AIMS Mathematics

AIMS Mathematics, 9(3):6238–6265.
DOI: 10.3934/math.2024304
Received: 14 December 2023
Revised: 21 January 2024
Accepted: 24 January 2024
Published: 04 February 2024

http://www.aimspress.com/journal/Math

Research article

Grey parameter estimation method for extreme value distribution of short-term wind speed data

Yichen Lv* and Xinping Xiao

School of Science, Wuhan University of Technology, Wuhan 430070, China

* **Correspondence:** Email: yichen_lyu2000@163.com.

Abstract: Accurate parameter estimation of extreme wind speed distribution is of great importance for the safe utilization and assessment of wind resources. This paper emphatically establishes a novel grey generalized extreme value method for parameter estimation of annual wind speed extremum distribution (AWSED). Considering the uncertainty and frequency characteristics of the parent wind speed, the generalized extreme value distribution (GEVD) is selected as the probability distribution, and the Weibull distribution is utilized as the first-order accumulation generating operator. Then, the GEVD differential equation is derived, and it is transformed into the grey GEVD model using the differential information principle. The least squares method is used to estimate the grey GEVD model parameters, and then a novel estimation method is proposed through grey parameters. A hybrid particle swarm optimization algorithm is used to optimize distribution parameters. The novel method is stable under different sample sizes according to Monte Carlo comparison simulation results, and the suitability for the novel method is confirmed by instance analysis in Wujiaba, Yunnan Province. The new method performs with high accuracy in various indicators, the hypothesis test results are above 95%, and the statistical errors such as MAPE and Wasserstein distance yield the lowest, which are 3.33% and 0.2556, respectively.

Keywords: probability distribution of wind speed; generalized extreme value distribution; parameter estimation; grey model; Monte Carlo simulation
Mathematics Subject Classification: 60G70, 62G32, 65C05

1. Introduction

1.1 Background and motivation

Extreme wind speed is of great significance in renewable energy [1] and construction engineering [2] fields. As a typical extreme climate, extreme wind speed has a prominent safety impact on wind turbines. Compared with the average wind speed, the change of extreme wind speed once a year will affect the life and safety performance of wind turbines more [3]. The 50-year maximum wind speed is one of the key indicators for unit selection and economic evaluation of wind farms [4]. In addition, extreme wind may also cause damage to engineering structures in civil engineering design. From the perspective of safety, we need to calculate the wind load in a certain period through the extreme wind speed [5]. With the improvement of wind power installed capacity and infrastructure construction, the study of extreme wind speed is particularly important.

The annual wind speed extremum is not perfectly independent and identically distributed. It was originally believed to follow the extreme value type II distribution in the analysis of actual wind speeds [6]. In subsequent statistical analysis, there was a conclusion that it was more appropriate to fit the extreme wind speed using the extreme value type I distribution when the parent wind speed (PWS) distribution obeyed the Weibull distribution [7]. In the development of wind power projects, the National Wind Energy Resources Evaluation Technical Regulations [8] also employed the extreme value type I distribution to calculate the maximum wind speed in fifty years. However, recent studies have shown that the extreme value type III distribution may give the best extreme wind speed estimations [9,10]. Lu et al. [11] also indicate that when the PWS obeys different distributions, its extreme distribution converges to different extreme value types. Jenkinson [12] synthesized these three distributions into GEVD in 1955, and there are investigations showing that the annual wind speed extremum also obeys GEVD well when the wind speed varies with different timescales [13,14]. The GEVD is selected for estimating parameters of annual extreme wind speed in this paper.

The main existing parameter estimation methods are Maximum Likelihood Estimation (MLE), Probability Weighted Moments (PWM) estimation and Maximum Product of Spacing (MPS) estimation. The nonlinear maximum likelihood function of MLE is usually difficult to solve, and its performance may be extremely erratic for small samples [3]. The PWM method is better than MLE (in terms of bias and mean square error) for parameter estimation in small samples, but it can only perform well for GEVD when the shape parameter is greater than 0 and less than 0.5 [15]. The MPS can be used as an alternative to the MLE in cases where the likelihood function fails to converge, but the MPS method has always been neglected in extreme value analysis [16]. The GEVD and limited sample sizes of annual wind speed extremum may restrict the use of these traditional parameter estimation methods in practice.

Grey system theory [17] is introduced for incomplete information and scarce data. Compared with the aforementioned methods, the grey estimation method can deal with small sample problems well [18]. Many scholars have conducted in-depth research on the grey model based on its characteristics [19,20]. This paper mainly focuses on the study of the grey GEVD parameter estimation method and its stability under small samples, providing accurate parameter estimations for annual wind speed extremum data. Next, this paper reviews the common methods of parameter estimation and related research status.

1.2 Literature review

The MLE method was proposed by Prescott and Walden [21] based on large sample theory, which provides inaccurate GEVD shape parameter estimation for small samples [22]. Papukdee et al. [23] analyzed the four-parameter kappa distribution, which is the generalized form of GEVD, and found that MLE with small sample sizes shows substantially poor performance in terms of a large variance. Cannon [24] explored rainfall extreme value estimates in Canada, and found that MLE performed the worst of the at-site GEVD estimations, especially for small sample sizes. Yang et al. [25] explored the parameter estimation of the three-parameter Weibull distribution, and concluded that the MLE results are insufficient in small samples. Meanwhile, Lin et al. [26] investigated the parameter estimation of generalized linear exponential distributions, and observed that the statistical errors of MLE increase as the sample size decreases in most situations.

The PWM estimation [27] is a promotion of the common moments of probability distributions, and Landwehr et al. [28] investigated the characteristics of PWM estimation based on the parameters and quantile of the Gumbel distribution and concluded that they performed better than traditional moment estimation and MLE in small sample sizes. However, Lu et al. [11] proposed that the parameter estimation using PWM estimation for GEVD is less stable, especially for extreme value type II and type III. Guan et al. [29] explored the best linear unbiased estimation of the location-scale parameters of the Beta-Exponential distribution, and found that PWM is not suitable for estimation under small samples. Shakeel et al. [30] explored the estimation of a flexible power function distribution, and they found that PWM performs better in the case of large sample size. Mahdia and Ashkar [31] used PWM to estimate the two-parameter Weibull distribution parameters, and also found that the RMSE would rise with the decrease in sample size, regardless of whether the shape parameter exceeds zero.

The MPS estimation is a new method proposed by Cheng and Amin [32], which compensates for the shortcomings of MLE that may fail in some cases, and it can provide better statistical estimators of robustness, consistency, and validity. El Gazar et al. [33] studied the statistical properties of the inverse power Ailamujia distribution, and found that the method of MPS gave the smallest MSE values followed by the MLE method in general for the majority of the cases. However, it is also known that the traditional numerical MPS method gives inefficient estimations when the sample size is less than 250 [34]. Yalçınkaya et al. [35] studied the parameter estimation of skew-normal distribution under doubly type II censoring, and found that all DEF and MSE values of the MPS method increase with the reduction of sample size. Similarly, Anis et al. [36] estimated the parameters of the Rayleigh distribution considering seven different methods, finding that the SE of the MPS method kept increasing as the sample size decreased and the bias and the MSE approached zero with the increase in sample size.

The current grey parameter estimation methods are mostly based on the Weibull distribution. Zheng et al. [37] used the GM (1,1) model to obtain parameters of the Weibull distribution, which applies to different sample sizes, especially for the case of small and medium samples. Zhao et al. [38]

organically combined the GM (1,1) model with Weibull to estimate the loss failure probability of chemical machinery and equipment, and found that the grey model had higher accuracy for limited failure data. Li [39] proposed a two-step approach for the parameter estimation of three-parameter Weibull distribution data combined with GM (1,1) and the MLE method, and found that the proposed method still performs well in small samples. Liu and Xie [40] proposed a two-stage hybrid method for the discrete grey Weibull model combined with the genetic algorithm, and the results showed that the proposed grey estimation method performs better than the modified MLE when the parameters and the sample size are all small. Gao et al. [41] estimated Weibull parameters by combining the grey model with the support vector machine. The simulation results showed that this model has obvious advantages in small samples, and can accurately obtain the three parameters of the Weibull distribution.

These parameter estimation methods combined with the grey model showed great adaptability for small samples, which is appropriate for the limited sample sizes of annual wind speed extremum. From the perspective of the limited annual wind speed extremum, this paper mainly focuses on the grey parameter estimation method based on GEVD.

1.3 Contributions

The main contributions are presented as follows:

1) A novel grey GEVD parameter estimation (G-GEVDPE) method is established for the small sample size of AWSED. The GEVD is selected considering the inaccurate parent wind speed distribution, and the Weibull distribution is introduced as the first-order accumulation operator to capture the distribution characteristic of wind.

2) The method is demonstrated to be stable through stability analysis of the GEVD differential equation and disturbance analysis of the grey GEVD model. The Monte Carlo comparison experiments are also used in verifying the compatibility under different sample sizes for the novel method compared with the traditional parameter estimation method.

3) The data set in Wujiaba, Yunnan province is used to demonstrate the validity of the G-GEVDPE method. The statistical errors and hypothesis testing are analyzed compared with MLE, PWM, and MPS, which shows the performance of parameter estimation and goodness of fitting distribution.

The remaining part of the paper proceeds as follows: A grey GEVD model is established for the parameter estimation in Section 2. Section 3 proposes a novel G-GEVDPE method for AWSED. The simulation verification and practical application are enumerated in Section 4. Section 5 is the summary and conclusion of the paper. Figure 1 provides the structural representation of the entire paper.

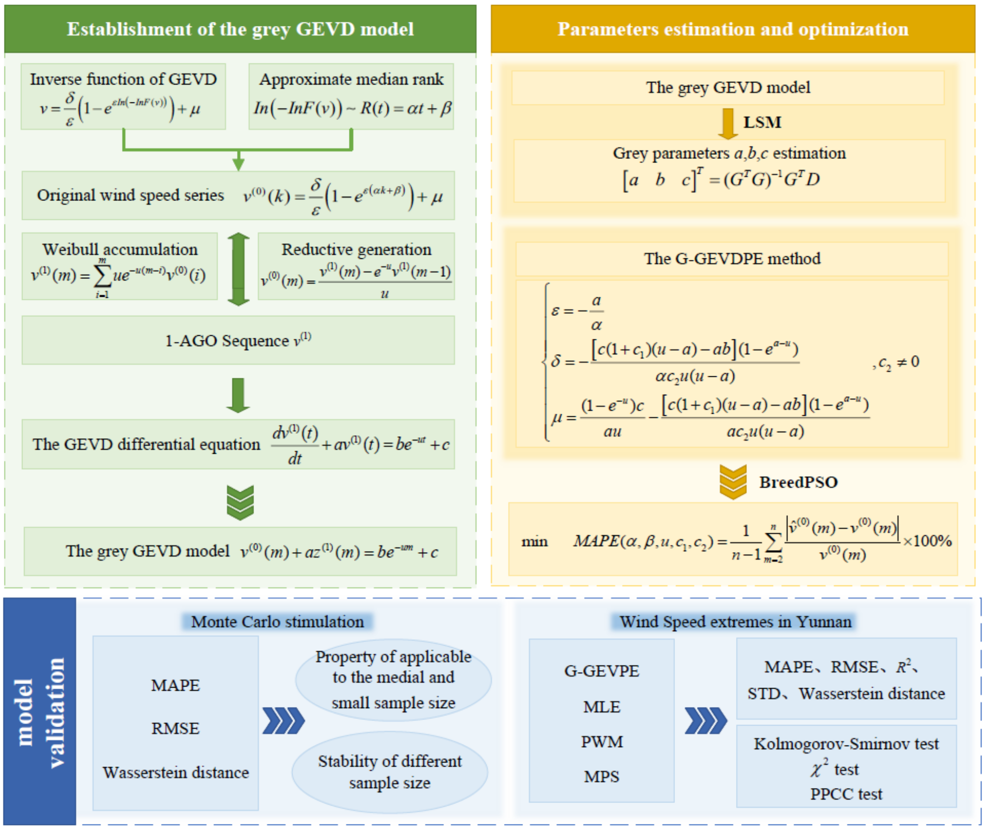

Figure 1. Entire paper structure.

2. The grey model for extreme wind speed

In this section, the GEVD and the Weibull accumulation generating operator are introduced to construct the GEVD differential equation, then translate it into the grey GEVD model, which provides a theoretical basis for the grey parameter estimation method for AWSED.

2.1. The generalized extreme value distribution

Extracting n extreme samples from wind speed samples with the same distribution $F(v)$ in a period T, which n are independent of each other and follow the same distribution, then v_n satisfies the distribution

$$P(v_n < v) = \left[F(v)\right]^n .$$

In the 1930s, Fisher and Tippett conducted a theoretical study that if $n \to \infty$, the distribution essentially belongs to the independent and identically distributed maximum asymptotic distribution, which is called the GEVD. The GEVD unifies three types of extreme value distributions and can be represented as

$$P(v_T < v) = e^{-(1-\varepsilon\frac{x-\mu}{\delta})^{\frac{1}{\varepsilon}}},$$

where μ, δ, and ε are the position, scale, and shape parameters.

This paper mainly studies the extreme wind speed over a one-year period *T*. The GEVD unifies these three distributions even when the parent distribution is ambiguous and the sample size is limited in accordance with the asymptotic theory.

2.2. Weibull accumulation generating operator

It is believed that the winds follow a Weibull distribution [7]. The Weibull distribution is introduced as the accumulation generating operator. Assuming $v^{(0)} = \left(v^{(0)}(1), v^{(0)}(2), \ldots, v^{(0)}(n)\right)$ is the original series, the first-order accumulation $v^{(1)}$ can be expressed as [42]

$$v^{(1)}(m) = \sum_{i=1}^{m} \frac{\theta}{\eta}\left(\frac{m-i}{\eta}\right)^{\theta-1} e^{-\left(\frac{m-i}{\eta}\right)^{\theta}} v^{(0)}(i), m = 1, 2, \ldots, n, \tag{1}$$

In particular, when $\theta = 1$, the Weibull distribution is an exponential distribution. Letting $u = \frac{1}{\eta}$, then Eq (1) can be simplified as

$$v^{(1)}(m) = \sum_{i=1}^{m} ue^{-u(m-i)} v^{(0)}(i).$$

Definition 1. The first-order Weibull accumulation is

$$v^{(1)}(m) = \sum_{i=1}^{m} ue^{-u(m-i)} v^{(0)}(i), \tag{2}$$

where u is the Weibull accumulation parameter, which is used to adjust the accumulation values of the series. Next, we explore the cumulative reduction and the stability of the reduction error of this operator.

Theorem 1. The accumulative reduction of the first-order Weibull accumulation generator is

$$v^{(0)}(m) = \frac{1}{u}\left[v^{(1)}(m) - \sum_{i=1}^{m-1} ue^{-u(m-i)} v^{(0)}(i)\right]. \tag{3}$$

Proof: In accordance with Eq (2), the first-order Weibull accumulation can be written as

$$
\begin{aligned}
&v^{(1)}(1) = uv^{(0)}(1) \\
&v^{(1)}(2) = ue^{-u}v^{(0)}(1) + uv^{(0)}(2) \\
&\qquad \vdots \\
&v^{(1)}(m) = ue^{-u(m-1)}v^{(0)}(1) + ue^{-u(m-2)}v^{(0)}(2) + \cdots + ue^{-u}v^{(0)}(2) + uv^{(0)}(m),
\end{aligned}
$$

We can conclude that $v^{(1)}(m) - e^{-u}v^{(1)}(m-1) = uv^{(0)}(m)$, and the accumulative reduction is

$$
\begin{aligned}
&v^{(0)}(1) = \frac{1}{u}v^{(1)}(1) \\
&v^{(0)}(2) = \frac{1}{u}\left[v^{(1)}(2) - ue^{-u}v^{(0)}(1)\right] \\
&\qquad \vdots \\
&v^{(0)}(m) = \frac{1}{u}\left[v^{(1)}(m) - \sum_{i=1}^{m-1} ue^{-(m-i)u}v^{(0)}(i)\right].
\end{aligned}
$$

Theorem 2. If $\left|v^{(1)}(m) - \hat{v}^{(1)}(m)\right| < \varepsilon, m = 1, 2, \ldots, n$, then $\left|v^{(0)}(m) - \hat{v}^{(0)}(m)\right| < \frac{1+e^{-u}}{|u|}\varepsilon$, where $v^{(1)}$ is the first-order accumulation of $v^{(0)}$, $\hat{v}^{(1)}$ is the fitted value of $v^{(1)}$, and $\hat{v}^{(0)}$ is the fitted value of $v^{(0)}$.

Proof: As demonstrated in the proof of Theorem 1, we hold that

$$
v^{(0)}(m) = \frac{v^{(1)}(m) - e^{-u}v^{(1)}(m-1)}{u}, \tag{4}
$$

Therefore,

$$
\begin{aligned}
\left|v^{(0)}(m) - \hat{v}^{(0)}(m)\right| &= \left|\frac{v^{(1)}(m) - \hat{v}^{(1)}(m)}{u} - \frac{e^{-u}\left(v^{(1)}(m-1) - \hat{v}^{(1)}(m-1)\right)}{u}\right| \\
&\le \frac{\left|v^{(1)}(m) - \hat{v}^{(1)}(m)\right|}{|u|} + \frac{e^{-u}\left|v^{(1)}(m-1) - \hat{v}^{(1)}(m-1)\right|}{|u|} \\
&< \frac{1+e^{-u}}{|u|}\varepsilon.
\end{aligned}
$$

Theorem 2 shows that the first-order Weibull accumulation parameter u can not only be used to adjust the accumulative values of the series but also to adjust the reduction error.

2.3. Establishment of the grey GEVD model

According to subsection 2.1, the cumulative distribution function of GEVD is

$$F(v) = e^{-(1-\varepsilon\frac{v-\mu}{\delta})^{\frac{1}{\varepsilon}}}. \tag{5}$$

Transforming Eq (5), we can obtain the inverse function of the distribution that

$$v = \frac{\delta}{\varepsilon}\left(1 - e^{\varepsilon In(-InF(v))}\right) + \mu. \tag{6}$$

Generally, the median rank is utilized to calculate the empirical distribution function of small samples. We can conclude from the calculation formula of some known approximate median rank [43] that the general expression for the median rank can be predicted as

$$F_n(v_t) = \frac{t+s}{n+q}, \tag{7}$$

where t is the rank of the original series, n is the total number, v_t refers to the tth wind speed extremum, and p and q are the median rank parameters to be determined. Then, Eq (6) can be expressed as

$$v(t) = \frac{\delta}{\varepsilon}\left(1 - e^{\varepsilon In\left(-In(\frac{t+s}{n+q})\right)}\right) + \mu. \tag{8}$$

Letting $R(t) = In\left(-In(\frac{t+s}{n+q})\right)$, it can be simplified in the actual study according to the annual wind speed extreme data, the wind data is selected from the National Environmental Information Centre (https://www.ncei.noaa.gov/data). The fitting results are presented in Figure 2.

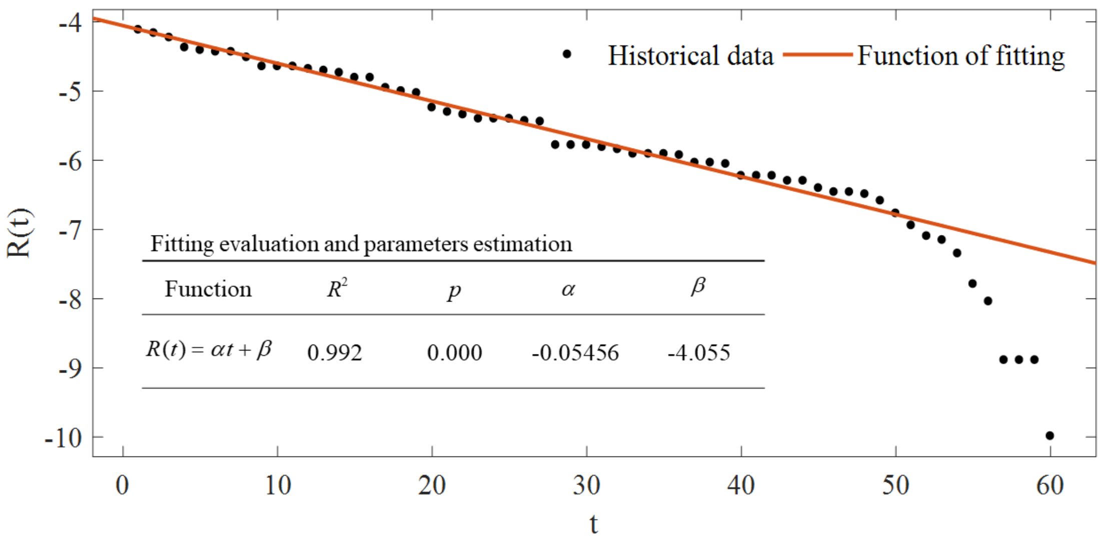

Figure 2. $R(t)$ fitting result.

The specific values of α and β in $R(t)$ are determined from historical wind speed data. The model's regression coefficients are significant (p value is below 0.01), and the fitting accuracy is high (R^2 is higher than 0.90). Therefore, the original wind speed data can be simplified as

$$v^{(0)}(t)=\frac{\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha t+\beta)}\right)+\mu . \tag{9}$$

According to Eq (4) about the Weibull accumulation reduction and the simplified original series Eq (9), then

$$\begin{aligned} v^{(1)}(t)-e^{-u}v^{(1)}(t-1)&=\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha t+\beta)}\right)+u\mu \\ v^{(1)}(t-1)-e^{-u}v^{(1)}(t-2)&=\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha(t-1)+\beta)}\right)+u\mu \\ &\vdots \\ v^{(1)}(1)-e^{-u}v^{(1)}(0)&=\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha+\beta)}\right)+u\mu \end{aligned},$$

Iterating the intermediate items, we hold that

$$\begin{aligned} v^{(1)}(t)-e^{-2u}v^{(1)}(t-2)&=\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha t+\beta)}\right)+u\mu+e^{-u}\left[\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha(t-1)+\beta)}\right)+u\mu\right] \\ v^{(1)}(t)-e^{-3u}v^{(1)}(t-3)&=\sum_{i=0}^{1}e^{-iu}\left[\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha(t-i)+\beta)}\right)+u\mu\right]+e^{-2u}\left[\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha(t-2)+\beta)}\right)+u\mu\right], \\ &\vdots \\ v^{(1)}(t)-e^{-tu}v^{(1)}(0)&=\sum_{i=0}^{t-1}e^{-iu}\left[\frac{u\delta}{\varepsilon}\left(1-e^{\varepsilon(\alpha(t-i)+\beta)}\right)+u\mu\right] \end{aligned}$$

For a given starting point $v^{(1)}(0)=v^{(0)}(0)$, the first-order Weibull accumulation $v^{(1)}(t)$ can be solved as

$$v^{(1)}(t)=u(1-e^{-u})^{-1}(\mu+\frac{\delta}{\varepsilon})(1-e^{-ut})-\frac{u\delta}{\varepsilon}(1-e^{-\alpha\varepsilon-u})^{-1}(e^{\varepsilon(\alpha t+\beta)}-e^{\varepsilon\beta-ut})+e^{-ut}v^{(0)}(0) . \tag{10}$$

Facilitating Eq (10), we hold that

$$v^{(1)}(t)=u(\mu+\frac{\delta}{\varepsilon})(1-(1+c_1)e^{-ut})(1-e^{-u})^{-1}-\frac{u\delta}{\varepsilon}(e^{\varepsilon(\alpha t+\beta)}+c_2e^{-ut})(1-e^{-\alpha\varepsilon-u})^{-1}, \tag{11}$$

where c_1 and c_2 are the constants.

Substituting Eq (11) into the differential equation, we can get

$$
\begin{aligned}
&\frac{dv^{(1)}(t)}{dt}-\alpha\varepsilon v^{(1)}(t)\\
&=u^2(\mu+\frac{\delta}{\varepsilon})(1+c_1)e^{-ut}(1-e^{-u})^{-1}-\frac{u\delta}{\varepsilon}(\alpha\varepsilon e^{\varepsilon(\alpha t+\beta)}-uc_2e^{-ut})(1-e^{-\alpha\varepsilon-u})^{-1}\\
&-u\alpha\varepsilon(\mu+\frac{\delta}{\varepsilon})(1-(1+c_1)e^{-ut})(1-e^{-u})^{-1}+\alpha u\delta(e^{\varepsilon(\alpha t+\beta)}+c_2e^{-ut})(1-e^{-\alpha\varepsilon-u})^{-1}\\
&=\left[u(\mu+\frac{\delta}{\varepsilon})(1+c_1)(1-e^{-u})^{-1}+\frac{c_2u\delta}{\varepsilon}(1-e^{-\alpha\varepsilon-u})^{-1}\right](\alpha\varepsilon+u)e^{-ut}-u\alpha\varepsilon(\mu+\frac{\delta}{\varepsilon})(1-e^{-u})^{-1}.
\end{aligned}
$$

Definition 2. The GEVD differential equation is defined as

$$\frac{dv^{(1)}(t)}{dt}+av^{(1)}(t)=be^{-ut}+c, \tag{12}$$

where

$$a=-\alpha\varepsilon,\ \ b=\left[u(\mu+\frac{\delta}{\varepsilon})(1+c_1)(1-e^{-u})^{-1}+\frac{c_2u\delta}{\varepsilon}(1-e^{-\alpha\varepsilon-u})^{-1}\right](\alpha\varepsilon+u),\ \ c=-u\alpha\varepsilon(\mu+\frac{\delta}{\varepsilon})(1-e^{-u})^{-1}.$$

The annual wind speed extremum is discrete with limited uncertain information or so-called grey information. According to the differential information principle, the grey GEVD model can be constructed as follows:

The change rate about $v^{(1)}(t)$ within $[t-1,t]$ can be approximated as

$$\left.\frac{dv^{(1)}(i)}{di}\right|_{i=t}\approx\frac{v^{(1)}(t)-v^{(1)}(t-1)}{t-(t-1)}=v^{(0)}(t). \tag{13}$$

In addition, the background values of the grey derivatives $z^{(1)}(m)$ can be replaced by the values on the interval according to the trapezoidal formula $z^{(1)}(m)=\frac{1}{2}\left[v^{(1)}(m)+v^{(1)}(m-1)\right]$, and we can define the following grey model.

Definition 3. Suppose $v^{(0)}=\left(v^{(0)}(1),v^{(0)}(2),\ldots,v^{(0)}(n)\right)$ is the original series and $v^{(1)}$ is a first-order Weibull accumulation generation of $v^{(0)}$, then the grey GEVD model can be expressed as

$$v^{(0)}(m)+az^{(1)}(m)=be^{-um}+c,m=1,2,\cdots,n. \tag{14}$$

Here, a refers to the development coefficient, and b and c are control coefficients. a,b,c are grey parameters about the GEVD parameters ε,δ,μ, and u is the parameter for the first-order Weibull accumulation.

3. Grey parameter estimation method

In this section, the grey GEVD model parameters are estimated using the least squares method (LSM), and the G-GEVDPE method is proposed by analyzing the grey parameters. Then, the optimal parameter estimation is obtained by the hybrid particle swarm optimization with breeding (BreedPSO) algorithm combined with the optimization model.

3.1. Estimation of the grey GEVD model

To facilitate the calculation, this subsection treats u as a known parameter, which can be calculated through an algorithm. Then, the grey parameters a, b, c of the grey GEVD model can be estimated using LSM.

Theorem 3. Assuming $P = \begin{bmatrix} a & b & c \end{bmatrix}^T$, the grey parameters satisfy that

$$\hat{P} = (G^T G)^{-1} G^T D, \tag{15}$$

where

$$G = \begin{bmatrix} -z^{(1)}(2) & e^{-2u} & 1 \\ -z^{(1)}(3) & e^{-3u} & 1 \\ \vdots & \vdots & \vdots \\ -z^{(1)}(n) & e^{-nu} & 1 \end{bmatrix}, D = \begin{bmatrix} v^{(0)}(2) \\ v^{(0)}(3) \\ \vdots \\ v^{(0)}(n) \end{bmatrix}.$$

Proof: The grey GEVD model can be expressed as $D = GP$. The goal is to minimize $\|D - GP\|^2$. The error sequence is expressed as $\varepsilon = D - SP$. Let

$$\begin{aligned} S = \varepsilon^T \varepsilon &= (D - GP)^T (D - GP) \\ &= \sum_{m=2}^{n} \left[v^{(0)}(m) + a z^{(1)}(m) - b e^{-um} - c \right]^2 . \end{aligned}$$

Then,

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2 \sum_{m=2}^{n} \left[v^{(0)}(m) + a z^{(1)}(m) - b e^{-um} - c \right] z^{(1)}(m) = 0 \\ \dfrac{\partial S}{\partial b} = -2\lambda \sum_{m=2}^{n} \left[v^{(0)}(m) + a z^{(1)}(m) - b e^{-um} - c \right] e^{-um} = 0 . \\ \dfrac{\partial S}{\partial c} = -2 \sum_{m=2}^{n} \left[v^{(0)}(m) + a z^{(1)}(m) - b e^{-um} - c \right] = 0 \end{cases}$$

We can obtain

$$G^T(D-GP)=0\,,$$

and therefore

$$P=(G^TG)^{-1}G^TP\,.$$

In accordance with Theorem 3, the parameter sequence $P=\begin{bmatrix}a & b & c\end{bmatrix}^T$ can be estimated if the original sequence $v^{(0)}$ is known. Then, the GEVD parameters ε,δ,μ can be estimated by analyzing the grey parameters a,b,c.

3.2. The G-GEVDPE method

According to Definition 2, the grey parameters a,b,c are represented by the GEVD parameters ε,δ,μ, and therefore the grey GEVD estimation method can be derived as follows:

Theorem 4. The GEVD parameters ε,δ,μ are solved as

$$\begin{cases}\varepsilon=-\dfrac{a}{\alpha}\\ \delta=-\dfrac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{\alpha c_2u(u-a)}\\ \mu=\dfrac{(1-e^{-u})c}{au}-\dfrac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{ac_2u(u-a)}\end{cases},c_2\neq 0\,. \tag{16}$$

Proof: From the expression for the grey parameter in Definition 2, we can obtain

$$\begin{cases}a=-\alpha\varepsilon\\ b=\left[u(\mu+\dfrac{\delta}{\varepsilon})(1+c_1)(1-e^{-u})^{-1}+\dfrac{c_2u\delta}{\varepsilon}(1-e^{-\alpha\varepsilon-u})^{-1}\right](\alpha\varepsilon+u)\\ c=-u\alpha\varepsilon(\mu+\dfrac{\delta}{\varepsilon})(1-e^{-u})^{-1}\end{cases}.$$

Solving the system of cubic equations about ε,δ,μ, it can be shown that

$$\begin{cases} \varepsilon = -\dfrac{a}{\alpha} \\ \delta = -\dfrac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{\alpha c_2 u(u-a)} \\ \mu = \dfrac{(1-e^{-u})c}{au} - \dfrac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{ac_2 u(u-a)} \end{cases}, c_2 \neq 0 .$$

Theorem 4 is also called the grey GEVD parameter estimation (G-GEVDPE) method. This theorem establishes the estimation method through the relationship between the grey GEVD model and GEVD. The stability of the G-GEVDPE method under disturbance is further explored as follows:

Definition 4. Equation (9) is said to be Hyers-Ulam stable if there exists a constant $\kappa > 0$ for every $\varepsilon > 0$ and every solution $\mu(t)$ of the inequality $\left|\dfrac{d\mu(t)}{dt} + a\mu(t) - be^{-ut} - c\right| \le \varepsilon, t \in [t_0, T]$, such that $\left|\mu(t) - v^{(1)}(t)\right| \le \kappa\varepsilon, t \in [t_0, T]$.

Theorem 5. The GEVD differential equation $\dfrac{dv^{(1)}(t)}{dt} + av^{(1)}(t) = be^{-ut} + c$ is Hyers-Ulam stable.

Proof: If $v_\xi(k)$ satisfies $\left|\dfrac{dv_\xi^{(1)}(t)}{dt} + av_\xi^{(1)}(t) - be^{-ut} - c\right| \le \xi$, which is also means that

$$be^{-ut} + c - \xi \le \frac{dv_\xi(t)}{dt} + av_\xi(t) \le be^{-ut} + c + \xi ,$$

and if $\dfrac{dv_\xi(t)}{dt} + av_\xi(t) = be^{-ut} + c + h(t)$, then $|h(k)| \le \xi$.

The solution can be expressed as

$$v_\xi(t) = e^{-at}\left[\frac{b}{a-u}e^{(a-u)t} + \frac{c}{a}e^{at} + \int_0^t h(s)e^{as}ds + C_\xi\right],$$

where C_ξ is any constant.

According to Eq (11), taking the initial condition that $v_\xi(0) = v^{(1)}(0)$, we can get

$$C_\xi = -\frac{u\delta e^{\varepsilon\beta}}{\varepsilon}(1-e^{-\alpha\varepsilon-u})^{-1},$$

The solution of the GEVD differential equation is

$$v^{(1)}(t) = u(\mu + \frac{\delta}{\varepsilon})(1-(1+c_1)e^{-ut})(1-e^{-u})^{-1} - \frac{u\delta}{\varepsilon}(e^{\varepsilon(\alpha t+\beta)} + c_2 e^{-ut})(1-e^{-\alpha\varepsilon-u})^{-1}.$$

Therefore,

$$\left|v_\xi(t)-v^{(1)}(t)\right|=\left|e^{-at}\left[\frac{b}{a-u}e^{(a-u)t}+\frac{c}{a}e^{at}+\int_0^t h(s)e^{as}ds+C_\xi\right]-e^{-at}\left[\frac{b}{a-u}e^{(a-u)t}+\frac{c}{a}e^{at}+const\right]\right|$$
$$=\left|e^{-ak}\int_0^t h(s)e^{as}ds\right|\le e^{-at}\xi\left|\int_0^t e^{as}ds\right|$$
$$=\frac{\xi}{|a|}\left|1-e^{-at}\right|\le\frac{\xi}{|a|}.$$

Letting $\kappa=\frac{1}{|a|}$, we then have $\left|v_\xi(t)-v^{(1)}(t)\right|\le\kappa\xi$.

Theorem 5 shows that the GEVD differential equation is Hyers-Ulam stable, which means there always exists a function that is infinitely close to the original model. The property proved in Theorem 5 ensures the existence of an approximate solution of the grey GEVD model, and we can conclude that the novel method is also stable for the parameter estimation of AWSED according to the derivation process of the parameter estimation method above.

3.3. Parameter optimization

According to Definition 2, the GEVD parameters ε,δ,μ of the wind speed extremum are in accordance with linearly fitting parameters α, β, accumulation parameter u, and parameters c_1, c_2. To enhance the estimation accuracy of the G-GEVDPE method, MAPE is taken as the optimal function, and the optimization model is constructed as follows:

$$\min\quad MAPE(\alpha,\beta,u,c_1,c_2)=\frac{1}{n-1}\sum_{m=2}^{n}\frac{\left|\hat{v}^{(0)}(m)-v^{(0)}(m)\right|}{v^{(0)}(m)}\times100\%$$
$$s.t.\begin{cases}z^{(1)}(m)=\frac{1}{2}\left[v^{(1)}(m)+v^{(1)}(m-1)\right]\\ G=\begin{bmatrix}-z^{(1)}(2)&e^{-2u}&1\\-z^{(1)}(3)&e^{-3u}&1\\\vdots&\vdots&\vdots\\-z^{(1)}(n)&e^{-nu}&1\end{bmatrix},D=\begin{bmatrix}v^{(0)}(2)\\v^{(0)}(3)\\\vdots\\v^{(0)}(n)\end{bmatrix}\\ \hat{P}=(G^TG)^{-1}G^TD\\ \varepsilon=-\frac{a}{\alpha},\delta=\frac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{\alpha c_2u(u-a)},\mu=\frac{(1-e^{-u})c}{au}+\frac{\left[c(1+c_1)(u-a)-ab\right](1-e^{a-u})}{ac_2u(u-a)}\end{cases}\tag{17}$$

The optimization model is solved using the BreedPSO algorithm in this paper, which draws on the concept of hybridization in genetic algorithms. The detailed process of this algorithm can be illustrated with the pseudocode in Figure 3.

algorithm 1 BreedPSO to caculate $\alpha, \beta, \lambda, c_1, c_2$ of the grey GEVD model

Input: fitness,initial$v^{(0)} = [v^{(0)}(1), v^{(0)}(2), ..., v^{(0)}(n)], N, c1, c2, w, bc, bs, M, D$

Output: The optimal value of $\alpha, \beta, \lambda, c_1, c_2$ and fitness

function $BreedPSO(v^{(0)})$
  Initialize the velocity and position of each particle
  fitness—function(MAPE=$\frac{1}{n-1}\sum_{k=1}^{n}\frac{\left|v^{(0)}(k)-\hat{v}^{(0)}(k)\right|}{v^{(0)}(k)} \times 100\%$)
  N—Initializes the number of community individuals
  $c1$,$c2$—learning factor 1,2
  w—inertial weight
  bc—hybridization probability
  bs—Size ratio of hybrid pool
  M—Maximum number of iterations
  D—Search space dimension
  xm—The argument when fitness function takes the minimum value
  fv—minimum target function
  for i=1 to N **do**
    $p(i) \leftarrow$ fitness of particle
    $p_{best} \leftarrow$ positions and fitness values of particle i
    $g_{best} \leftarrow$ the best position and fitness value of particle in p_{best}
  end for
  while $t \leq M$ **do**
    for i=1 to N **do**
      Update the displacement x_i and velocity v_i of particle
      $x_i(t+1) \leftarrow x_i(t) + v_i(t+1)$
      $v_i(t+1) \leftarrow w.v_i(t) + c_1.r_1[p_i - x_i(t)] + c_2.r_2[g_{best} - x_i(t)]$
      if $fitness(x_i) < g_{best}$ **then**
        $g_{best} \leftarrow x_i$ and $fitness(x_i)$
        $r_1 = rand()$
        if $r_1 < bc$ **then**
          Select the specified number of particles according to bc
          choose $parent_1$ and $parent_2$ randomly in the pool
          $pb = rand()$
          $child(x) \leftarrow pb.parent_1(x) + (1-pb).parent_2(x)$
          $child(v) \leftarrow \frac{parent_1(v)+parent_2(v)}{|parent_1(v)+parent_2(v)|}|parent_1(v)|$
        end if
      end if
    end for
  end while
  $xm \leftarrow g_{best}$
  $fv \leftarrow fitness(x_m)$
end function

Figure 3. Pseudocode of BreedPSO.

4. Algorithm verification and case analysis

In this section, the properties of the proposed G-GEVDPE method are analyzed theoretically through disturbance analysis, and then it is verified experimentally by Monte Carlo simulation. The efficacy of the method is also tested by estimating AWSED in Wujiaba, Yunnan Province.

4.1. Disturbance analysis and Monte Carlo simulation

Lemma 1. [44] Assume $G \in C^{n\times n}$, $\Delta G \in C^{n}$, $D \in C^{n\times n}$, $\Delta D \in C^{n}$, and vector parametrization $\|\bullet\|$ is compatible with matrix parameterization $\|\bullet\|$. If there exists $\|G^{-1}\|\|\Delta G\| < 1$ for some matrix parametrization $\|\bullet\|$ on $C^{n\times n}$, then the solution of the non-simultaneous linear equation $Gx = D$ and $(G+\Delta G)(x+\Delta x) = D+\Delta D$ satisfies

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{\kappa}{\gamma}\left(\frac{\|\Delta G\|}{\|G\|} + \frac{\|\Delta D\|}{\|D\|}\right),$$

where $\kappa = \|G^{-1}\|\|G\|, \gamma = 1-\kappa\frac{\|\Delta G\|}{\|G\|} > 0$.

Theorem 6. If $\hat{v}^{(0)}(m) = v^{(0)}(m)+\xi, m = 2,3,\cdots,n$ is perturbed separately, the corresponding *D* and *G* will both change, and the perturbation bound for $P = \begin{bmatrix} a & b & c \end{bmatrix}$ can be noted as

$$L_m[P] = |\xi|\frac{\kappa}{\gamma}\left(\frac{\sqrt{\sum_{i=0}^{n-m-1}\left(ue^{-iu}+ue^{-(i+1)u}\right)^2}}{2\|G\|} + \frac{1}{\|D\|}\right). \tag{18}$$

Proof: We can know from Definition 3 that

$$\begin{bmatrix} z^{(1)}(2) \\ z^{(1)}(3) \\ z^{(1)}(4) \\ \vdots \\ z^{(1)}(n) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} u & 0 & 0 & \cdots & 0 \\ ue^{-u} & u & 0 & \cdots & 0 \\ ue^{-2u} & ue^{-u} & u & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ ue^{-(n-1)u} & ue^{-(n-2)u} & ue^{-(n-3)u} & \cdots & u \end{bmatrix} \begin{bmatrix} v^{(0)}(1) \\ v^{(0)}(2) \\ v^{(0)}(3) \\ \vdots \\ v^{(0)}(n) \end{bmatrix}$$

1) Adding a perturbation ξ into $v^{(0)}(2)$, then

$$\Delta z=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}u & 0 & 0 & \cdots & 0\\ ue^{-u} & u & 0 & \cdots & 0\\ ue^{-2u} & ue^{-u} & u & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & 0\\ ue^{-(n-1)u} & ue^{-(n-2)u} & ue^{-(n-3)u} & \cdots & u\end{bmatrix}\begin{bmatrix}0\\ \xi\\ 0\\ \vdots\\ 0\end{bmatrix}=\begin{bmatrix}\frac{\xi}{2}u\\ \frac{\xi}{2}\left(u+ue^{-u}\right)\\ \frac{\xi}{2}\left(ue^{-u}+ue^{-2u}\right)\\ \vdots\\ \frac{\xi}{2}\left(ue^{-(n-3)u}+ue^{-(n-2)u}\right)\end{bmatrix}$$

Therefore,

$$\hat{G}=G+\Delta G=G+\begin{bmatrix}-\frac{\xi}{2}u & 0 & 0\\ -\frac{\xi}{2}\left(u+ue^{-u}\right) & 0 & 0\\ -\frac{\xi}{2}\left(ue^{-u}+ue^{-2u}\right) & 0 & 0\\ \vdots & \vdots & \vdots\\ -\frac{\xi}{2}\left(ue^{-(n-3)u}+ue^{-(n-2)u}\right) & 0 & 0\end{bmatrix},\quad \hat{D}=D+\Delta D=D+\begin{bmatrix}\xi\\ 0\\ 0\\ \vdots\\ 0\end{bmatrix},$$

$$\Delta G^{T}\Delta G=\begin{bmatrix}\frac{\xi^{2}}{4}\sum_{i=0}^{n-3}\left(ue^{-iu}+ue^{-(i+1)u}\right)^{2} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\quad \Delta D^{T}\Delta D=\xi^{2}.$$

The maximum characteristic roots are $\frac{\xi^{2}}{4}\sum_{i=0}^{n-3}\left(ue^{-iu}+ue^{-(i+1)u}\right)^{2}$ and ξ^{2}, respectively, so that

$\|\Delta G\|_{2}=\sqrt{\lambda_{\max}\left(\Delta G^{T}\Delta G\right)}=\frac{|\xi|}{2}\sqrt{\sum_{i=0}^{n-3}\left(ue^{-iu}+ue^{-(i+1)u}\right)^{2}}$, $\|\Delta D\|_{2}=\sqrt{\lambda_{\max}\left(\Delta D^{T}\Delta D\right)}=|\xi|$, and $GP=D$.

In accordance with Lemma 1, we have

$$L_{m}[P]=|\xi|\frac{\kappa}{\gamma}\left(\frac{\sqrt{\sum_{i=0}^{n-m-1}\left(ue^{-iu}+ue^{-(i+1)u}\right)^{2}}}{2\|G\|}+\frac{1}{\|D\|}\right).$$

2) Analogously, adding a perturbation ξ into $v^{(0)}(m), m=2,3,\cdots,n$, then

$$\Delta G=\begin{bmatrix} 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ -\frac{\xi}{2}u & 0 & 0 \\ -\frac{\xi}{2}\left(u+ue^{-u}\right) & 0 & 0 \\ \vdots & \vdots & \vdots \\ -\frac{\xi}{2}\left(ue^{-(n-k-1)u}+ue^{-(n-k)u}\right) & 0 & 0 \end{bmatrix}, \Delta D=\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \xi \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\Delta G^T\Delta G=\begin{bmatrix} \frac{\xi^2}{4}\sum_{i=0}^{n-k-1}\left(ue^{-iu}+ue^{-(i+1)u}\right)^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } \Delta D^T\Delta D=\xi^2.$$

The perturbation bound for its solution is

$$L_m\left[P\right]=\left|\xi\right|\frac{\kappa}{\gamma}\left(\frac{\sqrt{\sum_{i=0}^{n-m-1}\left(ue^{-iu}+ue^{-(i+1)u}\right)^2}}{2\left\|G\right\|}+\frac{1}{\left\|D\right\|}\right).$$

While other information is known, the larger the overall sample size, the larger the perturbation bound. Therefore, the G-GEVDPE method is more suitable for the smaller sample size data according to Theorem 6. In addition, a Monte Carlo simulation is used to verify the stability of the G-GEVDPE method, and the accuracy of the G-GEVDPE method can also be tested through it.

Taking $\varepsilon=1$, $\delta=6$, $\mu=12$ as an example and generating different sizes of random original series that follow the GEVD, a comparative Monte Carlo simulation is designed. Each simulation is repeated 200 times for a given sample size N, and then the average estimated results are calculated in Table 1. The optimal results are highlighted in bold. All the computations in this paper are implemented by MATLAB 2017.

Table 1. Monte Carlo simulation results of parameter estimation.

		ε	δ	μ	MAPE	RMSE	Wasserstein distance
N=10	G-GEVDPE	0.9115	6.004	12.028	**4.57%**	**0.0657**	**19.3353**
	MLE	1.301	5.324	12.087	21.04%	0.5269	271.8563
	PWM	0.556	9.298	12.893	53.40%	2.436	62.9466
	MPS	0.848	7.108	12.474	18.81%	0.8589	35.4661

Continued on next page

		ε	δ	μ	MAPE	RMSE	Wasserstein distance
N=50	G-GEVDPE	0.9458	6.0053	11.984	**2.82%**	**0.0401**	12.7810
	MLE	0.9679	5.833	12.085	3.35%	0.0693	**9.2841**
	PWM	0.709	9.123	12.478	42.57%	2.2434	34.9576
	MPS	0.768	6.514	12.177	16.62%	0.4179	35.5273
N=100	G-GEVDPE	0.8494	6.031	11.998	7.79%	**0.1087**	28.6548
	MLE	1.004	5.813	11.937	2.04%	0.1399	0.7439
	PWM	0.754	7.867	12.419	29.56%	1.3639	33.1273
	MPS	0.995	6.173	12.039	**1.843%**	0.1260	**0.6638**

It can be seen from Table 1 that the accuracy of MLE, PWM, and MPS decreases with the increase of sample size, and the estimation results perform poorly in small samples, while the G-GEVDPE method shows high accuracy in different samples specifically in small sample sizes. The Wasserstein distance [45] of 200 simulations is shown in Figure 4. Simulations show that the Wasserstein distance of traditional parameter estimation methods varies greatly in small samples, while the G-GEVDPE method varies moderately, which shows strong stability of the proposed method. The Monte Carlo simulation further demonstrates the G-GEVDPE method is stable and suitable for small samples, as is analyzed in Theorem 6, which is compatible with the size of the annual wind speed extremum completely.

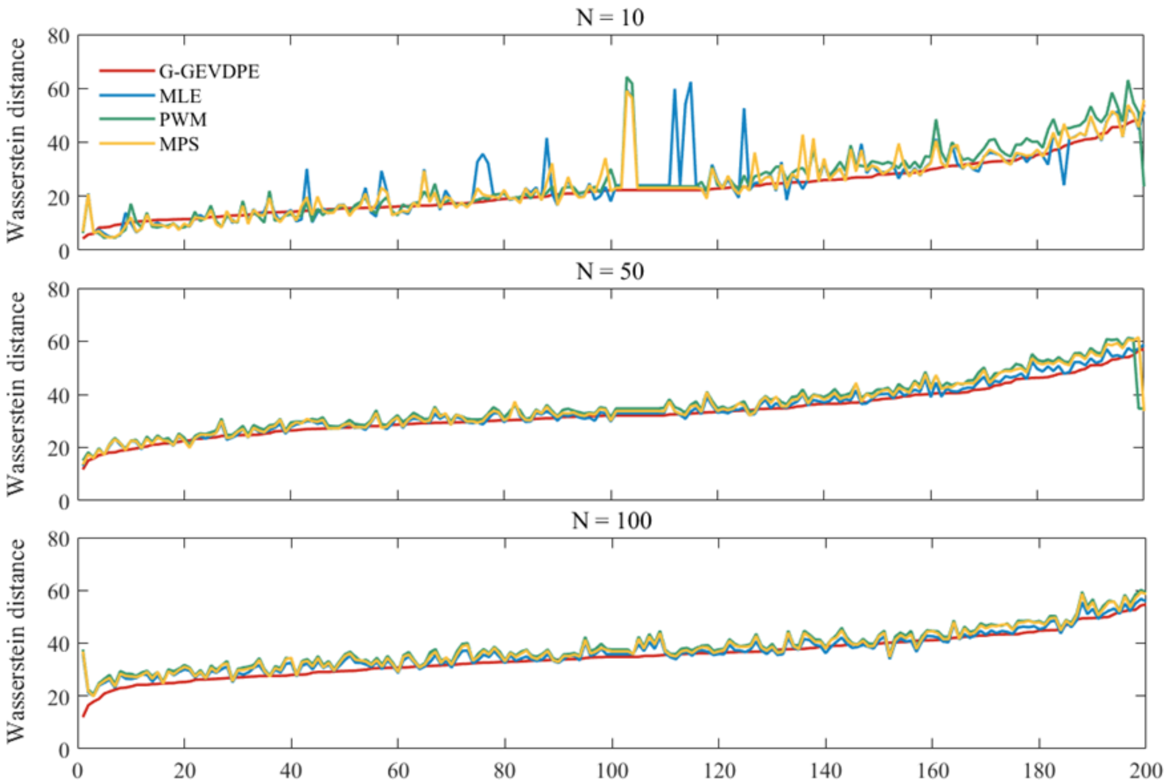

Figure 4. Wasserstein distances for 200 experiments with different sample sizes.

4.2. Case analysis

To further test the effectiveness of the method, the extreme wind speed in Wujiaba, Yunnan province is utilized as the case of empirical analysis in this subsection. The original monthly, seasonal, and annual wind speed extremum and distributions are shown in Figure 5, We can see from Figure 5 that the change of wind speed extremum has a strong random uncertainty with time, while the distributions of them show obvious regularity. The annual wind speed extremum is selected for the parameter estimation to explore the distribution law of wind speed extreme value in this paper. The validity of the method is verified compared with the MLE method, the PWM estimation, and the MPS estimation through statistical error, Wasserstein distance, and hypothesis testing of the parameter distributions [46].

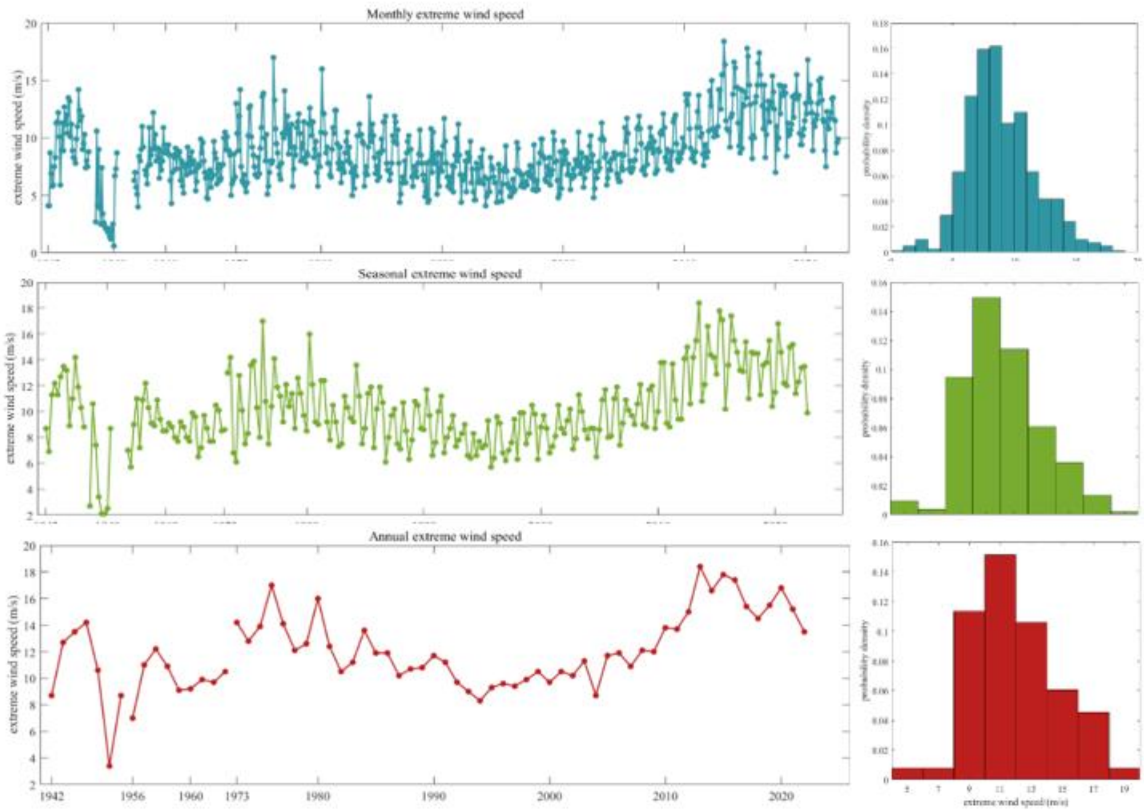

Figure 5. The original wind speed extremum and distributions.

Except for mean absolute percentage error (MAPE), four other metrics are selected for statistical error, namely absolute percentage error (APE), root mean square error (RMSE), coefficient of determination (R^2), and standard deviation (STD). These indicators for assessing model validity are presented below. The criterion of these metrics is that the smaller the value of MAPE, APE, RMSE, and STD, the more valid the model, and the larger the value of R^2, the more accurate the model.

$$\text{MAPE} = \frac{1}{n-1}\sum_{m=1}^{n}\frac{\left|\hat{v}^{(0)}(m)-v^{(0)}(m)\right|}{v^{(0)}(m)}, \tag{19}$$

$$\text{APE} = \frac{\left|\hat{v}^{(0)}(m)-v^{(0)}(m)\right|}{v^{(0)}(m)}, \tag{20}$$

$$\text{RMSE} = \sqrt{\frac{1}{n-1}\sum_{m=1}^{n}\left(\hat{v}^{(0)}(m)-v^{(0)}(m)\right)^2}, \tag{21}$$

$$R^2 = 1-\frac{\sum_{m=1}^{n}\left(\hat{v}^{(0)}(m)-v^{(0)}(m)\right)^2}{\sum_{m=1}^{n}\left(\hat{v}^{(0)}(m)-\overline{v}^{(0)}\right)^2}, \tag{22}$$

$$\text{STD} = \sqrt{\frac{1}{n}\sum_{m=1}^{n}\left(\frac{\left|\hat{v}^{(0)}(m)-v^{(0)}(m)\right|}{v^{(0)}(m)}-\text{MAPE}\right)^2}. \tag{23}$$

Table 2 shows the estimation results, and Figure 6 visualizes the fitness of the four contrastive methods. The results show that the G-GEVDPE method has optimal values for the MAPE metrics of 3.33%, 3.96%, 3.34%, and 3.50%, and the RMSE metrics of 2.092, 2.883, 2.171, and 2.377 for MLE, PWM, and MPS, respectively. Although R^2 and STD are not the best, they only differed from the optimal value by 0.0171 and 0.0256, respectively. Meanwhile, the correlation metrics of R^2 reached over 95%, and STD metrics were controlled within 10%, showing strong correlation and high accuracy, which also indicated that the annual extreme wind speed in the region obeyed the GEVD. The Wasserstein distance of 0.2556, 0.3522, 0.2652, and 0.2904 for these parameter estimations also indicate that the G-GEVDPE method has the best advantage in parameter estimation.

Table 2. Parameter estimation and comparative indicator results.

Methods of estimation	ε	δ	μ	MAPE	RMSE	R^2	STD	Wasserstein distance
G-GEVDPE	0.03134	2.2863	10.801	**3.33%**	**2.092**	0.95287	7.63%	**0.2556**
MLE	0.2685	2.7701	10.954	3.96%	2.883	0.96726	**5.07%**	0.3522
PWM	0.1357	2.5272	10.789	3.34%	2.171	0.96889	6.27%	0.2652
MPS	0.2007	2.6328	10.8703	3.50%	2.377	**0.96995**	5.71%	0.2904

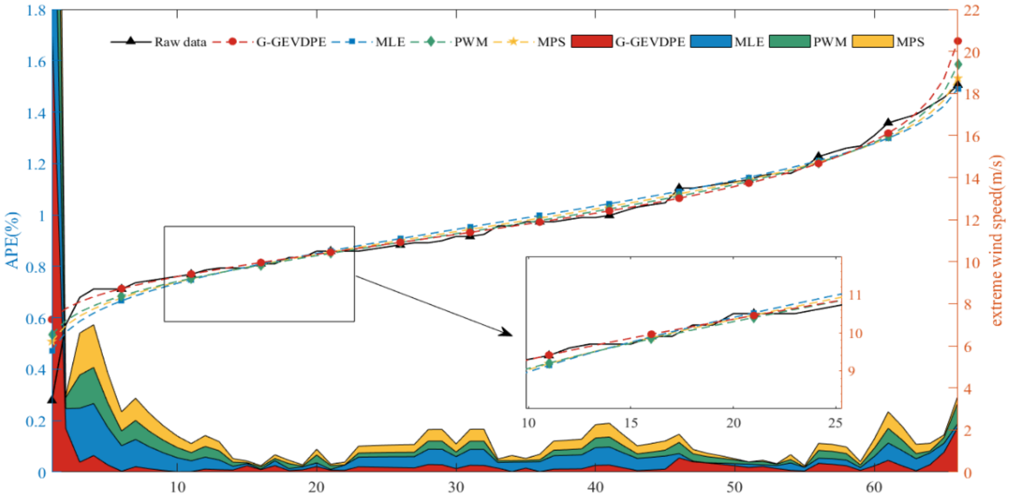

Figure 6. Comparison of four parameter estimation results and indicators.

The original and the estimated extreme wind speed distribution are displayed in Figure 7. The MAPE, RMSE, and STD are shown in the right side of the figure, and the scatter plots of four different methods are shown in the bottom. The G-GEVDPE method shows a competitive result among the four estimation methods.

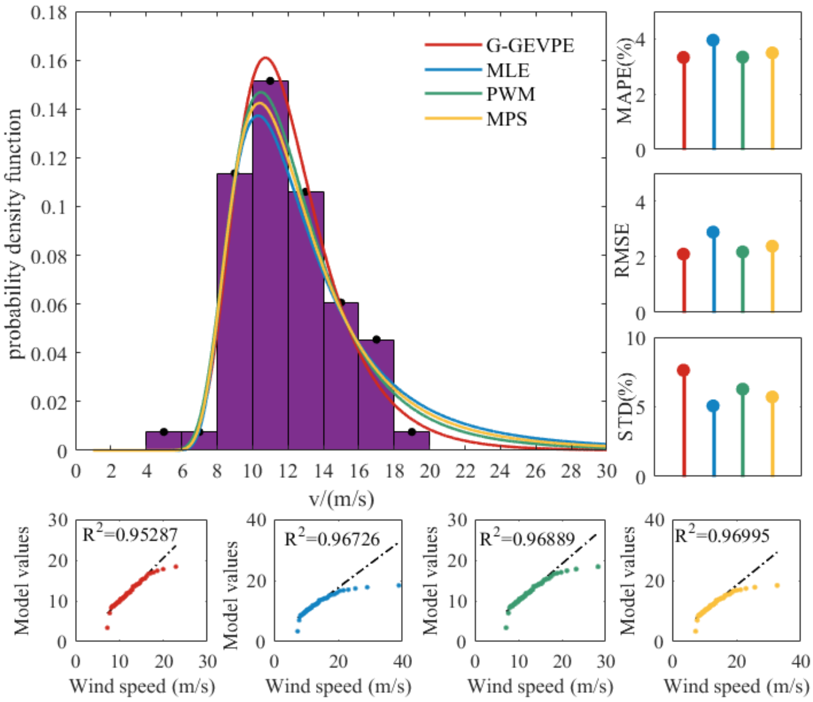

Figure 7. Fitting performance of four parameter estimation models.

Hypothesis testing is performed to test the significance of the four different parameter estimation methods using the K-S test [47], the χ^2 test [48], and the PPCC test [49]. The hypothesis testing results are shown in Table 3.

Table 3. Hypothesis testing results.

Methods of estimation	K-S (Kolmogorov-Smirnov) test			χ^2 test			PPCC test
	p	k	h	p	s	h	
G-GEVDPE	**0.99536**	0.0488	0	**0.99643**	0.3571	0	0.97799
MLE	0.23863	0.1244	0	0.44889	3.6947	0	0.98427
PWM	0.80114	0.0769	0	0.92092	1.4307	0	0.98514
MPS	0.50463	0.0991	0	0.77567	2.5055	0	**0.98596**

p is the probability that the original hypothesis holds. k and s are the test statistics of the K-S test and the χ^2 test, respectively. h is the test result: 0 indicates that the original data follows the estimated distribution whereas 1 indicates it does not. As is shown in Table 3, these four methods all passed the K-S test and χ^2 test at 5% significance. The G-GEVDPE method shows the highest significance for accepting the original hypothesis, which means the distribution estimated by the G-GEVDPE method is closest to the AWSED. The PPCC test results also show a strong correlation at coefficients that all the significance of the PPCC test reach above 95% (the maximum is 1), further confirming that the AWSED obeys the GEVD and the G-GEVDPE method has higher accuracy than other methods.

5. Conclusions

Wind energy is subject to fluctuations in wind speed and has great randomness, intermittency, and uncontrollability. Therefore, it is essential to select an appropriate wind speed extremum probability distribution model and estimate distributional parameters accurately for the rational utilization of clean wind resources. This paper focuses on establishing the AWSED model and the estimation algorithm of model parameters. The analysis leads to the following conclusions:

1) A novel G-GEVDPE method for AWSED is proposed through the GEVD differential equation and the grey GEVD model. The GEVD is selected considering the uncertainty of the PWS distribution, and the first-order Weibull accumulation operator is used to capture the characteristics of wind distribution. The GEVD differential equation is then established and transformed into the grey GEVD model. The G-GEVDPE method is specially designed for the uncertainty and distribution characteristics of short-term wind speed data.

2) The G-GEVDPE method is stable in parameter estimation and it is adaptive to the size characteristic of short-term wind speed data. The stability of the G-GEVDPE method is demonstrated through stability analysis of the GEVD differential equation and perturbation bound of the grey GEVD model. The Monte Carlo simulation was designed to further verify that it is compatible under

different sample sizes and it is suitable for the novel G-GEVDPE method to estimate the GEVD for annual wind speed extremum data.

3) The G-GEVDPE method still performs with high accuracy in the application of Wujiaba compared with MLE, PWM, and MPS. The statistical errors and hypothesis testing results show that it is reasonable to select the GEVD in estimating AWSED, and the G-GEVDPE method has a strong advantage over traditional methods. The novel G-GEVDPE method proposed in this paper provides an effective way of estimating the extreme value distribution for short-term wind speed data.

Although the G-GEVDPE method is adaptive to the characteristics of annual wind speed extremum and has strong theoretical significance, the application of the novel method requires further discussion.

1) The safety of wind power projects and the design of engineering structures are also closely related to the wind direction. With the improvement of wind speed measurement and recording tools, the techniques used to measure and record wind speed and direction are becoming increasingly advanced, resulting in a more comprehensive set of data for wind resource estimation. In subsequent work, it will be necessary to analyze extreme wind in combination with direction.

2) The novel method is applied to the parameter estimation of AWSED in this paper. It is suitable to describe extreme wind speed with the GEVD; meanwhile, it can also be applied to other extreme events, such as floods and earthquakes. The occurrence of these events has a certain complexity, and the reliability of estimating their distribution using the GEVD needs to be further studied. A more scientific parameter estimation method should be conducted by considering the description of mixed and multivariate extreme distributions to extreme events in subsequent work.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 72371194).

Conflict of interest

The authors confirm no conflicts of interest.

References

1. B. Kresning, M. Hashemi, A. Shirvani, J. Hashemi, Uncertainty of extreme wind and wave loads for marine renewable energy farms in hurricane-prone regions, *Renew. Energ.*, **220** (2024), 119570. https://doi.org/10.1016/j.renene.2023.119570

2. L. Da, Q. Yang, M. Liu, L. Zhao, T. Wu, B. Chen, Estimation of extreme wind speed based on upcrossing rate of mean wind speeds with Weibull distribution, *J. Wind Eng. Ind. Aerod.*, **240** (2023), 105495. https://doi.org/10.1016/j.jweia.2023.105495
3. Y. Charabi, S. Abdul-Wahab, A. M. Al-Mahruqi, S. Osman, I. Osman, The potential estimation and cost analysis of wind energy production in Oman, *Environ. Dev. Sustain.*, **24** (2022), 5917–5937. https://doi.org/10.1007/s10668-021-01692-7
4. International Standard IEC61400-1, International Electrotechnical Commission, 2005. Available from: https://webstore.iec.ch/preview/info_iec61400-1%7Bed3.0%7Den.pdf.
5. M. Abdelwahab, T. Ghazal, K. Nadeem, A. Haitham, E. Ahmed, Performance-based wind design for tall buildings: Review and comparative study, *J. Build. Eng.*, **68** (2023), 106103. https://doi.org/10.1016/j.jobe.2023.106103
6. H. C. S. Thom, Distributions of extreme winds in the United States, *Journal of the Structural Division*, **86** (1960), 11–24. https://doi.org/10.1061/TACEAT.0008243
7. E. Simiu, M. ASCE, J. Biétry, J. J. Filliben, Sampling errors in estimation of extreme winds, *Journal of the Structural Division*, **104** (1978), 491–501. https://doi.org/10.1061/JSDEAG.0004882
8. National Wind Energy Resources Evaluation Technical Regulations, 2004. Available from: http://www.nea.gov.cn/201201/04/c_131260326.html.
9. E. Simiu, M. ASCE, M. J. Changery, J. J. Filliben, Extreme wind speeds at 129 airport stations, *Journal of the Structural Division*, **106** (1980), 809–817. https://doi.org/10.1061/JSDEAG.0005399
10. E. Simiu, M. ASCE, J. J. Filliben, Probability distributions of extreme wind speeds, *Journal of the Structural Division*, **102** (1976), 1861–1877. https://doi.org/10.1061/JSDEAG.0004434
11. A. P. Lu, L. Zhao, Z. W. Guo, Y. J. Ge, A comparative study of extreme value distribution and parameter estimation based on the Monte Carlo method, *Journal of Harbin Institute of Technology*, **45** (2013), 88–95.
12. A. F. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q. J. Roy. Meteor. Soc.*, **81** (1955), 158–171. https://doi.org/10.1002/qj.49708134804
13. R. W. Katz, M. B. Parlange, P. Naveau, Statistics of extremes in hydrology, *Adv. Water Resour.*, **25** (2002), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8
14. H. Madsen, P. F. Rasmussen, D. Rosbjerg, Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling, *Water Resour. Res.*, **33** (1997), 747–757. https://doi.org/10.1029/96WR03848
15. E. Castillo, A. S. Hadi, Parameter and quantile estimation for the generalized extreme-value distribution, *Environmetrics*, **5** (1994), 417–432. https://doi.org/10.1002/env.3170050405
16. S. D. Nerantzaki, S. M. Papalexiou, Assessing extremes in hydroclimatology: A review on probabilistic methods, *J. Hydrol.*, **605** (2022), 127302. https://doi.org/10.1016/j.jhydrol.2021.127302
17. N. M. Xie, R. Z. Wang, A historic review of grey forecasting models, *The Journal of Grey System*, **29** (2017), 1–29.

18. W. Lei, S. Mao, Y. Zhang, Estimating China's CO2 emissions under the influence of COVID-19 epidemic using a novel fractional multivariate nonlinear grey model, *Environ. Dev. Sustain.*, **2023** (2023), 1–32. https://doi.org/10.1007/s10668-023-03325-7
19. Q. Z. Xiao, M. Y. Gao, L. Chen, J. C. Jiang, Dynamic multi-attribute evaluation of digital economy development in China: A perspective from interaction effect, *Tech. Econ. Dev. Eco.*, **29** (2023), 1728–1752. https://doi.org/10.3846/tede.2023.20258
20. S. W. Wang, X. P. Xiao, Q. Ding, A novel fractional system grey prediction model with dynamic delay effect for evaluating the state of health of lithium battery, *Energy*, **290** (2024), 130057. https://doi.org/10.1016/j.energy.2023.130057
21. P. Prescott, A. T. Walden, Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika*, **67** (1980), 723–724. https://doi.org/10.1093/biomet/67.3.723
22. E. S. Martins, J. R. Stedinger, Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, **36** (2000), 737–744. https://doi.org/10.1029/1999WR900330
23. N. Papukdee, J. S. Park, P. Busababodhin, Penalized likelihood approach for the four-parameter kappa distribution, *J. Appl. Stat.*, **49** (2022), 1559–1573. https://doi.org/10.1080/02664763.2021.1871592
24. A. J. Cannon, An intercomparison of regional and at-site rainfall extreme value analyses in southern British Columbia, Canada, *Can. J. Civil Eng.*, **42** (2015), 107–119. https://doi.org/10.1139/cjce-2014-0361
25. X. Yang, L. Xie, B. Zhao, X. Kong, N. Wu, An iterative method for parameter estimation of the three-parameter Weibull distribution based on a small sample size with a fixed shape parameter, *Int. J. Struct. Stab. Dy.*, **22** (2022), 2250125. https://doi.org/10.1142/S0219455422501255
26. C. T. Lin, Y. Liu, Y. W. Li, Z. W. Chen, H. M. Okasha, Further properties and estimations of exponentiated generalized linear exponential distribution, *Mathematics*, **9** (2021), 3328. https://doi.org/10.3390/math9243328
27. G. P. Sillitto, Interrelations between certain linear systematic statistics of samples from any continuous population, *Biometrika*, **38** (1951), 377–382. https://doi.org/10.2307/2332583
28. J. M. Landwehr, N. C. Matalas, J. R. Wallis, Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles, *Water Resour. Res.*, **15** (1979), 1055–1064. https://doi.org/10.1029/WR015i005p01055
29. R. Guan, W. Cheng, Y. Rong, X. Zhao, Parameter estimation of Beta-exponential distribution using linear combination of order statistics, *Commun. Math. Stat.*, **2023** (2023), 1–41. https://doi.org/10.1007/s40304-022-00306-6
30. M. Shakeel, M. A. Haq, I. Hussain, A. M. Abdulhamid, M. Faisal, Comparison of two new robust parameter estimation methods for the power function distribution, *PloS one*, **11** (2016), e0160692. https://doi.org/10.1371/journal.pone.0160692
31. S. Mahdi, F. Ashkar, Exploring generalized probability weighted moments, generalized moments and maximum likelihood estimating methods in two-parameter Weibull model, *J. Hydrol.*, **285** (2004), 62–75. https://doi.org/10.1016/j.jhydrol.2003.08.012

32. R. C. H. Cheng, N. A. K. Amin, Estimating parameters in continuous univariate distributions with a shifted origin, *J. R. Stat. Soc. B*, **45** (1983), 394–403. https://doi.org/10.1111/j.2517-6161.1983.tb01268.x
33. A. M. El Gazar, M. ElGarhy, B. S. El-Desouky, Classical and Bayesian estimation for the truncated inverse power Ailamujia distribution with applications, *AIP Adv.*, **13** (2023), 125122. https://doi.org/10.1063/5.0174794
34. Y. M. Kantar, B. Şenoğlu, A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter, *Comput. Geosci.*, **34** (2008), 1900–1909. https://doi.org/10.1016/j.cageo.2008.04.004
35. A. Yalçınkaya, U. Yolcu, B. Şenoğlu, Maximum likelihood and maximum product of spacings estimations for the parameters of skew-normal distribution under doubly type II censoring using genetic algorithm, *Expert Syst. Appl.*, **168** (2021), 114407. https://doi.org/10.1016/j.eswa.2020.114407
36. M. Z. Anis, I. E. Okorie, M. Ahsanullah, A review of the Rayleigh distribution: properties, estimation & application to COVID-19 data, *Bull. Malays. Math. Sci. Soc.*, **47** (2024), 6. https://doi.org/10.1007/s40840-023-01605-z
37. R. Y. Zheng, Z. Z. Qin, Grey estimations for the three parameters Weibull distribution, *Strength and Environment*, **4** (1989), 34–40.
38. D. F. Zhao, B. L. Li, S. Lu, J. Wang, Life estimation of chemical machinery based on linear regression and Grey-Weibull model, *Chemical Engineering and Machinery*, **38** (2011), 517–521. https://doi.org/10.3969/j.issn.0254-6094.2011.05.002
39. X. Y. Li, Research on estimation for the three-parameter Weibull distribution, PhD Thesis, Beijing Jiaotong University, 2012.
40. X. M. Liu, N. M. Xie, Grey-based approach for estimating Weibull model and its application, *Commun. Stat.-Theor. M.*, **52** (2023), 7601–7617. https://doi.org/10.1080/03610926.2022.2050397
41. B. Gao, K. Q. Cao, L. M. Hu, S. G. Li, 3-Parameter Weibull distribution estimation based on GM and SVM, *Journal of Mechanical Strength*, **40** (2018), 632–638. https://doi.org/10.16579/j.issn.1001.9669.2018.03.021
42. K. Li, P. Xiong, Y. Wu, Y. Dong, Forecasting greenhouse gas emissions with the new information priority generalized accumulative grey model, *Sci. Total Environ.*, **807** (2022), 150859. https://doi.org/10.1016/j.scitotenv.2021.150859
43. X. Ma, S. Zhang, L. Hu, R. Cao, P. Liu, L. Luo, et al., An improved rank assessment method for Weibull analysis of reliability data, *Chinese Journal of Nuclear Science and Engineering*, **27** (2007), 152–155. https://doi.org/10.3321/j.issn:0258-0918.2007.02.010
44. J. G. Sun, *Matrix perturbation analysis*, 2 Eds., Beijing: Sciences Press, 1987.
45. V. M. Panaretos, Y. Zemel, Statistical aspects of Wasserstein distances, *Annu. Rev. Stat. Appl.*, **6** (2019), 405–431. https://doi.org/10.1146/annurev-statistics-030718-104938
46. P. A. Shaw, M. A. Proschan, Estimation and hypothesis testing, *Principles and practice of clinical trials*, Cham: Springer, 2020, 1–16. https://doi.org/10.1007/978-3-319-52677-5_114-1
47. R. A. Olea, V. Pawlowsky-Glahn, Kolmogorov-Smirnov test for spatially correlated data, *Stoch. Environ. Res. Risk Assess.*, **23** (2009), 749–757. https://doi.org/10.1007/s00477-008-0255-1

48. R. Da, C. F. Lin, Sensitivity analysis of the state Chi-Square test, *IFAC Proceedings*, **29** (1996), 6596–6601. https://doi.org/10.1016/S1474-6670(17)58741-8
49. J. H. Heo, Y. W. Kho, H. Shin, S. Kim, T. Kim, Regression equations of probability plot correlation coefficient test statistics from several probability distributions, *J. Hydrol.*, **355** (2008), 1–15. https://doi.org/10.1016/j.jhydrol.2008.01.027