IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 1, 2019, accepted March 11, 2019, date of publication March 22, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906935

Study on Threshold Selection Methods in Calculation of Ocean Environmental Design Parameters

GUILIN LIU[1], ZHIKANG GAO[1], BAIYU CHEN[2], HANLIANG FU[3], SONG JIANG[3], LIPING WANG[4], AND YI KOU[5]

[1]College of Engineering, Ocean University of China, Qingdao 266100, China
[2]College of Engineering, University of California Berkeley, Berkeley, CA 94720, USA
[3]School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China
[4]School of Mathematical Sciences, Ocean University of China, Qingdao 266100, China
[5]Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA

Corresponding author: Baiyu Chen (baiyu@berkeley.edu)

This work was supported by the NSFC–Shandong Joint Fund through the study on the multidimensional discrete variables compound extremum distribution model and its engineering application under Grant E091001 and Grant E090904 and the study on the disaster-causing mechanism and disaster prevention countermeasures of multi-source storm surges under Grant U1706226, and in part by the Program of Promotion Plan for Postgraduates' Educational Quality (paying more attention to the study on the cultivation mode of mathematical modeling for engineering postgraduates) under Grant HDJG18007.

ABSTRACT In marine engineering design, the threshold selection is a basic and very important part for the analysis of measured data and subsequent acquisition of sample data for probability analysis. In this paper, threshold selection methods are deeply studied from the two perspectives of time domain and frequency domain, and the obtained calculation results with different thresholds are provided. First, the theory of mean residual life plot is used to quantify the threshold value of the measured tide levels of Hangzhou Bay. For the 100-year return periods design values calculated in the Gumbel distribution, the results under the threshold $u = 3.9$ m shows an increasing of 1.70% over those under the threshold $u = 3.75$ m; and the results under the threshold $u = 3.75$ m shows an increase of 0.97% over those under the threshold $u = 3.5$ m. The quantitative selection method of a threshold is simple and practical, and eliminates the tedious process of multiple trials based on empirical threshold determination. Besides, for the first time, the coherence spectrum threshold is used to explore the significant correlations of different tide level sequences in the frequency domain. Based on the threshold selection method used in the WOSA method-based coherence spectrum significance test, the coherence spectrum thresholds of different tide level in Hangzhou Bay are provided to effectively avoid the false identification of false peaks. It is suggested by the calculation results that, at the overlapping ratio of 50%, the segment number of three and the segment length of 16, and the spectral analysis shows that the extreme tide level of Hangzhou Bay has a vibration period of 21.3 years.

INDEX TERMS Coherence spectrum, return level, threshold.

I. INTRODUCTION

In marine engineering design, reasonably and accurately determining the design parameters of hydrological elements is of great significance for disaster prevention and reduction. Currently, based on Gumbel, Weibull and other probability distribution models, observed sample sequences have been used to determine the distribution parameters and predict the return levels, and the return level corresponding to a certain return period is obtained as the design criterion [1], [2]. Every finally derived return level relies on both the selected probability model for analysis of the data's statistical properties and the required original measured data for calculation of the undetermined parameters in the probability distribution function [3], [4]. How to utilize the measured data in a full, more reasonable and more effective way is the premise for a probability distribution model-based accurate estimation of a return level. Although it has only been explored in a

The associate editor coordinating the review of this manuscript and approving it for publication was Xiang Huang.

few studies, it is of high significance and necessity. After all, how to select the sample is the premise for probability analysis.

Currently, common sampling methods for calculation of marine environmental elements mainly include the annual maxima method, the process sampling method and the peak over threshold (POT) method. The annual maxima method is simple and convenient, but it requires long-term data to guarantee the reasonableness of an estimated value, since only a maximum value for a year is selected for a sample sequence [5]–[7]. The process sampling method involves the sample sequences of synchronous marine environmental elements at the occurrences of typhoon and hurricane. The method provides the design parameters to most likely cause risks under extreme marine conditions according to the probability model-based calculation of a return level, and the design parameters are of high engineering significance. However, due to the limitation of theoretical models and the difficulty in obtaining synchronous data, the used data in calculation is not really synchronous data. Therefore, the process sampling method remains in a theoretical improvement stage [8]–[14]. To maximize the quantity of used information in measured data, the threshold-value method has been proposed and developed. The peak over threshold (POT) method is a sampling method for probability analysis with the data reaching or exceeding a certain relatively large value as the sample [15]–[21]. In view of the small number of observation station in China and the insufficient length of the observation sequences, POT method has a high practical value. Mazas [22] believed that the data utilization is significantly better with the POT method than the annual extremum sampling method. Luo [23] has studied the threshold value method-based model from unidimensional and multidimensional aspects, and proposed the advantages of a super-threshold model in analyzing marine environmental parameters, and further explored how to select an optimal threshold. However, the application of the peak over threshold method has been explored in the above mentioned researches merely in the time domain.

The key of the peak over threshold method is how to select a “ threshold ”. Currently, most users determine a threshold empirically, showing high randomness. Different thresholds lead to different samples and affect the estimation of distribution parameters, the selection of a probability model and the determination of a return level, and there will be several design criteria whose advantages and disadvantages can barely be distinguished [24]–[29]. When Wang and Liu [30] analyzed the wave data of a certain station in the South China Sea, the samples were selected with 2.0m, 4.0 m, 5.0m and 6.0 m as the thresholds to calculate the 100-year return period design wave heights. The minimum and the maximum predicted values were 8.3 m and 9.8 m respectively, and the variation is 1.5 m; it is suggested that the impacts of different thresholds on the results were relatively significant. If the threshold is excessively high, the size of the over-threshold data is small, the conclusion is not stable enough, and some information data is wasted; if the threshold is excessively low, the approach fails to meet the theoretical requirements of the extremum value model. In order to minimize the subjectivity of the threshold selection, Godes [31] recommended the application of the mean residual life plot for determination of a threshold value according to the judge of starting and ending positions of the straight line segment in the mean residual life plot. When the impacts of astronomical tides are analyzed, Yangtze River floods and storm surge elevations on the flood control of a certain city, Liu et al. [32] used the over-threshold method for sampling, utilized the mean residual life plot method to obtain the three-component thresholds of the astronomical tide, Yangtze River flood and storm surge elevation, and obtained favorable effects and the literature work. It can be seen that the application of the mean residual life plot method for determination of a threshold can avoid the uncertainty of empirical threshold determination, effectively improve the practical value of the threshold method, and effectively increase the information amount of the data for an available little data to some extent.

Currently, researches considering thresholds for the calculation of return levels are mostly limited to measured data. That is, discussion has only been carried out in the time domain. In fact, hydrological time series can also be analyzed from the perspective of the frequency domain to further mine data information. Due to the overlapping randomness of astronomical tide and storm surge as well as the periodic variation of gravitational tide and weather system, the time series of the annual extreme tide level are characterized by both randomness and periodicity, so the spectral analysis method can be used to study tide sequences. Commonly, the used spectral analysis methods are power spectrum and coherence spectrum methods [33]–[38]. How to accurately locate the storm surge intensity using quantitative analysis method based on the existing data, so as to reduce the loss caused by the storm surge disaster, is a key technical link for ocean disaster rescuing and evaluating. Spectrum can be used to analyse the energetic changes within the tide level. Comparing with other spectral analysis, by use of the method of coherence spectrum, the relationships between different tide level samples in the frequency domain can be studied. The method of power spectrum can be used to show the powers of vibrations at different frequencies, i.e. the volumes of energy contributions. In this way, the main vibration and its corresponding period can be determined from the spectral peak in the spectral curve. Currently, the relationships between different tide level samples in the frequency domain have been rarely studied. In-depth exploration on the relationship between the two can explain marine phenomena in a more accurate way and analyze the activity characteristics of extreme tidal levels. The coherence spectrum method is a powerful tool for analyzing the correlation of two sequences in the frequency domain. When the coherence spectrum value is greater than the threshold of the significance test, the two sequences are believed to be significantly correlated at this frequency [39]–[42]. In practical applications, the core problem is how to determine a proper threshold. Empirical determination of a threshold is usually insufficient in its accuracy, and is inclined to cause the false identification of false peaks and the neglect of true peaks. Gallet and Julien [43] studied the WOSA method-based quantitative determination of a coherence spectrum threshold, and has achieved favorable results and applied the achievements in multiple fields.

This paper attempts to study a threshold selection method from both time domain and frequency domain, for the first time. A threshold selection method based on the mean residual life plot theory is introduced to determine the threshold of hydrological sequence quantitatively. This method is used to determine the threshold value of the measured tidal level data in Hangzhou Bay, and the sampling uncertainty existing in the calculation of the designed water level is studied and compared. Furthermore, in this paper, the characteristics of tide level activity are analyzed, for the first time, the coherence spectrum; the threshold selection method used in the WOSA method-based coherence spectrum significance test is adopted. This method is used to calculate the coherence spectrum threshold of different tide sequences in Hangzhou Bay, the false peak identification can be effectively avoided, so as to better analyze the correlation of different tide sequences and further analyze the activity characteristics of extreme tide levels with the power spectrum.

II. THEORY INTRODUCTION

A. THE MEAN RESIDUAL LIFE PLOT

In the extreme value prediction of marine environmental conditions, when the continuous observation data is relatively short, the sample data can be expanded by the POT method, so as to make full use of the expensive measured data. In order to overcome artificiality of the threshold selection, scholars have been studying how to select threshold objectively and quantitatively through mathematical methods. Currently, the main method to determine the threshold is the judge of starting and ending positions of the straight line segment in the mean residual life plot. The method has been further improved in theory, according to its application and research in recent years. The method is presented below.

If the expression of generalized extreme value distribution (GEV) is simplified to $G(x)$, Eq.(1) can be obtained as,

$$W(x) = 1 + \ln G(x) \tag{1}$$

where, $W(x) \in (0, 1)$. The expression of generalized Pareto distribution (GPD) can be written as,

$$W(x) = 1 - \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi} \tag{2}$$

The generalized extreme value distribution (GEV) has the characteristics of maximum stability, that is, the maximum value distribution of the generalized extreme value distribution is still the same type of GEV. If the distribution of x, satisfying $x > u$, can be expressed by GPD asymptotically, the Eq.(3) can be obtained from the properties of the generalized Pareto distribution.

$$E(X - u|X > u) == \frac{\sigma - \xi\mu}{1-\xi} + \frac{\xi}{1-\xi}u \tag{3}$$

where, u is the threshold, $E(x - u|x > u)$ is the expected value of the data over threshold. Because the expected value is approximately the average value, the scatter distribution of the threshold u and the average values of the observed value over threshold $(x - u|x > u)$, can be made based on Eq.(3). When the shape parameter ξ is stabilized, the plot is approximately a straight line. Namely, when the horizontal axis represents threshold u and the vertical axis represents the average value, the slope and intercept of the line are $\xi/(1-\xi)$ and $(\sigma - \xi u)/(1-\xi)$, respectively. The trend lines' flat parts with less volatility can be obtained in the plot and its corresponding x-coordinate interval can be taken as the selectable range of the threshold. It is worth noting that if the threshold value is too small, it will violate the theoretic demand of the extremum model and lead to a large deviation.

In the mean residual life plot, it is highly subjective to judge a line as "straight" or "curved". That is, there is still some arbitrariness in the selection of threshold. In order to obtain the representative threshold, the "Criterion for stability of parameter estimation" can be combined to further determine the threshold accurately [23]. The main idea of this method is to seek the stability of parameter estimation by fitting generalized Pareto distribution in a certain threshold range. If the data over the initial threshold u_0 obeys the generalized Pareto distribution, then the the data over the larger threshold u also obeys GPD and has the same shape parameter.

B. COHERENCE SPECTRUM AND ITS THRESHOLD ESTIMATION METHOD

The theory of coherence spectrum analysis has been introduced in many literatures. Coherence spectrum (MSC) is a classical tool to identify the correlation in the frequency domain between two time series in engineering, oceanography and other fields. It commonly estimated by the Welch method (WOSA), and the method is introduced as follows.

$x(t)$ and $y(t)$ are two time series of length N, and each series is decomposed into n_s segments of length L, shifted by a fixed delay D. The corresponding mathematical expressions can be obtained as,

$$\begin{aligned} x_i(t) &= x[(i-1)D + t] \\ y_i(t) &= y[(i-1)D + t] \end{aligned} \tag{4}$$

where, $i = 1, 2, \ldots, n_s, t = 1, 2, \ldots, L, p = 100(1-D/L)\%$ corresponds to the overlapping ratio.

Each segment multiplied by a window function $w(t)$, and the spectra are calculated using Fourier transform. Coherence spectrum can be expressed as,

$$\hat{\gamma}_{xy}^2(f) = \frac{\left|\sum_{i=1}^{n_s} F[w(t)x_i(t)]^* F[w(t)y_i(t)]\right|^2}{\sum_{i=1}^{n_s} |F[w(t)x_i(t)]|^2 \sum_{i=1}^{n_s} |F[w(t)y_i(t)]|^2} \tag{5}$$

where F represents Fourier transform, * represents conjugation.

The estimated values of coherence spectrum is on the interval [0, 1]. The value is close to 1, the correlation between them is stronger in the corresponding frequency component. In the fact, the estimated values are usually larger than zero, but it does not mean that series are significantly correlated at all frequency points. Thus, it is necessary to seek a threshold. If the estimated value exceeds the threshold, significant correlation can be considered at these frequency points.

Gallet and Julien [43] gave the calculation formula for the threshold at the significance level α, which was based on WOSA method:

$$c = 1 - \alpha^{1/(n_{c1}-1)} \tag{6}$$

where,

$$n_{c1} = \frac{n_s}{c_w(D)},$$

$$c_w(D) = 1 + 2\sum_{j=1}^{n_s} \frac{n_s - j}{n_s}\rho^2(jD),$$

$$\rho(M) = \frac{\sum_{t=1}^{L-M} w(t)w(t+M)}{\sum_{t=1}^{L} w^2(t)},$$

$w(t)$ represents a Hanning window.

The method considers the influence of overlapping ratio and window function. When the coherence spectrum value is greater than the threshold of the significance test, the two sequences are believed to be significantly correlated at this frequency point. The frequency intervals of significant correlation between the two sequences can be obtained by using the correlation analysis of the coherence spectrum.

III. EXAMPLES OF THE APPLICATIONS IN OCEAN ENGINEERING

A. UNCERTAINTY ANALYSIS OF THE THRESHOLD SELECTION

Taking the tide level data measured between 1981 to 2006 at the Zhapu Hydrologic station (30° 42’N, 121° 01’E) in the Hangzhou Bay as an example, this section discussed how data samples with different thresholds affects the results of probability analysis for design parameters. Concrete analysis is as follows.

The method based on the mean residual life plot theory is firstly used to determine the threshold. Considering that if the threshold value is too small, it will violate the theoretic demand of the extremum model. And thus the region 2-3 in Fig. 1 is not suitable. In the mean residual life plot (see Fig. 1), the trend lines’ flat parts have less volatility and its corresponding x-coordinate interval (3.2, 4.0) can be regarded as the selectable range of the threshold. In order to select the most representative threshold, 150 values are uniformly selected as thresholds within the range of threshold estimation interval [3.2, 4.0], and a series of parameter values are obtained by maximum likelihood estimation method. If the parameter estimation values of the GPD are stable near the selected threshold, then the threshold can be seen as basically reasonable.

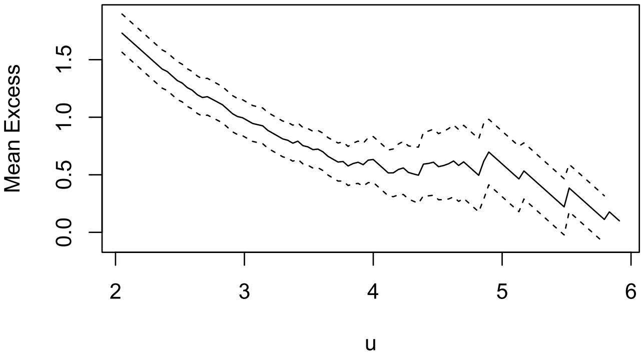


FIGURE 1. **The mean residual life plot.**

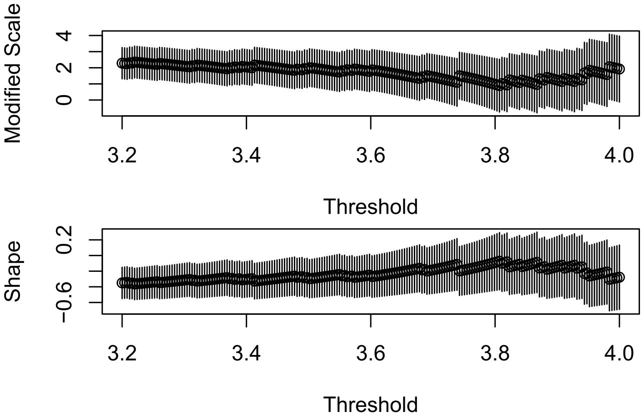


FIGURE 2. **Parameter estimation of the GPD under different thresholds.**

Fig. 2 shows that the parameter estimated values of the GPD model are approximately stable when the threshold u is less than 3.74; the parameter estimated values fluctuate apparently when u is greater than 3.74. To ensure the independency of data, the selected threshold should be as large as possible, thus, $u = 3.74$ is selected as the threshold. In order to further study the influence of different thresholds on the calculated results, $u = 3.5$ and $u = 3.9$ are taken as thresholds, respectively. Therefore, the observed tide level data is used, the whole data sample is segmented into 4 groups based on different standard. Group A26 is a dataset with annual maximal value, B26 is a dataset where the values are above the threshold $u = 3.74$m, C26 and D26 are datasets where the values are above $u = 3.5$m and $u = 3.9$m, respectively. The time span of the above four groups is all from1981 to 2006.

The scatter plot of measured tide level between 1981 and 2006 is shown in Fig. 3. The circles in the figure represent all the observation points, and the solid circles represent the annual maximal tide level. The horizontal line represents the threshold 3.74m, and the data points beyond the threshold is above the line. It can be seen that the POT method can

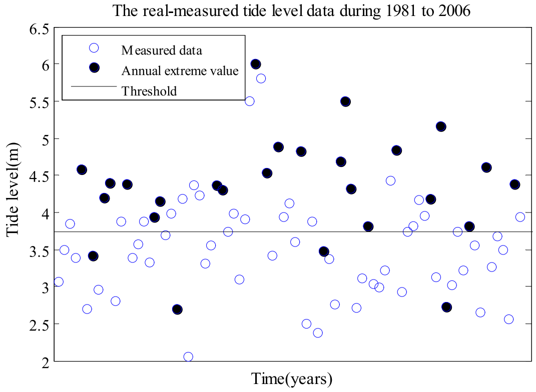


FIGURE 3. Scatter plot of tide level.

make full use of precious measured tide level data, which is especially significant under the condition that data are relatively absence.

Fig. 4 to Fig. 7 show diagnostic tests including charts of probability, quantile, and return level and a density histogram. The circles represent data points, and the straight lines or curves are theoretical curves in different coordinate systems. The logarithm or histogram of measured data are taken respectively, and the corresponding theoretical curves are drawn in the coordinate plane. The results of the corresponding diagnostic tests show that observation data comply with the theoretical extreme value model, and thus, can be used as an analysis sample of the extreme value distribution.

As specified in the ''Technical Specifications for Harbor Design'', the Gumbel distribution method is used to make possibility prediction for probable extreme tide level. This method is in accord with the characteristics of typhoon and coastal in China, and easy to be used. But this method is not very accurate in predicting the return periods of probable extreme tide level in all regions. The Weibull distribution is a common frequency distribution curve in hydrological analysis, which is a great addition to the Gumbel. Therefore, the two distributions are used for probability analysis of above four datasets and the predicted results are compared.

TABLE 1. K-S test of the Gumbel distribution.

Data Group	A26	B26	C26	D26
Test value D_n	0.1464	0.1343	0.1028	0.1290
Critical value $D_0(0.05)$	0.2591	0.2006	0.1866	0.2275
Comparison	$D_n<D_0$	$D_n<D_0$	$D_n<D_0$	$D_n<D_0$
Testing results	Accept	Accept	Accept	Accept

Table 1 and 2 describe the results of the K-S tests for the fittings of Gumbel distribution and Weibull distribution on the tide level series A26, B26, C26 and D26, respectively. The 95% confidence intervals of estimation of Gumbel distribution and Weibull distribution of the parameters is shown in Table 3. The predicted values for 10-, 20-, 50-, 100-, 200-, 500-, and 1000-year return periods calculated using the Gumbel and Weibull distributions are shown in Table 4 and 5.

TABLE 2. K-S test of the Weibull distribution.

Data Group	A26	B26	C26	D26
Test value D_n	0.1234	0.1071	0.1708	0.2053
Critical value $D_0(0.05)$	0.2591	0.2006	0.1866	0.2275
Comparison	$D_n<D_0$	$D_n<D_0$	$D_n<D_0$	$D_n<D_0$
Testing results	Accept	Accept	Accept	Accept

TABLE 3. Parameter calculation of distribution models.

	Gumbel		Weibull	
Data group	Parameter	Interval estimation	Parameter	Interval estimation
A26	μ=4.6705 σ=0.7156	[4.3787, 4.9622] [0.5459, 0.9381]	a=4.6138 b=6.5676	[4.3373, 4.9079] [4.9462, 8.7205]
B26	μ=4.6449 σ=0.6852	[4.4289, 4.8608] [0.5616, 0.8360]	a=4.5950 b=7.0939	[4.3955, 4.8036] [5.7923, 8.6879]
C26	μ=4.5530 σ=0.7094	[4.3454, 4.7606] [0.5901, 0.8528]	a=4.4990 b= 6.7392	[4.3080, 4.6985] [5.5824, 8.1356]
D26	μ=4.7899 σ=0.6534	[4.5557, 5.0241] [0.5194, 0.8219]	a=4.7451 b=7.5966	[4.5265, 4.9743] [6.0180, 9.5893]

From Table 4 and 5, it can be seen that the design tide level of the multiyear return period derived from the different distributions differs a little, even based on the same sample. Taking the 1000-year return period in the A26 dataset as an example, the Weibull distribution is 2.30% higher than the Gumbel distribution's standard. The tables also show that the derived design values by the same distribution model are different when based on the different samples. Different selected thresholds will produce different samples, which affects the estimation of distribution parameters and the calculation of return level. The selected threshold is larger, the corresponding return level will be higher. As shown in Table 4, for the 100-year return periods design values calculated in the Gumbel distribution, the results under the threshold u = 3.9m shows an increase of 1.70% over those under the threshold u=3.75m; the results under the threshold u = 3.75m shows an increase of 0.97% over those under the threshold u = 3.5m. For the 500-year return periods design values, the similar results can be obtained. These results show that the calculated design values using the generated sample by the annual maxima method is close to that calculated values by POT method. But the predicted values based on POT method are more stable on the premise of reasonable threshold selection.

Fig. 8 shows that the relationship between predicted values and threshold selection. As threshold level increases, the predicted results increases as well. When the threshold u = 3.75m, the predicted results are between those in the thresholds u = 3.5m and u = 3.9m, which indicates that POT method based on the mean residual life plot theory is more reasonable, and the selected threshold has certain representation. All of these results show that if the threshold value is too small, it will violate the assumptions of the model and lead to a large deviation. If the threshold value is too high, only little data will be generated and result in a large variance.

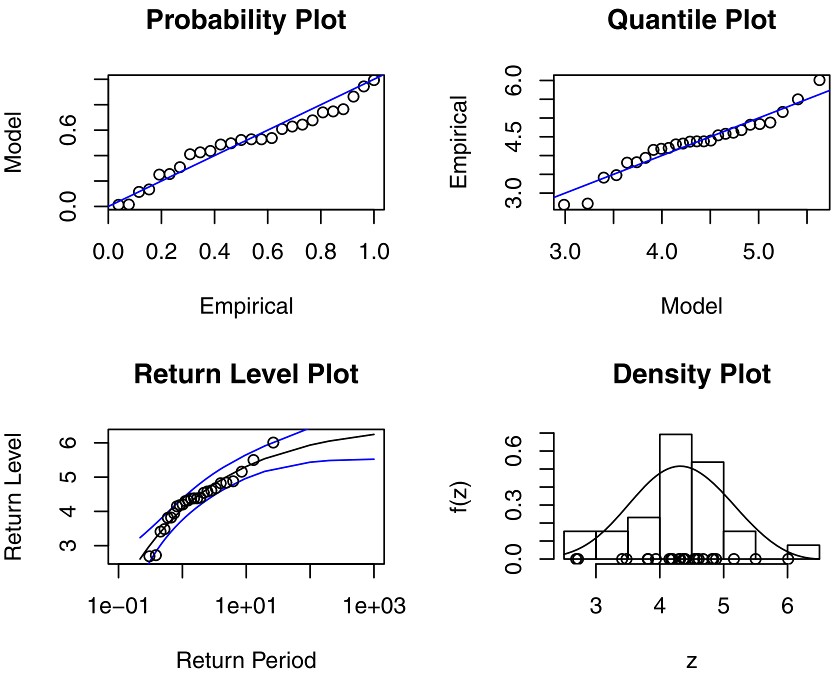


FIGURE 4. A26 statistical description plot (a. probability; b. quantiles; c. return level; d. density).

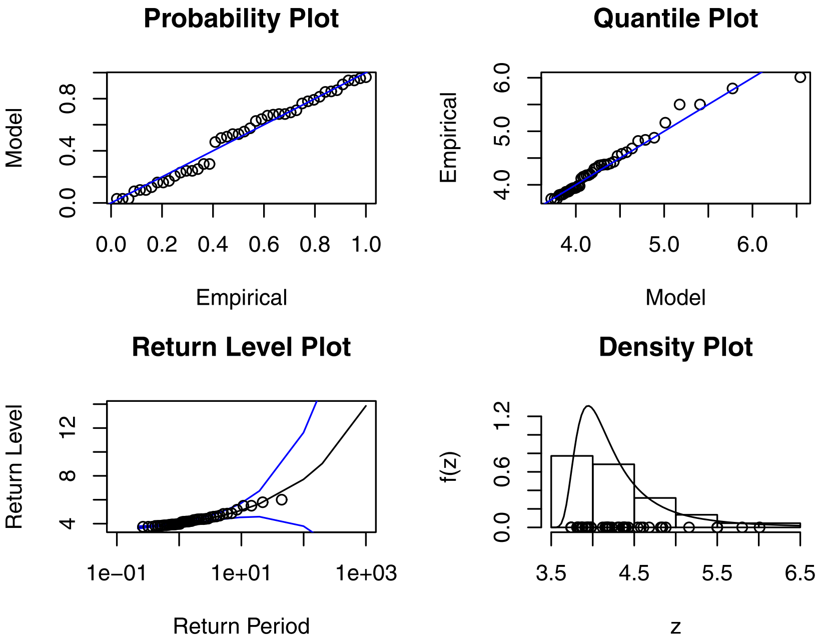


FIGURE 5. B26 statistical description plot (a. probability; b. quantiles; c. return level; d. density).

The method presented in this section gives us a quantitative method to determine the threshold value, which makes the POT method more practical.

B. ANALYSIS RESULTS OF COHERENCE SPECTRUM AND POWER SPECTRUM

The process sampling method breaks the limit that annual maxima method only takes one extreme value every year, which reflects the probability characteristics of occurrence frequency of typhoon- or hurricane-induced extreme sea environments, and the long-term distribution law of the ocean environment elements. Based on process sampling method, a sample is extracted from the original data, then tide levels of the sample in one year is averaged. The final average values are denoted as the annual average tide level. The annual maximal tide level of each year is selected and denoted as the annual extreme tide level. In this section, two time series of the annual average tide level and the annual

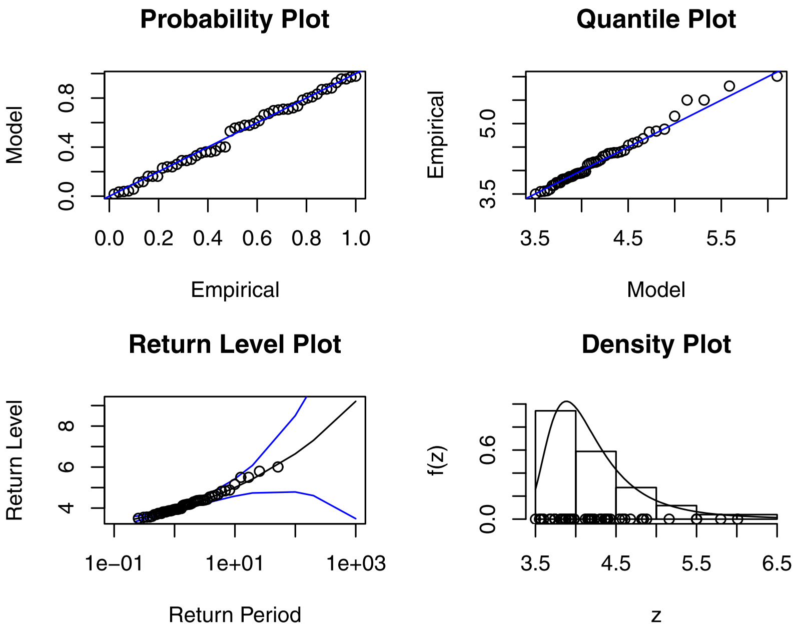


FIGURE 6. C26 statistical description plot (a. probability; b. quantiles; c. return level; d. density).

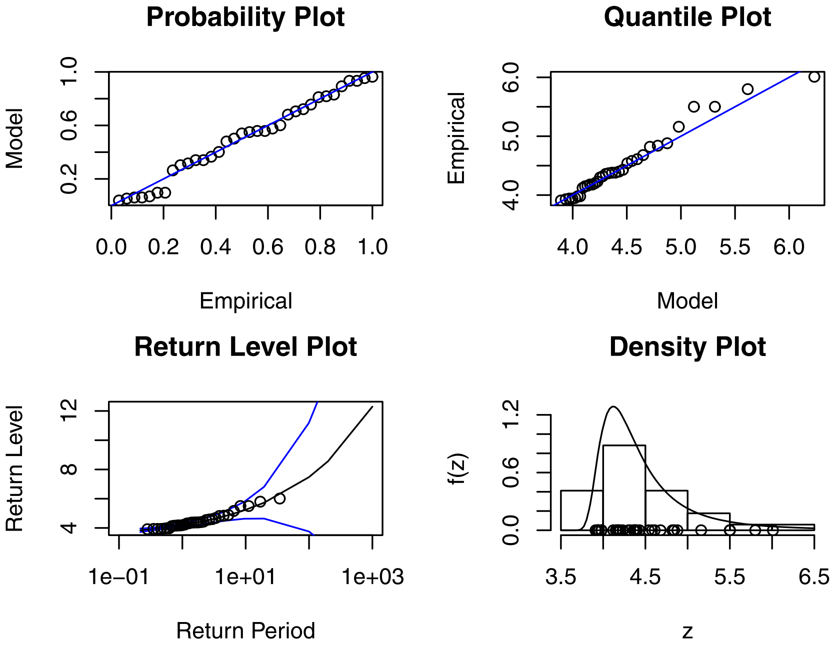


FIGURE 7. D26 statistical description plot (a. probability; b. quantiles; c. return level; d. density).

extreme tide level that are extracted from the data measured between 1971 to 2006 at the Zhapu Hydrologic station as examples, coherence and continuous power spectrum methods are used to analyze the characteristics of extreme tide level in Hangzhou Bay. Firstly, we use coherence spectrum to analyze the correlation between the two different tide level time series. In order to facilitate the following calculation and analysis, Table 6 gives the significance thresholds for coherence, which are obtained by using Eq. (6) under different overlapping rates, segment numbers and segment lengths.

From Fig. 9 to 16, it can be seen that the coherence spectrum charts and power spectrum charts of corresponding annual extreme tide level are different under different segment numbers, segment lengths and overlapping rates. As the power spectrum charts shown, that the main vibration period of the annual extreme tide level has a peak in the

TABLE 4. Predicted values for the Gumbel distribution based on different samples.

Sampling method		Data group	Design tide level values /m						
			10a	20a	50a	100a	200a	500a	1000a
Annual maxima method		A26	5.2673	5.4557	5.6466	5.7634	5.8637	5.9779	6.0535
POT	u=3.5	C26	5.1447	5.3313	5.5207	5.6364	5.7358	5.8490	5.9240
	u=3.7	B26	5.2163	5.3966	5.5795	5.6912	5.7873	5.8966	5.9691
	u=3.9	D26	5.3349	5.5068	5.6812	5.7878	5.8794	5.9836	6.0527

TABLE 5. Predicted values for the Weibull distribution based on different samples.

Sampling method		Data group	Design tide level values /m						
			10a	20a	50a	100a	200a	500a	1000a
Annual maxima method		A26	5.2386	5.4527	5.6789	5.8217	5.9473	6.0935	6.1924
POT	u=3.5	C26	5.0917	5.2945	5.5084	5.6433	5.7620	5.9000	5.9933
	u=3.7	B26	5.1683	5.3636	5.5693	5.6988	5.8126	5.9447	6.0340
	u=3.9	D26	5.2958	5.4824	5.6784	5.8017	5.9098	6.0352	6.1198

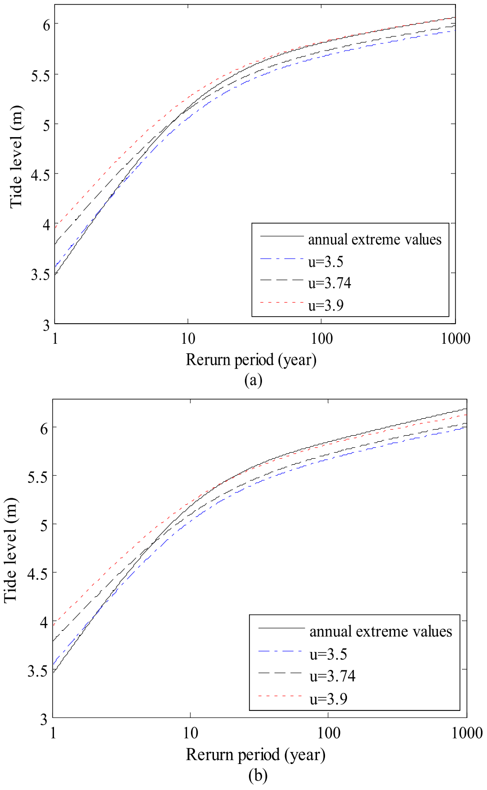


FIGURE 8. Comparison of the predicted values based on different samples (a.Gumbel; b.Weibull).

low frequency and high frequency parts, respectively. The coherence spectrum can intuitively show the relationship between two sequences within a certain frequency range.

TABLE 6. Thresholds for the significance test of the coherence spectrum.

p	25%	50%						75%
n_s	3	3	3	3	3	4	5	3
L	12	12	14	16	18	12	12	12
N	30	24	28	32	36	30	36	18
c	0.78	0.81	0.81	0.80	0.80	0.67	0.56	0.92

* p: overlapping rate; n_s: segment numbers;
L: segment lengths; N:data sizes;
c: thresholds for the significance test of the coherence spectrum

In the coherence spectrum charts, the solid lines represent the change curves of the spectrum value with the frequency, and the dotted line represent the threshold for the significance test of the coherence spectrum. If the spectrum value is above the dotted line, it indicates that the two sequences are significantly correlated in this frequency. Through the intersection points of the two lines, we can determine the frequency interval of significant correlation, and draw the frequency interval in the corresponding power spectrum charts (indicated by the dash dot lines in the figures).

Fig. 9 to 11 show that at the overlapping ratio of 50% and the segment number of 3, the coherence spectrum and the corresponding power spectrum's charts under the segment length of 14, 16 and 18. The peak value of the high-frequency part of the power spectrum is within the interval, and the main periods of the high-frequency part are 27.8, 21.3 and 23.8 years, respectively. As the segment length increases, the frequency interval of significant correlation increases as

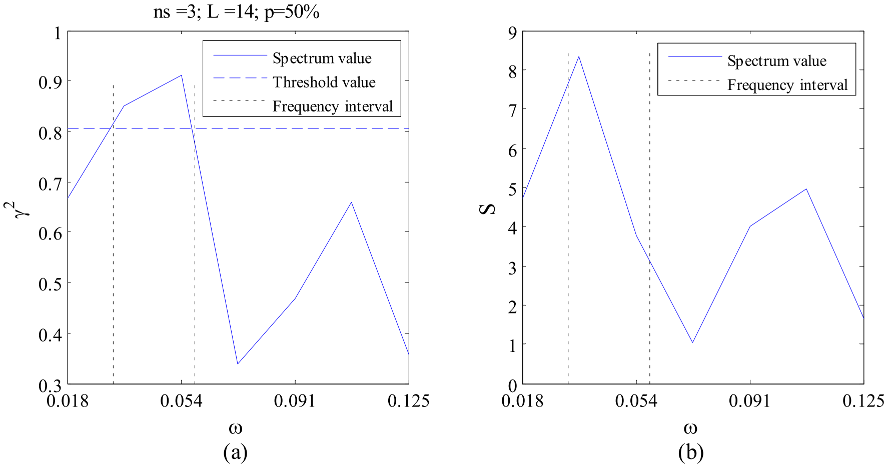


FIGURE 9. Spectrum analysis: n_S=3, L=14, p=50% (a, coherence spectrum; b power spectrum).

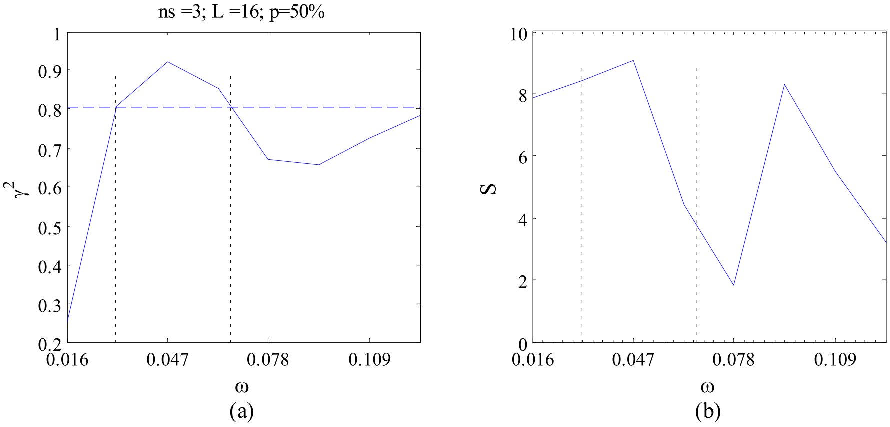


FIGURE 10. Spectrum analysis: n_S=3, L=16, p=50% (a, coherence spectrum; b power spectrum).

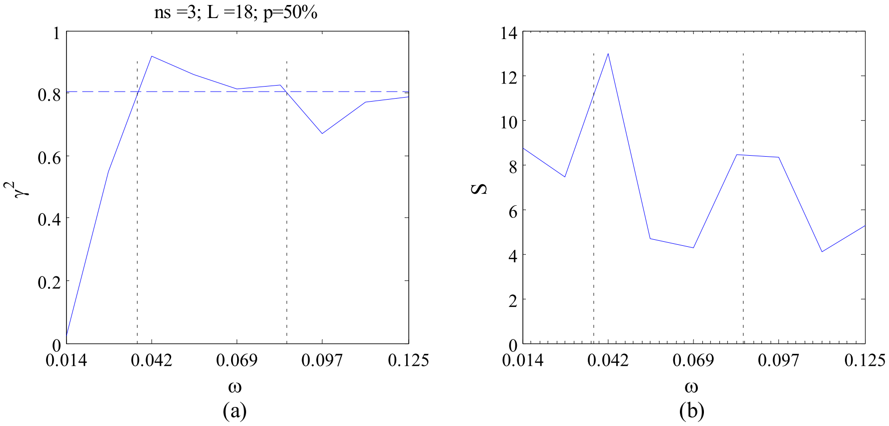


FIGURE 11. Spectrum analysis: n_S=3, L=18, p=50% (a, coherence spectrum; b power spectrum).

well. The interval under the segment length L=16 shows an increase of 32.75% over that under L=14; the interval under the segment length L=18 shows an increasing of 40.2% over that under L=16. When the segment length is 14, there is too little information above the threshold. When the segment length is 18, too many spectral values exceed the threshold, and the quality of spectral estimation needs to be improved. The spectral estimation attains a good effect

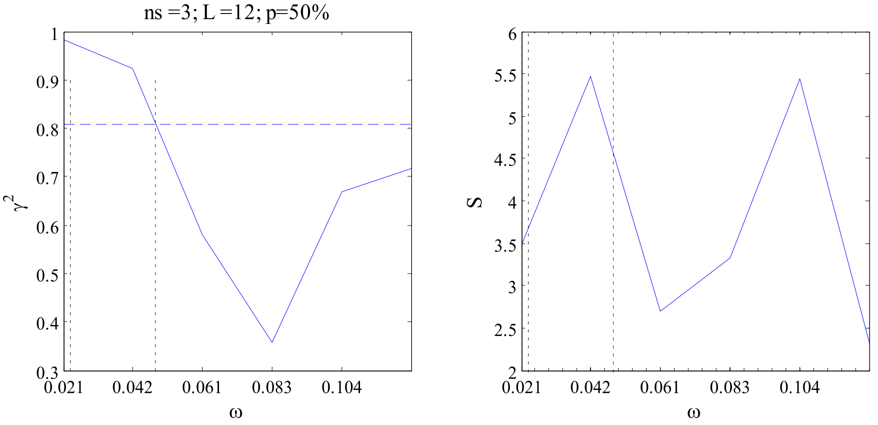


FIGURE 12. **Spectrum analysis: n_s=3, L=12, p=50% (a, coherence spectrum; b power spectrum).**

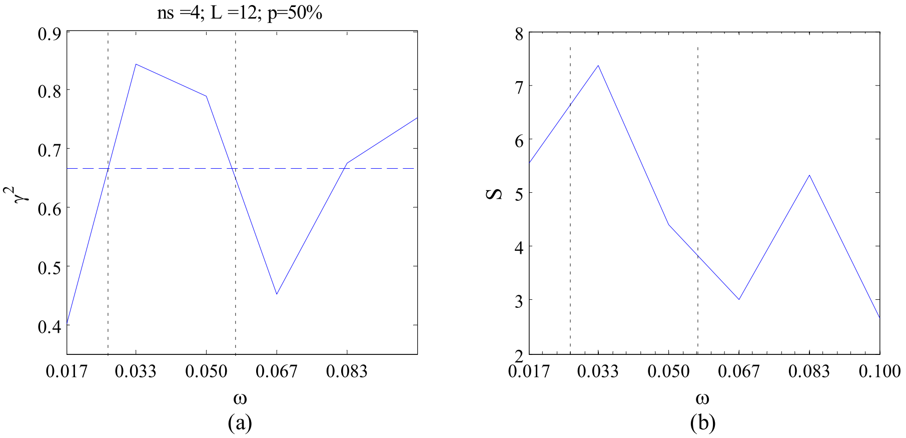


FIGURE 13. **Spectrum analysis: n_s=4, L=12, p=50% (a, coherence spectrum; b power spectrum).**

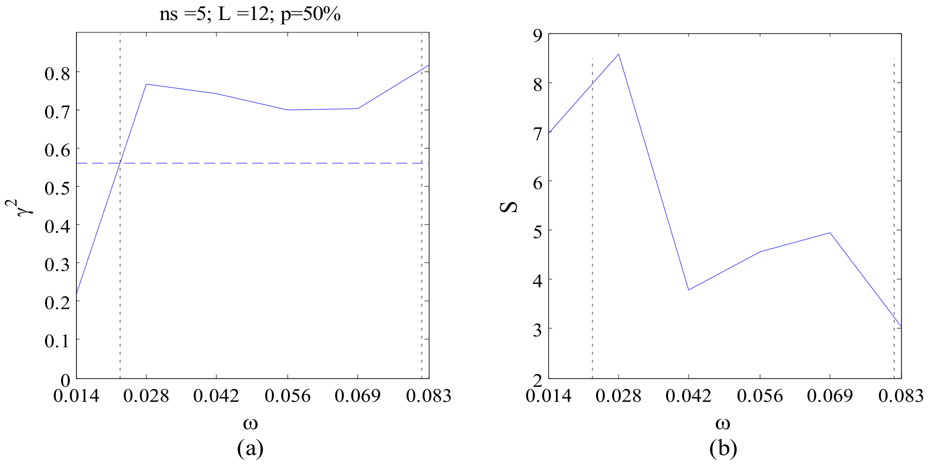


FIGURE 14. **Spectrum analysis: n_s=5, L=12, p=50% (a, coherence spectrum; b power spectrum).**

at the segment length of 16, the corresponding threshold is representative.

Fig. 12 to 14 show that at the overlapping ratio of 50% and the segment length of 12, the coherence spectrum and the corresponding power spectrum's charts under the segment number of 3, 4 and 5. As the segment number increases, the thresholds for the significance test of the coherence spectrum decrease obviously, and the spectrum shape changes

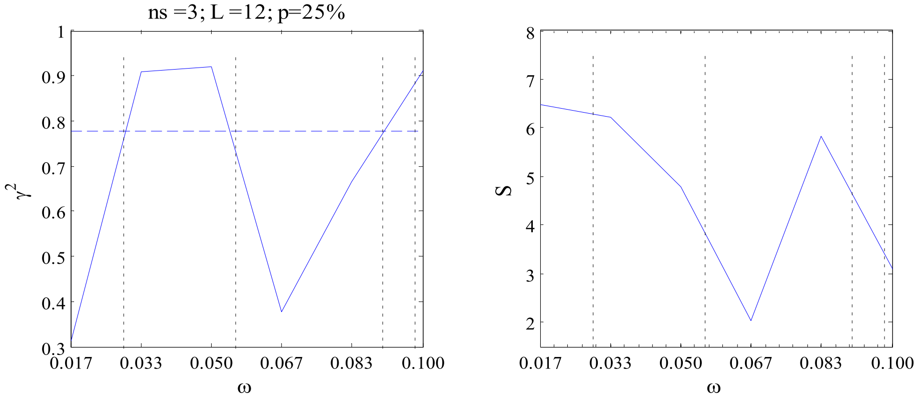


FIGURE 15. Spectrum analysis: n_S=3, L=12, p=25% (a,coherence spectrum; b power spectrum).

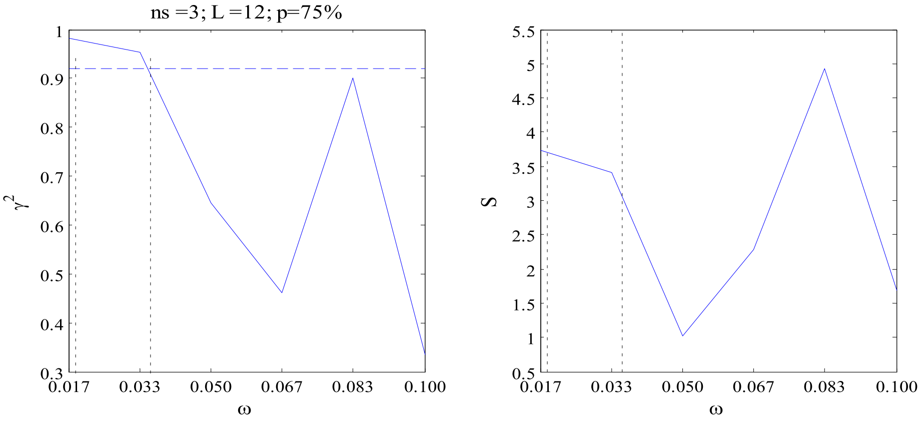


FIGURE 16. Spectrum analysis: n_S=3, L=12, p=75% (a, coherence spectrum; b power spectrum).

as well. When the segment number is 5, the most of the coherence spectrum values are above threshold. For the power spectrum, the peaks of the high-frequency part are all within the intervals, and the corresponding main periods are 23.8, 30.3 and 35.7 years, respectively.

Fig. 15, 12 and 16 show that at the segment number of 3 and the segment length of 12, the coherence spectrum and the corresponding power spectrum's charts under the overlapping ratio of 25%, 50%, 75%, respectively. When the overlapping ratio is 25% and 75%, the peaks of power spectrum are not within the interval of correlation frequency, the performance of the thresholds are not ideal, and the quality of spectrum estimation needs to be improved. It is suggested by all those results that, when the spectrum estimation is carried out by using different segment number, segment length and overlapping ratio in a reasonable range, the two sequences are closely related. That is, the samples generated based on two sampling methods imply relatively consistent fluctuation characteristics of tide level time series. Considering that China's hydrological sequence is generally not long, we take the spectrum estimation at the segment length of 3, the segment number of 16 and the overlapping ratio of 50% as the result of better quality, and the main period of the high-frequency part of the power spectrum is 21.3 years.

IV. CONCLUSION

In this paper, for the first time, the methods used to determine the threshold value of measured data in the calculation of design wave height and design water level are studied from both time domain and frequency domain. The sensitivity of data density and data sampling process for the calculation of environmental design parameters is discussed. In the time domain, a quantitative theoretical method is introduced to determine the threshold of hydrological sequence. In the frequency domain, a estimation method of the significance threshold for coherence is given to better analyze the spectrum. Conclusions obtained from this study are as follows:

(1) Uncertainty of sampling method and threshold selection exist in data sample's generation, which leads to the uncertainty of prediction result of design water level. In order to reduce the uncertainty, it is necessary to reasonably select the threshold value of the measured data, so as to obtain the results of the optimization probability analysis and reduce the risk of the project in the design stage.

(2) This paper introduces a method of quantitative selection threshold based on the mean residual life plot. This method has the sufficient mathematical theory foundation, at the same time, it is easy to operate and effectively raise the use value of the POT method.

(3) In order to explore the energy period distribution of tide level and predict the development of tide level sequence, so as to make more reasonable use of the original sequence, this paper attempts to study the activity characteristics of tide level in Hangzhou Bay by means of coherence and power spectrum. In this paper, an estimation method of the significance threshold for coherence based on WOSA method is proposed to better analyze the correlation between different tide sequences. The calculation results show that the tide level in Hangzhou Bay has a 21.3-year vibration period, and its formation mechanism remains to be studied. At present, there are few study on the exploration of extreme tide level through spectral analysis, and it is worth further research.

REFERENCES

[1] L.-P. Wang, B.-Y. Chen, C. Chen, Z.-S. Chen, and G.-L. Liu, "Application of linear mean-square estimation in ocean engineering," *China Ocean Eng.*, vol. 30, no. 1, pp. 149–160, 2016.

[2] B. Chen *et al.*, "Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles," in *Proc. 36th IEEE Int. Perform. Comput. Commun. Conf.*, San Diego, CA, USA, Dec. 2017, pp. 1–6.

[3] L. Wang, X. Xu, G. Liu, B. Chen, and Z. Chen, "A new method to estimate wave height of specified return period," *Chin. J. Oceanol. Limnol.*, vol. 35, no. 5, pp. 1002–1009, 2017.

[4] L. Kang *et al.*, "Study on dye wastewater treatment of tunable conductivity solid-waste-based composite cementitious material catalyst," *Desalination Water Treatment*, vol. 125, pp. 296–301, Jan. 2018.

[5] L. Wang, G. Liu, B. Chen, and L. Wang, "Typhoon based on the principle of maximum entropy waters affect the design wave height calculation method," China Patent 2010 10 595 815, Dec. 20, 2010.

[6] L. Kang, Y. J. Zhang, L. Zhang, and K. Zhang, "Preparation, characterization and photocatalytic activity of novel CeO_2 loaded porous alkali-activated steel slag-based binding material," *Int. J. Hydrogen Energy*, vol. 42, pp. 17341–17349, Jul. 2017.

[7] H. Zhao, R. Yao, L. Xu, G. Li, and W. Deng, "Study on a novel fault damage degree identification method using high-order differential mathematical morphology gradient spectrum entropy," *Entropy*, vol. 20, no. 9, p. 682, 2018.

[8] L.-P. Wang, B. Chen, J.-F. Zhang, and Z. Chen, "A new model for calculating the design wave height in typhoon-affected sea areas," *Natural Hazards*, vol. 67, no. 2, pp. 129–143, 2013.

[9] B. Chen, G. Liu, and L. Wang, "Predicting joint return period under ocean extremes based on a maximum entropy compound distribution model," *Int. J. Energy Environ. Sci.*, vol. 2, no. 6, pp. 117–126, 2017.

[10] Z. R. Ren, R. Skjetne, and Z. Gao, "A crane overload protection controller for blade lifting operation based on model predictive control," *Energies*, vol. 12, no. 1, p. 50, 2019.

[11] L. Wang, G. Liu, B. Chen, and L. Wang, "Typhoon influence considered method for calculating combined return period of ocean extreme value," China Patent 2010 10 595 807, Mar. 20, 2013.

[12] A. Barrs and B. Chen. (2018). *How Emerging Technologies Could Transform Infrastructure*. Accessed: Mar. 10, 2018. [Online]. Available: http://www.governing.com/commentary/col-hyperlane-emerging-technologies-transform-infrastructure.html,

[13] B. Chen, S. Escalera, I. Guyon, V. Ponce-López, N. Shah, and M. O. Simón, "Overcoming calibration problems in pattern labeling with pairwise ratings: Application to personality traits," in *Proc. Workshops Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, 2016, pp. 419–432. doi: 10.1007/978-3-319-49409-8_33.

[14] V. Ponce-López *et al.*, "ChaLearn LAP 2016: First round challenge on first impressions—Dataset and results," in *Proc. Comput. Vision-ECCV Workshops*. Amsterdam, The Netherlands: Springer, 2016, pp. 400–418. doi: 10.1007/978-3-319-49409-8_32.

[15] G. Liu, B. Chen, S. Jiang, H. Fu, L. Wang, and W. Jiang, "Double entropy joint distribution function and its application in calculation of design wave height," *Entropy*, vol. 21, no. 1, p. 64, 2019.

[16] B. Chen, G. Liu, L. Wang, K. Zhang, and S. Zhang, "Determination of water level design for an estuarine city," *J. Oceanol. Limnol.*, vol. 2, pp. 1–11, Jan. 2018. doi: 10.1007/s00343-019-8107-z.

[17] K. Chen, L. Zhang, S. J. Wang, and G. Liu, "Generalized extreme value-Pareto distribution function and its applications in ocean engineering," *Chin. Ocean Eng.*, to be published.

[18] H. Fu, Z. Li, Z. Liu, and Z. Wang, "Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization," *Sustainability*, vol. 10, no. 7, p. 2488, 2018.

[19] X. Liu, Y. He, H. Fu, B. Chen, M. Wang, and Z. Wang, "How environmental protection motivation influences on residents' recycled water reuse behaviors: A case study in Xi'an city," *Water*, vol. 10, no. 9, p. 1282, 2018.

[20] G. Liu, B. Chen, L. Wang, S. Zhang, K. Zhang, and X. Lei, "Wave height statistical characteristic analysis," *J. Oceanol. Limnol.*, vol. 36, pp. 1–13, May 2018. doi: 10.1007/s00343-019-8006-3.

[21] H. J. Escalante *et al.*, "ChaLearn joint contest on multimedia challenges beyond visual analysis: An overview," in *Proc. IEEE 23rd Int. Conf. Pattern Recognit.*, Cancun, Mexico, Dec. 2016, pp. 67–73.

[22] F. Mazas and L. Hamm, "A multi-distribution approach to POT methods for determining extreme wave heights," *Coastal Eng.*, vol. 58, no. 5, pp. 385–394, 2011.

[23] Y. Luo, "Study on theories and methods return values calculation of extreme environmental condition with POT method for ocean engineering," South China Univ. Technol., Tech. Rep., 2013.

[24] S. Zhang, W. Shen, D. Li, X. Zhang, and B. Chen, "Nondestructive ultrasonic testing in rod structure with a novel numerical Laplace based wavelet finite element method," *Latin Amer. J. Solids Struct.*, vol. 15, no. 7, pp. 1–17, 2018.

[25] B. Chen and B. Wang, "Location selection of logistics center in e-commerce network environments," *Amer. J. Neural Netw. Appl.*, vol. 3, no. 4, pp. 40–48, 2017. doi: 10.11648/j.ajnna.20170304.11.

[26] J. Song, Q. Feng, X. Wang, H. Fu, W. Jiang, and B. Chen, "Spatial association and effect evaluation of CO_2 emission in the Chengdu–Chongqing urban agglomeration: Quantitative evidence from social network analysis," *Sustainability*, vol. 11, no. 1, p. 1, 2019.

[27] S. Jiang, M. Lian, C. Lu, S. Ruan, Z. Wang, and B. Chen, "SVM-DS fusion based soft fault detection and diagnosis in solar water heaters," *Energy Explor. Exploitation*, to be published. doi: 10.1177/0144598718816604.

[28] S. Jiang, M. Lian, C. Lu, Q. Gu, S. Ruan, and X. Xie, "Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope," *Complexity*, vol. 8, Aug. 2018, Art. no. 1048756. doi: 10.1155/2018/1048756.

[29] S. Jiang, M. Lian, C. Lu, Q. Gu, S. Ruan, and X. Xie, "Prediction of the death toll of environmental pollution in China's coal mine based on metabolism-GM (1, n) Markov model," *Proc. EKOLOJI*, vol. 26, no. 101, pp. 17–23, 2017.

[30] C. Wang and D. Liu, "Undefined analysis of selecting design wave factors," *Acta, Ceanol. Sinica*, vol. 13, no. 4, pp. 874–881, 1991.

[31] S. Coles, "An introduction to statistical modeling of extreme values," *Technometrics*, vol. 44, no. 4, p. 397, 2001.

[32] D. Liu, L. Wang, Y. Song, and L. Pang, "Multivariate compound extreme value distribution and its application," *Periodical Ocean Univ. China*, vol. 34, no. 5, pp. 893–902, 2004.

[33] W. Deng, H. Zhao, L. Zou, G. Li, X. Yang, and D. Wu, "A novel collaborative optimization algorithm in solving complex optimization problems," *Soft Comput.*, vol. 21, no. 15, pp. 4387–4398, 2017.

[34] A.-Y. Yang, X.-L. Yang, J.-C. Chang, B. Bai, F.-B. Kong, and Q. Ran, "Research on a fusion scheme of cellular network and wireless sensor for cyber physical social systems," *IEEE Access*, vol. 6, pp. 18786–18794, 2018. doi: 10.1109/ACCESS.2018.2816565.

[35] H. M. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, p. 14, 2017.

[36] W. Deng, H. Zhao, X. Yang, J. Xiong, M. Sun, and B. Li, "Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment," *Appl. Soft Comput.*, vol. 59, pp. 288–302, Oct. 2017.

[37] K. P. McKone, "Multiple methods of spectral analysis with applications to the Florida current," Ph.D. dissertation, Dept. Mar. Sci., Univ. Southern Mississippi, Hattiesburg, MS, USA, 2003.

[38] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, no. 1, pp. 35042–35056, 2018.

[39] H. Fu and X. Liu, "A study on the impact of environmental education on individuals' behaviors concerning recycled water reuse," *EURASIA J. Math. Sci. Technol. Educ.*, vol. 13, no. 10, pp. 6715–6724, 2017.

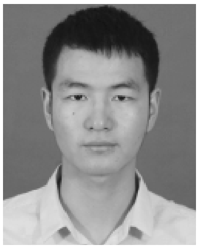
[40] H. Fu and X. Liu, "Research on the phenomenon of chinese residents' spiritual contagion for the reuse of recycled water based on SC-IAT," *Water*, vol. 9, no. 11, p. 846, 2017.

[41] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Comput.*, vol. 23, no. 7, pp. 2445–2462, 2017. doi: 10.1007/s00500-017-2940-9.

[42] W. Deng, J. Xu, and H. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE Access*, vol. 7, pp. 20281–20292, 2019. doi: 10.1109/ACCESS.2019.2897580.

[43] C. Gallet and C. Julien, "The significance threshold for coherence when using the Welch's periodogram method: Effect of overlapping segments," *Biomed. Signal Process. Control*, vol. 6, no. 4, pp. 405–409, 2011.

GUILIN LIU received the bachelor's and Ph.D. degrees in coastal and offshore engineering from the Ocean University of China. He has authored three books, more than 20 journals, and conference articles. His research interests include joint probability for environmental loads of coastal and ocean engineering, hydrodynamic research in marine engineering, and digital intelligent construction in coastal and ocean engineering.

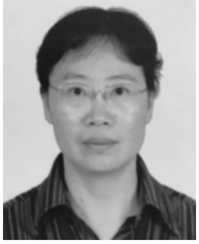


ZHIKANG GAO received the bachelor's degree from the Changsha University of Science and Technology. He is currently pursuing the master's degree with the College of Engineering, Ocean University of China, Qingdao, China. His research interests include the joint probability analysis of extreme ocean environmental elements, hydrodynamic research in marine engineering, and fluid dynamics.

BAIYU CHEN received the B.S. and M.S. degrees in civil engineering and the M.S. degree in electrical engineering and computer science and from the University of California Berkeley. He has authored two books, more than 10 journals, and conference articles. His research interests include multiple disciplines, with a focus on intelligent transportation system control, machine learning, and risk control.

HANLIANG FU received the Ph.D. degree from the School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi, China, in 2017, where he currently holds a Postdoctoral position. His research interests include big data, and environmental management and bibliography.

SONG JIANG received the M.S. degree in mine engineering from the School of Management, Xi'an University of Architecture and Technology, Shaanxi, China, in 2012, where he is currently pursuing the Ph.D. degree. He has authored two books and more than 20 articles. His research interests include mining system engineering, big data, and mining management. He has a strong interest in the exploration of interdisciplinary fields of computers and mine engineering.



LIPING WANG received the Ph.D. degree in coastal and offshore engineering from the Ocean University of China, where she is currently a Professor of mathematical sciences. She has authored two books and more than 20 articles. Her research interests include multiple disciplines, with a focus on analysis of statistical characteristics of random ocean wave, forecast warning of storm surge, the applications of mathematical model in marine and financial engineering, and the proposal of big data products and solutions.

YI KOU received the degree from Medical College, Peking University, and the Ph.D. degree in biochemistry from The University of Texas at Austin. His research interests involve from both bench work and computational modeling across various fields. His main research interests include X-ray crystallography, anti-cancer drug development, DNA related structure modeling, protein-substrate simulation and modeling, mutagenesis and carcinogenesis studies, machine learning on protein mutagenesis and 3D genome organization, and 3D model designing.