Journal of Sea Research 192 (2023) 102354

Contents lists available at ScienceDirect

Journal of Sea Research

journal homepage: www.elsevier.com/locate/seares

Extreme value prediction with modified Enhanced Monte Carlo method based on tail index correction

Siyuan Yu [a], Wenhua Wu [a,b,*], Arvid Naess [c]

[a] State Key Laboratory of Structure Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China
[b] Ningbo Institute of Dalian University of Technology, Ningbo 315000, China
[c] Department. of Mathematical Sciences and Centre for Ships and Ocean Structures (CeSOS), Norwegian Univ. of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Enhanced Monte Carlo method
Extreme value prediction
Hill-type estimator
Tail behavior analysis

ABSTRACT

As one of the main branches of extreme statistics, extreme value theory is widely used in marine engineering. Due to its special application in real ocean environmental states, the scale of the obtained monitoring data is limited. Aiming at the problem of extreme value prediction of different return periods with medium-scale data, this paper proposes a modified Enhanced Monte Carlo (EMC) extreme value prediction method based on the tail index correction. First, the Hill-type estimator is introduced to quantitatively evaluate the tail behavior of the data. Tail behavior analysis is then performed for sample data of various typical distribution functions, and a modified EMC method based on the tail index is proposed. Furthermore, tail estimator extrapolation is performed for a situation where the estimator does not converge to improve the engineering applicability of the proposed method. Based on a series of numerical and engineering examples, the extreme value prediction performance of the proposed method is compared with the classical extreme value prediction methods. The results show that the modified EMC extreme value prediction method proposed in this paper can provide useful guidance for the extreme value analysis of marine environmental loads and structural responses. At the same time, the method proposed in this paper introduces the slow-varying assumption in the classical EMC method, and the limitations caused by the assumption are also discussed.

1. Introduction

As one of the main branches of extreme statistics, extreme value theory is widely used in marine engineering (Lu and Wang, 2020; Galiatsatou et al., 2021; Sun, 2021). The assessment of the extreme environmental loads and structural responses for a given return period is of crucial importance in the preliminary design stage of any floating structure (Raed et al., 2020). Compared with onshore structures, marine structures are subjected to long-term environmental loads and are often affected by harsh conditions, such as typhoons. Therefore, accurate extreme value prediction methods are of great significance to the operation and maintenance of marine structures.

Marine structures usually face a much harsher environment than onshore structures and are subject to both long-term external alternating environmental loads and corrosion behavior throughout their life cycle. Sensors must operate in complicated service conditions to conduct field monitoring of marine environmental loads and structural responses, which include high humidity, high salinity, and underwater pressure. As a result, it is much more difficult to ensure the long-term stability of the power supply and network systems required for sensor data acquisition, storage, and transmission (Du, 2016; Zhao et al., 2021). In addition, the response of marine engineering structures is mostly dominated by low frequencies (Du, 2016; Gao et al., 2021; Zhong, 2021). These factors limit the scale of marine engineering field monitoring data. To the best of the authors' knowledge, a medium-sized data volume (10^3–10^4 orders of magnitude) is the most common case for marine engineering field measurement data.

Classical extreme value theory mainly includes the generalized extreme value (GEV) distribution and peak over threshold (POT) method (Coles, 2001). It is based on asymptotic theory in statistics and assumes that the extreme values from observed data obey a specific form of asymptotic distribution. Because the classical extreme value theory has a sufficient theoretical basis and is often conservative in practical applications, it is still widely used in the extreme value prediction of environmental loads (Bruserud et al., 2018; Shao et al., 2018; Feng et al., 2021) and structural responses (Stanisic et al., 2018; Su, 2020) in marine

* Corresponding author at: State Key Laboratory of Structure Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China.
E-mail address: lxyuhua@dlut.edu.cn (W. Wu).

https://doi.org/10.1016/j.seares.2023.102354
Received 20 October 2022; Received in revised form 19 January 2023; Accepted 20 January 2023
Available online 21 January 2023

engineering. However, the asymptotic theory assumes that the sample size tends to infinity, while the limited data size leads to an inevitable deviation of this assumption from the actual situation. Since it is difficult to predict the real distribution of extreme value samples, the extreme value prediction error caused by the asymptotic distribution assumption is also hard to evaluate accurately (Liu, 2018).

Aiming to solve the problem of the approximate description of the distribution characteristics of random variables, Winterstein (1985, 1988) proposed the Hermite series model of standard Gaussian random variables to establish an explicit nonlinear relationship between non-Gaussian random variables and standard Gaussian random variables (Gao, 2019). The method could approximate the data of interest with a standard Gaussian variable distribution sample. The determined parameters could then be estimated by ensuring that the statistical moments of the original sample and the fitted sample were approximately equal. Hermite models have been applied to the field of extreme value prediction (Gao et al., 2018; Gao, 2019; Zhang et al., 2019; Lin et al., 2020). Among them, Gao et al. (2018) focused on the effect of Hermite models of different orders on the extreme value prediction of log-normally distributed sample data. However, there is still a lack of research on the accuracy and stability of Hermite models for predicting the extreme values of samples with different tail behaviors. Moreover, the amount of data used to study the performance of this method in the literature (Gao et al., 2018; Gao, 2019) reaches 10^5 magnitudes, while the corresponding discussion of medium-sized data is missing.

In recent years, Naess and Gaidai (2009) proposed a more flexible extreme value analysis technique called the Average Conditional Exceedance Rate (ACER) method. ACER introduced the conditional exceedance rate into the Enhanced Monte Carlo (EMC) method (Naess et al., 2009) to consider the temporal correlation of the data series. In recent years, the ACER method has been gradually developed and applied to the extreme value prediction of marine environmental variables (Teng, 2018; Yu et al., 2020) and structural response variables (Su, 2020). However, it was also found that the calculation accuracy of the ACER method decreased in extreme value prediction with long return periods (Liu, 2018; Gao et al., 2018; Gao, 2019), limiting the application of EMC and ACER methods in extreme value prediction.

Medium-scale monitoring data are often processed in practical applications of marine engineering. However, as the current extreme value prediction methods may have inaccurate and unstable prediction results, this paper proposes a modified EMC extreme value prediction method based on the tail index. First, the Hill-type estimator is introduced to evaluate the tail behavior of the data quantitatively. Then, the tail behavior of random sequences with various typical distribution functions is analyzed. A modified EMC method based on the tail index is presented, where an extrapolation strategy is proposed for the situation that the estimator does not converge to improve the engineering applicability of the proposed method. In addition, the extreme value prediction performance of the proposed method is compared with the classical extreme value prediction methods based on a series of numerical and engineering examples. The results indicate that the proposed modified EMC method is more accurate and stable in predicting extreme values. The method is also applicable to the correction of the ACER method, which can capture the tail characteristics of the sample distribution while considering the temporal correlation of the sample data series. At the same time, the method proposed in this paper introduces the slow-varying assumption in the classical EMC method, and the limitations caused by the assumption are also discussed.

2. Extreme value prediction by the Enhanced Monte Carlo method

In response to the problem of the high computational cost of the Monte Carlo (MC) method, Naess et al. (2009) proposed an enhanced MC method, i.e., the EMC method. This technique has been applied to the fields of building structures and marine engineering in recent years, demonstrating high computational efficiency and engineering practicality.

Let the resistance and load of a considered component or structural system be R and S, respectively, and the safety margin eq. $M = R - S$. Failure behavior occurs when $M < 0$, and μ_M denotes the mean value of M. The safety margin equation can then be extended to the following form:

$$M(\lambda) = M - \mu_M(1-\lambda),\ 0 \le \lambda \le 1 \tag{1}$$

where λ is the safety margin reduction factor. When $\lambda = 1$, Eq. (1) is reduced to the classical safety margin equation.

Naess et al. (2009) presented the failure probability P_f as a function of the parameter λ based on the case where R and S are normal variables and extended this expression to the general reliability problem case as follows:

$$P_f(\lambda) \approx q(\lambda) exp\{-a(\lambda - b)^c\},\ \lambda \to 1 \tag{2}$$

where a, b, and c are positive constants. In the EMC method, $q(\lambda)$ is assumed to be a slow-varying function, which is defined as: for a measurable function $L: (0,+\infty) \to (0,+\infty)$, if for all $k > 0$, $\lim_{x\to\infty} L(kx)/L(x) = 1$, then the function L is named as a slow-varying function (Beirlant et al., 2004; Minkah et al., 2021).

Naess et al. (2009) proposed that since the function $q(\lambda)$ is assumed to be slowly varying for values of λ close to 1.0, a suitable constant value, q, is proposed to replace $q(\lambda)$ for tail values of λ. That leads to a simplified form of Eq. (2) is used as:

$$P_f(\lambda) \approx q exp\{-a(\lambda - b)^c\},\ \lambda_l \le \lambda \le \lambda_u \tag{3}$$

where a, c, $q \in (0,+\infty)$, $b \in (0, \lambda_l)$. In contrast to Eq. (2), q in Eq. (3) is treated as a constant value, and the range of applicability of the extrapolation function is determined through the tail bounds λ_l and λ_u within a certain range. Naess et al. (2009) concluded that within a certain range, the extrapolation results are insensitive to the values of the tail bounds λ_l and λ_u.

Given a series of parameter values λ_i ($0 \le \lambda_i < 1$), the corresponding weakened systems are defined by Eq. (1), and the failure probability values P_{fi} can be obtained by the MC method. The variance of interest is the failure probability of the original system, which corresponds to the case of $\lambda = 1$, so this is a mathematical extrapolation problem, and Eq. (2) and its simplified form Eq. (3) are the extrapolation functions. Since the failure probability of the weakened component or system is higher than that of the original one, the number of simulations will be significantly reduced, thus improving the computational efficiency. Based on each given λ value and its corresponding empirically estimated value of failure probability, the undetermined parameters in Eq. (3) can be obtained using the Levenberg–Marquardt least squares estimation method.

The EMC method can not only predict the failure probability of a given component or system but also give confidence intervals (CI) for the prediction results. Assuming that N-times MC simulations are performed for each given value of λ, and the failure frequency corresponding to each λ is obtained as $N_f(\lambda)$, the corresponding empirical estimate of the failure probability can be obtained as:

$$\widehat{P}_f(\lambda) = N_f(\lambda)/N \tag{4}$$

The coefficient of variation of the failure probability is:

$$\mathrm{Cov}\left[\widehat{P}_f(\lambda)\right] = \sqrt{\frac{1 - P_f(\lambda)}{P_f(\lambda)N}} \tag{5}$$

The upper and lower bounds of the 95% CI are:

$$CI^{\pm}(\lambda) = \widehat{P}_f(\lambda)\cdot\left\{1 \pm 1.96\mathrm{Cov}\left[\widehat{P}_f(\lambda)\right]\right\} \tag{6}$$

Based on the derived CI for the failure probability, the objective function of the Levenberg–Marquardt least-squares estimation can be

modified as:

$$F(q,a,b,c)=\sum_{i=1}^{M} w_i\left[log\widehat{P}_f(\lambda_i)-logq+a(\lambda_i-b)^c\right]^2 \tag{7}$$

where λ_i and $\widehat{P}_f(\lambda_i)$ are the selected reduction factors and the corresponding empirically estimated failure probabilities, respectively, and $w_i=[\log CI^+(\lambda_j)-\ \log\ CI^-(\lambda_j)]^{-2}$ is a weighting factor defined through the CI.

The extreme value prediction problem can be regarded as an inverse reliability problem of finding the corresponding extreme value of a given failure probability. In the EMC method, the failure probability of the original component or system at $\lambda=1$ can be obtained through the extrapolation expression (Eq. (3)). On the other hand, the value of λ for a given failure probability can also be calculated based on the extrapolation expression. Therefore, if a suitable reliability problem can be established, the extreme value of a given data series can be predicted by solving the problem.

The steps for solving the extreme value problem based on the EMC method are as follows:

1. Take the safety margin equation as $M=R-S$, where S is the variable corresponding to the sample data of the extreme value to be found, and R is a large value from the data (significantly greater than the mean of S) to ensure that the reliability problem established is valid. For example, consider S as a log-normal variable and R as the 99.99% quantile of a sample of S;
2. Solve the above reliability problem by the EMC method to calculate the optimal parameters of the extrapolated expression (Eq. (3));
3. For a given failure probability P_{f0} (calculated based on sampling interval and given return period, see Section 4.1), the corresponding λ_0 is obtained based on the extrapolated expression;
4. Consider the definition of the weakened system (Eq. (1)), where:

$$M(\lambda_0)=R-S-(1-\lambda_0)(R-\mu_S)=\lambda_0 R+(1-\lambda_0)\mu_S-S \tag{8}$$

Thus, the extreme value to be found is $\lambda_0 R+(1-\lambda_0)\mu_S$.

It's noted that although the form of the safety margin equation used above is the same as in the reliability calculation, the definitions of variables R and S, as well as the value range of parameter λ have all changed. This is because despite a large value is selected for R, it comes from a sample of S, and it is usually impossible to be higher than the extreme values to be calculated. As a result, the failure probability corresponding to R is higher than that of extreme values (for example, the return period of the variable S is required to be the extreme value of 100 years, while R usually cannot reach the 100-yr extreme level). This means that in extreme value prediction, the original system will be strengthened instead, and the value of λ_0 is usually >1.

3. Modified Enhanced Monte Carlo method

3.1. Hill-type estimator of Weibull-type distribution

The EMC method is concerned with the tail behavior of the underlying distribution of the sample data. A quantitative evaluation of the tail behavior will help improve the computational accuracy. For this reason, it is necessary to study the relationship between the parameters in the EMC extrapolation function (Eq. (3)) and the tail behavior of the distribution of sample data.

For the weakened system described by Eq. (1), the probability of failure is:

$$P_f(\lambda)=\mathrm{P}(M(\lambda)<0)=\mathrm{P}(M-\mu_M(1-\lambda)\langle 0)\tag{9}$$

where P(·) represents probability. For the aforementioned reliability problem established based on the extreme value prediction problem ($M=R-S$, where S is the variable corresponding to the sample data, and R is a constant), Eq. (9) becomes:

$$P_f(\lambda)=\mathrm{P}(\lambda R+(1-\lambda)\mu_S<S)\tag{10}$$

According to the definition of cumulative distribution function (CDF), we have:

$$P_f(\lambda)=1-\mathrm{P}(S\le\lambda R+(1-\lambda)\mu_S)=1-F_S(\lambda R+(1-\lambda)\mu_S)\tag{11}$$

where $F_S(s)$ is the CDF of the variable S. Let $x=\lambda R+(1-\lambda)\mu_S$, And substituting Eq. (3) into Eq. (11), we can get:

$$1-F_S(x)=qexp\left\{-a\left(\frac{x-\mu_S}{R-\mu_S}-b\right)^c\right\}=qexp\{-a_S(x-b_S)^c\},x>b_S \tag{12}$$

where $a_S=-a\left(\frac{1}{R-\mu_S}\right)^c$, $b_S=(1+b)\mu_S+bR$. Eq. (12) establishes the relationship between the underlying distribution of sample data and the parameters q, a, b, and c in the EMC extrapolation function. According to the above derivation, the tail distribution of the sample data under the framework of the EMC method is obtained, and the distribution function is same in form as the extrapolation function assumed by the EMC method.

According to Eq. (12), the CDF and probability density function (PDF) of the underlying tail distribution of the sample data can be obtained, where the subscript S is omitted for the sake of brevity:

$$F(x)=1-qexp\{-a(x-b)^c\},x>b \tag{13}$$

$$f(x)=qacexp\{-a(x-b)^c\}\cdot(x-b)^{c-1},x>b \tag{14}$$

where a, c, $q\in(0,+\infty)$, $b\in(\mu_S,(1+\lambda_l)\mu_S+\lambda_l R)$. Eq. (13) can be transformed into the following form:

$$1-F(x)=q\cdot exp\{-a[(x-b)^c-x^c]\}\cdot exp\{-ax^c\}=Q(x)\cdot exp\{-ax^c\} \tag{15}$$

where $Q(x)=q\cdot\ exp\{-a[(x-b)^c-x^c]\}$ satisfies the definition of the aforementioned slow-varying function. Therefore, both the scale parameter q and the location parameter b do not affect the tail behavior of the underlying distribution of sample data. Let $b=0$ and $q=1$, and the above CDF and PDF can be simplified as:

$$F(x)=1-exp(-ax^c) \tag{16}$$

$$f(x)=acexp(-ax^c)\cdot x^{c-1} \tag{17}$$

where the support and the ranges of parameters a and c are the same as Eq. (13). That is, the tail distribution of the sample data can reduce to a Weibull distribution, or the tail distribution is a Weibull-type distribution. From the properties of the Weibull distribution, it is known that only parameter c affects the tail behavior of the underlying distribution of sample data (de Wet et al., 2016). This also means that the parameter c in the EMC model can be determined by studying the tail behavior of the sample data. After the value of parameter c is determined, least-squares estimation is performed on the remaining three parameters q, a, and b in the EMC model, which is expected to effectively reduce the uncertainty of the results.

A commonly used tail index in the analysis of distribution function tail behavior is the Hill estimator (Hill, 1975), which is widely used to evaluate the tail behavior of distribution functions with power function tails. As for a Weibull-type distribution function, the tail behavior is described as:

$$1-F(x)\to exp(-Cx^{\alpha}),\ \ x\to\infty \tag{18}$$

where $1/\alpha$ is called the Weibull tail-coefficient. In this paper, due to the properties of the EMC extrapolation function, the "tail index" that appears in the following refers to the Weibull tail-coefficient.

Let X_1, X_2, …, X_n be a sequence of independent and identically distributed random variables with cumulative distribution function F, and $X_{(1,n)} \le X_{(2,n)} \le \cdots \le X_{(n,n)}$ be the corresponding order statistics associated with the sample, where n is the sample size (Minkah et al., 2021). Referring to the classical form of the Hill estimator, the estimator for the Weibull-type distribution function can be derived as (Girard, 2004):

$$\widehat{\theta}_n^H = \frac{\sum_{i=1}^{k(n)-1} \left[log\left(X_{(n-i+1,n)}\right) - log\left(X_{(n-k(n)+1,n)}\right) \right]}{\sum_{i=1}^{k(n)-1} \{ log[log(n/i)] - log[log(n/k(n))] \}} \tag{19}$$

where the threshold $k(n)$ is a positive integer.

Eq. (19) is an estimator of the Weibull tail-coefficient $1/\alpha$ in Eq. (18), which is known as a Hill-type estimator. In practice, the positive integer $k(n)$ is given first, and then the corresponding estimator $\widehat{\theta}_n^H$ is calculated. The "tail estimator" or "Hill-type estimator" that appears in the following refers to the reciprocal of $\widehat{\theta}_n^H$, so that it corresponds directly to the shape parameter c in the EMC model.

Fig. 1 plots the Hill-type estimator of a Weibull sample with exponents $\alpha = 1$ and $\alpha = 2$, respectively, with a sample size of 10^4. It can be observed that as $k(n)$ increases, the tail estimators gradually converge to their true values (the red dashed line in the figure). When $k(n)$ is 5% to 10% of the sample size, the estimator reaches convergence. With the further increase of $k(n)$, the estimator changes slowly and can be regarded as constant.

Gardes and Girard (2008) extended the classical Hill-type estimator by introducing a weight function to give better results in terms of bias and mean square error:

$$\widehat{\theta}_n^G = \frac{\sum_{i=1}^{k(n)-1} \alpha_{i,n} \left[log\left(X_{(n-i+1,n)}\right) - log\left(X_{(n-k(n)+1,n)}\right) \right]}{\sum_{i=1}^{k(n)-1} \alpha_{i,n} \{ log[log(n/i)] - log[log(n/k(n))] \}} \tag{20}$$

where $\alpha_{i,\,n}$ are weight functions. In this paper the Zipf estimator proposed by Gardes and Girard (2008) is used:

$$\widehat{\theta}_n^Z = \frac{\sum_{i=1}^{k(n)-1} \{ log[log(n/i)] - \zeta_n \} log\left(X_{(n-i+1,n)}\right)}{\sum_{i=1}^{k(n)-1} \{ log[log(n/i)] - \zeta_n \} log[log(n/i)]} \tag{21}$$

where

$$\zeta_n = \frac{1}{k(n)-1} \sum_{i=1}^{k(n)-1} log[log(n/i)]. \tag{22}$$

Fig. 2 shows the simulated values of the above classical Hill-type and Zipf-type tail estimators of a Weibull sample with exponents $\alpha = 1$ and a sample size of 10^4. It can be found that both estimators can quantitatively describe the tail behavior of the Weibull distribution function. For other alternative estimators, see He et al. (2020) and Caeiro et al. (2022) for further details.

3.2. Modified EMC method based on tail index

For data samples obeying a Weibull-type distribution, the Hill-type estimator can be used to describe the tail behavior of the distribution quantitatively, and these estimators will be used for the modification of the EMC method. In practice, the tail estimator is obtained by analyzing the sample data and is then used as the shape parameter c of the EMC extrapolation function. The remaining three parameters in the EMC model are estimated by least squares.

However, as the data usually do not follow the Weibull distribution exactly, it is necessary to judge the convergence of the tail estimator. In addition to the Hill plot shown in Fig. 1, the Weibull probability plot commonly used in engineering can also be used as a convenient and intuitive method to discriminate the convergence of the tail estimator. The CDF of the Weibull distribution can be linearized as:

$$log[-log(1-F)] = k_w log x + log\lambda_w \tag{23}$$

That is, the log[− log (1 − F)] and log (x) of a Weibull distribution are in a linear relationship, and the slope is the Weibull exponent. The empirical relationship curve of log[− log (1 − F)] and log (x) is the Weibull probability plot. It is worth noting that the tail estimator is an estimate of the slope of the Weibull probability plot, as can be seen from the definition of the Hill-type estimator. Therefore, the three estimators, i.e., the Hill-type estimator, the slope of the Weibull probability plot, and the shape parameter c of the EMC extrapolation function, are different empirical representations of the same mathematical quantity.

A straightforward approach to the case of non-convergence of the tail estimator of the data samples is to take the estimator corresponding to a $k(n)$ close to 1. However, a too small $k(n)$ will lead to a large variance of the Hill-type estimator (Németh and Zempléni, 2020). Therefore, in this paper, the extrapolation of the Weibull slope corresponding to moderate $k(n)$ values toward the tail of the distribution is considered to obtain estimates of the EMC shape parameter c.

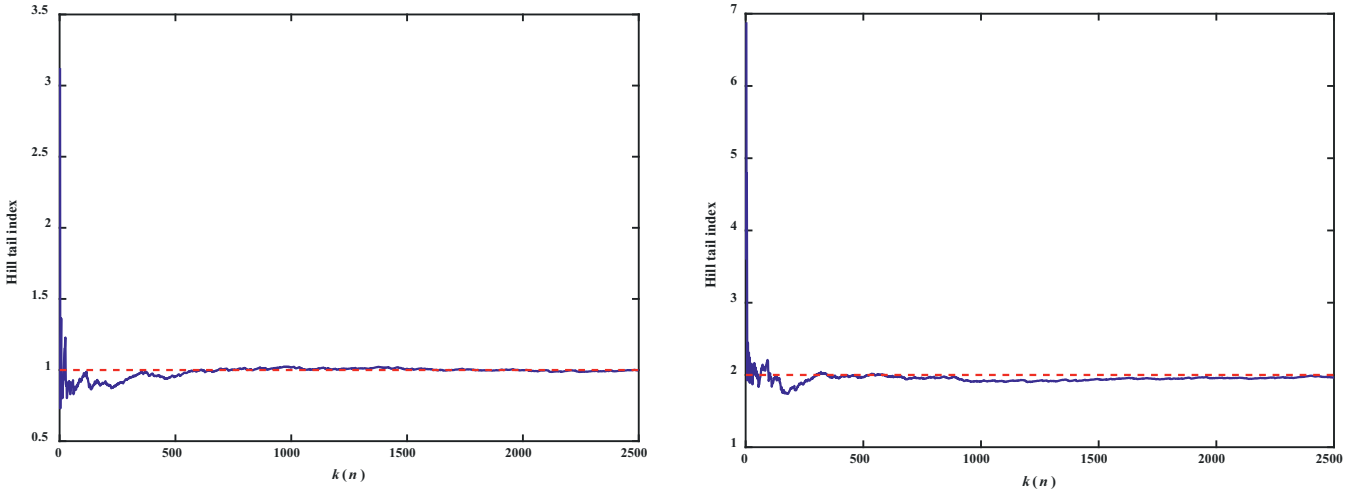


Fig. 1. Simulated values of Hill-type estimator for Weibull distribution function.

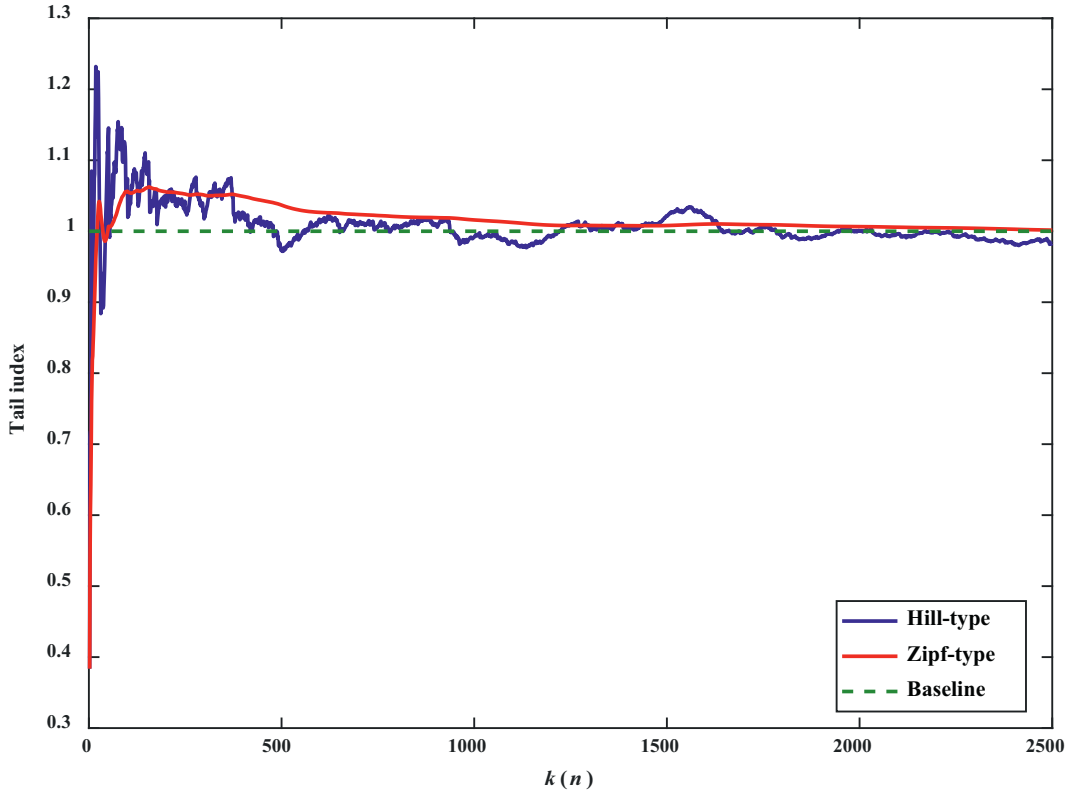


Fig. 2. Simulated values of Hill-type and Zipf-type estimators for Weibull distribution function with exponent $\alpha = 1$.

We perform the extrapolation operation based on the empirical cumulative probabilities, i.e., the relationship between the shape parameter c and the variable $\log[-\log(1-F)]$, which is the vertical axis variable of the Weibull probability plot. For the sake of computational simplicity and to avoid introducing too many undetermined parameters (to ensure the convenience of engineering applications and minimize the additional uncertainty introduced by the extrapolation process), the functional relationship is assumed to be linear. In the practical calculation, a series of high cumulative probability values denoted as F_1, F_2, …, F_m can be selected, and each cumulative probability value corresponds to a $k(n)$ value, i.e.

$$k_i(n) = \lceil n(1-F_i) \rceil, \quad i = 1, 2, \ldots, m \tag{24}$$

where $\lceil \cdot \rceil$ denotes upward rounding. For each $k(n)$ value, the value of the corresponding shape parameter c_i can be calculated, and then a linear fit and extrapolation can be achieved by the obtained sequence of F_i and c_i.

The steps of the proposed modified EMC extreme value prediction method are summarized as follows:

1. Obtain a Hill or Weibull probability plot based on the data and observe the convergence of the tail estimator;
2. Based on the observation results, perform the following treatments, respectively:
   (i) If the tail estimator can converge quickly and remains approximately constant as $k(n)$ increases (as in the case of the Weibull distribution), the Hill-type estimator can be directly applied as an estimate of the EMC shape parameter c. The threshold $k(n)$ can ensure the reasonableness of the estimator as long as it is not too small (i.e., only the lower bound of $k(n)$ should be concerned);
   (ii) If the tail estimator can converge but will change significantly as $k(n)$ continues to increase it means that the sample distribution can be approximated as an EMC extrapolation function form only in the tail range. In this case, the Hill-type estimator can still be directly applied as an estimate of the EMC shape parameter c, but attention should be paid to the reasonable range of $k(n)$ (i.e., both the lower and upper bounds of $k(n)$ should be considered);
   (iii) If the tail estimator does not converge, the estimator corresponding to a certain cumulative probability should be obtained by an extrapolation operation (which can be combined with other auxiliary methods, such as qualitative analysis of the tail (Rojo, 1996; Rojo and Ott, 2010) to reduce the extrapolation uncertainty);
3. Calculate the remaining three parameters in the EMC extrapolation expression by least-squares fitting, and finally obtain the extreme value prediction results (refer to Section 1).

The proposed method is also applicable to the modification of the ACER method, which can be realized to capture the tail characteristics of the sample distribution while considering the temporal correlation of a sample data series to provide more accurate and reliable prediction results.

4. Numerical studies of the modified Enhanced Monte Carlo method

Section 3 presents the computational principles of the proposed modified EMC method. In this section, two kinds of numerical examples are considered: one includes random samples with typical distribution functions, and the other uses simulated wind speed samples with frequency-domain features. Based on the samples with different tail characteristics, the accuracy and stability of the proposed modified EMC method for extreme value prediction are tested by comparing with the Hermite model and the classical EMC method.

4.1. Examples of random sequences obeying specific distributions

4.1.1. Weibull distribution

The PDF of the Weibull distribution is:

$$f(x) = \frac{k_w}{\lambda_w}\left(\frac{x}{\lambda_w}\right)^{k_w-1} e^{-(x/\lambda_w)^{k_w}} \tag{25}$$

Only k_w affects the tail behavior of the distribution and is also the standard value of the aforementioned tail estimator.

Different shape parameters k_w and scale parameter $\lambda_w = 1$ are taken to generate the corresponding random numbers as sample data for extreme value prediction. If there are no special instructions, 100 simulations are performed for each distribution function case in this paper. The sample size is chosen as 10^4 with a sampling interval of 1 h, which corresponds to a duration of 1.14 years for each sample data to fit the common medium-scale data situation in engineering.

According to the expressions of Hill-type and Zipf-type tail estimators (Eq. (19) and Eq. (21)), it can be found that these statistics depend on the selection of the threshold $k(n)$. As mentioned earlier, a too small threshold $k(n)$ will lead to a large variance, but at the same time, too large $k(n)$ will also increase the bias of the tail estimators. In the calculation, the $k(n)$ when convergence is reached in the Hill plot can be selected, or the bootstrap method can be used (Danielsson et al., 2001; Németh and Zempléni, 2020). The positive integer $k(n)$ is taken as 1/10 of the sample size in this example, i.e., $k(n) = 10^3$.

To test the accuracy and stability of the proposed modified EMC method, in addition to the classical EMC method, the model with higher prediction accuracy among the Hermite central moment and L-moment models is also selected. Considering the accuracy, stability, and computational efficiency, the order of the Hermite model is set to 4. Table 1 shows the extreme value prediction errors of the proposed method, where the error distribution parameters in each cell are the mean and standard deviation of the percentage of prediction errors, respectively. In calculation, for a time series with sampling interval Δt (in seconds), the relationship between failure probability P_{f0} and return period T_0 is:

$$P_{f0} = \frac{\Delta t}{T_0 \cdot 365 \cdot 24 \cdot 3600} \tag{26}$$

Fig. 3 shows the boxplot of the proposed modified EMC model and the models for comparison, with the vertical axis showing the percentage error of the extreme value prediction results with a return period of 10 years. The red dashed lines are used to separate the cases with different shape parameters k_w, and the green dashed line is the zero error percentage standard line. The horizontal coordinates Hermite-4c, EMC, and Hill-EMC represent the fourth-order central moment Hermite model, the classical EMC model, and the proposed modified EMC model, respectively. It can be observed that the proposed model avoids overestimation of the extreme values of the Weibull distribution, and the stability is significantly improved. In addition, it is found that the selection of Hill-type or Zipf-type tail estimators has no significant effect on the predicted extreme value.

Table 1
Modified EMC method extreme value prediction error distribution parameters (Weibull distribution case).

	Error parameters of 1-yr extreme value	Error parameters of 10-yr extreme value	Error parameters of 100-yr extreme value
$k_w = 2$	0.138, 2.963	0.235, 3.834	0.327, 4.570
$k_w = 5$	0.095, 1.161	0.134, 1.458	0.169, 1.712
$k_w = 8$	0.071, 0.637	0.089, 0.806	0.105, 0.951

As mentioned earlier, the tail behavior described by the EMC extrapolation function is influenced only by the shape parameter c. Classical EMC methods obtain the parameter c through a least-squares fit, leading to higher uncertainty in this parameter and the extreme value prediction result. The proposed method, on the other hand, allows for a more accurate estimation of the parameter c. For the case of $k_w = 5$, the histograms of parameter c obtained by the two methods are plotted in Fig. 4, respectively. As illustrated, the parameter c estimated by the classical EMC method has a high degree of dispersion, and the value is not accurately estimated even once. The proposed modified EMC method can estimate the shape parameter more accurately, which effectively improves the extreme value prediction accuracy.

4.1.2. Normal distribution

The PDF of the normal distribution is:

$$f(x) = \frac{1}{\sigma_n\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu_n}{\sigma_n}\right)^2\right] \tag{27}$$

Its skewness and kurtosis are constant, and the tail behavior does not change with the distribution parameters.

The Weibull probability plot of a sample of a normal distribution with $\mu_n = 0$, $\sigma_n = 1$ is shown in Fig. 5. As illustrated, the slope of the plot tends to increase gradually with the cumulative probability, i.e., the tail estimator does not converge, so extrapolation of the estimator is considered to obtain an estimate of the EMC shape parameter c.

The normal distribution parameters $\mu_n = 0$, and $\sigma_n = 1$ are taken with sample sizes of 10^4 and 10^5, respectively, and the corresponding random numbers are generated as sample data. The tail range of 75% to 95% quantiles F_i is taken, and the corresponding values of c_i are calculated and then extrapolated to a certain return period cumulative probability. Table 2 shows the prediction errors for the extrapolation to the 10-year cumulative probability.

4.1.3. Log-normal distribution

The PDF of the log-normal distribution is:

$$f(x) = \frac{1}{x\sigma_{lo}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{lnx-\mu_{lo}}{\sigma_{lo}}\right)^2\right] \tag{28}$$

Only the parameter σ_{lo} affects the tail behavior of the log-normal distribution.

The Weibull probability plot of a log-normal sample with $\mu_{lo} = 0$, $\sigma_{lo} = 0.5$ is shown in Fig. 6, which indicates that the slope of the plot tends to decrease from the left tail to the right tail.

Different parameters σ_{lo} and $\mu_{lo} = 0$ are taken and the corresponding random numbers are generated as sample data. Taking the tail range of the 75% to 95% quantile, the corresponding c_i are calculated and then extrapolated to a 10-year return period cumulative probability. Table 3 and Fig. 7 show the extreme value prediction errors of the proposed method and the boxplot of the percent errors of predicted extreme values for different models with a return period of 10 years, respectively. It is found that the modified EMC method has higher stability than the classical EMC method but does not perform as well as the L-moment Hermite model.

4.2. Examples of wind speed series

The accuracy and stability of the proposed method are investigated in Section 4.1 using random samples of several typical distributions with different tail characteristics. However, the real ocean engineering data have both distribution and frequency domain features, and the above random sequence cannot truly reflect the actual situation of the field monitoring data. In this section, numerical samples satisfying both the given PDF and the given power spectral density (PSD) are generated,

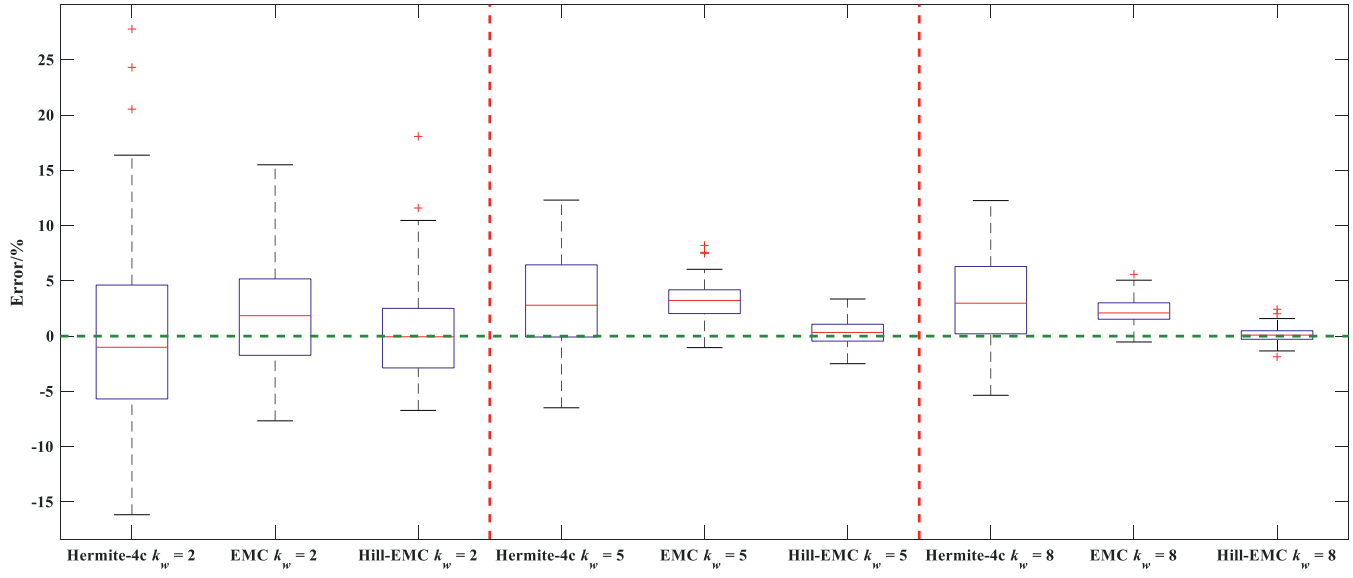


Fig. 3. Boxplot of extreme value prediction errors of the modified EMC model and models for comparison (Weibull distribution case, return period = 10 yrs).

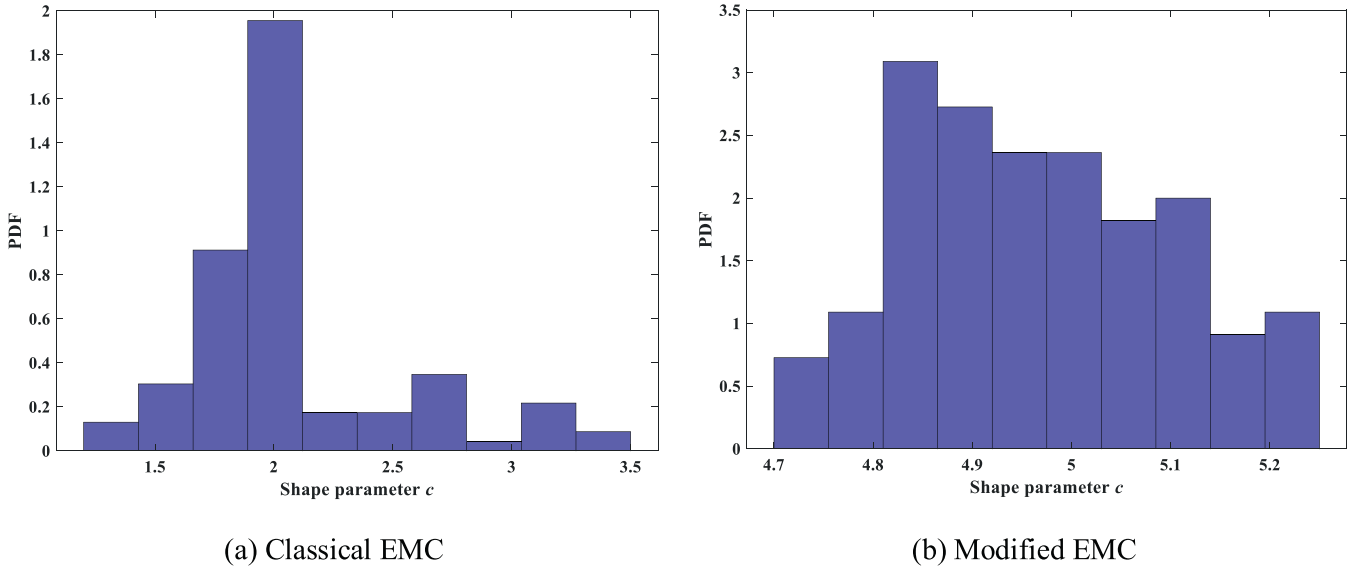


Fig. 4. Histogram of simulated values of classical EMC and modified EMC for Weibull distribution shape parameter c (standard value $k_w = 5$).

and the extreme value prediction performance of the proposed method is evaluated on this basis.

The artificial wind speed samples employed in this work are obtained using the generation method proposed by Ambrosio et al. (2020) as follows:

1. Generate a random sequence $X_{n,pdf}$ based on a given PDF and a random sequence $X_{n,psd}$ based on a given PSD.
2. Calculate the amplitude spectrum $A_{n,psd}$ of the sequence $X_{n,psd}$.
3. Reorder the sequence $X_{n,pdf}$ according to $X_{n,psd}$, that is, the minimum value of the sequence $X_{n,pdf}$ must be located at the same location as the minimum value of the sequence $X_{n,psd}$, the location of the second smallest element of the sequence $X_{n,pdf}$ must also be located at the same location as $X_{n,psd}$, and so on, to obtain a new sequence $\widetilde{X}_{n,pdf}$.
4. Calculate the phase spectrum $\varphi_{n,psd}$ of the new sequence $\widetilde{X}_{n,pdf}$ by Fourier transform. Next, apply the inverse Fourier transform to the frequency domain signal formed by the combination of the amplitude spectrum $A_{n,psd}$ and phase spectrum $\varphi_{n,psd}$, and take the real part of the obtained complex series as the new sequence $\widetilde{X}_{n,psd}$.
5. Repeat steps 2–4 until the final sequence $\widetilde{X}_{n,pdf}$ and $\widetilde{X}_{n,psd}$ converge, and the desired sequence is obtained.

In this section, short-term wind speed sequences with a sampling frequency of 1 Hz and a sample size of 10^4 are considered, corresponding to a duration of about 3 h. The empirical Ochi and Shin spectrum is applied for design of offshore structures (DNV GL, 2017):

$$\frac{fS(f)}{u^{*2}} = \begin{cases} 583f_* & \text{for } 0 \le f_* \le 0.003 \\ \dfrac{420f_*^{0.7}}{\left(1+f_*^{0.35}\right)^{11.5}} & \text{for } 0.003 < f_* \le 0.1 \\ \dfrac{838f_*}{\left(1+f_*^{0.35}\right)^{11.5}} & \text{for } f_* > 0.1 \end{cases} \tag{29}$$

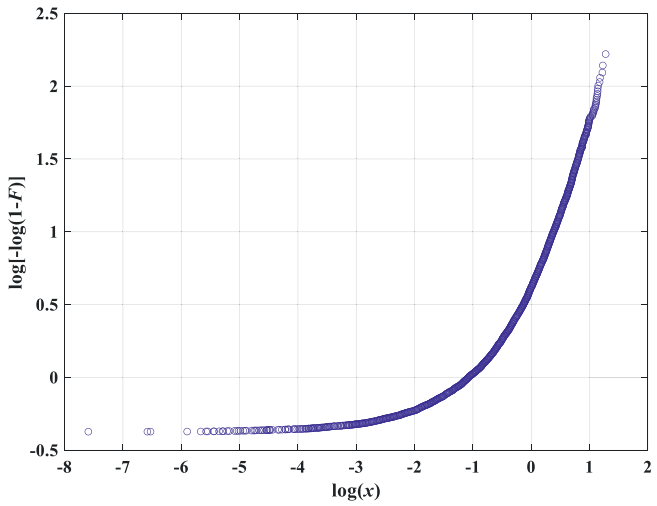


Fig. 5. Weibull probability plot for a normal distribution sample.

Table 2
Modified EMC method extreme value prediction error distribution parameters (normal distribution case, extrapolated to the cumulative probability of 10-yr return period).

Sample size	Error parameters of 1-yr extreme value	Error parameters of 10-yr extreme value	Error parameters of 100-yr extreme value
10^4	0.852, 3.426	1.414, 4.305	1.947, 5.059
10^5	1.140, 0.994	1.696, 1.205	2.209, 1.384

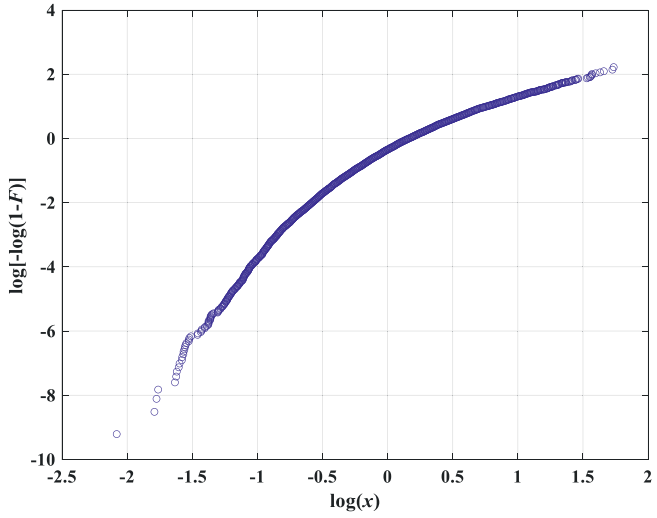


Fig. 6. Weibull probability plot for a log-normal distribution sample.

where $f_* = \frac{f \cdot z}{U_{10}(z)}$, z is the height above sea level, U_{10} is the 10-min mean wind speed, u^* is the friction velocity. In this paper, the 1-h wind speed mean value $U_0 = 10$ m/s at 10 m altitude is taken so that $z = 10$ m.

Three distribution functions commonly used in ocean engineering are selected for the wind speed PDF, namely, Weibull distribution, Gamma distribution, and Gumbel distribution. The parameters of the distributions are determined by the sample mean and variance. The wind speed sample means are obtained by discounting through a logarithm model (DNV GL, 2017):

$$\frac{U_{pdf}}{U_0} = \frac{1 + 0.137 ln\frac{z}{H} - 0.047 ln\frac{T}{T_{10}}}{1 + 0.137 ln\frac{z}{H} - 0.047 ln\frac{T_0}{T_{10}}} \tag{30}$$

where U_{pdf} is the sample mean to be sought, $H = 10$ m, $T_0 = 1$ h, $T_{10} = 10$ min, and T is the sample duration. The wind speed sample variance is chosen as the zero-order spectral moment of the wind speed PSD, where the spectral moment is defined as:

$$m_j = \int_0^\infty f^j S_U(f) \mathrm{d}f \tag{31}$$

where f is the frequency, $S_U(f)$ is the given PSD, and j is the order. The above definition of the sample mean and variance ensures that the sequences generated based on the given PSD and PDF have consistent statistical information. Fig. 8 shows the artificial series and its histogram with PSD for a sample obeying the Gumbel distribution obtained by the above method.

Table 4 and Fig. 9 show the extreme value prediction errors of the proposed method and the boxplot of the percentage of prediction errors of different prediction models with a return period of 24 h, respectively. As illustrated, although there is a certain degree of underestimation in the method proposed for the Gumbel examples, the relative errors are always controlled within 10%. Besides, for all the three distributions, the proposed method has the highest stability. Therefore, the proposed modified EMC model has improved computational accuracy and stability and can better meet the requirements of short-term extreme value prediction in marine engineering.

5. Applications of the modified Enhanced Monte Carlo method

This section will show the potential engineering applications of the proposed method through two engineering cases. The first case is the extreme value prediction of the current speed in each layer based on the measured current profile data. The second case concerns the hot-spot stress of catenary mooring lines on floating platforms, where the tension is based on the catenary equation and the measured six-degree-of-freedom displacement of the platform. Through the above two cases, the performance of the proposed modified EMC method in extreme value prediction of environmental loads and structural responses is investigated.

5.1. Current extreme value prediction

The Liuhua 11–1 oilfield is located in the Pearl River Mouth Basin (East 115°42′, North 20°49′) and is located 220 km southeast of Hong Kong. The in situ monitoring platform is a semi-submersible drilling platform operating in the Liuhua 11–1 oilfield (Yu et al., 2020), as shown in Fig. 10.

The measured data used in this paper come from the prototype monitoring system established on the platform by the research group (Du, 2016). A set of acoustic Doppler current profiler (ADCP) equipment is utilized to collect current profile information. The obtained current profile data are divided into 14 layers according to the water depth

Table 3
Modified EMC method extreme value prediction error distribution parameters (log-normal distribution case).

	Error parameters of 1-yr extreme value	Error parameters of 10-yr extreme value	Error parameters of 100-yr extreme value
$\sigma_{lo} = 0.25$	−0.964, 3.869	−2.137, 5.509	−3.615, 7.004
$\sigma_{lo} = 0.5$	−1.269, 7.439	−3.480, 10.867	−6.370, 13.818
$\sigma_{lo} = 0.65$	−0.282, 11.800	−2.226, 17.737	−5.058, 22.779

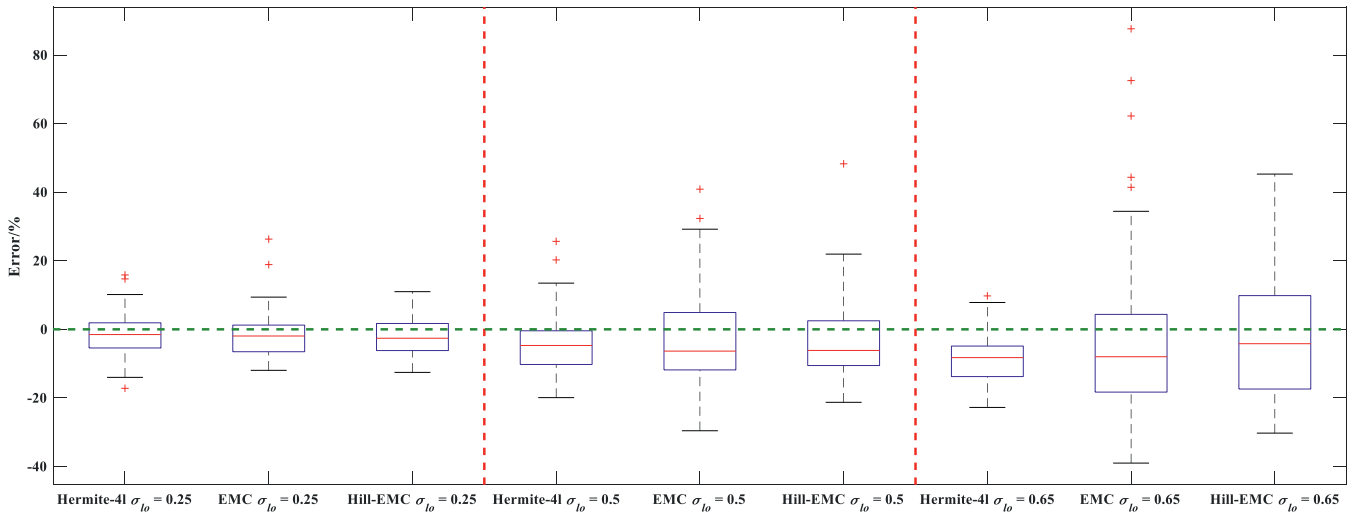
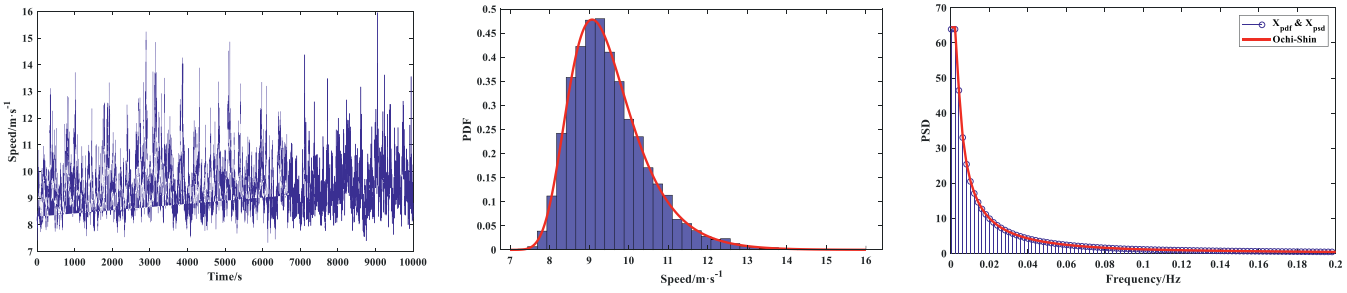


Fig. 7. Boxplot of extreme value prediction errors of the modified EMC model and models for comparison (log-normal distribution case, return period = 10 yrs).



(a) Sample time history (b) Sample histogram (c) Sample PSD

Fig. 8. Short-term simulation time history, statistical and frequency domain information of an artificial wind speed series obeying Gumbel distribution.

Table 4
Modified EMC method extreme value prediction error distribution parameters (artificial wind speed series).

Distribution	Error parameters of 12-h extreme value	Error parameters of 24-h extreme value	Error parameters of 72-h extreme value
Weibull	-4.60×10^{-4}, 0.544	−0.001, 0.574	−0.003, 0.619
Gamma	−0.926, 1.365	−1.085, 1.475	−1.348, 1.641
Gumbel	−2.161, 2.296	−2.507, 2.518	−3.069, 2.861

(30–121 m) with an interlayer depth of 7 m. The current profile dataset has a total of 8205 current profiles, and the sampling interval is 1 h. The speed and direction distribution of the first layer is shown in Fig. 11.

Since the measurement duration of the measured current data is about 1 yr, the prediction results of a long-term return period such as 100 yrs. may have great uncertainty. Therefore, this paper only focuses on the return periods of 1 yr and 10 yrs. Table 5 shows the design values of the extreme current speeds in the sea area where the platform is located. For the water depths not involved in the table, the design speeds are obtained by linear interpolation.

Due to the time correlation characteristics of the measured ocean current data series, the ACER method is used instead of the classical EMC method (Liu et al., 2018). The method proposed in this paper is also adjusted accordingly, that is, applying the tail index correction strategy to the ACER method. In addition, the GEV and POT models are selected as comparison models. Table 6 lists the prediction results and 95% CI of the above methods for the extreme speed of the first layer, where the 95% CIs for the extreme values of the POT are calculated by a bootstrap method (Karpa and Naess, 2013).

Fig. 12 shows the extreme value prediction results and their 95% CI of the above models for the speed of each layer with a 10-yr return period. The blue curve in each subplot represents the current speed design values, the red solid line represents the predicted values, and the red dotted lines represent the 95% CI. Each subplot has the same dimensions and axis ranges. It can be found that the spatial distribution of extreme speeds of the proposed method is most consistent with the design values. On the other hand, there is a big difference between the spatial distribution characteristics of the design values and the extreme speeds of the GEV method and the POT method, and the POT method gives an overly wide CI.

5.2. *Mooring hot-spot stress extreme value prediction*

The mooring system of the aforementioned platform is a multi-point catenary mooring system, which adopts a mooring method in which 11 mooring chains are arranged asymmetrically (Shen et al., 2020), as shown in Fig. 13. Since the focus of this paper is not on the accurate analysis of the mooring responses, the catenary equation is used to calculate the top tension of mooring line 1#, ignoring the effect of current loads and other dynamic effects, including the effect of drag force on the mooring line and the influence of vortex-induced vibration (DNV GL, 2017; Yu et al., 2020).

Each mooring line consists of four segments. From the platform to the seabed, they are the platform anchor chain, the suspension cable, the

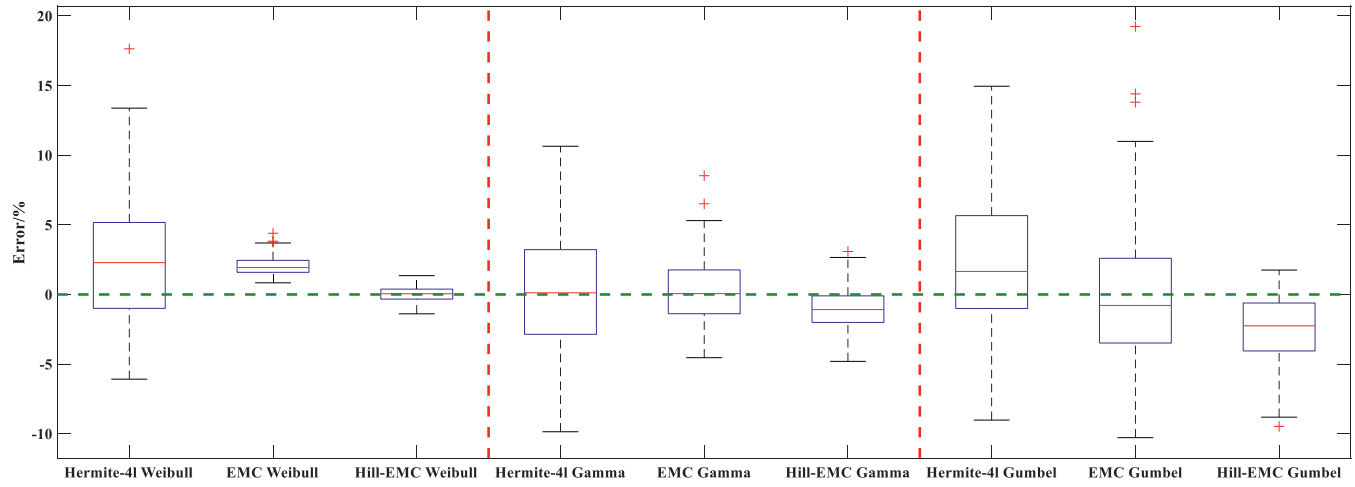


Fig. 9. Boxplot of extreme value prediction errors of the modified EMC model and models for comparison (artificial wind speed series case, return period = 24 h).

Fig. 10. In situ monitoring platform.

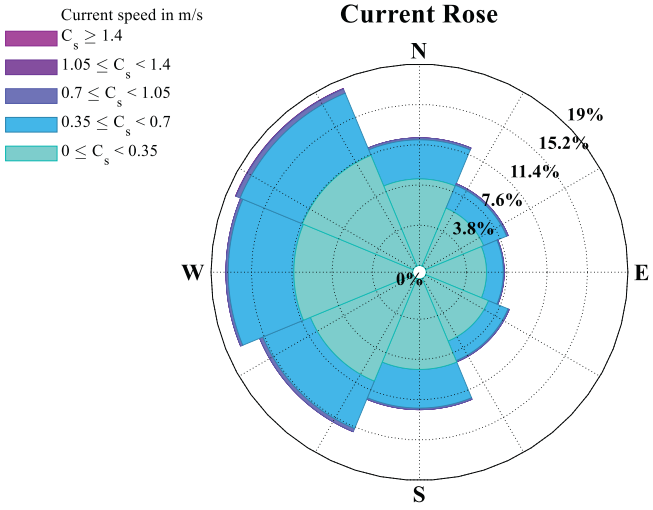


Fig. 11. Measured current speed and direction distribution of layer 1.

Table 5
Design values of current speeds in the platform sea area.

Return period	Design values					
1 yr	Depth(m)	0	23	68	113	159
	Speed(m/s)	1.46	1.30	1.00	0.86	0.76
10 yrs	Depth (m)	0	23	68	113	159
	Speed (m/s)	1.73	1.57	1.27	1.05	0.91

Table 6
Predicted extreme values for the current speed of the first layer by the proposed method, GEV, and POT.

Method	Return period/yr	Extreme speed/m·s^{-1}	95% CI/m·s^{-1}
Proposed method	1	1.17	(1.04, 1.25)
GEV	1	1.23	(1.14, 1.32)
POT	1	1.38	(1.17, 1.60)
Proposed method	10	1.36	(1.19, 1.46)
GEV	10	1.43	(1.30, 1.56)
POT	10	1.53	(1.24, 1.87)

lying anchor chain, and the anchor cable. The theoretical design lengths of each segment of mooring line 1# are 220.98 m, 502.92 m, 524.26 m, and 121.92 m, respectively, and the weight parameters of each segment are shown in Table 7 (Shen et al., 2020). The mooring line material (used in each segment) is the R4 grade anchor chain steel introduced in DNVGL-OS-E302 (DNV GL, 2018), with a yield strength of 580 MPa and a tensile strength of 860 MPa (both are minimum mechanical properties).

The six-degree-of-freedom motion response information of the floating structure of the platform is monitored by the Differential Global Positioning System (DGPS) and the Inertial Navigation System (INS) (Du, 2016). The DGPS data include longitude, latitude, and altitude, and surge, sway and heave information can be obtained based on the calculation of ellipsoid meridian arc length and the coordinate transformation (Li, 1995; Jiang and Yan, 1998), as well as the initial position coordinates of the platform and the position of the center of gravity. The INS data include roll angle, pitch angle, and heading angle, and the roll, pitch, and yaw information can be obtained based on the initial state of the platform.

The sampling frequencies of the DGPS and INS data are 1 Hz and 10 Hz, respectively, and the total data duration is the same as that of the aforementioned current data. The calculated six-degree-of-freedom data are divided into short-term series according to the 3-h duration, and the maximum value of the absolute value is taken for each short-term time

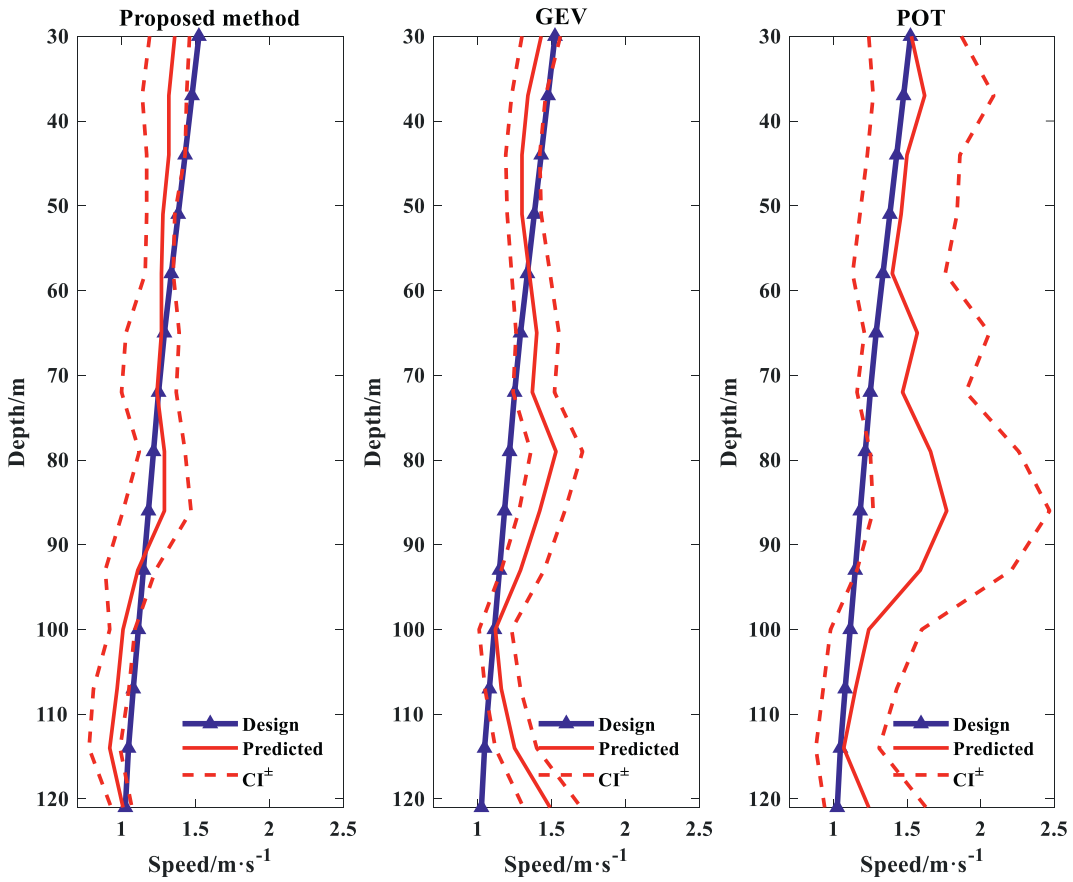


Fig. 12. Extreme current speeds with a 10-yr return period and comparison of ocean environmental design index in Liuhua sea area.

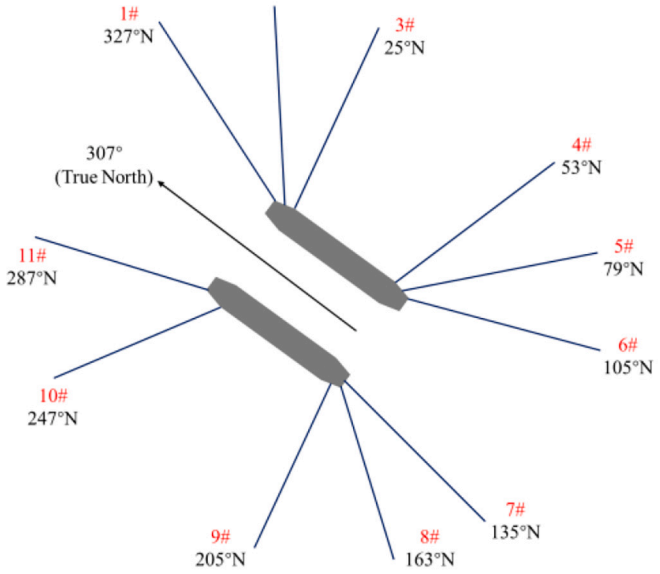


Fig. 13. Schematic diagram of mooring system layout.

course to obtain extreme value samples of each degree of freedom. The GEV distribution is used to fit the extreme values of each degree of freedom, and the corresponding distribution type and the extreme values predicted by the GEV model are obtained as shown in Table 8.

In this paper, based on the above distributions, a six-degree-of-freedom responses dataset is generated. The sampling interval is set to 3 h, and the data duration is 1 yr, so the length of each sequence is 2920.

Table 7
Weight parameters of each mooring line.

	Platform anchor chain	Suspension cable	Lying anchor chain	Anchor cable
Dry weight (10^3 kg/m)	0.27461	0.07072	0.36909	0.07072
Wet weight (10^3 kg/m)	0.31639	0.08415	0.41437	0.08415

Table 8
Top tension at the equilibrium position of each mooring line.

Degree of freedom	Distribution	1-yr extreme value	10-yr extreme value	100-yr extreme value
Rolling	Weibull	2.84°	2.87°	2.88°
Pitching	Weibull	2.65°	2.86°	3.00°
Yawing	Fréchet	4.22°	6.58°	10.04°
Surging	Gumbel	5.45 m	6.76 m	8.10 m
Swaying	Gumbel	5.40 m	6.82 m	8.30 m
Heaving	Weibull	1.74 m	1.86 m	1.93 m

Based on the catenary equation, the governing equation of the aforementioned four-component mooring line is constructed (Wu et al., 2016). Combined with the top displacement calculated from the motion data, the time series of the top tension of the 1# mooring line is obtained. The hot spot stress sequence is then calculated by finite element simulation, and the stress approximately obeys the Weibull distribution, as shown in Fig. 14.

Table 9 lists the prediction results and 95% CI of the hot spot stress of the 1# mooring line of the proposed method, GEV, and POT method. It

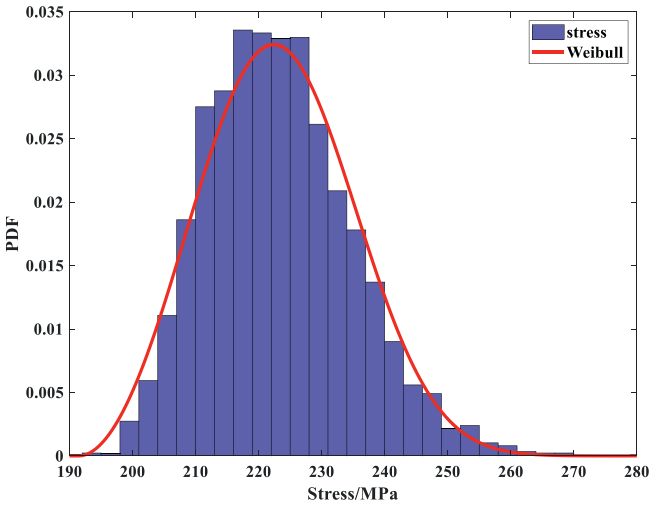


Fig. 14. Hot spot stress distribution of 1# mooring line.

can be found that the extreme value and CI predicted by the proposed method are close to the results of the GEV model, while the POT method significantly overestimates the extreme value.

It should be noted that when predicting the extreme value of the structural response, the rationality of the method needs to be evaluated in advance. For this engineering example, the hot spot stress is significantly below the yield limit of the material. Considering that mooring lines are frequently replaced, complex mechanical behaviors such as creep and yielding of materials during service can be ignored.

6. Discussion

According to the aforementioned calculation process and results, the proposed modified EMC method has insufficient stability in the heavy-tailed distribution prediction results. This is mainly because of the assumption introduced by the EMC method, that is, the function $q(\lambda)$ in Eq. (2) is assumed to be a slow-varying function. For heavy-tailed distributions, this assumption usually does not hold, leading to insufficient applicability of EMC, ACER, and the proposed method to heavy-tailed distribution data.

In the case that the sample size cannot be increased, the treatment of long-tailed distribution samples should be considered. Logarithmic transformation is a common data processing tool in statistical analysis, which can effectively reduce the skewness and kurtosis of the distribution and the non-Gaussianity of samples with significant super-Gaussian characteristics (Nam and Hong, 2021). Typical long-tailed distributions include log-normal and Pareto distributions. It is well known that taking the logarithm of a log-normal sample will make the new sample obey a normal distribution, while a similar operation on a Pareto sample can make the new sample obey exponential distribution. Therefore, the logarithm operation can theoretically handle the extreme value prediction problem of long-tailed distribution.

For the practical calculation, S refers to the sample and R refers to a large sample value. The safety margin equation is then taken as $M = \log(R) - \log(S)$. Based on the proposed extreme value prediction model in Section 3, the logarithmic extreme value is calculated, and then the extreme value to be sought is obtained through the exponent operation.

The PDF of the Pareto distribution is:

$$f(x) = \frac{\alpha_p x_p^{\alpha_p}}{x^{\alpha_p+1}} \tag{32}$$

where α_p and x_p are the shape and scale parameters, respectively. Only α_p affects the tail behavior of the distribution, and a larger α_p corresponds to a smaller skewness and kurtosis.

Here, we take the logarithm of a Pareto distribution sample with $x_p = 1$, $\alpha_p = 5$, and the Weibull probability plot of the new sample is shown in Fig. 15. As good linearity can be observed in the figure, the aforementioned Hill-type estimator is directly applied as an estimate of the EMC shape parameter c.

Different shape parameters α_p and scale parameter $x_p = 1$ are then taken to generate the corresponding random numbers as sample data, where $k(n) = 10^3$. Table 10 and Fig. 16 show the extreme value prediction errors of the proposed method and the boxplot of errors for extreme values with a return period of 10 years for different models, respectively. According to the results, the modified EMC method after the logarithmic operation has the highest stability among the three models and avoids the overall underestimation of extreme values. However, for lower α_p, the stability of the prediction results of the proposed method is not satisfactory. After further inspection, it is found that the proposed method has good prediction accuracy for the logarithm of estimated extreme values, but the exponentiation operation leads to a significant increase in the uncertainty of the final results.

In summary, for sample data with a non-converging tail estimator and long-tailed characteristics, the data can be processed by logarithmic transformation. The extreme value prediction can then be performed based on the proposed modified EMC method. However, the accuracy and stability of the proposed method may still not meet the engineering requirements. While this aspect is beyond the scope of this paper, we suggest choosing other appropriate functions instead of the original EMC extrapolation function (Eq. (3)), such as using the Pareto distribution function or its extended form (Lee and Kim, 2019, Andria, 2022). However, to the best of the authors' knowledge, the measured data of marine environmental loads and structural responses rarely show severe long-tail effects. Thus, the proposed method has sufficient accuracy and stability for extreme value prediction in the usual marine engineering application scenarios.

Table 9
Predicted extreme values for the hot spot stress of 1# mooring line by the proposed method, GEV, and POT.

Method	Return period/yr	Extreme value/MPa	95% CI/ MPa
Proposed method	1	271.67	(266.23, 276.26)
GEV	1	270.25	(265.17, 275.91)
POT	1	285.43	(270.52, 304.14)
Proposed method	10	281.79	(274.19, 288.15)
GEV	10	277.63	(271.13, 285.04)
POT	10	294.00	(272.10, 324.59)

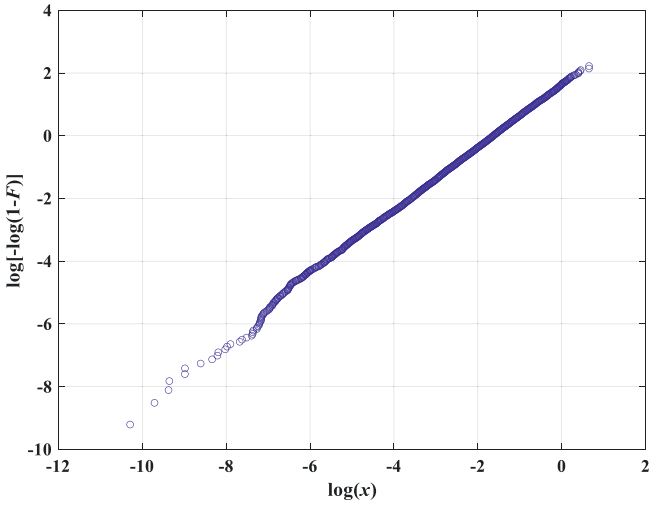


Fig. 15. Weibull probability plot for a log-transformed Pareto distribution sample.

Table 10
Modified EMC method extreme value prediction error distribution parameters (Pareto distribution case).

	Error parameters of 1-yr extreme value	Error parameters of 10-yr extreme value	Error parameters of 100-yr extreme value
$\alpha_p = 3$	1.527, 11.246	2.586, 16.546	4.008, 22.498
$\alpha_p = 5$	1.950, 7.297	3.087, 10.858	4.440, 14.804
$\alpha_p = 10$	0.171, 4.414	0.380, 6.596	0.655, 9.027

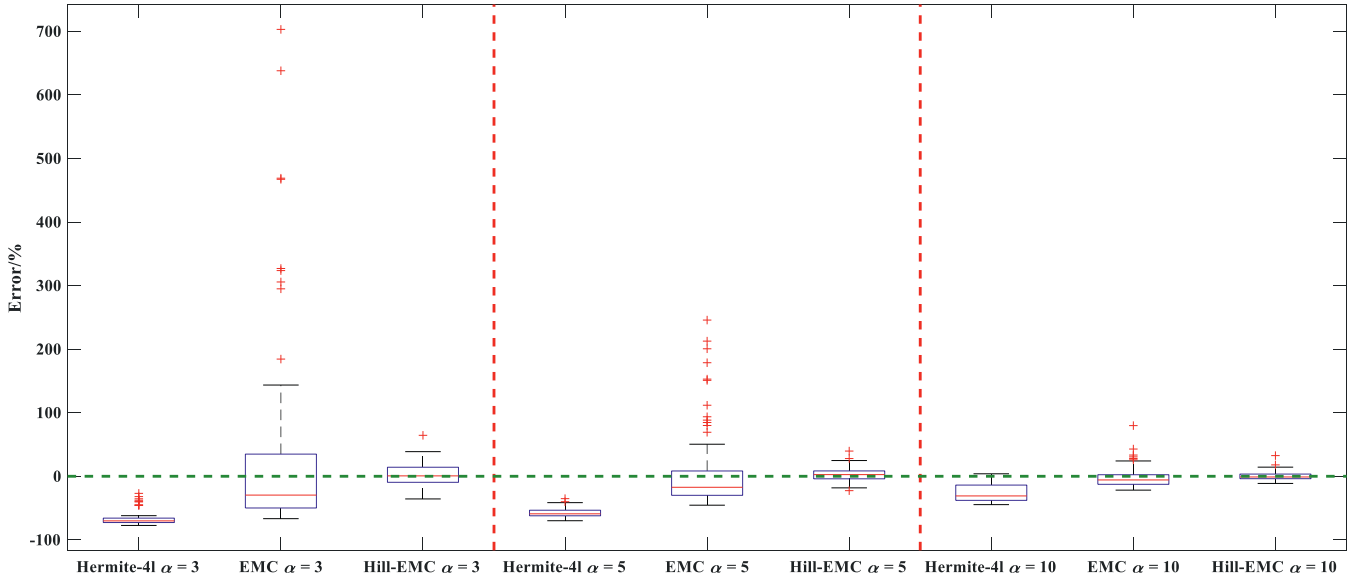


Fig. 16. Boxplot of extreme value prediction errors of the modified EMC model and models for comparison (Pareto distribution case, return period = 10 yrs).

7. Conclusions

As one of the most important disciplines of statistics of extremes in applied science, extreme value theory is widely used in the field of marine engineering. Due to its special application in real ocean environmental states, the scale of the monitoring data obtained is limited. Aiming to solve the problem of extreme value prediction based on medium-scale data, this paper proposes a modified EMC extreme value prediction method based on the tail index correction, and the conclusions are as follows:

1. The Hill-type estimator was adopted to quantitatively evaluate the tail behavior of data samples. This modified the EMC method, which could also be used as a quantitative index to evaluate the tail behavior of data.
2. Compared with the classical EMC method, the modified EMC method proposed in this paper had higher accuracy and stability in extreme value prediction, and the prediction error of the proposed method was also more suitable than the Hermite model for the engineering requirements for most of the distribution functions under study.
3. For some cases where the tail estimator could not reach convergence, an extrapolation method for the estimator was proposed. It was easy to operate and insensitive to the selected extrapolated cumulative probabilities, further improving the engineering applicability of the proposed method.

In conclusion, this paper investigated the ability of the Hill-type estimator to quantitatively assess the tail behavior of medium-scale data, and on this basis, a modification of the classical EMC method was achieved. The method was also applicable to the correction of ACER extreme value prediction results, which could capture the tail characteristics of the sample distribution while considering the temporal correlation of the sample data series. The proposed method provides useful guidance for the extreme value analysis of marine environmental loadings and structural responses.

CRediT authorship contribution statement

Siyuan Yu: Methodology, Validation, Formal analysis, Investigation, Writing – original draft. **Wenhua Wu:** Supervision, Writing – review & editing, Funding acquisition, Project administration. **Arvid Naess:** Supervision, Writing – review & editing.

Declaration of Competing Interest

None.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was financially supported by the National Key Research and Development Program of China (No.2021YFA1003501), the Key R&D Program of Shandong Province (2019JZZY010801), the Central Guidance on Local Science and Technology Development Fund of Shenzhen (2021Szvup021), the Fundamental Research Funds for the Central Universities (DUT22ZD209，DUT21ZD209), These supports are gratefully acknowledged.

References

Ambrosio, D.D., Schoukens, J., De Troyer, T., et al., 2020. Synthetic wind speed generation for the simulation of realistic diurnal cycles. J. Phys. Conf. Ser. 1618, 062019 https://doi.org/10.1088/1742-6596/1618/6/062019.

Andria, J., 2022. A computational proposal for a robust estimation of the Pareto tail index: an application to emerging markets. Appl. Soft Comput. 114, 108048 https://doi.org/10.1016/j.asoc.2021.108048.
Beirlant, J., Goegebeur, Y., Segers, J., et al., 2004. Statistics of Extremes: Theory and Applications. John Wiley & Sons, Hoboken, NJ.
Bruserud, K., Haver, S., Myrhaug, D., 2018. Joint description of waves and currents applied in a simplified load case. Mar. Struct. 58, 416–433. https://doi.org/10.1016/j.marstruc.2017.12.010.
Caeiro, F., Henriques-Rodrigues, L., Gomes, M.I., 2022. The use of generalized means in the estimation of the Weibull tail coefficient. Comput. Math. Methods 2022, 7290822. https://doi.org/10.1155/2022/7290822.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.
Danielsson, J., de Haan, L., Peng, L., et al., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivar. Anal. 76 (2), 226–248. https://doi.org/10.1006/jmva.2000.1903.
de Wet, T., Goegebeur, Y., Guillou, A., et al., 2016. Kernel regression with Weibull-type tails. Ann. Inst. Stat. Math. 68 (5), 1135–1162. https://doi.org/10.1007/s10463-015-0531-z.
DNV GL, 2017. Recommended Practice — DNVGL-RP-C205 Environmental Conditions and Environmental Loads.
DNV GL, 2018. Recommended Practice — DNVGL-OS-E302 Offshore Mooring Chain.
Du, Y., 2016. Full-Scale Testing Method and Prototype Monitoring Technology for Deep-Water Floating Platforms. Dalian University of Technology, Liaoning.
Feng, X., Li, H.C., Feng, X.B., et al., 2021. Dependence of ocean wave return levels on water depth and sampling length: a focus on the South Yellow Sea. Ocean Eng. 219, 108295 https://doi.org/10.1016/j.oceaneng.2020.108295.
Galiatsatou, P., Makris, C., Krestenitis, Y., et al., 2021. Nonstationary extreme value analysis of nearshore sea-state parameters under the effects of climate change: application to the greek coastal zone and port structures. J. Mar. Sci. Eng. 9 (8), 817. https://doi.org/10.3390/jmse9080817.
Gao, S., 2019. Non-Gaussian Response Extrema and Fatigue of Marine Structures. Dalian University of Technology, Liaoning.
Gao, S., Zheng, X.Y., Huang, Y., 2018. Hybrid C-and L-moment–based Hermite transformation models for non-Gaussian processes. J. Eng. Mech. 144 (2), 04017179. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001408.
Gao, S.J., Liu, F.S., Jiang, C.Y., 2021. Improvement study of modal analysis for offshore structures based on reconstructed displacements. Appl. Ocean Res. 110, 102596 https://doi.org/10.1016/j.apor.2021.102596.
Gardes, L., Girard, S., 2008. Estimation of the Weibull tail-coefficient with linear combination of upper order statistics. J. Stat. Plan. Infer. 138 (5), 1416–1427. https://doi.org/10.1016/j.jspi.2007.04.026.
Girard, S., 2004. A Hill type estimate of the Weibull tail-coefficient. Commun. Stat.-Theory Methods 33 (2), 205–234. https://doi.org/10.1081/STA-120028371.
He, F., Wang, H.J., Tong, T., 2020. 2020. Extremal linear quantile regression with Weibull-type tails. Stat. Sin. 30 (3), 1357–1377. https://doi.org/10.5705/ss.202018.0073.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Stat. 3 (5), 1163–1174. https://doi.org/10.1214/aos/1176343247.
Jiang, C.G., Yan, G.F., 1998. A novel method for calculation of meridional arc length of ellipsoid. Eng. Survey. Map. 4, 38–42.
Karpa, O., Naess, A., 2013. Extreme value statistics of wind speed data by the ACER method. J. Wind Eng. Ind. Aerodyn. 112, 1–10. https://doi.org/10.1016/j.jweia.2012.10.001.
Lee, S., Kim, J.H.T., 2019. Exponentiated generalized Pareto distribution: properties and applications towards extreme value theory. Commun. Stat.-Theory Methods 48 (8), 2014–2038. https://doi.org/10.1080/03610926.2018.1441418.
Li, X.X., 1995. Conversion of GPS measurement results. Railway Investigat. Survey. 2, 29–35.
Lin, Q., Liu, M., Yang, Q.S., et al., 2020. A comparative study on moment-based translation process methods for the peak factor estimation of non-gaussian wind pressures. Eng. Mech. 37 (4), 78–86. https://doi.org/10.6052/j.issn.1000-4750.2019.04.0201.
Liu, M., 2018. Current Load Model Orient to Failure Modes of Underwater Structures. Dalian University of Technology, Liaoning.
Liu, M., Wu, W., Tang, D., et al., 2018. Current profile analysis and extreme value prediction in the LH11-1 oil field of the South China Sea based on prototype monitoring. Ocean Eng. 153, 60–70. https://doi.org/10.1016/j.oceaneng.2018.01.064.
Lu, J.J., Wang, Y.G., 2020. Prediction of long-term loads of jacket offshore wind turbine in nonlinear mixed sea states. Shipbuild. China 61 (3), 47–57.
Minkah, R., de Wet, T., Ghosh, A., 2021. Robust estimation of Pareto-type tail index through an exponential regression model. Commun. Stat.-Theory Methods 1–21. https://doi.org/10.1080/03610926.2021.1916530.
Naess, A., Gaidai, O., 2009. Estimation of extreme values from sampled time series. Struct. Saf. 31 (4), 325–334. https://doi.org/10.1016/j.strusafe.2008.06.021.
Naess, A., Leira, B.J., Batsevych, O., 2009. System reliability analysis by enhanced Monte Carlo simulation. Struct. Saf. 31 (5), 349–355. https://doi.org/10.1016/j.strusafe.2009.02.004.
Nam, Y., Hong, S., 2021. Growth mixture modeling with nonnormal distributions: implications for data transformation. Educ. Psychol. Meas. 81 (4), 698–727. https://doi.org/10.1177/0013164420976773.
Németh, L., Zempléni, A., 2020. Regression estimator for the tail index. J. Stat. Theory Pract. 14 (3), 1–23. https://doi.org/10.1007/s42519-020-00114-7.
Raed, K., Teixeira, A.P., Soares, C.G., 2020. Uncertainty assessment for the extreme hydrodynamic responses of a wind turbine semi-submersible platform using different environmental contour approaches. Ocean Eng. 195, 106719 https://doi.org/10.1016/j.oceaneng.2019.106719.
Rojo, J., 1996. On tail categorization of probability laws. J. Am. Stat. Assoc. 91 (433), 378–384. https://doi.org/10.2307/2291417.
Rojo, J., Ott, R.C., 2010. Testing for Tail Behavior using Extreme Spacings. arXiv preprint. arXiv:1011.6458.
Shao, Z.X., Liang, B.C., Li, H.J., et al., 2018. Study of sampling methods for assessment of extreme significant wave heights in the South China Sea. Ocean Eng. 168, 173–184. https://doi.org/10.1016/j.oceaneng.2018.09.015.
Shen, H., Zhu, W.Q., Wang, H.P., et al., 2020. Strength assessment of mooring system life extension in FPS Nanhai Tiaozhan. China Offshore Platform 35 (2), 66–73.
Stanisic, D., Efthymiou, M., Kimiaei, M., et al., 2018. Design loads and long term distribution of mooring line response of a large weathervaning vessel in a tropical cyclone environment. Mar. Struct. 61, 361–380. https://doi.org/10.1016/j.marstruc.2018.06.004.
Su, Q., 2020. Coupled Thermo–Mechanical Analysis and Dynamic Extreme Response Prediction of Strong-Electricity Composite Umbilical Cable Structure. Dalian University of Technology, Liaoning.
Sun, L.L., 2021. Motion Distribution Analysis and Extreme Prediction of Semi-Submersible Platform Based on Prototype Monitoring Data. Dalian University of Technology, Liaoning.
Teng, H.Y., 2018. Research and Applications on the Prediction Method of Extreme Value of Marine Environment. Jiangsu University of Science and Technology, Jiangsu.
Winterstein, S.R., 1985. Non-normal responses and fatigue damage. J. Eng. Mech. 111 (10), 1291–1295. https://doi.org/10.1061/(ASCE)0733-9399(1985)111:10(1291).
Winterstein, S.R., 1988. Nonlinear vibration models for extremes and fatigue. J. Eng. Mech. 114 (10), 1772–1790. https://doi.org/10.1061/(ASCE)0733-9399(1988)114:10(1772).
Wu, W.H., Feng, J.G., Xie, B., et al., 2016. Development and sensing properties study of underwater assembled water depth-inclination sensors for a multi-component mooring system, using a self-contained technique. Sensors 16 (11), 1925. https://doi.org/10.3390/s16111925.
Yu, S.Y., Wu, W.H., Xie, B., et al., 2020. Extreme value prediction of current profiles in the South China Sea based on EOFs and the ACER method. Appl. Ocean Res. 105, 102408 https://doi.org/10.1016/j.apor.2020.102408.
Zhang, X.Y., Zhao, Y.G., Lu, Z.H., 2019. Unified Hermite polynomial model and its application in estimating non-Gaussian processes. J. Eng. Mech. 145 (3), 04019001. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001577.
Zhao, T.C., Xu, M.Y., Xiao, X., et al., 2021. Recent progress in blue energy harvesting for powering distributed sensors in ocean. Nano Energy 88, 106199. https://doi.org/10.1016/j.nanoen.2021.106199.
Zhong, H., 2021. Preliminary Study on Numerical Simulation of Hydrodynamic Characteristics of Power and Aquaculture Integrated Platform. Dalian University of Technology, Liaoning.