

Determination of water level design for an estuarine city*

CHEN Baiyu¹, LIU Guilin^{2,**}, WANG Liping^{3,**}, ZHANG Kuangyuan⁴, ZHANG Shuaifang⁵

¹ College of Engineering, University of California Berkeley, Berkeley 94720, USA

² College of Engineering, Ocean University of China, Qingdao 266000, China

³ College of Mathematical Science, Ocean University of China, Qingdao 266000, China

⁴ Department of Economics, Penn State University, State College 16801, USA

⁵ Department of Mechanical Engineering, University of Florida, Gainesville 32611, USA

Received Apr. 24, 2018; accepted in principle Aug. 8, 2018; accepted for publication Sep. 6, 2018

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract Based on the extreme value theory, self-affinity, and scale invariance, we studied the temporal and spatial relationship and the variation of water level and established a model of Gumbel-Pareto distribution for designed flood calculation. The model includes the previous extreme value models, the over-threshold data, and the fractal features shared by previous extreme value models. The model was simplified into a logarithmic normal distribution and a Pareto distribution for specific parameter values, and was used to calculate the designed flood values for the Shanghai Wusong Station in 100- and 1 000-year return periods. The calculated results show that the value of the designed flood height calculated in the Gumbel-Pareto distribution is between those in the Gumbel and Pearson-III distributions. The designed flood values in the 100- and 1 000-year return periods of the model were 0.03% and 0.11% lower, respectively, than the Gumbel distribution and 0.06% and 1.54% higher, respectively, than the Pearson-III distribution. Compared to the traditional model based solely on extreme probability, the Gumbel-Pareto distribution model could better describe the probabilistic characteristics of extreme marine elements and better use the data.

Keyword: self-affinity; scale invariance; extreme value distribution

1 INTRODUCTION

Located in the coast of the East China Sea by the Changjiang (Yangtze) River estuary, Shanghai is affected by marine environmental factors (e.g., winds, waves, currents, tides, and typhoons). However, the flood discharge of the Changjiang River is also an important factor on the calculation of the flood level design for Shanghai. In January 1963, the Bureau of Municipal Construction promulgated the first urban flood control standards that required the top elevation of the flood control walls near Huangpu Park in Shanghai to be at least 4.94 m. In November 1974, the Municipal Flood Control Headquarters promulgated a new flood control standard that required the defense flood volume around Huangpu Park to be 5.30 m and the top elevation of the flood control walls to be 5.80 m. On September 1, 1981, a tide with a measured level of 5.22 m (the measured tide level at Wusong was 5.74 m) occurred in the area of Huangpu Park. Around 1984, the 1 000-year return period annual

extreme flood volume of 5.86 m was established as the standard for the reconstruction of the protective walls. Within the context of a warming climate, rising sea level, and increased frequency of tropical storms, the *Comprehensive Treatment Planning Report of Huangpu River in Shanghai* was prepared by the Shanghai Survey and Design Institute of the Ministry of Water Resources and Electrical Power in December 1990. It proposed that the long-term flood prevention standard of Shanghai should be raised to resist the 10 000-year return period peak flood volume. Correspondingly, the flood volumes of Huangpu Park and Wuchang Station have been set at 6.34 and

* Supported by the NSFC-Shandong Joint Fund “Study on the Disaster-Causing Mechanism and Disaster Prevention Countermeasures of Multi-Source Storm Surges” (No. U1706226) and 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” (No. HDJG18007)

** Corresponding authors: liuguilin73@ouc.edu.cn; wlpjsh@163.com

6.81 m, respectively. However, the principal problem in calculating the design flood for a multiyear return period is how best to infer the statistical properties of a large-scale hydrological element, as accurately as possible, based on the statistical properties of a small-scale hydrological element.

Traditionally, calculation of the design flood of a multiyear return period involves selecting a probability distribution model as the probability distribution of the annual extreme flood volume, establishing probability distribution parameters based on the observed annual extreme flood volume, and determining the design flood of a multiyear return period based on the cumulative rate (Wang et al., 2013, 2016; Chen et al., 2017a; Liu et al., 2018). Examples of distribution models commonly used for such purposes are the Gumbel, Weibull, Pearson-III, and maximum entropy distributions (Liu et al., 2006; Wang et al., 2017a). However, when a flood volume is greater than a certain value, the probability density functions of the tails of these models tend to have the form of a power function. This power function has a fractal feature when the upper end is censored, which means the extremal model can describe the sharp peaks of the function well but it cannot fit the tail data (Wang et al., 2010; Wang and Wang, 2011; Liu et al., 2015). The Pareto distribution has scale and shape parameters, and a fractal feature when the lower end is censored, and it fits the tail data well (Chen et al., 2017b). Unfortunately, the peak of the distribution function does not perform satisfactorily.

The amount of observed data directly affects the accuracy of calculated hydrological parameters. What is usually obtained is the small-scale statistical variation rule of data due to various factors. The ability to transform data and produce accurate calculations in accordance with objective facts for different timescales requires study of the relationship among the different scales. Generally, large-scale change rules are calculated from small-scale measured data. For example, 30-year measured data can be used to calculate a design wave height in 100-year return period based on probability statistical analysis and related probability models, which uses the statistical self-affinity of the changes of marine hydrological elements. This is precisely the essential component of fractal theory. Harold Edwin Hurst (1880–1978), a famous hydrologist, proposed the *R/S* analysis method and the Hurst exponent based on studies of problems related to the Nile River and lakes/reservoirs, and accordingly analyzed the long-term correlation and

fractal features of the time series data (Hurst, 1951). Yang et al. (2017) first used the *R/S* method in trend analysis and mutation diagnosis of a hydrological time series of a river basin in North China, and verified its feasibility through empirical analysis. Currently, the concept of fractals has been abstracted into theory (Chen et al., 2016, 2017c; Escalante et al., 2016; Ponce-López et al., 2016; Chen and Wang, 2017; Fu et al., 2018). Research based on fractal data analysis has shown positive results because fractal analysis can provide a new perspective and method with which complex systems can be studied and analyzed, linking local details and the holistic properties of the system (Deng et al., 2017a; Fu and Liu, 2017; Geng et al., 2017; Zhang et al., 2017, 2018a; Jiang et al., 2017, 2018a, b).

This paper presents a model for designed flood calculation in combination with extreme value theory and fractal theory, i.e., an extreme-value Pareto distribution. The novelty of the model is that when the proposed model is used to calculate a design flood, it contains the previous extremal model or over-threshold data information. Furthermore, the model contains fractal features that were shared by extremal models in the past. In addition, it introduces the concepts of self-affinity and scale invariance of fractal theory, and it selects observational data with as many statistical features as possible and applies them to the calculation of a design flood. Thus, in comparison with the previous simplex extreme value probability model, the proposed model describes the probabilistic characteristics of extreme marine environmental elements more completely, and it constitutes a new probability distribution model that uses information more fully.

2 HURST LAW

Morphologically, a fractal is an object for which the integrity of the body and its parts are in some way similar. Mathematically, a fractal is a set in which the Hausdorff dimension is strictly greater than the topological dimension. A fractal has two key characteristics: self-affinity and scale invariance.

Definition 1: Let us assume a steady-state incremental random process of continuous time $\{X(t)\}$, if

$$\begin{aligned} X(t) &= \alpha^{-H} X(\alpha t), t \in T, \\ \forall \alpha > 0, 0 < H < 1, \end{aligned} \quad (1)$$

then, $X(t)$ is a single fractal, where H is the Holder index that does not change over time and does not

include the local properties at a certain moment. The equal sign indicates that the finite dimensional distributions are equal.

Fractal self-affinity means that the structures of fractal objects are similar irrespective of scale. This similarity is reflected between the part and the whole system, or between any individual parts. Scale invariance means that for any part of a fractal object, whether reduced or enlarged, any new diagram maintains its characteristics and properties.

The fractal dimension is an important parameter for describing the measurement and irregularity of fractal objects. It indicates the roughness of a fractal, which is usually not an integer. If a fractal object is a time series, the fractal dimension describes how uneven the data are; if it is a space, the fractal dimension describes how the space is filled (Mao et al., 2016; Wang et al., 2017b).

Studies have shown that most natural phenomena follow the ‘random walk with deviation’ law (Wang et al., 2017a). The general expression of Hurst’s law is as follows.

Definition 2: When a time series $\{x(t)\}$, $t=1, 2, \dots$ is considered, for any positive integer τ , there is empirically

$$R/S \sim \tau^H, \tag{2}$$

where H is a constant that is located in the $[0, 1]$ interval, known as the Hurst exponent, and the symbol “ \sim ” represents the direct ratio. The logarithms are taken on both sides of Eq.2 to obtain the Hurst exponent with the least squares’ method. Here, R and S represent the range and the mean square deviation, respectively, which are defined as

$$R(\tau) = \max D(t, \tau) - \min D(t, \tau), \quad 1 \leq t \leq \tau, \tag{3}$$

and

$$S(\tau) = \left\{ \frac{1}{\tau} \sum_{t=1}^{\tau} \left\{ x(t) - \overline{x(\tau)} \right\}^2 \right\}^{\frac{1}{2}}, \tag{4}$$

where $D(t, \tau)$ is the accumulated deviation.

According to Eqs.3 and 4, R/S is a dimensionless empirical statistical characteristic quantity. Therefore, the Hurst law, as expressed in Eq.2, can relate the observation results of different timescales. Specifically, it can infer large-scale statistical rules from small-scale observation results. If the Hurst exponent values are different, the trends corresponding to the time series are different. For example, when, $H=1/2$, the sequence belongs to the random walk process and the sequences are unrelated. Conversely, when $1/2 < H < 1$ and $0 < H < 1/2$, the sequence trends are

enhanced or diminished, respectively, i.e., the sequences are related and belong to the random walk with deviation process. The Hurst exponent has been used widely in research into the measurement of sequence correlation and trend intensity.

Definition 3: If there is a random process $B_H(t)$, it is continuous and it meets the criterion:

$$B_H(t) - B_H(t_0) \sim \varepsilon |t - t_0|^H, \tag{5}$$

where t and t_0 are two different moments, H is a parameter, and $0 \leq H \leq 1$. Then, the random process $B_H(t)$ is called the fractional Brownian motion and the symbol “ \sim ” represents the direct ratio.

Fractional Brownian motion is one of the Markov random processes, which can be generated by developing a random walk model (Li and Burgueño, 2010; Cai and Shi, 2016; Cai et al., 2016a, b, c; Barrs and Chen, 2018; Deng et al., 2018, 2017b, c; Kang, 2018a, b). A parameter value of $H=0.5$ reflects ordinary Brownian motion. The difference between Brownian motion and fractional Brownian motion is that the increment in Brownian motion is independent, whereas the increment in fractional Brownian motion is not.

Fractional Brownian motion is characterized by using a statistical law of a previous moment to relate a future changing trend (Cai et al., 2016b; Zhang et al., 2018b). Without loss of generality, when $B_H(0)=0$, the relevant function $c(t)$ is:

$$\begin{aligned} c(t) &= \frac{E\{[B_H(0) - B_H(-t)][B_H(t) - B_H(0)]\}}{E[B_H(t) - B_H(0)]^2} \\ &= \frac{-E[B_H(-t)B_H(t)]}{E[B_H(t)]^2} \\ &= \frac{1}{2} \frac{E\{[B_H(-t)]^2 + [B_H(t)]^2 - [B_H(-t) - B_H(t)]^2\}}{E[B_H(t)]^2} \\ &= \frac{1(-t)^{2H} + t^{2H} - (-2t)^{2H}}{2t^{2H}} = 2^{2H-1} - 1. \end{aligned} \tag{6}$$

According to the relationship between parameter H and correlation coefficient $c(t)$, there are three types of time series motion: (i) when $H=1/2$ and $c(t)=0$, the sequence is independent and it is a random process with an independent increment; (ii) when $1/2 < H < 1$ and $c(t) > 0$, the sequence is positively correlated; and (iii) when $0 < H < 1/2$ and $c(t) < 0$, the sequence is negatively correlated. In an empirical problem, R/S analysis of a time series is conducted to determine whether it belongs to fractional Brownian motion

through the Hurst index value, and to study further the correlation between the sequences.

3 GENERALIZED EXTREME VALUE PARETO DISTRIBUTION MODEL

In risk assessment for flood control projects in coastal estuary cities, selecting the appropriate design flood distribution model is crucial. In this section, we established the Gumbel-Pareto distribution (GPD), i.e., a new model for design flood calculation, and we provided an explicit analytical expression of the GPD for practical applications in design flood projects.

Theorem: We assume random variable T is a continuous random variable that is defined in the interval $[a, b]$ and which uses $r(t)$ as a probability density function. When $-\infty \leq a < b \leq \infty$, x is a random variable with $f(x)$ as the density function, the distribution function is $F(x)$, and the distribution function $G(x)$ exists, which can be expressed as:

$$G(x) = \int_0^{-\log(1-F(x))} r(t) dt. \tag{7}$$

Proof: Alzaatreh provided a method for the generation of a continuous probability distribution family (Alzaatreh et al., 2013). When discussing the distribution of random variable X , it uses random variable T as an auxiliary variable, which shows the influence of random variable T on random variable X . First, the distribution function is defined for a new distribution function family:

$$G(x) = \int_a^{W(F(x))} r(t) dt, \tag{8}$$

where $W(F(x))$ is a function of the distribution function $F(x)$, which is differentiable, monotonous, and non-decreasing with $W(F(x))$ $[a, b]$. $F(x)$ is the distribution function of random variable x . When $x \rightarrow -\infty$, $W(F(x)) \rightarrow a$, and when $x \rightarrow +\infty$, $W(F(x)) \rightarrow b$. The distribution function $G(x)$ can be written as $G(x) = R\{W(F(x))\}$, which is a composite function, where $R(t)$ is the distribution function of random variable T . Apparently, $G(x)$ is the distribution function of a distribution function family, and its density function is:

$$g(x) = \left\{ \frac{d}{dx} W(F(x)) \right\} r\{W(F(x))\}, \tag{9}$$

where the random variable X can be discrete or continuous. Obviously, the density function $r(t)$ of random variable T is “transformed” into a new

distribution function $G(x)$ of random variable X through the “transfer function” $W(F(x))$, which can be called the “ T - X ” distribution. $G(x)$ shows in different forms with different $W(F(x))$ expressions. Without loss of generality, assuming $\alpha=0$ when the range of T is $[0, \infty)$, $W(F(x))$ can be defined as $-\log(1-F(x))$, $F(x)/(1-F(x))$, $-\log(1-F^\alpha(x))$ and $F^\alpha(x)/(1-F^\alpha(x))$, and $\alpha > 0$.

Specifically, when $W(F(x)) = -\log(1-F(x))$

$$G(x) = \int_0^{-\log(1-F(x))} r(t) dt = R\{-\log(1-F(x))\}, \tag{10}$$

where $R(t)$ is the distribution function of random variable T . The corresponding density function of Eq.10 is:

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

where $h(x)$ is risk function of random variable X .

In the selection of different functions according to the actual project $r(t)$, the corresponding $g(x)$ can have different expressions.

Corollary 1: When $r(t) = \theta e^{-\theta t}$, $g(x) = \theta f(x)(1-F(x))^{\theta-1}$

Corollary 2: When $r(t) = \frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}}$,

From Eq.7, we know that when discussing the statistical characteristics of random variable X , it takes random variable T as an auxiliary variable, which can reflect the influence of random variable T on the statistical characteristics of random variable X .

Corollary 3: If random variable T obeys the generalized extreme value distribution:

$$r(t) = \frac{1}{\sigma_1} \exp \left\{ - \left[\left(1 + \frac{\xi t}{\sigma_1} \right)^{\frac{-1}{\xi}} \right] \right\} \tag{11}$$

then,

$$g(x) = \frac{1}{\sigma_1} \frac{f(x)}{1-F(x)} \exp \left\{ - \left[\left(1 + \frac{\xi(-\log(1-F(x)))}{\sigma_1} \right)^{\frac{-1}{\xi}} \right] \right\} \left\{ \left[\left(1 + \frac{\xi(-\log(1-F(x)))}{\sigma_1} \right)^{\frac{-1}{\xi}} - 1 \right] \right\}, \tag{12}$$

where σ_1 and ξ are the parameters and $t \geq 0$.

Proof: We brought Eq.11 into Eq.7 and obtained the following:

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

The probability density function $g(x)$ of Eq.12 was derived from the derivation of $G(x)$ as an explicit expression.

Corollary 4: If random variable X obeys the Pareto distribution $f(x) = \alpha \sigma^\alpha x^{-\alpha-1}$, $\alpha > 0$, $x \geq \sigma > 0$ where α is the shape parameter and σ is the scale parameter, which is also the threshold, Eq.12 becomes

$$\begin{aligned} g(x) &= \frac{1}{\sigma_1} \alpha x^{-1} \exp\left\{-\left[1 + \frac{\xi(-\log(\frac{x}{\sigma})^{-\alpha})}{\sigma_1}\right]^{\frac{-1}{\xi}}\right\} \\ &\quad \left\{-\left[1 + \frac{\xi(-\log(\frac{x}{\sigma})^{-\alpha})}{\sigma_1}\right]^{\frac{-1}{\xi}-1}\right\} \\ &= \frac{\alpha}{\sigma_1} x^{-1} \exp\left\{-\left[1 + \log\left(\frac{x}{\sigma}\right)^{\frac{\alpha_\xi}{\sigma_1}}\right]^{\frac{-1}{\xi}}\right\} \\ &\quad \left\{-\left[1 + \log\left(\frac{x}{\sigma}\right)^{\frac{\alpha_\xi}{\sigma_1}}\right]^{\frac{-1}{\xi}-1}\right\}. \end{aligned} \tag{13}$$

If we make $\alpha/\sigma_1 = \beta$, Eq.13 is simplified to:

$$\begin{aligned} g(x) &= \beta x^{-1} \exp\left\{-\left[1 + \log\left(\frac{x}{\sigma}\right)^{\beta_\xi}\right]^{\frac{-1}{\xi}}\right\} \\ &\quad \left\{-\left[1 + \log\left(\frac{x}{\sigma}\right)^{\beta_\xi}\right]^{\frac{-1}{\xi}-1}\right\}. \end{aligned} \tag{14}$$

Definition 4: If X is a random variable subject to the Pareto distribution, $f(x) = \alpha \sigma^\alpha x^{-\alpha-1}$, $\alpha > 0$, $x \geq \sigma > 0$, Eq.14 is called the generalized extreme value Pareto distribution, which is denoted $GPD(\beta, \sigma, \xi)$.

Corollary 5: When random variable T is in the Gumbel distribution:

$$r(t) = b \exp(-bt) \exp(-\exp(-bt)), \tag{15}$$

then,

$$\begin{aligned} g(x) &= b \frac{f(x)}{1-F(x)} \exp(b \log(1-F(x))) \\ &\quad \exp(-\exp(b \log(1-F(x))))), \end{aligned} \tag{16}$$

where b is the parameter and $t \geq 0$. Equation 15 is brought into Eq.10 and the derivation of $G(x)$ is undertaken.

Corollary 6: When random variable T is in the Gumbel distribution and random variable X is subject to the Pareto distribution:

$$\begin{aligned} g(x) &= b \alpha x^{-1} \exp\left(b \log\left(\left(\frac{x}{\sigma}\right)^{-\alpha}\right)\right) \\ &\quad \exp\left(-\exp\left(b \log\left(\left(\frac{x}{\sigma}\right)^{-\alpha}\right)\right)\right) \\ &= b \alpha x^{-1} \left(\frac{x}{\sigma}\right)^{-b\alpha} \exp\left(-\left(\frac{x}{\sigma}\right)^{-b\alpha}\right) \\ &= \frac{b\alpha}{\sigma} \left(\frac{x}{\sigma}\right)^{-1-b\alpha} \exp\left(-\left(\frac{x}{\sigma}\right)^{-b\alpha}\right). \end{aligned} \tag{17}$$

Definition 5: Equation 17 indicates that random variable X obeys the GPD, which is denoted $X \sim GPD(b, \alpha, \sigma)$.

The characteristics of Eq.17 under different parameters are shown in Figs.1–4.

According to the derivation process of the GPD model theory, we can see that the generalized extreme-value Pareto distribution function contains the extreme-value information of the past functional models and it combines over-threshold data, while reflecting the fractal features. The Gumbel-Pareto distribution model was applied in the engineering practice as follow.

4 EXAMPLES OF THE APPLICATION OF THE GPD MODEL

Most major floods in Shanghai City have been related to storm surges under the influence of typhoons and they tend to be accompanied by an upstream flood peak anstau. Therefore, we used flood volume data measured during 1970–1990 at the Wusong Hydrographic Station (Shanghai) in the Changjiang River estuary and flood observation data from the Datong Hydrographic Station 624 km upstream of Wusong. The GPD model was used to calculate the design flood for multiyear return periods, and the calculated results were compared with the value of

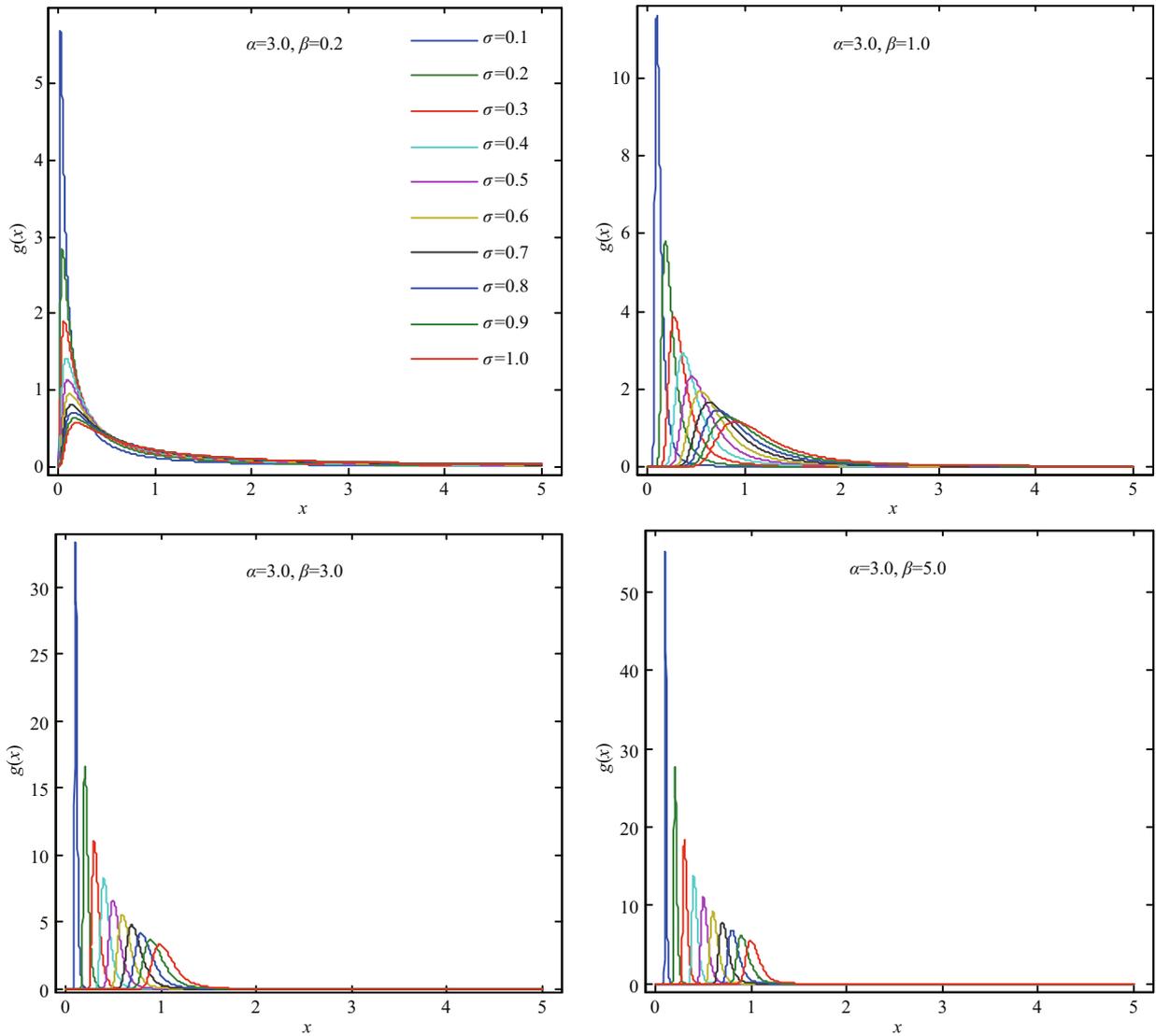


Fig.1 Probability density plots of the GPD ($\alpha=3.0, \beta=0.2; 1.0; 3.0; 5.0$)

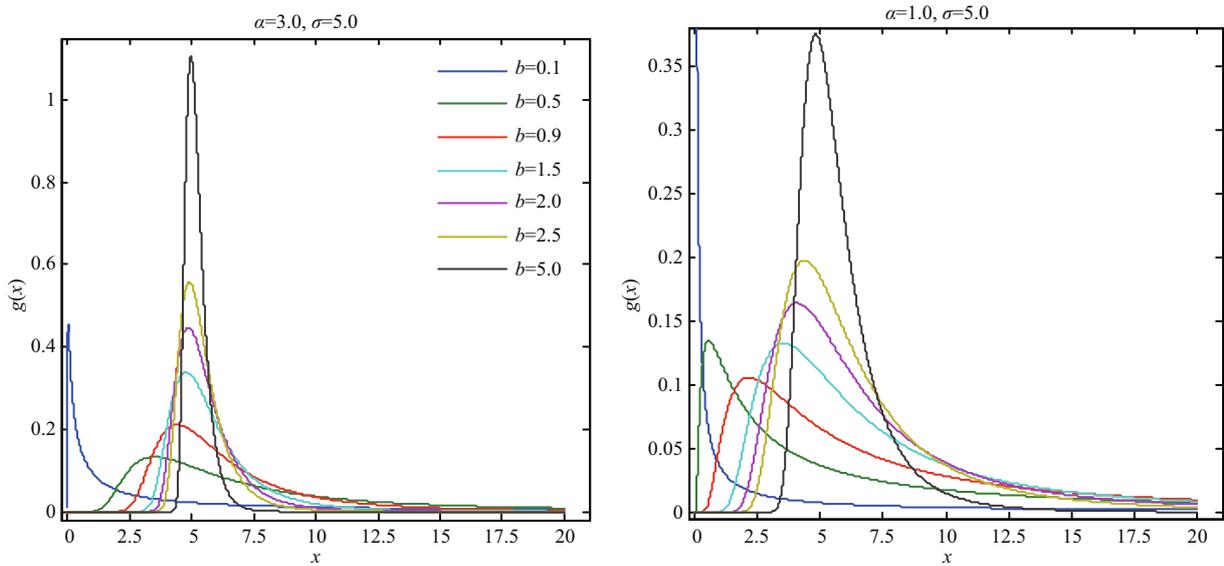


Fig.2 Probability density plots of the GPD ($\alpha=3.0, 1.0; \sigma=5.0$)

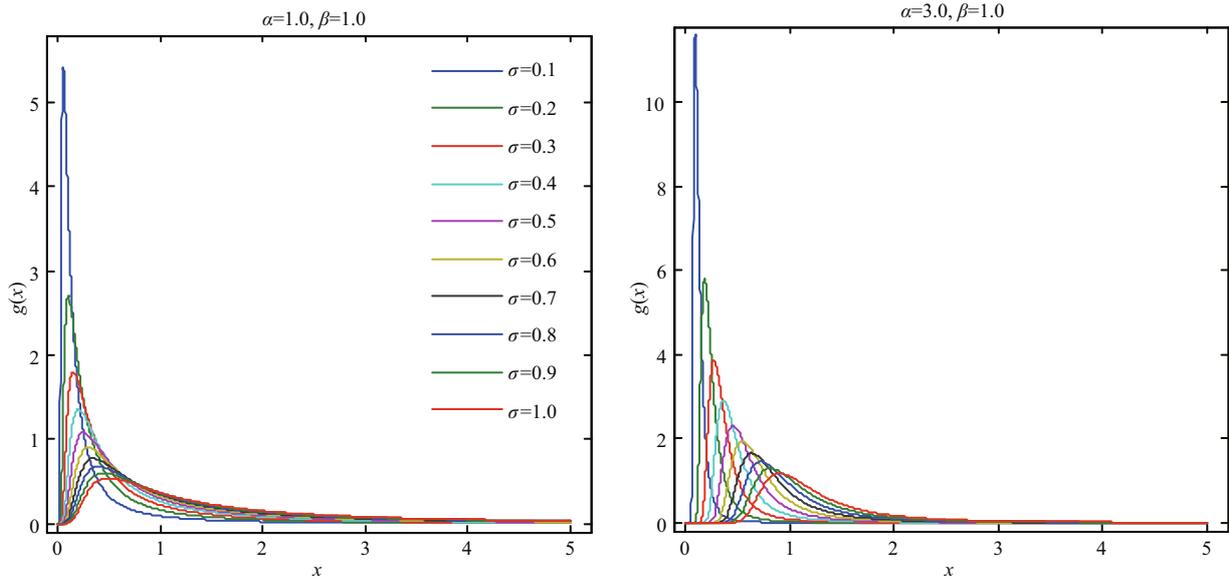


Fig.3 Probability density plots of the GPD ($\alpha=1.0, 3.0; \beta=1.0$)

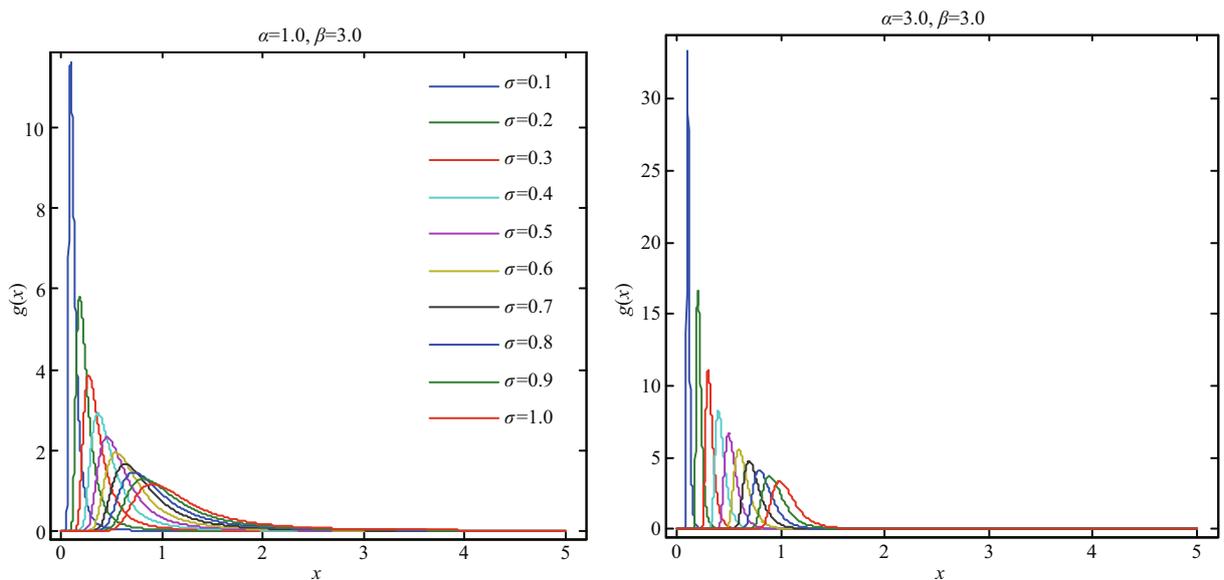


Fig.4 Probability density plots of the GPD ($\alpha=1.0, 3.0; \beta=3.0$)

the designed flood height calculated in the Gumbel and Pearson-III distributions.

Figures 5 and 6 show diagnostic tests of the generalized extreme value distribution mode including charts of probability, quantile, and return level and a density histogram. The circles in the figures represent data points and solid lines represent model curves. It can be seen that the quantiles of the observation points are in good agreement with the model. The return level plot shows that when the theoretical confidence interval is 95%, the curve relationship $x_p - \log(-\log(F(x_p)))$ is obtained by the generalized extreme-value distribution model, which

shows the distribution of the data fully within the confidence interval. The data points in the figures are distributed fully within the return level of the 95% confidence interval of the model. The density curve shows the distribution of the data intuitively, in that the observed data points (histogram) are in good agreement with the model curve.

The results of the above diagnostic tests show that the storm surge elevation and the corresponding upstream flood-peak anstau observation data comply with the generalized extreme value distribution and thus, can be used as an analysis sample of an extreme value distribution.

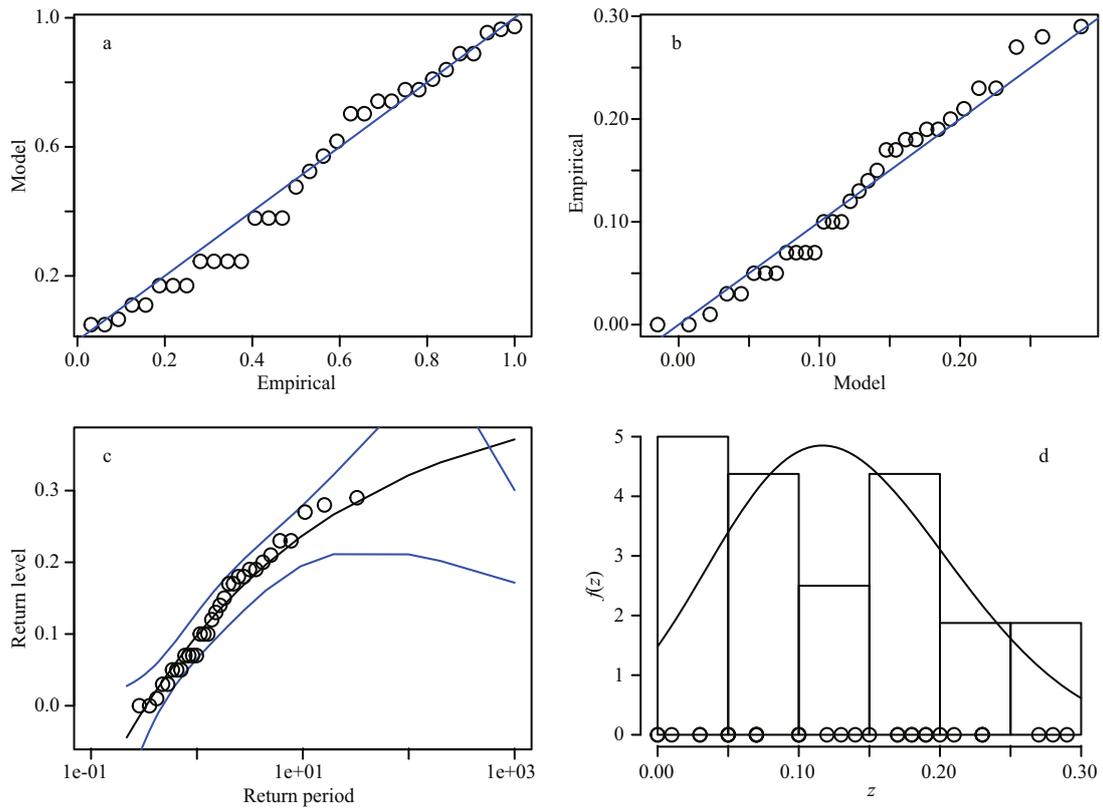


Fig.5 Flood peak anstau

a. probability; b. quantile; c. return level; d. density.

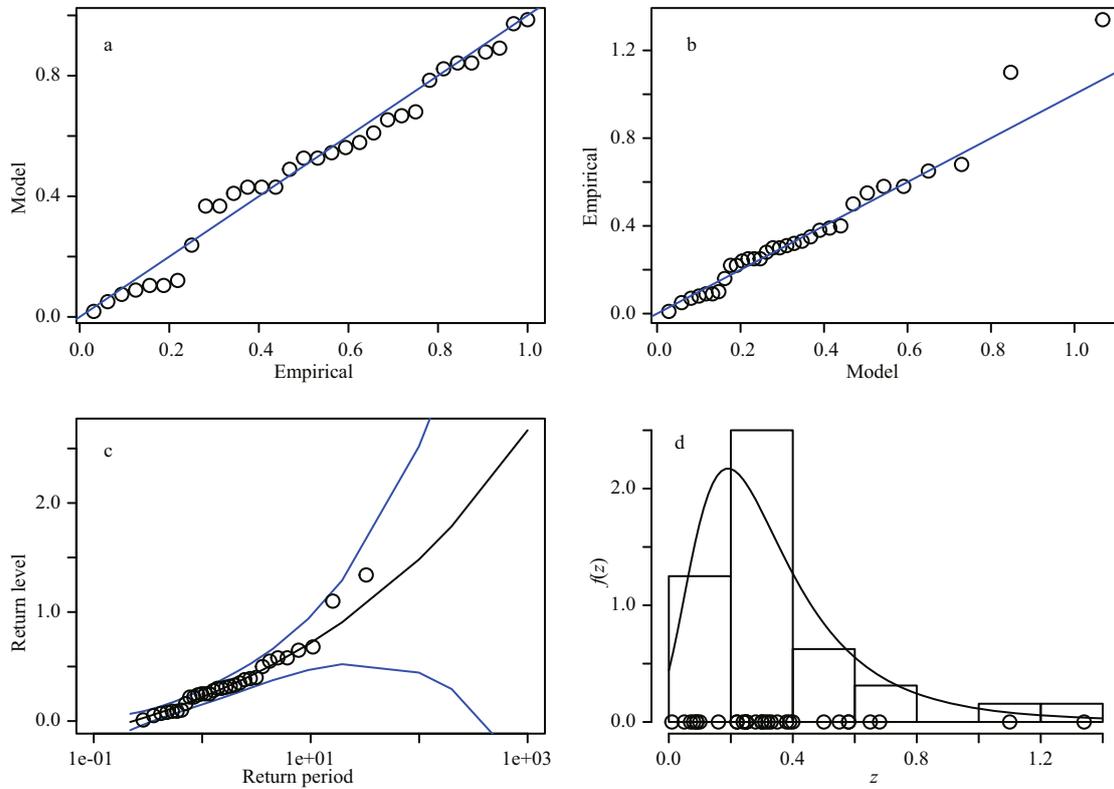


Fig.6 Storm surge elevation

a. probability; b. quantile; c. return level; d. density.

Table 1 Statistical characteristics

Statistical characteristics	Storm surge elevation	Upstream flood peak anstau
$E(X)$	2.410	0.163
Deviation	2.382	0.198

Table 2 R/S analysis and calculation values

	Storm surge elevation	Upstream flood peak anstau
H	0.761	0.812
$c(t)$	0.431	0.517
R^2	0.976	0.960

Table 3 K-S test of Gumbel distribution

Data set	Storm surge elevation	Flood
Test value D_n	0.14	0.16
	0.07	0.10
Critical value $D_0(0.05)$	0.22	0.27
Comparison	$D_n < D_0$	$D_n < D_0$
Test result	Accept	Accept

D_0 : the critical value of K-S test with a significance level of 0.05.

Table 1 shows the statistical values of a storm surge elevation and the flood peak corresponding upstream anstau time series. The distribution of the surge and flood sequences was normal and biased, with an extended right tail. The R/S relation was obtained using the least squares method, and the specific values of the Hurst index H , correlation function $c(t)$, and correlation coefficient R^2 are shown in Table 2. It can be seen that the storm surge elevation and flood time series are random processes with a certain trend. A fractal feature analysis of the storm surge and flood time series can be conducted using the R/S method. According to the relevance of the fractional Brownian motion trajectories, when $c(t) \geq 0$, the closer the value of H is to 1, and the stronger the long-range correlation of the time series; thus, the long-term memory of the storm surge elevation and flood time series can be obtained. In other words, if there has been a trend of increase in the past, there will be a trend of increase in the future. Conversely, if there has been a trend of decrease in the past, there will be a corresponding trend of decrease in the future. Tables 3 and 4 display the results of the K-S tests for the Gumbel and Pareto distributions of the storm surge elevation and upstream flood peak anstau. Tables 5 and 6 present the parameters and interval estimates of the 95% confidence levels when using the Gumbel distribution and Pearson-III fitting of the storm surge elevation and flood sequences.

Table 4 K-S test of Pareto distribution

Data set	Storm surge elevation	Flood
Test value D_n	0.17	0.16
	0.07	0.09
Critical value $D_0(0.05)$	0.23	0.25
Comparison	$D_n < D_0$	$D_n < D_0$
Test result	Accept	Accept

Table 5 Parameter calculation of Gumbel distribution

Data set	Storm surge elevation	Flood
μ	2.435 4	0.169 6
Standard deviation of μ	0.017 65	0.004 6
σ	0.138 0	0.023 0
Standard deviation of σ	0.016 5	0.005 2

μ : the shape parameter; σ : the scale parameter of the Gumbel distribution.

Table 6 Parameter calculation of Pearson-III distribution

Data set	Storm surge elevation	Flood
a	2.403	0.101
Standard deviation of a	0.011	0.002
b	0.002	0.013
Standard deviation of b	0.001	0.001

a : the shape parameter; b : the scale parameter of the Pearson-III distribution.

The design flood values for 100-, 200-, 400-, 500-, 700-, and 1 000-year return periods calculated using the Gumbel, Pearson-III, and GPD distributions are shown in Table 7. It can be seen that the design flood of the multiyear return period derived from the new GPD distribution model differs a little from that derived using the common distribution. The derived design floods are generally lower than the Gumbel distribution and larger than the Pearson-III distribution. Taking the 100- and 1 000-year return periods as examples, it can be seen that the new model is 0.06% and 1.54%, respectively, higher than the Pearson-III distribution standard and 0.03% and 0.11%, respectively, lower than the Gumbel distribution standard. These results show that the proposed model based on fractal and extreme value theories has certain value as a reference in relation to a design flood with a 100-year return period.

5 CONCLUSION

1) Based on fractal and extreme value theories, this study proposed a design flood calculation model based on the GPD, and the related properties of the new model were discussed.

Table 7 Design flood values for multiyear return period

Return period	Storm surge elevation	Gumbel		Storm surge elevation	Pearson-III		Storm surge elevation	GPD	
		Flood	Flood volume		Flood	Flood volume		Flood	Flood volume
100	1.599 5	0.486 4	6.225 9	1.599 0	0.486 0	6.225 0	1.599 3	0.486 3	6.225 6
200	1.930 9	0.575 4	6.646 3	1.930 1	0.570 0	6.640 1	1.930 0	0.575 1	6.645 1
400	2.248 0	0.694 3	7.082 3	2.247 0	0.694 3	7.081 3	2.247 5	0.694 2	7.081 7
500	2.493 0	0.725 1	7.358 1	2.487 8	0.724 7	7.352 5	2.490 0	0.724 7	7.354 7
700	2.869 2	0.800 1	7.809 3	2.849 9	0.788 0	7.777 9	2.858 7	0.788 6	7.787 3
1000	3.038 4	0.869 0	8.048 7	3.037 1	0.853 8	8.030 9	3.037 5	0.868 8	8.046 3

Astronomical tide value is 4.14 m; Unit.: m.

2) For specific parameter values, the GPD distribution was simplified into a logarithmic normal distribution and a Pareto distribution. Therefore, the GPD can include properties of both the Gumbel and Pareto distributions, highlighting a primary advantage of the new model.

3) For the calculation of a design flood, analysis methods such as fractal theory, an extreme value model, and an over-threshold model were used. However, many problems that will require further study remain. For example, the flood peak anstau and the storm surge elevation were considered as independent factors. In fact, many different combinations can constitute a 100-year return period. Therefore, a multivariable joint probability under certain controlling factor conditions will be the focus of future research.

6 DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

- Alzaatreh A, Lee C, Famoye F. 2013. A new method for generating families of continuous distributions. *Metron*, **71**(1): 63-79.
- Barrs A, Chen B Y. 2018. How emerging technologies could transform infrastructure. <http://www.governing.com/commentary/col-hyperlane-emerging-technologies-transform-infrastructure.html>. Accessed on 2018-3-10.
- Cai W, Huang L, Wang S. 2016a. Class D power amplifier for medical application. *International Journal of Informatics Engineering*, **4**(2): 9-15.
- Cai W, Huang L, Wen W J. 2016b. 2.4GHZ Class AB power amplifier for healthcare application. *International Journal of Biomedical Engineering and Science*, **3**(2): 1-6.
- Cai W, Huang L, Wen W J. 2017. SI-Based Class AB Power Amplifier for Wireless Medical Sensor Network. *Journal of Image and Graphics*, **5**(1): 34-38.
- Cai W, Shi F. 2016. High performance SOI RF switch for healthcare application. *International Journal of Enhanced Research in Science, Technology & Engineering*, **5**(10): 23-28.
- Chen B Y, Escalera S, Guyon I, Ponce-López V, Shah N, Simón M O. 2016. Overcoming calibration problems in pattern labeling with pairwise ratings: application to personality traits. In: Hua G, Jégou H eds. European Conference on Computer Vision (ECCV 2016) Workshops. Springer, Amsterdam, Netherlands. p.419-432, https://doi.org/10.1007/978-3-319-49409-8_33.
- Chen B Y, Liu G L, 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, Zhang J F. 2017b. Estimation method of designed wave height, capable of embodying influences of three typhoon factors: China Patent, CN201610972118. 2017-08-29. (in Chinese)
- Chen B Y, 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, <https://doi.org/10.11648/j.ajna.20170304.11>.
- Chen B Y, Yang Z Y, Huang S Y, Du X Z, Cui Z W, Bhimani J, Xie X, Mi N F. 2017c. Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles. In: Proceedings of the 36th IEEE International Performance Computing and Communications Conference. IEEE, San Diego, California, USA. p.1-6.
- Deng W, Yao R, Zhao H M, Yang X H, Li G Y. 2017a. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Computing*, (2-4): 1-18.
- Deng W, Zhang S J, Zhao H M, Yang X H. 2018. A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing. *IEEE Access*, **6**(1): 35 042-35 056.
- Deng W, Zhao H M, Yang X H, Xiong J X, Sun M, Li B. 2017b. Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment. *Applied Soft Computing*, **59**: 288-302.
- Deng W, Zhao H M, Zou L, Li G Y, Yang X H, Wu D Q. 2017c. A novel collaborative optimization algorithm in solving complex optimization problems. *Soft Computing*, **21**(15): 4 387-4 398.

- 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, Larson M. 2016. ChaLearn joint contest on multimedia challenges beyond visual analysis: an overview. *In: Proceedings of the 23rd International Conference on Pattern Recognition*. IEEE, Cancun, Mexico. p.67-73.
- Fu H L, Li Z X, Liu Z J, 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): 2 488.
- Fu H L, 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.
- Geng Y Y, Zhang G H, Li W Z, Gu Y, Liang R Z, Liang G Y, Wang J B, Wu Y B, Patil N, Wang J Y. 2017. A novel image tag completion method based on convolutional neural transformation. *In: Proceedings of the 26th International Conference on Artificial Neural Networks*. Springer, Alghero, Italy. p.539-546.
- Hurst H E. 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**(1): 770-799.
- Jiang S, Lian M J, Lu C W, Gu Q H, Ruan S L, Xie X C. 2017. Prediction of the Death Toll of Environmental Pollution in China's Coal Mine Based on Metabolism-GM (1, n) Markov Model. *Ekoloji*, **26**(101): 17-23.
- Jiang S, Lian M J, Lu C W, Gu Q H, Ruan S L, Xie X C. 2018a. Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope. *Complexity*, **2018**: Article ID 1048756, <https://doi.org/10.1155/2018/1048756>.
- Jiang S, Lian M J, Lu C W, Ruan S L, Wang Z, Chen B Y. 2018b. SVM-DS fusion based soft fault detection and diagnosis in solar water heaters, Energy Exploration and Exploitation, <https://doi.org/10.1177/0144598718816604>.
- Kang L, Du H L, Du X, Wang H.T, Ma W L, Wang M L, Zhang F B. 2018a. Study on dye wastewater treatment of tunable conductivity solid-waste-based composite cementitious material catalyst. *Desalination and Water Treatment*, **125**: 296-301.
- Kang L, Du H L, Zhang H, Ma W L. 2018b. Systematic research on the application of steel slag resources under the background of big data. *Complexity*, 2018, <https://doi.org/10.1155/2018/6703908>.
- Li Z, Burgueño R. 2010. Using soft computing to analyze inspection results for bridge evaluation and management. *Journal of Bridge Engineering*, **15**(4): 430-438.
- Liu D F, Wang L P, Pang L. 2006. Theory of multivariate compound extreme value distribution and its application to extreme sea state prediction. *Chinese Science Bulletin*, **51**(23): 2 926-2 930.
- Liu G L, Chen B Y, Wang L P, Zhang S F, Zhang K Y, 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, Wang C. 2015. Power-type wave absorbing device and using method thereof: China Patent, CN105113452A. 2015-12-02. (in Chinese)
- Mao Y, Wang J Y, Sheng B. 2016. Mobile message board: location-based message dissemination in wireless ad-hoc networks. *In: Proceedings of International Conference on Computing, Networking and Communications*. IEEE, Kauai, HI, USA. p.1-5.
- Ponce-López V, Chen B Y, Oliu M, Corneanu C, Clapés A, Guyon I, Baró X, Escalante H J, Escalera S. 2016. ChaLearn LAP 2016: first round challenge on first impressions-dataset and results. *In: Computer Vision — ECCV 2016 Workshops*. Springer, Amsterdam, The Netherlands, https://doi.org/10.1007/978-3-319-49409-8_32.
- Wang J Y, Wang T, Yang Z Y, Mao Y, Mi N F, Sheng B. 2017a. SEINA: a stealthy and effective internal attack in Hadoop systems. *In: Proceedings of 2017 International Conference on Computing, Networking and Communications*. IEEE, Santa Clara, CA, USA. p.525-530.
- Wang L P, Chen B Y, Chen C, Chen Z S, 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, 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, Liu G L, Chen B Y, Wang L. 2010. New method for calculating typhoon-influenced sea area designed wave height based on maximum entropy principle: China Patent, CN201010595815. 2010-12-20. (in Chinese)
- Wang L P, Wang L. 2011. Typhoon influence considered method for calculating combined return period of ocean extreme value: China Patent, CN102063527A. 2011-05-18. (in Chinese)
- Wang L P, Xu X, Liu G L, Chen B Y, Chen Z S. 2017b. A new method to estimate wave height of specified return period. *Chinese Journal of Oceanology and Limnology*, **35**(5): 1 002-1 009.
- Yang J Y, Zhao C, Liu G S, Xu Y. 2017. Mann-Kendall and R/S methods combined-analysis on changing trend of hydrological series—a case of Suzhou. *Water Resources and Hydropower Engineering*, **48**(2): 27-30, 137. (in Chinese with English abstract)
- Zhang G H, Liang G Y, Li W Z, Fang J, Wang J B, Geng Y Y, Wang J Y. 2017. Learning convolutional ranking-score function by query preference regularization. *In: Proceedings of the 18th International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Guilin, China. p.1-8.
- Zhang K Y, Olawoyin R, Nieto A, Kleit A N. 2018b. Risk of commodity price, production cost and time to build in resource economics. *Environment, Development and Sustainability*, **20**(6): 2 521-2 544.
- Zhang S F, Shen W, Li D S, Zhang X W, Chen B Y. 2018a. 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): e48.