_μ being the total number of exceeding the selected threshold μ and N representing the time span of the recorded data. Note that a significant technique, namely de-clustering process, has been extensively used in POT-GPD method, which contributes to the collection of peaks within the clusters of successively exceeding a specified threshold[1]. Specifically, cluster maxima with distance less than 48 h is regarded as attributing to the same cluster in the present paper[18].

Polynomial approximation method

Two mainstream methods of extreme value analysis have been introduced in the previous sections. Now a new polynomial approximation method is discussed in this section.

In order to conduct the extreme value analysis, the polynomial approximation method includes the construction of an analytical approximation $F_{ap}(H)$, aiming at its extrapolation beyond the maximum observation H_M[refs. 1,16]. First, the probability provision function $F(H)$ based on the histogram (the discrete interval of histogram is ΔH) of the considered time series of Hneeds to be calculated. Then the domain of $F(H)$ used for approximation $F_{ap}(H)$ is determined by the condition

$$H_1 \leq H \leq H_h \leq H_M \quad \dots (9)$$

Where, H_1 and H_h represent the lower and upper margins of the domain of $F(H)$ used for forming an approximate estimation, respectively. In general, due to the fact that fewer data exist at the tail of $F(H)$, the approximation $F_{ap}(H)$ should be constructed under the logarithmic coordinates frame work for the purpose of providing importance to the tail values. The approximation $F_{ap}(H)$ in the form of a polynomial of the degree, n is considered; the value of degree can vary and in general $n = 2$ or 3 are selected. The changeable n allows for a much more accurate approximation $F_{ap}(H, n)$ in comparison with the fixed distributions (such as AMM-GEV method and POT-GPD method).

The statistical distribution with the provision function is given as follows:

$$F_{ap}(H) = \exp\left[\sum_{i=0}^{i=n} a_k H^k\right] \quad \dots (10)$$

Hence, when the function $F_{ap}(H)$ is obtained, the return value can be deduced from the following equation

$$F(H_R) = \Delta t/(8760 \cdot N_R), \quad \dots (11)$$

Where, Δt stands for the time interval of recorded data. $\Delta t = 6$ hours is used in the present paper.

However, the polynomial approximation needs to regulate the credibility of its extrapolation, when the order $n > 1$ could cause twists and extreme. Fortunately, this problem can be resolved or avoided by varying the order of polynomial n and the parameters N_S, N_T. The number of points (N_M) considered in the histogram is $H_M/\Delta H$ and N_S and are defined as

$$N_S = (H_M - H_h)/\Delta H \quad \dots (12)$$

The number of points (N_T) used for constructing approximation $F_{ap}(H)$ is defined as

$$N_T = (H_h - H_1)/\Delta H + 1 \quad \dots (13)$$

It is worth noting that the optimal selection of N_S, N_T and n, allows optimizing the estimate of return values. This is a significant advantage that other known methods do not have, which would have significant impact on the estimation of the return values.

Results and Discussion

The impact of estimated parameters in fitting the sampled Hs data for the Haikou area to the GEV distribution is presented in Figure 3(a), where the level of fitting of the empirical CDF with the GEV

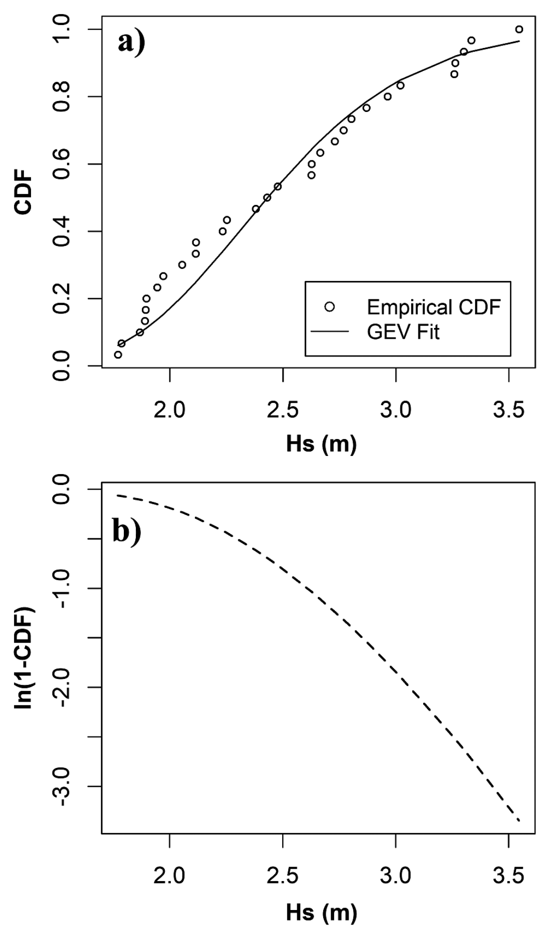

Fig. 3 — a) Variation of GEV model fit in normal coordinates for Hs data at Haikou, and b) Variation of tail for GEV model fitting in logarithmic coordinates for Hs data at Haikou

distribution can be displayed. It can be seen that the fitting effect is better, the discrepancy between the normal coordinates in their fitting and empirical CDF is insignificant. Next, Figure 3(b) reveals the variation in tail estimates of empirical CDF for GEV model at the logarithmic scale.

Quantile–quantile (QQ) plots are usually considered as a powerful tool to identify the estimated parameters for the GPD model. In Figure 4(a), the QQ plot for return values of Hs in Zhongjian Island is shown. It can be observed that the fitting effect of the empirical CDF with the GPD distribution is good. Then the plot about return level is presented in Figure 4(b). The return level plot with 95 % confidence intervals for the sampled Hsdata at Zhongjian Island reveals the extreme wave height for corresponding return periods such as 30, 50 and 100 years.

From the above figures, we can find that a better estimate of the parameters can be obtained by using the AMM-GEV and POT-GPD method. These two methods have been widely used in extreme value analysis, so much importance has been paid to the polynomial approximation method in the present paper. The excellent feasibility of the polynomial approximation method to the real performance of the tails for provision functions $F(H)$(ref. 16) can be observed. Taking the Hs of the ERA-Interim data in Haikou location as an example, the optimized parameters obtained are $N_S = 1, N_T = 6$, and $n = 3$ would bring about the optimal return value as shown in Figure 5. Obviously, Figure 5 shows that the polynomial approximation method is applicable to the actual behavior of the tails for the provision functions of Hs. Hence, one can see that the polynomial approximation present a better method to fit the probability provision function $F(H)$, and extrapolation is also more reliable in practical work.

The eventual estimations of the return value for 30 and 100 years obtained by the three approaches are presented in Table 2. At the same time, the observed maximum values for 30-years of observations at these locations are also listed in Table 2. Variation and deviation of these estimated outcomes from the maximum measurements will validate the behavior of the estimation models used in experiments statistically.

From Table 2, when the polynomial approximation method is used to estimate the return values, it can be known that the 30-years return values for both Hs and wind speed have higher consistency with the 30-years

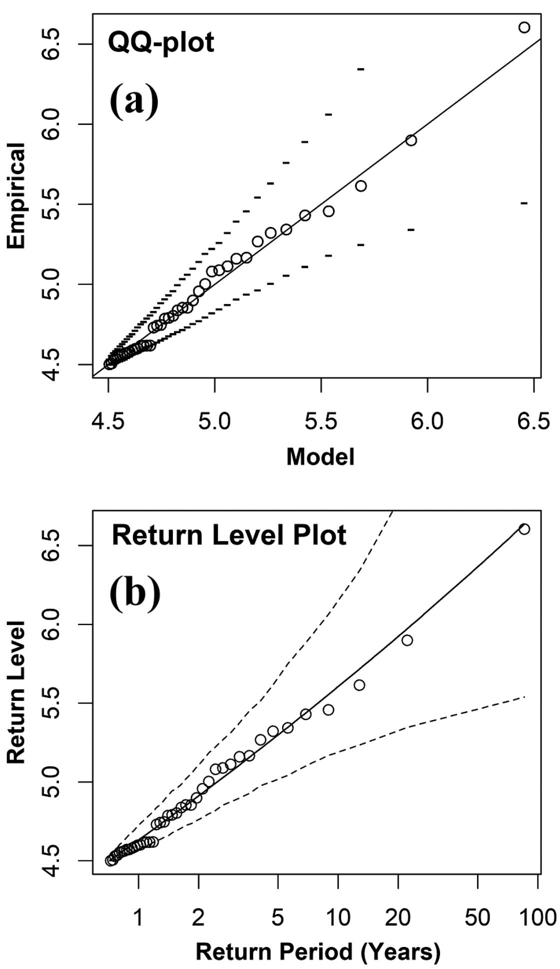

Fig. 4 — a) Quantile–quantile plot of GPD model for Hs data at Zhongjian Island, and b) Return level plot with 95 % confidence limits for Hs data at Zhongjian Island

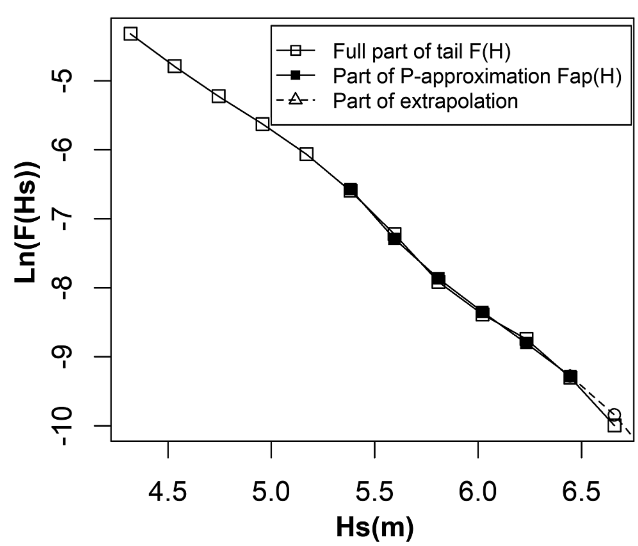

Fig. 5 — Illustration of polynomial approximation method applied for Hs series at Haikou with parameters: $N_S = 1, N_T = 6$ and $n = 3$

Table 2 — The return values for wind speed (Wi, m/s) and Hs(m)

Location	Variable	Time interval (h)	Observed maximum value	30 years			100 years		
				P-app	GEV	GPD	P-app	GEV	GPD
Haikou	Hs	6	3.5	3.5	3.6	3.4	3.7	3.9	3.6
	Wi	6	17.3	17.5	16.8	16.6	18.2	18.2	17.6
Macau	Hs	6	4.3	4.3	4.1	4.1	4.4	4.9	4.7
	Wi	6	19.7	20.9	19.9	19.3	22.0	23.5	21.7
Zhongjian Island	Hs	6	6.6	6.3	6.1	6.1	6.6	6.5	6.7
	Wi	6	22.1	22.2	22.0	21.4	23.0	24.5	22.9
Taiping Island	Hs	6	6.2	6.3	6.0	5.9	6.8	6.9	6.6
	Wi	6	23.5	21.0	21.0	20.4	22.0	25.3	22.0
Sansha	Hs	6	6.7	6.9	6.5	6.4	7.2	7.1	6.8
	Wi	6	22.3	22.5	22.6	21.9	23.1	24.5	22.3
Pratas Island	Hs	6	8.3	8.4	8.0	7.9	8.8	9.4	9.3
	Wi	6	24.1	24.9	24.6	23.9	25.4	26.8	25.3

maximum observations compared with the other estimation methods. The 30-years return values of the Hs and wind speed generated by GEV and GPD are almost lower than the 30-years maximum measurements, and even the 100-years return value is partially smaller than the 30-years observed maximum value. The underestimation caused by using the AMM-GEV and POT-GPD method is been widely realized. Samayam *et al.*[19] also stated and discussed the disadvantage of underestimation for using the AMM-GEV and POT-GPD methods. This is not a good feature because they ignored the tails of provision functions, accepted in GEV and GPD methods. Therefore, this behavior results in underestimation of the return values in practice. In addition, the 30-years return values produced by the polynomial approximation method show better return value estimation and significantly overcome the shortcomings of the underestimation of 30-years return value.

It is also found that the return values generated by the POT-GPD method were slightly less than those from the AMM-GEV method. Because the AMM-GEV method only considers the highest value in the year, which could lead to higher return values than those of the POT-GPD method. Generally, the AMM-GEV method can be conducted with ease, however, the POT-GPD method is a more desirable method, because it has a superior performance in the locations with multiple storm events occurred in a single year.

Conclusion

To better achieve the extreme value analysis statistically, the ERA-Interim data were used to estimate the return values in the present paper, because they are able to provide constantly long and regular time series on a global ocean scale.

Extreme events have an important impact on the coastal and offshore regions. This study focuses on the estimation of the return values for Hs and wind speed and desires to find a better method to use in the South China Sea. Three different methods have been considered to obtain the return values: the AMM-GEV method, the POT-GPD method and the polynomial approximation method. Then their performances have been compared in different locations.

There are advantages and shortcomings of these methods. From the results shown, the major limitation of the AMM-GEV method and the POT-GPD method is that they tend to underestimate the return values in practice. The disadvantage of the polynomial approximation method based on extrapolation of the provision functions tail is that, the validity of its extrapolation needs to be taken into account, as the order higher than 1 could cause twists and extreme. Meanwhile, the return values from polynomial approximation overcome the underestimating return values that emerged in the other method.

However, the polynomial approximation method would arrive at the optimum return value, when the optimized parameters N_S, N_T and n are obtained. Therefore, how to choose the optimal parameters is a tricky problem. Compared with the other two methods, however, the polynomial approximation method remains a good method for estimating the return values. From the data used in the present paper and the results of the experiments, the polynomial approximation is a better estimation method which can be adopted in the South China Sea.

Acknowledgements

The authors appreciate the National Key Research and Development Project of China (No. 2018YFC1406200), the Shandong Provincial Natural Science Foundation (No. ZR2020MD060) and the Fundamental Research Funds for the Central Universities (No.19CX05003A-5).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization and software: JCW; Methodology: JCW & SDZ; Validation and writing-review & editing: YJJ & YHZ; Formal analysis: JCW & YJJ; Writing-Original Draft: SDZ.

References

1 Polnikov V G, Sannasiraj S A, Satish S, Pogarskii F A & Sundar V, Estimation of extreme wind speeds and wave heights along the regional waters of India, *Ocean Eng*, 146 (2017) 170-177.

2 Young I R, Vinoth J, Zieger S & Babanin A V, Investigation of trends in extreme value wave height and wind speed, *J Geophys Res-Oceans*, 117 (2012) p. C00J06.

3 Godoi V A, Bryan K R, Stephens S A & Gorman R M, Extreme waves in New Zealand waters, *Ocean Model*, 117 (2017) 97-110.

4 Dentale F, Furcolo P, Pugliese C E, Reale F, Contestabile P, *et al*., Extreme Wave Analysis by Integrating Model and Wave Buoy Data, *Water*, 10 (2018) p. 373.

5 Caires S & Sterl A, 100-Year Return Value Estimates for Ocean Wind Speed and Significant Wave Height from the ERA-40 Data, *J Climate*, 18 (2005) 1032-1048.

6 Alves J H & Young I R, On estimating extreme wave heights using combined Geosat, Topex/Poseidon and ERS-1 altimeter data, *Appl Ocean Res*, 25 (2003) 167-186.

7 Vinoth J & Young I R, Global Estimates of Extreme Wind Speed and Wave Height, *J Climate*, 24 (2011) 1647-1665.

8 Holthuijsen L, *Waves in Oceanic and Coastal Waters*, (Cambridge Univ, New York), 2007, pp. 387.

9 Coles S, Threshold Models, In: *An Introduction to Statistical Modeling of Extreme Values*, (Springer), 2001, pp. 74-91.

10 Teena N V, Sanil Kumar V, Sudheesh K & Sajeev R, Statistical analysis on extreme wave height, *Nat Hazards*, 64 (2012) 223-236.

11 Salcedo-Castro J, da Silva N P, de Camargo R, Marone E & Sepúlveda H, Estimation of extreme wave heights return period from short-term interpolation of multi-mission satellite data: application to the South Atlantic, *Ocean Sci*, 14 (4) (2018) 1-11.

12 Polnikov V G & Gomorev I A, On estimating the return values for time series of wind speed and wind wave height, *Russ Meteorol Hydro*, 40 (2015) 820-827.

13 Hosking J R & Wallis J R, Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, 29 (1987) 339-349.

14 Sankaran P & Kumar D, Pareto-weibull quantile function, *J Appl Probab*, 13 (2018) 81-95.

15 Smith R L, Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, 72 (1985) 67-90.

16 Samayam S, Laface V, Annamalaisamy S, Arena F, Vallam S, *et al*., Assessment of reliability of extreme wave height prediction models, *Nat Hazard Earth Sys*, 17 (2017) 409-421.

17 Davison A C & Smith R L, Models for Exceedances over High Thresholds, *J R Stat Soc*, 52 (1990) 393-442.

18 Caires S, Extreme Value Analysis: Wave Data, JCOMM Technical Report No. 57, In: *Technical Report*, (World Meteorological Organization), 2011, pp. 1-32.

19 Samayam S, Laface V, Annamalaisamy S, Arena F & Sundar V, Statistical methodology for reliability analysis of wave height return value estimation models, In: *Pianc-Copendec*, (Rio de Janerio, Brazil), 2016, pp. 1-12.