China Ocean Eng., 2019, Vol. 33, No. 2, P. 127–136
DOI: 10.1007/s13344-019-0013-9, ISSN 0890-5487
http://www.chinaoceanengin.cn/ E-mail: coe@nhri.cn

Generalized Extreme Value-Pareto Distribution Function and Its Applications in Ocean Engineering

CHEN Bai-yu[a], ZHANG Kuang-yuan[b], WANG Li-ping[c], JIANG Song[d], LIU Gui-lin[e,*]

[a]College of Engineering, University of California Berkeley, Berkeley, USA
[b]Department of Energy and Mining Engineering, Pennsylvania State University, University Park, PA, USA
[c]School of Mathematical Sciences, Ocean University of China, Qingdao 266100, China
[d]School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China
[e]College of Engineering, Ocean University of China, Qingdao 266100, China

Received September 14, 2018; revised October 30, 2018; accepted December 2, 2018

Abstract

In this paper, we establish a generalized extreme Value-Pareto distribution model and derive an analytical expression of Weibull–Pareto distribution model. Based on a data sample of 26-year wave height, we adopt the new model to estimate the design wave height for 500, 700 and 1000-year return periods. Results show that the design wave height from Weibull–Pareto distribution is between that of the Weibull distribution and that of the Pearson-III distribution. For the 500-year return period design wave height, the results from the new model is 1.601% lower than those from the Weibull distribution and 1.319% higher than those from the Pearson-III distribution. The Weibull–Pareto distribution innovatively considers the fractal features, extreme-value statistics and the truncated data in the derivation process. Therefore, it is a more holistic and practical model for estimating the design parameters in marine and coastal environments.

Key words: fractal, Pareto distribution, design wave height

Citation: Chen, B. Y., Zhang, K. Y., Wang, L. P., Jiang, S., Liu, G. L., 2019. Generalized extreme value-pareto distribution function and its applications in ocean engineering. China Ocean Eng., 33(2): 127–136, doi: 10.1007/s13344-019-0013-9

1 Introduction and Background

In area of designing coastal and marine structures, the estimation of parameters of the marine environment such as wave height and water level is one of the critical factors. Proper and accurate calculation of the wave height or water level is very important for the research on safety and building cost of engineering structures, especially under severe marine conditions (Wang et al., 2017; Xian et al., 2018). The marine environment can be very complicated. Offshore projects need to withstand the damage caused by natural disasters such as strong waves and typhoons. When designing coastal structures and assessing the relevant risk, if the designing security's standard is not high enough, extreme marine impacts such as typhoons can easily result in immeasurable losses. On the other hand, if the designing security standard is too high, there may be issues like over-investment and financial waste. Therefore, the estimation of a proper and accurate design standard, which can not only withstand the natural impacts but also guarantee the security and stability of the designed marine structures, is required. The essence of the design lies in the creation of the most reasonable and appropriate model to estimate marine environment design parameters.

Traditionally, univariate extreme value models, such as Gumbel, Weibull, Pearson-III type distributions and the maximum entropy distribution have been adopted for modeling wave height and water level under extreme conditions (Liu et al., 2006; Wang et al., 2014). As the multivariate statistical theory develops, multivariate models like Logistic model, Hüsler-Reiss model, Beta model and other mixture models have been adopted in an accumulation of literature, and compound extreme value distribution under extreme conditions has been introduced recently (Chen et al., 2017a; Liu G.L. et al., 2015, 2018; Liu X.J. et al., 2018; Wang and Wang, 2013; Wang et al., 2013, 2016). Traditionally, the main idea of deriving design wave height (or wave height for a specific return period) is to select a certain type of probability distribution (such as, Weibull, Pearson-III, Gumbel distribution) to model the distribution of the maximum annual wave height. Then observed data of maxim-

Foundation item: This work was supported by the NSFC-Shandong Joint Fund (Grant No. U1706226), and Graduate Education Reform and Research Fund (Grant No. HDJG18007).
*Corresponding author. E-mail: liuguilin73@ouc.edu.cn

um annual wave height and frequency is used to determine the model parameters. After the calculation of parameters, the cumulative density function is used to finally determine the design wave height for a specific return period. Specifically, this modeling process is based on the following assumptions: 1) The annual maximum wave height follows the same probability distribution both in the short-term and the long-term; 2) the observed statistical characteristics in a relatively short period of time are strictly self-similar to those in the long-term (100-year return period or longer); 3) due to constrains in the distribution function, only the annual maximum wave height (only one data point per year) is used in the models of wave height and water level, instead of the majority of observed data.

The data volume and quality are fundamental for the modeling and analysis and the quantity of the data often directly influences the quality of the results. Due to practical constraints in many cases, only data over a limited time range (For example, 20 or 30 years) is available. Therefore, it is crucial to make full use of all obtained data and extract as much useful information as possible.

The probabilistic analysis on ocean environmental elements such as wave height and water level with Pearson-III, Weibull distribution and multivariate extreme value distribution is designed to explore the changing laws of marine environmental factors on different time scales (i.e., 100 years, 500 years) and locations, and the expansion and inferences to and from different scales. In previous extreme value models, if the random variable is larger than a certain threshold, the distribution function shows the form of a power function, which has a fractal characteristic if the large-value tail is truncated. Therefore, these distribution functions can describe the power functions at large values very well. However, the tail of the distribution function converges to 0 so rapidly that the distribution function fits the data poorly near the tail. The Pareto distribution shows fractal characteristics at the low-end of the tail and has superior performance in fitting the tailing data. However, its performance for large values is mediocre.

This research proposes a generalized extreme value – Pareto distribution function to make full use of the marine data based on the extreme value statistical theory and fractal theory. The new model contains the commonly used extreme value distribution that shows the statistical characteristics of random variables in extreme conditions, as well as the fractal characteristics that utilize data efficiently. If the observed data are from a rather short period of time, the data can be expanded by selecting an appropriate threshold. The new model presents both the interactive and hierarchical structures in explicit form. The new model can be simplified to obtain the original extreme value distribution model. It is a novel model that covers larger amounts of information with flexibility and has broader scope of applications.

The motivation of this research lies in four folds. First, most distribution functions of extreme values/statistics have the form of a power function. Taking advantage of the fractal characteristics of power function when the large-value tail is truncated, this research discusses the changing laws of marine environment elements over a shorter time period (20 to 50 years). Then, using self-similarity of data statistics, this paper calculates the design wave height over an extended period of time (100 or 500 years). Second, when using above-threshold data to derive design wave height, the tail of the distribution function approaches 0, a poor fit of tail data. Therefore, this research introduces the Pareto distribution that exhibits fractal characteristics when low-value data is truncated. Third, this research combines generalized extreme distribution function (de Haan and Ferreira, 2006) and Pareto distribution function, to take advantage of their complementary strengths. Fourth, in Section 3, we derived the generalized extreme value – Pareto distribution model, and gave the explicit express of Weibull–Pareto distribution, where the steps are: first create a function family; then, prove when random variable T follows the generalized extreme value distribution, and X follows Pareto distribution, the expression for Weibull–Pareto distribution can be found.

2 Fractal characteristics of the function

Fractal refers to a system with irregularity and complexity, but with similarities between parts and the whole. There is randomness in the formation of fractal with a certain regularity in the randomness. The most important fundamental feature of fractal is self-similarity and scale-invariance. Self-similarity means that an entity complies to the same statistical distribution locally and globally. The scale-invariance feature refers to the entity's structure and remains unchanged on different scales.

Dimension is another important quantity to describe the geometric entity. In the traditional Euclidean geometry, a single point is considered to be zero-dimension. A straight line is one-dimensional, plane is two-dimensional, and space is three-dimensional. The fractal dimension, in contrast to the Euclidean geometry of the line, plane and space, has the dimension that may not only be an integer, but can also be a fraction, which is different compared with the traditional dimensions in Euclidean geometry. So far, fractal has been abstracted into a theory. Data analysis based on fractal has shown promising results (Barrs and Chen, 2018; Cai et al., 2011; Escalante et al., 2016; Liu G.L. et al., 2018; Zhang S.F. et al., 2018). Fractal has become a new and perspective method for complex system analysis, bridging micro and macro analysis (Cai et al., 2016; Chen and Wang, 2017; Chen et al., 2016; Fu et al., 2018; Ponce-López et al., 2016; Wang et al., 2016).

Definition 1 In a Euclidean space, if the Hausdorff dimension of a set is always larger than its topological dimension, such as:

$$D_H > D_T, \quad (1)$$

then the collection is called a fractal set and referred to as fractal. This mathematical definition can be difficult to understand. Several features of a fractal are listed as follows by the British mathematician Falconer to provide more details.

(1) Fractal has a very detailed structure, with an infinitely small-scale proportion of details.

(2) The fractal shows irregularity, and the irregularities of the part and the whole have infinite levels.

(3) A fractal generally has statistical self-similarity.

(4) A fractal's dimension is generally larger than its topological dimension.

(5) A fractal can generally be iterated by simple rules that are not complex.

In practice, the complex features of natural phenomena represent the universality of many entities in nature. Therefore, fractal geometry can better describe the natural world than the traditional Euclidean geometry (Cai et al., 2016; Chen et al., 2017b; Fu and Liu, 2017; Wang et al., 2013; Zhang and Kleit, 2016; Zhang K.Y. et al., 2018).

Definition 2 If the research entity's size is changed properly, the local and global fractal of any research entity are the same. If following formula holds,

$$f(br) = bf(r), \quad (2)$$

the entity is self-similar fractal or linear fractal.

Definition 3 Self-similarity becomes less strict if scaling ratios of similar mappings in all directions are not exactly the same, which is,

$$f(br) = bf^{H}(r), \quad (3)$$

then, the entity is called self-affine fractal, where H is between 0 and 1. It is obvious that if $H = 1$, the self-affine fractal becomes a self-similar fractal. Self-affine fractal is not as symmetric and regular as self-similar fractal. In nature, a self-affine fractal has different scaling across dimensions (Chen et al., 2018; Jiang et al., 2018; Ponce-López et al., 2016; Liu X.J. et al., 2018; Wang et al., 2015). Take stock price and index for example. The change scale is no longer simple expansion or shrinkage, but rather expansion in different directions with different factors.

Fractal characteristics of some common distributions are discussed below. It can be concluded that the fractal characteristics of these distributions are in the form of power functions, which is consistent with the form of extreme value models' power function in certain conditions.

Definition 4 If the probability distribution function of a random variable X is,

$$f(x) = \alpha\sigma^{-\alpha}x^{\alpha-1},\ \alpha > 0,\ 0 \leqslant \alpha \leqslant \sigma, \quad (4)$$

then, the random variable X is subject to power function distribution. Where σ is the scale parameter and α is the feature index. From Eq. (4), the mathematical expectation and variance for X are,

$$F(x) = P(X \leqslant x) = \left(\frac{x}{\sigma}\right)^{\alpha},\ \alpha > 0,\ 0 \leqslant \alpha \leqslant \sigma;$$
$$E(x) = \alpha\sigma(\alpha+1)^{-1},\ \alpha > 0;$$
$$V(x) = \alpha\sigma^{2}(\alpha+2)^{-1}(\alpha+1)^{-2},\ \alpha > 0. \quad (5)$$

In the truncated tailing of the data, $0 < x \leqslant \sigma_1 \leqslant \sigma$, following formulas can be derived,

$$P(X \leqslant x | X \leqslant \sigma_1) = \left(\frac{x}{\sigma}\right)^{\alpha}\left(\frac{\sigma_1}{\sigma}\right)^{-\alpha} = \left(\frac{x}{\sigma_1}\right)^{\alpha},\ 0 < x \leqslant \sigma_1 \leqslant \sigma \quad (6)$$

and, Eq. (6) shows that under upper truncation, power functions exhibit similar expression as $P(X \leqslant x)$. Or in other words, fractal expression remains the same, which is fractal characteristic. Since,

$$P(X \leqslant cx | X \leqslant c\sigma_1) = \left(\frac{cx}{c\sigma}\right)^{\alpha}\left(\frac{\sigma_1}{\sigma}\right)^{-\alpha} = \left(\frac{x}{\sigma_1}\right)^{\alpha},\ 0 < x \leqslant \sigma_1 \leqslant \sigma \quad (7)$$

the following can be derived,

$$P(X \leqslant x | X \leqslant \sigma_1) = P(X \leqslant cx | X \leqslant c\sigma_1). \quad (8)$$

Note that in Eq. (8), the power function has the property of scale invariance; where, σ_1 is the scale parameter, and c is a positive constant. Then,

Proposition 1 Power function distribution in the high-end tailing complies to a fractal distribution with a constant scale parameter.

Definition 5 If the probability distribution function of random variable X is

$$f(x):\ \alpha\sigma^{-\alpha}x^{\alpha-1},\ \alpha > 0,\ x \to 0, \quad (9)$$

then, the random variable X is subject to an asymptotic power function distribution. Where σ is the scale parameter and α is the feature index. The asymptotic power cumulative distribution function is,

$$F(x) = P(X \leqslant x):\left(\frac{x}{\sigma}\right)^{\alpha},\ \alpha > 0,\ x \to 0. \quad (10)$$

From Eq. (10), as x approaches to 0, if $P(X \leqslant x)$ and $x_i (i = 1, 2, 3 \cdots n)$ are plotted on a surface, they will converge to a straight line with the slope α. Then,

Proposition 2 The asymptotic power distribution in the high-end tailing is a fractal distribution with a constant scale parameter.

Definition 6 If the probability distribution function of random variable X is

$$f(x) = \alpha\sigma^{\alpha}x^{-\alpha-1},\ \alpha > 0,\ x \geqslant \sigma > 0, \quad (11)$$

then the random variable X is subject to a Pareto distribution. Where σ is the scale parameter and X is the feature index. From Eq. (11) it can be found that the mathematical expectation and variance for X are,

$$F(x) = P(X \leqslant x) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha},\ \alpha > 0, x \geqslant \sigma > 0;$$
$$E(x) = \alpha\sigma(\alpha-1)^{-1},\ \alpha > 1;$$
$$V(x) = \alpha\sigma^{2}(\alpha-2)^{-1}(\alpha-1)^{-2},\ \alpha > 2. \quad (12)$$

Mandelbrot argues that in the lower tailing truncation, Pareto distribution shows a similar form with $P(X \leqslant x)$ and fractal characteristics,

$$P(X \geqslant x|X \geqslant \sigma_1) = \left(\frac{x}{\sigma}\right)^{-\alpha}\left(\frac{\sigma_1}{\sigma}\right)^{\alpha} = \left(\frac{x}{\sigma_1}\right)^{-\alpha}, \ x \geqslant \sigma_1 \geqslant \sigma > 0. \quad (13)$$

In addition, as following holds

$$P(X \geqslant cx|X \geqslant c\sigma_1) = \left(\frac{cx}{c\sigma}\right)^{-\alpha}\left(\frac{c\sigma_1}{c\sigma}\right)^{\alpha} = \left(\frac{x}{\sigma_1}\right)^{-\alpha}, \quad (14)$$

then,

$$P(X \geqslant x|X \geqslant \sigma_1) = P(X \geqslant cx|X \geqslant c\sigma_1), x \geqslant \sigma_1 \geqslant \sigma > 0. \quad (15)$$

Eq. (15) shows that Pareto distribution carries a constant scaling parameter. Where, σ_1 is the scale parameter, and c is a positive constant. In summary, we obtain:

Proposition 3 Pareto distribution has a constant scaling parameter in the lower tailing truncation.

Definition 7 If the probability distribution function of random variable X is,

$$f(x): \alpha\sigma^{\alpha}x^{-\alpha-1}, \ \alpha > 0, \ x \to \infty, \quad (16)$$

the random variable X is then subject to an asymptotic Pareto distribution. Where σ is the scale parameter and α is the feature index. The function of the asymptotic Pareto distribution is,

$$F(x) = P(X \leqslant x): 1 - \left(\frac{x}{\sigma}\right)^{-\alpha}, \ \alpha > 0, \ x \to \infty. \quad (17)$$

Similarly,

Proposition 4 The asymptotic Pareto distribution is a fractal distribution with constant scale in lower tailing truncation.

Definition 8 If a transformed random variable $Y = \ln X$ follows the normal distribution, then the non-negative random variable X obeys a logarithmic normal distribution. If the density function of variable X is,

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \ x > 0, \quad (18)$$

then the random variable X follows a lognormal distribution. From Eq. (18), the distribution function, expectation and the variance of X are,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^x \frac{1}{x}\exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]\mathrm{d}x;$$

$$E(x) = \exp\left(\mu + \frac{\sigma}{2}\right);$$

$$D(x) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]. \quad (19)$$

Denote $X' = cX$, and c is a positive constant, and the density function for X' is,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[\frac{-(\ln x - \mu - \ln c)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left[\frac{-(\ln x - \mu')^2}{2\sigma^2}\right], \ x > 0, \quad (20)$$

where, $\mu' = \mu + \ln c$. Eq. (20) shows that X and X' follow the same distribution (lognormal distribution). A random variable's distribution formation remains unchanged by multiplying a positive constant number, which means that the lognormal distribution function holds the multifractal properties.

Proposition 5 The lognormal distribution function holds a fractal property.

3 Generalized extreme value-Pareto distribution function

In the field of marine engineering and study of coastal disaster prevention, Weibull and Gumbel extreme value distributions are often adopted to estimate the occurrence frequency of wave height's extreme value. Nonetheless, the generalized extreme value distribution is used for the modeling, which is helpful to understand these distributions in the macro scope. This section establishes the generalized extreme-Pareto distribution model based on the fractal theory and provides the analytical solution to the Weibull-Pareto distribution.

Theorem 1 Suppose random variable T follows the generalized extreme value distribution, random variable X can be either a discrete or continuous, then exists family of generalized distribution $G(x)$, and its density function is $g(x)$,

$$g(x) = \frac{1}{\sigma_1}\frac{f(x)}{1 - F(x)}\exp\left[-\left(1 + \frac{\xi\{-\log[1 - F(x)]\}}{\sigma_1}\right)^{\frac{-1}{\xi}}\right] \left[-\left(1 + \frac{\xi\{-\log[1 - F(x)]\}}{\sigma_1}\right)^{\frac{-1}{\xi}-1}\right], \quad (21)$$

where ξ, σ, β are parameters, and $t \geqslant 0$.

Alzaatreh et al. (2013) proposed a family of generalized distribution and uses random variable T as a supporting variable when deriving the impacts of T on the distribution of X.

Proof: construct a distribution function as follows:

$$G(x) = \int_0^{-\log[1-F(x)]} r(t)\mathrm{d}t = R\{-\log[1 - F(x)]\}, \quad (22)$$

where $F(X)$ is the cumulative distribution function for X, and $r(t)$ is the density function for the random variable T, and $T \in [0, \infty)$. $R(t)$ is the cumulative density function of the random variable T. The corresponding probability density function associated with Eq. (22) is,

$$g(x) = \frac{f(x)}{1 - F(x)}r\{-\log[1 - F(x)]\} = h(x)r\{-\log[1 - F(x)]\}, \quad (23)$$

where $h(x)$ is the hazard function for the random variable X with the cumulative density function of $F(x)$. The quantile function, $Q(\lambda)$, $0 < \lambda < 1$, for the $T - X$ family of distributions can be computed by using the formula,

$$Q(\lambda) = F^{-1}[1 - \mathrm{e}^{-R^{-1}(\lambda)}]. \quad (24)$$

Suppose the random variable T follows generalized extreme value distribution,

$$r(t)=\frac{1}{\sigma_1}\exp\left[-\left(1+\frac{\xi t}{\sigma_1}\right)^{\frac{-1}{\xi}}\right]\left[-\left(1+\frac{\xi t}{\sigma_1}\right)^{\frac{-1}{\xi}-1}\right],\tag{25}$$

where σ_1 and ξ are parameters, $t\geqslant 0$. By merging Eq. (25) into Eq. (22), the following can be obtained,

$$G(x)=\int_0^{-\log[1-F(x)]}\frac{1}{\sigma_1}\exp\left[-\left(1+\frac{\xi t}{\sigma_1}\right)^{\frac{-1}{\xi}}\right]\left[-\left(1+\frac{\xi t}{\sigma_1}\right)^{\frac{-1}{\xi}-1}\right]\mathrm{d}t\tag{26}$$

Take a derivative on $G(x)$ with respect to x,

$$g(x)=\frac{1}{\sigma_1}\frac{f(x)}{1-F(x)}\exp\left[-\left(1+\frac{\xi\{-\log[1-F(x)]\}}{\sigma_1}\right)^{\frac{-1}{\xi}}\right]\left[-\left(1+\frac{\xi\{-\log[1-F(x)]\}}{\sigma_1}\right)^{\frac{-1}{\xi}-1}\right].\tag{27}$$

Inference 1 If X follows Pareto distribution, then

$$g(x)=\beta x^{-1}\exp\left\{-\left[1+\log(x/\sigma)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\left\{-\left[1+\log(x/\sigma)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\right\}.\tag{28}$$

Proof: take $f(x)=\alpha\sigma x^{-\alpha-1}$, $\alpha>0$, $x\geqslant\sigma>0$, into Eq. (21),

$$g(x)=\frac{1}{\sigma_1}\alpha x^{-1}\exp\left[-\left(1+\frac{\xi[(-\log(x/\sigma)^{-\alpha})]}{\sigma_1}\right)^{\frac{-1}{\xi}}\right]\left[-\left(1+\frac{\xi[-\log(x/\sigma)^{-\alpha}]}{\sigma_1}\right)^{\frac{-1}{\xi}-1}\right]=\frac{\alpha}{\sigma_1}x^{-1}\exp\left\{-\left[1+\log(x/\sigma)^{\frac{\alpha\xi}{\sigma_1}}\right]^{\frac{-1}{\xi}}\right\}\left\{-\left[1+\log(x/\sigma)^{\frac{\alpha\xi}{\sigma_1}}\right]^{\frac{-1}{\xi}-1}\right\}.\tag{29}$$

Assume $\frac{\alpha}{\sigma_1}=\beta$, it can be simplified to

$$g(x)=\beta x^{-1}\exp\left\{-\left[1+\log(x/\sigma)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\left\{-\left[1+\log(x/\sigma)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\right\}.\tag{30}$$

Definition 9 If a random variable T follows generalized extreme value distribution, random variable X follows Pareto distribution, $f(x)=\alpha\sigma x^{-\alpha-1}$, $\alpha>0$, $x\geqslant\sigma>0$, then Eq. (30) is the generalized extreme value-Pareto distribution, and denoted as *GPD*(β,σ,ξ). In Eq. (30), when ξ takes values, random variable X follows Pareto distribution, then Inference 2 exists.

Inference 2 If $\frac{-1}{\xi}=1$, Eq. (30) can be simplified to

$$g(x)=\beta x^{-1}\exp\left\{-\left[1+\log(x/\sigma)^{-\beta}\right]\right\}=\frac{\beta}{x}\exp\left\{-\left[\log\mathrm{e}(x/\sigma)^{-\beta}\right]\right\}=-\mathrm{e}\beta\sigma^{\beta}x^{-1-\beta},\tag{31}$$

which is similar to a Pareto distribution. In Eq. (30), when σ takes values that would result in random variable X follows Pareto distribution, we have Inference 3 below. Eq. (30) gives explicit expression of Eq. (32).

Inference 3 If $\sigma=1$, Eq. (30) can be simplified to,

$$g(x)=\beta x^{-1}\exp\left\{-\left[1+\log(x)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\left\{-\left[1+\log(x)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\right\}=\beta x^{-1}\exp\left[-(\log\mathrm{e}x^{\beta\xi})^{\frac{-1}{\xi}}\right]\left[-(\log\mathrm{e}x^{\beta\xi})^{\frac{-1}{\xi}-1}\right].\tag{32}$$

The above can be simplified to a Logarithm-Weibull distribution.

We can deduct the distribution function of Eq. (30) in Eq. (33), which means $g(x)$ is the density function of $G(x)$.

Inference 4 If distribution function $G(x)$ is known:

$$G(x)=\exp\left\{-\left[1+\log(x/\sigma)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}.\tag{33}$$

And take the derivative on Eq. (29) respect to x.

$$\begin{aligned}g(x)=&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}'_x=\\&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\times\left(-\frac{1}{\xi}\right)\times\\&\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\right\}\times\left[1+\log(x/\alpha)^{\beta\xi}\right]'_x=\\&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\times\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\times\\&\frac{1}{\xi}\times\left[\log(x/\alpha)^{\beta\xi}\right]'_x=\\&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\times\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\times\\&\frac{1}{\xi}\times\frac{1}{(x/\alpha)^{\beta\xi}}\times\left[(x/\alpha)^{\beta\xi}\right]'_x=\\&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\times\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}\times\\&\frac{1}{\xi}\times\frac{1}{(x/\alpha)^{\beta\xi}}\times\left(\frac{1}{\alpha}\right)^{\beta\xi}\times\left(x^{\beta\xi}\right)'_x=\frac{\beta}{x}\times\\&\exp\left\{-\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}\times\left[1+\log(x/\alpha)^{\beta\xi}\right]^{\frac{-1}{\xi}-1}.\end{aligned}\tag{34}$$

In summary, if the parameters take specific values, the *GPD* distribution is reduced to the form of a logarithmic normal distribution and a Pareto distribution. Therefore, the *GPD* can include both the properties of the Weibull distribution and those of the Pareto distribution, which highlights the advantages of the newly proposed model. When random variable T follows Weibull distribution, then Inference 5 exists. The expression in the theorem is the explicit form of Eq. (39).

Inference 5 If a random variable T follows a Weibull distribution,

$$r(t)=\left(\frac{c}{\gamma}\right)\left(\frac{t}{\gamma}\right)^{c-1}\exp\left[-\left(\frac{t}{\gamma}\right)^{c}\right], \tag{35}$$

where, c and γ are parameters, and $t\geqslant 0$. Then, the Weibull-Pareto distribution can be explicitly written as,

$$g(x)=\frac{\beta c}{x}\left[\beta\log\left(\frac{x}{\sigma}\right)\right]^{c-1}\exp\left\{\left[\beta\log\left(\frac{x}{\sigma}\right)\right]^{c}\right\},$$
$$x\geqslant\sigma>0,\ \ c>0,\ \ \beta>0 \tag{36}$$

Proof: assume that T follows a Weibull distribution,

$$r(t)=\left(\frac{c}{\gamma}\right)\left(\frac{t}{\gamma}\right)^{c-1}\exp\left[-\left(\frac{t}{\gamma}\right)^{c}\right], \tag{37}$$

where, c and γ are parameters, and $t\geqslant 0$. By merging Eq. (37) with Eq. (22), the density function of $G(x)$ can be obtained:

$$g(x)=\frac{\alpha c}{\gamma x}\left[\frac{\alpha}{\gamma}\log\left(\frac{x}{\sigma}\right)\right]^{c-1}\exp\left\{-\left[\frac{\alpha}{\gamma}\log\left(\frac{x}{\sigma}\right)\right]^{c}\right\},\ x\geqslant\sigma>0. \tag{38}$$

Denote $\beta=\frac{\alpha}{\gamma}$, above derivation can be simplified to,

$$g(x)=\frac{\beta c}{x}\left[\beta\log\left(\frac{x}{\sigma}\right)\right]^{c-1}\exp\left\{\left[\beta\log\left(\frac{x}{\sigma}\right)\right]^{c}\right\},$$
$$x\geqslant\sigma>0, c>0,\beta>0. \tag{39}$$

Definition 9 Eq. (39) assumes X follows a Weibull–Pareto Distribution, and denoted as $X: WPD(c,\beta,\sigma)$.

Inference 6 If $\sigma=1$, then Eq. (39) can be written as:

$$g(x)=\frac{\beta c}{x}[\beta\log(x)]^{c-1}\exp\{[-\beta\log(x)]^{c}\}. \tag{40}$$

Eq. (40) represents a logarithmic Weibull.

If $c=1$ then Weibull–Pareto distribution can be written as:

$$g(x)=\frac{\beta}{x}\exp\left[-\beta\log\left(\frac{x}{\sigma}\right)\right]=\frac{\beta}{x}\left(\frac{x}{\sigma}\right)^{-\beta}=\beta\sigma^{\beta}x^{-1-\beta}. \tag{41}$$

Eq. (41) shows that Pareto distribution is a special case of Weibull–Pareto distribution. Moreover, to simplify Eq. (33), if $\frac{-1}{\xi}=1$ then

$$G(x)=\exp\left\{-\left[1+\log\left(\frac{x}{\sigma}\right)^{\beta\xi}\right]^{\frac{-1}{\xi}}\right\}=$$
$$\exp\left\{-\left[1+\log\left(\frac{x}{\sigma}\right)^{-\beta}\right]\right\}=$$
$$\exp\left[-\log\left(\frac{x}{\sigma}\right)^{-\beta}\right]=\frac{1}{\mathrm{e}}\left(\frac{x}{\sigma}\right)^{-\beta}. \tag{42}$$

Eq. (42) appears to be a power function with two parameters, which is the simplest fractal model. In the following section, the Weibull–Pareto distribution is applied to an engineering example. Based on the theories and the statistical characteristics of Weibull distribution and Pareto distribution, the designed wave height and water level are estimated, which reflect both the extreme value feature as well as the features of the data beyond the threshold. Together, they somewhat cover the fractal features.

In engineering practice, the main problem is: given R, where $0<R<1$, solve:

$$G(x)=T. \tag{43}$$

It is usually expressed as $T=1-R$, where T is the design frequency. Then,

$$N=\frac{1}{T}=\frac{1}{1-G(x)}. \tag{44}$$

We call N the return period of wave height random variable X. If the equation is satisfied, which means $G(x)=R$, then we call it N-year return period. In practical engineering problems, we utilize the fractal properties at the tail of data, using statistical analysis of shorter time period of wave height data, and extend it to calculate design wave height over longer period of time.

4 An engineering case study based on WeibullPareto model

Since the model derived in this paper is extreme value statistic characteristic model, it can also be used in the calculation of other marine environment design parameters, like design water level. This section includes a Weibull–Pareto distribution case study of wave height data measured between 1963 to 1988 in the Chaolian Island (35°53.6′N 120°53.1′E) in Shandong Province, China (the data in the year 1976 is missing). The whole data sample is segmented into 4 groups: with the observed wave height data, the whole data sample is segmented into 4 groups based on different time periods. Group A26 is a dataset between 1963 and 1988, with annual extreme value; Group A16 is a subset of A26, where the data are between 1973 and 1988; B26 is a dataset between 1963 and 1988 where the values are above a threshold, which will be further explained in the next paragraphs; B16 is a subset of B26, where the time span is from 1973 to 1988. Average wave height A26 is 3.727 m, while A16 is 3.768 m.

Assume that Chaolian Island's wave height is a random variable X and firstly a rescaled range (*R/S*) analysis is conducted on X. The autocorrelation in X's time series is examined based on the Hurst index test, and it is shown that there are fractal characteristics in the wave height's time series. Thus, the generalized extreme value Pareto distribution function can be used to capture its statistical characteristics.

The data beyond threshold method is used to determine the thresholds of the data groups B16 and B26. In the remaining life map, the trend lines' flat parts have less volatility and the data beyond thresholds can be obtained for B26 and B16 respectively, as shown in Fig. 1. The curves in Fig. 1 consist of observed data points. In Fig. 2 and Fig. 3, 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 is taken respectively, and the corresponding theoretical curves are drawn in the coordinate plane. Figs. 2 and 3 are obtained respectively,

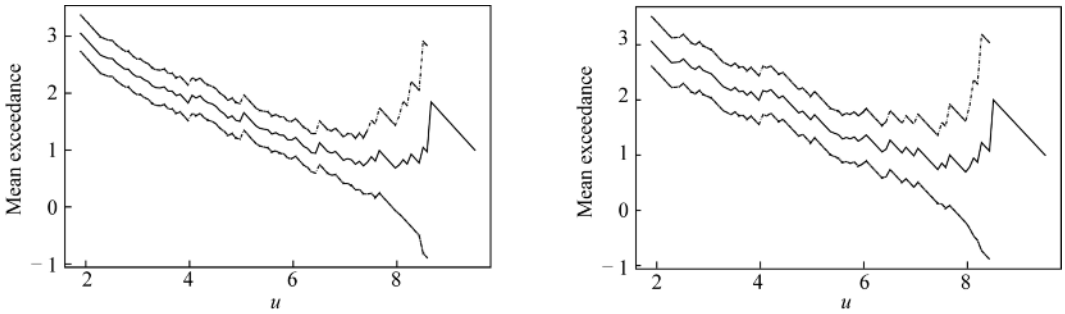

Fig. 1. Average remaining life plots for data groups B26 and B16.

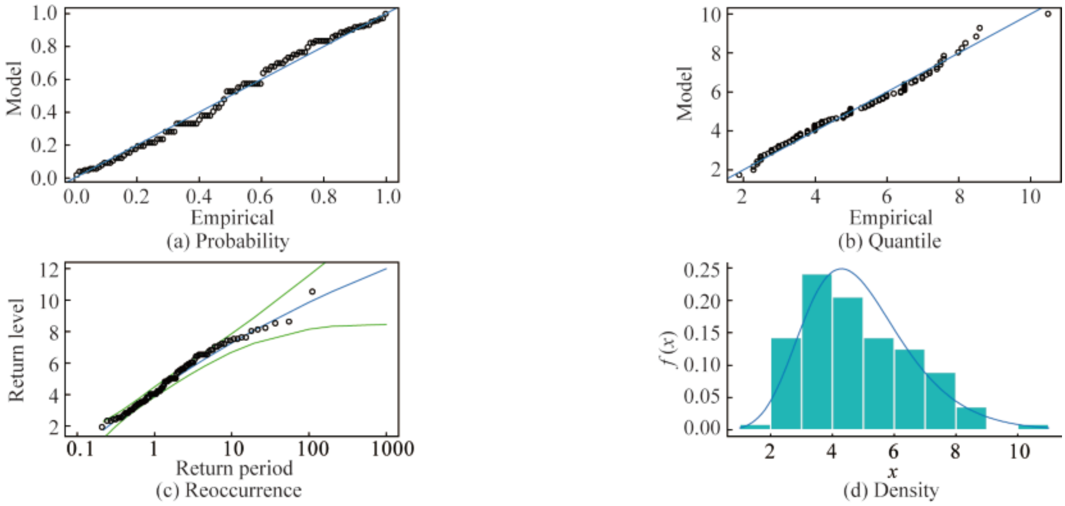

Fig. 2. B26 statistical description plot (a. Probability; b. Quantiles; c. Reoccurrence; d. Density).

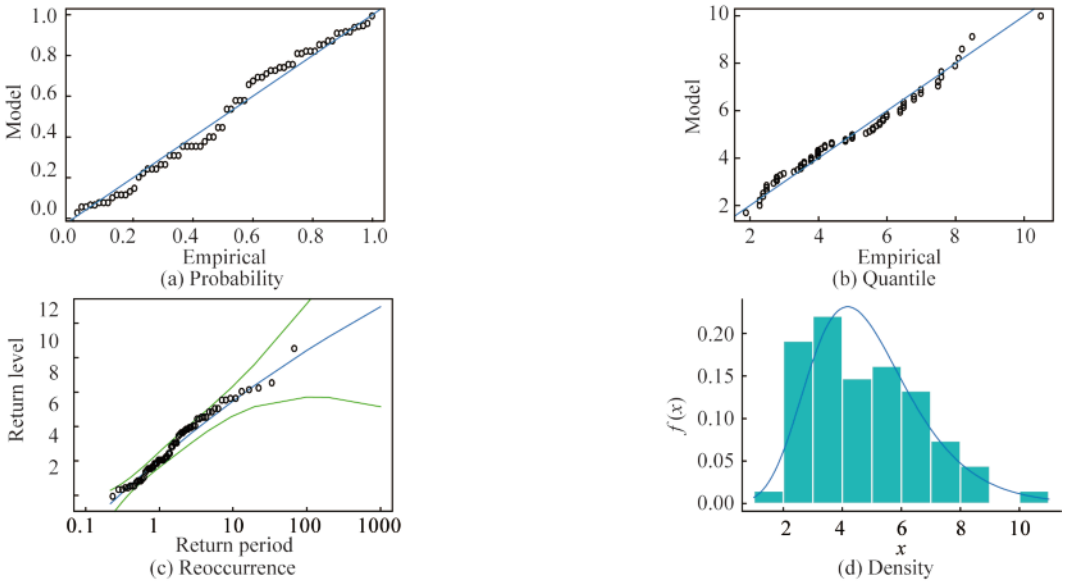

Fig. 3. B16 statistical description plot (a. Probability; b. Quantiles; c. Reoccurrence; d. Density).

which shows that it is reasonable to fit the measured data with the theoretical extreme value model.

The statistical description plot in Fig. 2 shows that for the sample group B26, the selected threshold sample fits very well with the generalized extreme value distribution within the confidence level of 95%. Similar results are shown in Fig. 3 for sample group B16. The density curve also fits well, which shows that the selected threshold is appropriate. Thus, the sample groups B26 and B16 can be used for modeling in the extreme value distribution. It is known that the distribution of wave height's annual extreme value is normally distributed with certain bias. The shape of the distribution can be captured by a spiky peak and long fat tail.

Table 1 shows the statistical characteristics for the wave height time series. The fractal characteristics of wave height time series are further discussed. It is a stochastic process with a deterministic trend. Table 2 analyzes the wave height's time series with R/S method with least square method, and R/S formula, Hurst index H, transition function $c(t)$, and R^2. The Hurst index H is a constant between 0 and 1. R and S are the range and mean square deviation respectively, which can reflect the type of data.

Table 1 Statistical characteristics of wave height data

Data group	A26	A16	B26	B16
$E(X)$	3.72	3.70	3.62	3.64
Skewness	3.54	3.57	3.60	3.53

According to the correlation of the motion trajectories of fractional Brownian motion, if $c(t) \geqslant 0$, as H approaches to 1, the stronger the longer-range correlation of the time series is, the longer memory embedded in the wave height time series will be. In other words, if the time series is trending up historically, there will also be an increasing trend in the future. Contrarily, if the time series is trending down historically, there will also be a decreasing trend in the future. The double logarithmic plots of wave height time series A26, A16, B26, and B16 are shown in Fig. 4.

From Tables 2 and 3, it is reasonable to set wave height observed data curve to fit Weibull distribution and Pareto distribution. Table 3 and Table 4 list the Kolmogorov–Smirnov test (K-S testing) results for the fittings of Weibull distribution and Pareto distribution on the wave height series A26, A16, B26 and B16, respectively. The 95% confidence intervals of estimation of Weibull distribution and Pearson-III of the parameters are shown in Table 5 and Table 6. After checking the theoretical curves of wave height data, the WPD distribution model, Weibull distribution and Pearson-III distribution can be used to calculate the design wave heights of 100, 500, 700 and 1000 years respectively. The estimated wave height extreme values for 100, 500, 700, and 1000 years for WPD distribution, Weibull distribution, and Pearson-III distribution are shown in Table 7.

From Table 8, it can be seen that the designed wave height with a recurrence within differentiated years by the GPD distribution model is quite close to those of some common distributions. The estimated designed wave height is generally lower than the Weibull distribution and larger than the Pearson-III distribution. For instance, the 500-year and 1000-year in the A26 dataset, the newly proposed model is 1.601% and 3.222% lower than the Weibull distribution's standards, respectively, and 1.319% and 2.170% higher than the Pearson-III distribution standards, respectively. The designed wave height from the new model can provide a benchmark. As the new model is based on rigorous mathematical derivation, the design wave height derived from it can be applied to providing scientific support in the design

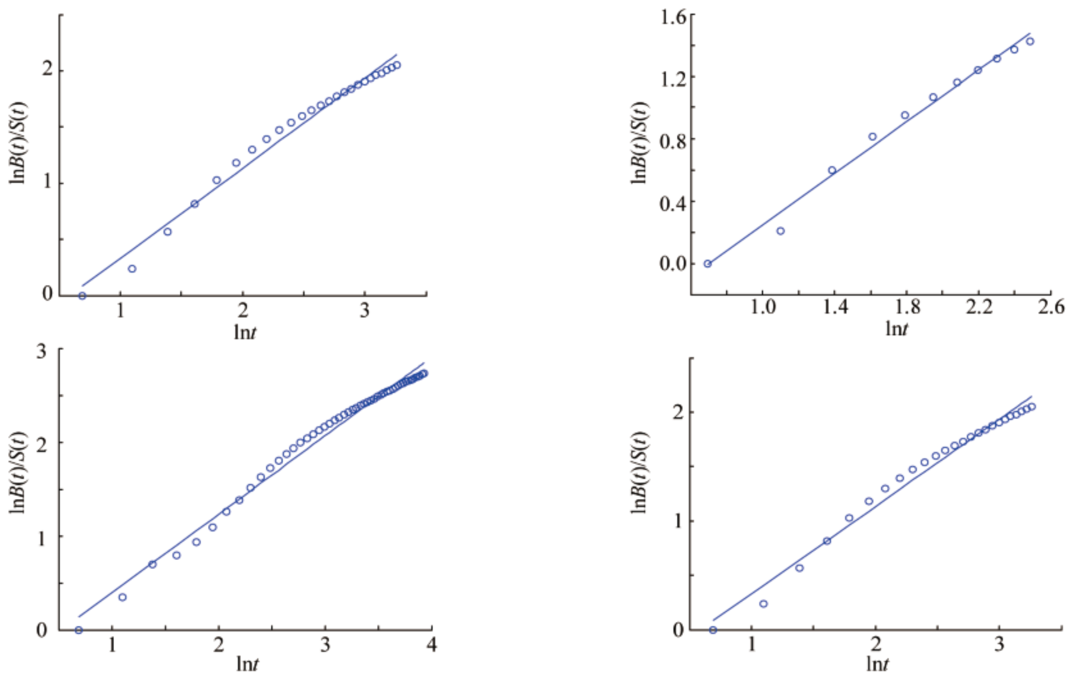

Fig. 4. Double logarithmic plot of wave height time series.

Table 2 Estimation of the *R/S* analysis

Data group	A26	A16	B26	B16
H	0.80	0.80	0.84	0.83
$c(t)$	0.52	0.51	0.60	0.54
R^2	0.98	0.96	0.98	0.97

Table 3 K–S testing results for the Weibull distribution

Data group	A26	A16	B26	B16
Test value D_n	0.15	0.17	0.16	0.20
	0.06	0.10	0.07	0.10
Threshold value D_0 (5%)	0.24	0.29	0.24	0.27
Comparison	D_n<D_0	D_n<D_0	D_n<D_0	D_n<D_0
Testing result	Accept	Accept	Accept	Accept

Table 4 K–S Testing results for the Pareto distribution

Data group	A26	A16	B26	B16
Test value D_n	0.1411	0.1509	0.1367	0.1477
	0.07	0.11	0.06	0.10
Threshold value D_0 (5%)	0.2347	0.2634	0.2334	0.2573
Comparison	D_n<D_0	D_n<D_0	D_n<D_0	D_n<D_0
Testing result	Accept	Accept	Accept	Accept

standard of breakwaters and levees.

5 Conclusion

This article studies a new calculation model for marine environment design parameters. The generalized extreme-Pareto distribution model based on fractal distribution and the analytic expression of Weibull–Pareto distribution model are provided. It is also shown that the commonly used function distributions exhibit fractal characteristics. For the new model, when it is used to estimate the designed wave height, the result is higher than that of the Pearson-III distribution, and lower than that of the extreme distribution. This indicates that when the new model is designed to calculate the wave height's annual extreme values, higher standards are needed than the Pearson-III distribution, i.e. the designed wave height's extreme value with 500-year return period. The new model's results are 1.601% lower than those of the Weibull distribution and 1.319% higher than those of the Pearson-III distribution.

Table 5 Estimated Weibull distribution parameters

Data group	A26	A16	B26	B16
a	40.11	40.76	38.26	40.24
Estimation interval	[36.45, 44.30]	[38.58, 43.73]	[36.75, 41.04]	[37.11, 43.00]
b	6.69	7.20	6.98	7.00
Estimation interval	[3.52, 6.23]	[5.96, 8.91]	[5.68, 8.76]	[8.23, 8.77]

Table 6 Estimated Pearson-III distribution parameters

Data group	A26	A16	B26	B16
a	35.46	36.12	34.67	35.23
Estimation interval	[19.42, 64.74]	[20.33, 64.20]	[20.59, 59.81]	[19.53, 62.01]
b	1.05	1.14	1.02	1.07
Estimation interval	[0.57, 2.01]	[0.45, 12.09]	[0.47, 1.68]	[0.62, 1.96]

Table 7 Designed wave height values for different return periods

Return period	Weibull distribution				Pearson-III distribution				GPD distribution			
	A26	A16	B26	B16	A26	A16	B26	B16	A26	A16	B26	B16
100	5.566	5.601	5.570	5.588	5.316	5.344	5.172	5.200	5.523	5.536	5.075	5.118
500	5.935	5.936	5.840	5.859	5.764	5.806	5.570	5.652	5.840	5.891	5.722	5.727
700	6.002	6.120	5.870	5.929	5.852	5.890	5.647	5.718	5.872	5.892	5.907	5.904
1000	6.070	6.201	5.890	5.981	5.943	6.014	5.728	5.756	6.072	6.179	5.953	5.958

Table 8 Comparison results in percentage

	W_{100}	W_{500}	W_{700}	W_{1000}	P_{100}	P_{500}	P_{700}	P_{1000}
A26	–0.772	–1.601	–2.166	–3.222	3.893	1.319	0.342	2.170
A16	–1.161	–0.758	–3.725	–0.355	3.593	1.464	0.034	2.743
B26	–8.887	–2.021	–0.630	–1.070	–1.876	2.729	4.604	3.928
B16	–8.411	–2.253	–0.422	–0.385	–1.577	1.327	3.200	3.509

Note: W_* represents the design wave height with return period of * years under the Weibull distribution; P_* represents the design wave height with return period of * years under the Pearson-III distribution.

The model for calculating the marine environment design parameters established in this research considers the fractal features, extreme-value statistics and beyond-threshold data. However, due to the impacts of various external environmental factors on the wave heights, the emerging mechanism can be complicated. Factors, such as the typhoon-induced wave and surface water level anomalies, pose challenges to determining the designed wave height in offshore engineering and are worth further investigation.

References

Alzaatreh, A., Lee, C. and Famoye, F., 2013. A new method for generating families of continuous distributions, *METRON*, 71(1), 63–79.

Barrs, A. and Chen, B.Y., 2018. *How Emerging Technologies Could Transform Infrastructure*, http://www.governing.com/gov-institute/voices/col-hyperlane-emerging-technologies-transform-infrastructure.html [2018-03-06].

Cai, W., Chan, J. and Garmire, D., 2011. 3-axes MEMS hall-effect sensor, *Proceedings of 2011 IEEE Sensors Applications Symposium*, IEEE, San Antonio, TX, USA, pp. 141–144.

Cai, W., Huang, L. and Wen, W.J., 2016. 2.4 GHz class AB power amplifier for wireless medical sensor network, *International Journal of Enhanced Research in Science, Technology & Engineering*, 5(4), 94–98.

Chen, B.Y. and Wang, B.Y., 2017. Location selection of logistics center in e-commerce network environments, *American Journal of Neural Networks and Applications*, 3(4), 40–48.

Chen, B.Y., Escalera, S., Guyon, I., Ponce-López, V., Shah, N. and Simón, M.O., 2016. Overcoming calibration problems in pattern labeling with pairwise ratings: application to personality traits, *Proceedings of Computer Vision-ECCV 2016 Workshops*, Springer, Amsterdam, The Netherlands, pp. 419–432.

Chen, B.Y., Liu, G.L. and Wang, L.P., 2017a. Predicting joint return period under ocean extremes based on a maximum entropy compound distribution model, *International Journal of Energy and Environmental Science*, 2(6), 117–126.

Chen, B.Y., Liu, G.L., Wang, L.P., Zhang, K.Y. and Zhang, S.F., 2018. Determination of water level design for an estuarine city, *Journal of Oceanology and Limnology*, 37(1), 20–29.

Chen, B.Y., Yang, Z.Y., Huang, S.Y., Du, X.Z., Cui, Z.W., Bhimani, J., Xie, X. and Mi, N.F., 2017b. Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles, *Proceedings of the 2017 IEEE 36th International Performance Computing and Communications Conference*, IEEE, San Diego, CA, USA.

de Haan, L. and Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer, New York.

Escalante, H.J., Ponce-López, V., Wan, J., Riegler, M.A., Chen, B.Y., Clapés, A., Escalera, S., Guyon, I., Baró, X., Halvorsen, P., Müller, H. and Larson, M., 2016. Chalearn joint contest on multimedia challenges beyond visual analysis: An overview, *Proceedings of the 2016 23rd International Conference on Pattern Recognition* (*ICPR*), IEEE, Cancun, Mexico, pp. 67–73.

Fu, H.L. and Liu, X.J., 2017. Research on the phenomenon of Chinese residents' spiritual contagion for the reuse of recycled water based on SC-IAT, *Water*, 9(11), 846.

Fu, H.L., Li, Z.X., Liu, Z.J. and Wang, Z.L., 2018. Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization, *Sustainability*, 10(7), 2488.

Jiang, S., Lian, M.J., Lu, C.W., Gu, Q.H., Ruan, S.L. and Xie, X.C., 2018. Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope, *Complexity*, 2018, 1048756.

Liu, D.F., Wang, L.P. and Pang, L., 2006. Theory of multivariate compound extreme value distribution and its application to extreme sea state prediction, *Chinese Science Bulletin*, 51(23), 2926–2930.

Liu, G.L., Chen, B.Y., Wang, L.P., Zhang, S.F., Zhang, K.Y. and Lei, X., 2018. Wave height statistical characteristic analysis, *Journal of Oceanology and Limnology*, 36(4), 1–13.

Liu, G.L., Zheng, Z.J., Wang, L.P., Chen, B.Y., Dong, Y.J., Xu, P.Y., Wang, J. and Wang, C., 2015. *Power-Type Wave Absorbing Device and Using Method Thereof*, CN 105113452B. (in Chinese)

Liu, X.J., He, Y.Q., Fu, H.L., Chen, B.Y., Wang, M.M. and Wang, Z.L., 2018. How environmental protection motivation influences on residents' recycled water reuse behaviors: A case study in Xi'an City, *Water*, 10(9), 1282.

Ponce-López, V., Chen, B.Y., liu, M.O., Corneanu, C., Clapés, A., Guyon, I., Baró, X., Escalante, H.J. and Escalera, S., 2016. ChaLearn LaP 2016: First round challenge on first impressions - dataset and results, *Proceedings of Computer Vision–ECCV 2016 Workshops*, Springer, Amsterdam, The Netherlands, pp. 400–418.

Wang, L.P. and Wang, L., 2013. *Method for Calculating Combined Return Period of Ocean Extreme Value Considering Typhoon Influence*, CN 102063527B. (in Chinese)

Wang, L.P., Chen, B.Y., Chen, C., Chen, Z.S. and Liu, G.L., 2016. Application of linear mean-square estimation in ocean engineering, *China Ocean Engineering*, 30(1), 149–160.

Wang, L.P., Chen, B.Y., Zhang, J.F. and Chen, Z.S., 2013. A new model for calculating the design wave height in typhoon-affected sea areas, *Natural Hazards*, 67(2), 129–143.

Wang, L.P., Huang, G.L., Chen, Z.S., Liang, B.C. and Liu, G.L., 2014. Risk analysis and assessment of overtopping concerning sea dikes in the case of storm surge, *China Ocean Engineering*, 28(4), 479–487.

Wang, L.P., Liu, G.L., Chen, B.Y. and Wang, L., 2015. *New Method for Calculating Typhoon-Influenced Sea Area Designed Wave Height Based on Maximum Entropy Principle*, CN 102063564B. (in Chinese)

Wang, L.P., Xu, X., Liu, G.L., Chen, B.Y. and Chen, Z.S., 2017. A new method to estimate wave height of specified return period, *Chinese Journal of Oceanology and Limnology*, 35(5), 1002–1009.

Xian, S.Y., Feng, K.R., Lin, N., Marsooli, R., Chavas, D., Chen, J. and Hatzikyriakou, A., 2018. Brief communication: Rapid assessment of damaged residential buildings in the Florida Keys after Hurricane Irma, *Natural Hazards and Earth System Sciences*, 18(7), 2041–2045.

Zhang K.Y. and Kleit, A.N., 2016. Mining rate optimization considering the stockpiling: A theoretical economics and real option model, *Resources Policy*, 47, 87–94.

Zhang, K.Y., Olawoyin, R., Nieto A. and Kleit, A.N., 2018. Risk of commodity price production cost and time to build in resource economics, *Environment, Development and Sustainability*, 20(6), 2521–2544.

Zhang, S.F., Shen, W., Li, D.S., Zhang, X.W. and Chen, B.Y., 2018. Nondestructive ultrasonic testing in rod structure with a novel numerical laplace based wavelet finite element method, *Latin American Journal of Solids and Structures*, 15(7), 1–17.