Journal of Marine Science and Technology

Volume 23 | Issue 5 Article 15

QUANTIFICATION OF STATISTICAL UNCERTAINTIES IN PERFORMING THE PEAK OVER THRESHOLD METHOD

Yi Zhang
School of Civil and Environmental Engineering, Nanyang Technological University, Singapore., zhang_yi87@163.com

Yingyi Cao
College of Preschool Education, Zibo Normal College, China.

Jian Dai
Department of Civil and Environmental Engineering, National University of Singapore, Singapore.

Recommended Citation
Zhang, Yi; Cao, Yingyi; and Dai, Jian (2015) "QUANTIFICATION OF STATISTICAL UNCERTAINTIES IN PERFORMING THE PEAK OVER THRESHOLD METHOD," *Journal of Marine Science and Technology*: Vol. 23: Iss. 5, Article 15.
DOI: 10.6119/JMST-015-0604-1
Available at: https://jmstt.ntou.edu.tw/journal/vol23/iss5/15

Journal of Marine Science and Technology, Vol. 23, No. 5, pp. 717-726 (2015)
DOI: 10.6119/JMST-015-0604-1

QUANTIFICATION OF STATISTICAL UNCERTAINTIES IN PERFORMING THE PEAK OVER THRESHOLD METHOD

Yi Zhang[1], Yingyi Cao[2], and Jian Dai[3]

Key words: offshore engineering, extreme value distribution, peak over threshold, random set.

ABSTRACT

Peak over threshold (POT) method is frequently used in the modeling of extreme values in offshore engineering. In this paper, the POT method is examined in terms of epistemic uncertainties in practical usage. Real observed ocean data with specific considerations to extreme events are analyzed. In particular, the procedures including the use of de-clustering and threshold in POT method are addressed. A key element in the application of probability and statistical theories is the estimation of model parameters. The performance of these estimation methods is tested in the context of epistemic uncertainty. This is done through a numerical simulation study for considering data samples having different tail behavior, sample size and noise conditions. The annual maximum method and the rth largest order statistic method in establishing the extreme value model are also included in the comparative study. Main focus is put on the critical issues and uncertainties that might be resulted in the established extreme value models.

I. INTRODUCTION

Occurrences of offshore extreme events are the main reason for the failure of constructed marine facilities. The prediction of ocean extreme values is, thus, an important component in the engineering design. Here we focus on coastal and offshore structures in this context. Many design codes or standards have specifications with regards to the design values which are consistent with the design life (long term), often extrapolated from short term data based on statistical concepts (DNV, 2012; ABS, 2013). The most common and simple approach follows the annual maximum method. The generalized extreme value distribution is fitted to the annual maximas (Gumbel, 1958). However, discarding data other than extremes within the year is not an efficient use of available information, especially in the case of scarce data. The established model may incorrectly characterize the true extreme values. Usually, the time duration of field data is limited to several decades which are considered short compared with the design life of structures. Consequently, many other techniques have been proposed to utilize more available data extensively, such as bootstrap method (Naess and Clausen, 2001), *r* largest order-statistics method (Guedes Soares and Scotto, 2004) and block maxima method (Muraleedharan et al., 2007).

Among these, the peak over threshold (POT) method has attracted the most attention and has been widely applied in various applications (Smith, 2001; Jonathan and Ewans, 2013; Petrov et al., 2013). The threshold approach is quite useful in treating and effectively utilizing time series data (Ferreira and Guedes Soares, 1998). It is suited for dealing with realizations of a stochastic process which is approximately stationary or can be split into stationary parts (Kyselý et al., 2010). However, when applying the POT method to model ocean environment data, numerous factors which affect the accuracy of the results, such as the number of data available, the criteria used to identify extremes, the choice of threshold and serial dependency effects (Mackay et al., 2011). These factors need particular treatments to arrive at a realistic model. Quantifying these uncertainties in the established extreme value model is quite necessary and critical for practical cases. In practice, real collected data are used to establish the model in relation to a specific application, and, undoubtedly, uncertainties exist in various forms depending on how sophisticated the model or situation is (Zhang, 2015a). For example, in offshore engineering, how will these uncertainties propagate to the results, namely, establishing the return level value which is corresponding to a specified level of reliability over the design life of offshore structure is of primary interest (Zhang, 2015b). This is of significant importance in deriving a consistent, acceptable and optimal design value which is leading to a safe

Paper submitted 05/27/14; revised 05/22/15; accepted 06/04/15. Author for correspondence: Yi Zhang (e-mail: zhang_yi87@163.com).
[1] *School of Civil and Environmental Engineering, Nanyang Technological University, Singapore.*
[2] *College of Preschool Education, Zibo Normal College, China.*
[3] *Department of Civil and Environmental Engineering, National University of Singapore, Singapore.*

and economical structure.

Many issues in POT based extreme value statistical applications have been addressed previously (Deidda and Puliga, 2009; Ribereau et al., 2011). These studies examined comprehensively the uncertainty conditions associated with POT. However, the characteristics inherent existed in the collected environmental data, such as serial correlations have not yet been part of these investigations so far. These characteristics could influence the performance of POT method quite significantly, e.g. through potentially inappropriate use of parameters and threshold. The uncertainties regarding the prediction of long term structural performance based on POT established extreme value model has been given in Cheng et al. (2003). The practical influences of these uncertainties are also reviewed by Bitner-Gregersen et al. (2014). Recent developments on the uncertainty quantifications regarding the POT based extreme value modeling with implementation of imprecise probability can be found in Zhang and Cao (2015).

In this paper, the study is focusing on the analysis of various types of uncertainties that may be created in performing the POT method. The motivation is to study the uncertainties related to an established extreme value model and target to have an improved understanding of the importance of the methods selected for the construction of an extreme value model. Based on practical issues, this paper wishes to quantify the uncertainties in such a way that could provide enough guidelines and brief ideas to the design engineers when they are facing some design problems involving extreme values. The problems in terms of very practical issues would be discussed.

Recognizing that, the paper content is organized as following. First, the basic concepts and relevant key elements of extreme value theories and POT method are reviewed. A numerical simulation based study is then conducted to address the aforementioned concerns about establishing extreme value model with regard to practices for modeling observed environmental data. In particular, the quality and performance of different parameter estimators are discussed. The issues of different types of uncertainties including tail behavior, noise and range of serial dependencies associated with the time series data are investigated. The issues of extrapolation in the construction of a statistical model is discussed. Finally, the obtained conclusions from this study on the uncertainty quantification applied to POT are summarized. The conclusion contributes to a base for establishing a robust extreme value model.

II. REVIEW OF PREVIOUS THEORIES ON EXTREME VALUE MODEL

The classical extreme value theory is based on the statistical behavior of block maximas (Gumbel, 1958)

$$M_n = \max\{Y_1, ..., Y_n\}, \tag{1}$$

where $\{Y_1, \ldots, Y_n\}$ is a collection of independent random variables following the same probability distribution. When the size of the block approaches infinity, $n \to \infty$, the probability of M_n tends to a stable function asymptotically such that

$$\Pr\{M_n \le x\} \to G(x) \text{ as } n \to \infty, \tag{2}$$

where G is a non-degenerate distribution function which can be expressed as the following *Generalized Extreme Value* (GEV) distribution

$$G(x) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \xi \ne 0 \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\} & \xi = 0 \end{cases}, \tag{3}$$

defined on the set $\{1 + \xi((z - \mu)/\sigma)\} > 0$, where the location parameter μ, scale parameter σ and shape parameter ξ satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. Three types of tail behaviors, namely exponentially tail, heavy tail and light tail, can be defined corresponding to $\xi = 0$, $\xi > 0$ and $\xi < 0$ respectively. Based on extreme value theory, the most convenient approach to establish an extreme value model is based on the bock maximas which models the maximum values within a defined time unit (a block). For example, the annual maximum method, which has a block size of one year leading to the GEV distribution, is well-advocated by researchers (Winterstein et al., 2001; Zhang, 2013).

III. PEAK-OVER-THRESHOLD (POT) METHOD

Compared to the traditional approaches, the POT method can utilize more information from the data set. Moreover, it does not require the time series data to be strictly stationary (Méndez et al., 2006; Jonathan and Ewans, 2011). It is developed based on the statistical properties of data sample that is following the Pareto distribution.

1. Pareto Family

Consider a set of data extracted from an original set of data, which is following probability distribution F, such that their values are above a certain threshold value of u. By following the asymptotic rules as given in Eq. (2), the cumulative probability function for the exceedances can be expressed as

$$G(x) = \begin{cases} 1-\left[1+\xi\left(\dfrac{x-u}{\tilde{\sigma}}\right)\right]^{-1/\xi} & \xi \ne 0 \\ 1-\exp\left[-\left(\dfrac{x-u}{\tilde{\sigma}}\right)\right] & \xi = 0 \end{cases}$$

$$\textit{for } x \ge 0 \textit{ and } 1+\xi\left(\frac{x-u}{\tilde{\sigma}}\right) > 0, \tag{4}$$

where ξ is the shape parameter, u is the threshold and $\tilde{\sigma}$ is the scale parameter which has a relationship with other parameters in GEV model (e.g. $\tilde{\sigma} = \sigma + \xi(u-\mu)$). Eq. (4) belongs to the family of *Generalized Pareto Distributions* (GPD). The concept is similar to GEV in the modeling of maxima which includes the classification of the tail behaviors in types I, II or III. 2004.

2. Poisson-GPD Model

In practice, the peaks over a sufficient high threshold of time series data are usually rare and memoryless events. As such, their occurrences can be appropriately modeled as a Poisson process. For example, for a set of time series data if the number, N, of exceedances $x_1, \ldots, x_n$ over the threshold u in any one year has the inter arrival time following a Poisson distribution with mean λ. The probability of the annual maxima less than X can be calculated as

$$\begin{aligned} \Pr\left\{\max_{1\le i\le N} x_i \le X\right\} &= \Pr\{N=0\} + \sum_{n=1}^{\infty}\Pr\{x_1 \le X, \cdots, x_n \le X\} \\ &= e^{-\lambda} + \sum_{n=1}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\left\{1-\left(1+\xi\frac{x-u}{\sigma}\right)^{-1/\xi}\right\}^n \\ &= \exp\left\{-\lambda\left(1+\xi\frac{x-u}{\sigma}\right)^{-1/\xi}\right\} \end{aligned} \tag{5}$$

where the exceeding values $x_1, \ldots, x_n$ over the threshold are assumed to follow the GPD model with the corresponding statistical parameters σ and ξ as given in Eq. (4). Thus the concepts of Poisson process and GPD models could be convoluted. From a mathematical point of view, this is the basic property of extremes in a stationary process, which shows that under very general conditions, the magnitudes of the exceedances can be modeled in a Pareto distribution while the occurrence rates are approximately following the Poisson process.

In POT method, the exceedances must be regarded as independent and identically distributed variables. For some real events, the extremes may have some degree of clustering, leading to the issue of dependency between exceedances above the threshold. To resolve this issue, *declustering* has been suggested, which is a process to filter the dependent exceedances to obtain a set of threshold excesses that are approximately independent (Coles, 2001). One possible way to identify the peak values within each cluster is to choose a time span Δt (e.g. 1 day, 3 days or 1 week), such that the extreme events separated by less than this period of time are considered as one "event", and the highest value is identified as a peak value (Morton et al., 1997). The selection of an appropriate time span will be discussed in the later part of this paper.

3. Parameter Estimate Method

The importance of parameter estimations in POT method cannot be underestimated as they may create errors in estimating the high quantiles. There are numerous parameter estimation methods available in the literature, such as likelihood-moment estimations (Zhang and Stephens, 2007), least-squares error method (Moharran et al., 1993) as well as empirical percentile method (Castillo and Hadi, 1997). However, most of these methods may not be easily implemented and some require intensive computations. Some of the better known GPD model parameter estimation methods are briefly summarized herein.

1) Method of Moments

The simplest method in estimating the statistical parameters in POT method could be the method of moments (MOM). The basic idea is to equate the sample mean and variance to the theoretical population mean and variance. Based on the statistical relationships within GPD model, the MOM estimates are given by

$$\xi = \frac{1}{2}\left(1-\frac{\overline{x}^2}{\overline{s}^2}\right), \quad \sigma = \frac{1}{2}\overline{x}\left(\frac{\overline{x}^2}{\overline{s}^2}+1\right), \tag{6}$$

where $\overline{x}$ and $\overline{s}^2$ stand for the sample mean and variance. However, the application of MOM requires a limiting value in the shape parameter. For example, a heavy tail GPD model may not have an estimate in the moment (the estimate of mean will be infinity for shape parameter ξ larger than 1).

2) Probability Weighted Moments Method

Based on a similar idea of the MOM, the probability weighted moments (PWM) method utilizes the sample PWM in estimating the parameters in the GPD model. The PWM is originally defined as

$$M_{p,q,r} = E\left[x^p\left(F(x)\right)^q\left(1-F(x)\right)^r\right], \tag{7}$$

where the p, q and r are determined coefficients. By linking with the Pareto distribution, the PWMs can be expressed in terms of the model parameters as

$$\alpha_s = E\left[x\left(1-F(x)\right)^s\right] = \frac{\sigma}{(s+1)(s+1-\xi)}, \qquad \textit{for}\ \ \xi<1,\ s=0,1,2,\ldots \tag{8}$$

By using the first two PWMs, the GPD model parameters can be easily estimated as

$$\xi = 2-\frac{\alpha_0}{\alpha_0-2\alpha_1} \quad \text{and} \quad \sigma = \frac{2\alpha_0\alpha_1}{\alpha_0-2\alpha_1}, \tag{9}$$

where α_0 and α_1 are estimated from Eq. (8). The α_s value can be calculated from the sample data by

$$\alpha_s = \frac{1}{n}\sum_{i=1}^{n} x_{i:n}\left(1 - p_{i:n}\right)^s , \tag{10}$$

where $x_{i:n}$ is the ith smallest value in n sample data, $p_{i:n}$ is the plotting position which is a general approximation to the true value of 1-F. An unbiased estimate is $p_{i:n} = (i\text{-}0.5)/n$ (usually we call it probability weighted moments unbiased (PWMU) method), while for other cases, various expressions are available. For example, a biased estimate of $p_{i:n} = (i\text{-}0.35)/n$ is given in Mackay et al. (2011) (usually we call it probability weighted moments biased (PWMB) method).

3) Goodness-of-Fit Method

Other than utilizing the statistical properties of GPD model in estimating the parameter values, the goodness-of-fit method estimates the statistical parameters in the most obvious way, from a plot of the data. The result of a fitted parametric model should give the least sum of squares and must be visually compared against the empirical data plot, for example, quantile-quantile (QQ) plot. In the model test statistics, the null hypothesis is H_o: $F(x) = F_o(x)$ where F is the empirical CDF and F_o is the distribution being tested. Two of these well known statistics are

Kolmogoriv-Smirnov (KS) statistic:

$$D_n = \max\left\{\max_{1\le i\le n}\left\{\frac{i}{n} - F_o\left(x_i\right)\right\}, \max_{1\le i\le n}\left\{F_o\left(x_i\right) - \frac{i-1}{n}\right\}\right\}. \tag{11}$$

Anderson-Darling (AD) statistic:

$$A_n^2 = -\sum_{i=1}^{n}\frac{2i-1}{n}\left\{\log F_o\left(x_i\right) + \log\left(1 - F_o\left(x_{n+1-i}\right)\right)\right\} - n . \tag{12}$$

The KS test measures the maximum discrepancy between the theoretical model and the empirical data whereas the AD test places more weight or discriminating power on the tails of the distribution. Theoretically, the smaller the statistic is, the better is of the fit. Thus, the estimators for the GPD model parameters could be obtained by minimizing these statistics.

IV. EXPERIMENT DESIGN OF UNCERTAINTY ASSESSMENT

In selecting the POT method to model the ocean data, numerous uncertainty factors should be considered, such as the number of data available, the criteria used to identify peak values, the choice of threshold and serial correlation effects.

The uncertainties associated with POT method is quantified herein. The performance of POT method is examined using Monte Carlo simulations for considering different parameter estimation methods, sample size, tail effects and noise. The investigations include the method of moments (MOM), maximum likelihood method (MLE), unbiased probability weighted moments method (PWMU), biased probability weighted moments method (PWMB), Anderson-Darling test based goodness-of-fit method (AD), and the Kolmogoriv-Smirnov test based goodness-of-fit method (KS). The maximum likelihood method refers to the classic statistical parameter estimation method which determines the parameter value based on maximizing the likelihood function value. The effect of sample size on the determination of GPD model parameters are investigated using simulated data with sample sizes of n = 10, 20, 30, 50, 80, 100, 150 and 200. For each data sample size n, the simulation will be repeated 10,000 times and their average estimated parameter values is used as a mean for comparisons. Detailed steps in performing this uncertainty assessment are provided as followings:

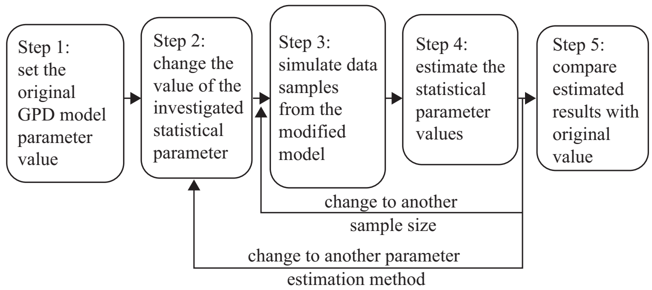


Fig. 1. Flowchart of performing the uncertainty assessment.

Step 1: set the original statistical parameter values for GPD model.
Step 2: based on the problem of concern, change the value of the investigated statistical parameter value.
Step 3: simulate a random data sample from the modified GPD model based on a chosen sample size.
Step 4: select a parameter estimation method and use it to estimate the parameter values from the simulated data set.
Step 5: repeat the procedures from Step 2 to Step 4 for different sample sizes and parameter estimation methods.
Step 6: compare the estimated results with their original values, calculate the associated bias results. The flow of this calculation process is illustrated in Fig. 1.

A brief explanations about the random simulations are given herein. At the beginning of the work, the GPD model parameters (scale, shape and location parameters) are defined. Based on these parameter values, we randomly simulate a group of data which is following the GPD model with the defined model parameters. Based on the simulated data sample, we use selected parameter estimation method to estimate the GPD model parameters. And then we compare the estimated parameter results with the original values. All the steps will be iterated to consider using other parameter estimation methods. Finally, all the results will be compared.

The interested results are the shape parameter, scale pa-

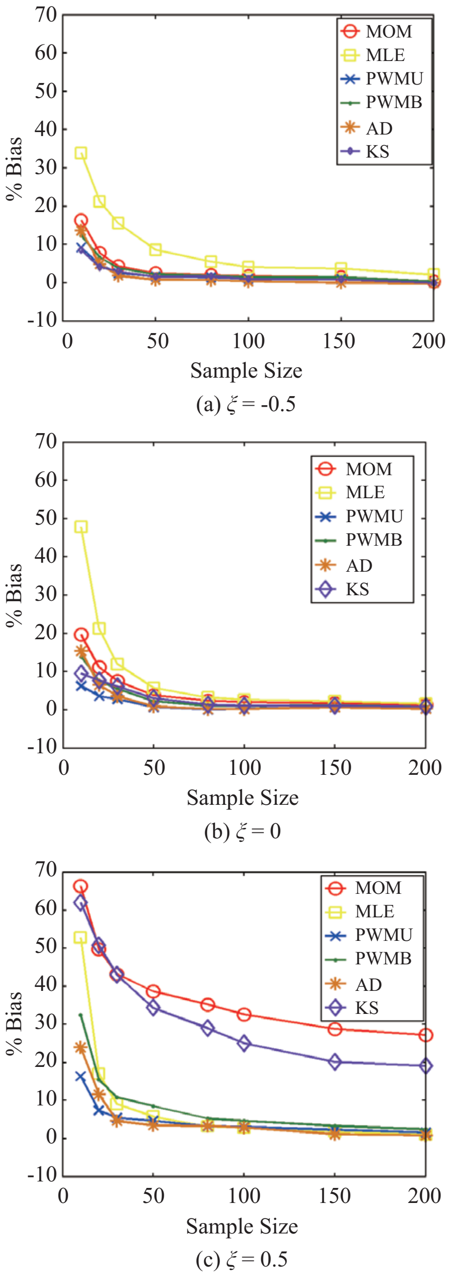


Fig. 2. Bias of estimated scale parameter for different tails.

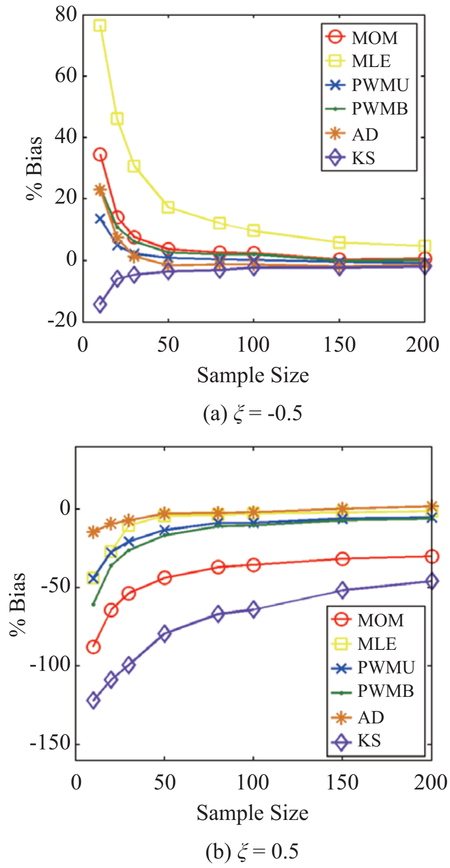


Fig. 3. Bias of estimated shape parameter for different tails.

rameter and a high percentile (for this purpose, a non-exceedance probability of 0.99 is used). The accuracy of the estimators is compared using the relative bias as a normalized measure of deviation from the theoretical value.

V. COMPARISONS AND DISCUSSIONS

1. Effects of Tail Behavior

The tail characteristics, or the value of ξ, of a GPD model can critically influence the parameter estimations, which in turn will affect the expected return values. Theoretically, the GPD is valid for any value of ξ. However, not all the estimation methods will yield reasonable estimates that can cover the entire range of possible values of ξ in a GPD model.

To investigate this issue, the simulated data based on GPD model for ξ = -0.5, -0.25, 0, 0.25 and 0.5 with $\sigma = 2$ and $u = 1$ are used in the numerical study. This range of values in $\xi \in [-0.5, 0.5]$ is commonly observed for environmental variables, such as significant wave height. The computed results of relative bias in the shape and scale parameters with respect to sample sizes from n = 10 to 200 using various estimated methods are presented in Figs. 2 and 3 respectively. The findings based on the simulated results are summarized as followings:

- Generally, the relative bias of estimated parameter values decreases with increasing sample size. However, for the heavy tails, the bias in the cases of KS and MOM for estimating shape and scale parameters are still large even the sample size is increased to 200.
- For all the estimators the relative biases in the shape and scale parameters are greater for heavy tails than for light tails.
- Amongst the estimators, the MLE is the most sensitive estimator to sample size. It produces the largest bias estimate for all the sample sizes considered.
- MOM estimator is the most sensitive estimator to the tail

behavior. For a heavy tail that has a value of ξ around 0.5, the bias is about 30% and does not significantly be improved by increasing the sample size.

- PWMU and PWMB show consistently better parameter estimates for different tail behaviors compared to the other estimators. However, compared to PWMB, PWMU is slightly less sensitive to the effects of tail behavior and sample size.
- AD gives a low bias in estimating shape and scale parameters and is not sensitive to the tail behavior. However, AD is quite sensitive to the sample size. The bias can go up to 20% for a sample size of around 10.
- KS estimators give large bias results in estimating the shape and scale parameters for $\xi > 0$. But for $\xi < 0$, the bias is relatively small. The performance of KS estimator is very poor for a data set that has a heavy tail or small sample size.

Obviously, it is a compromise to find the most suitable estimator covering all conditions shown by the results. However, for a data set with $\xi < 0$, MOM, AD, PWMU and PWMB are reasonably good estimators as the relative bias in the estimates is fairly small (< 10%) even for a sample size of 20. If the sample size is greater than 100, the MLE is a suitable alternative.

However if $\xi > 0$, both PWMB and PWMU stand out as the best estimation methods. If $n > 100$, the AD and MLE estimators can be adopted in view of their small bias for a large sample size.

2. Effects of Noise

Another contribution to uncertainty arises from noises in the collected data. For example, as most of the environmental data collected at a site is not enough, the data collected at a nearby site may also be utilized together for the same statistical analysis. This combination introduces some nonstationary data into the data group and causes some noises in the statistical parameters of GPD model.

The effect of noise on the parameter estimates in GPD model is investigated in this study by polluting the simulated data with Gaussian noise. Noise is firstly added to the parameters of the GPD having $\xi = -0.5$, $\sigma = 2$ and $u = 1$ before the data are simulated. The following cases of noises are simulated:

- Noise in location parameter: $u = 1 + N(0, \varepsilon^2)$, $\sigma = 2$, $\xi = -0.5$ for $\varepsilon = 0.1, 0.3, 0.5$.
- Noise in scale parameter: $u = 1$, $\sigma = 2 + N(0, \varepsilon^2)$, $\xi = -0.5$ for $\varepsilon = 0.2, 0.6, 1.0$.
- Noise in shape parameter: $u = 1$, $\sigma = 2$, $\xi = -0.5 + N(0, \varepsilon^2)$ for $\varepsilon = 0.05, 0.15, 0.25$.

where $N(0, \varepsilon^2)$ is a value drawn from a standard Gaussian distributed random number generator having a mean of 0 and variance equals to ε^2. Three noises intensities ε^2 are chosen, corresponding to coefficients of variation of 0.1, 0.3 and 0.5. The noise is firstly generated and then added to the GPD statistical parameters. Based on the "new" GPD model parameter, which combines the initial setting value and the noises, the data sample are randomly simulated.

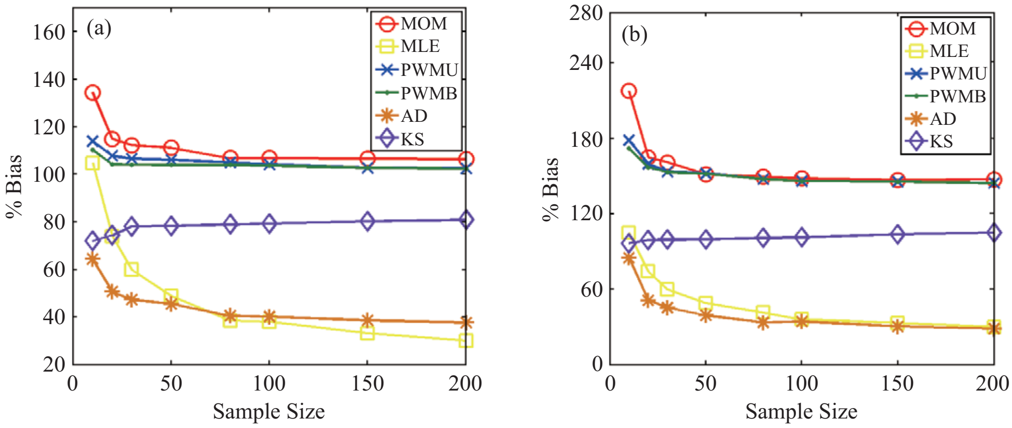
The calculated biases for the shape and scale parameters are presented in Figs. 4-6. Comparison of the results yields the following conclusions:

- The noise in the location parameter yields the largest bias results compared to the noise in the scale and shape parameters. While the effect of the noise in scale and shape parameters can be reduced by increasing the sample size, the bias for large noise in location parameter cannot be reduced, at least between $n = 20$ and 100.
- All the parameters in GPD model experience increase in relative bias with increase in noise intensity irrespective of the parameter estimation methods, with the location parameter being the most affected. The estimates are very sensitive to the noises especially for the shape parameter.
- MLE gives the largest relative bias when noise is present in scale and shape parameters. However, MLE is relatively the best estimator when noise occurs in location parameter. But the accuracy of MLE estimator is highly sensitive to the sample size.
- MOM, PWMU and PWMB estimators produce similar results, giving large relative bias with noise in location parameter but low bias with noise in scale and shape parameters. None of the parameter estimation methods is able to give reliable results when the noise in location parameter is very high.
- Among all the estimators, AD shows the best performance with noise in location parameter. However, for noises in shape parameter, AD gives a large bias in the 99th percentile estimate.
- KS method gives a negative bias in estimating shape parameter with noise in scale and shape parameters. However, the quantity of this bias estimate is not large.

The effects of noises are clearly not insignificant and the parameter estimation methods need to be carefully selected in this context. For noise in location parameter, AD would be the most suitable method in estimating the parameters in GPD model. If the intensity of noise in location parameter is high ($\varepsilon > 0.5$ in this study) and sample size is not small ($n > 100$), the MLE method is another good choice. However, one should note that MLE and AD, when the noise intensity is large, still lead to large relative bias (> 30%). For noise in shape and scale parameters, MOM, PWMU, PWMB and AD are all applicable as long as sample size is not too small ($n > 20$). However, if the noise in shape parameter is high ($\varepsilon > 0.25$ in this study), the estimations may still produce a large bias ($\approx$ 10%) even though the sample size is 200.

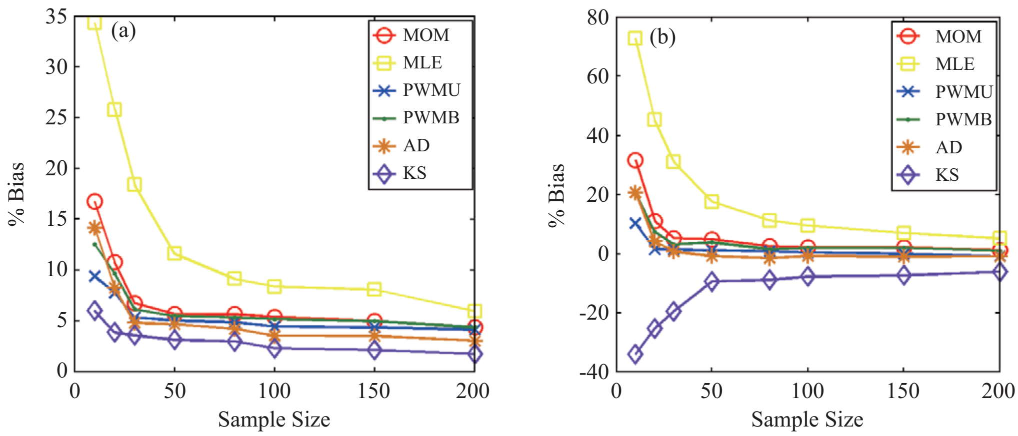
3. Effects of Range of Dependency

Dependencies between data points in a time series become



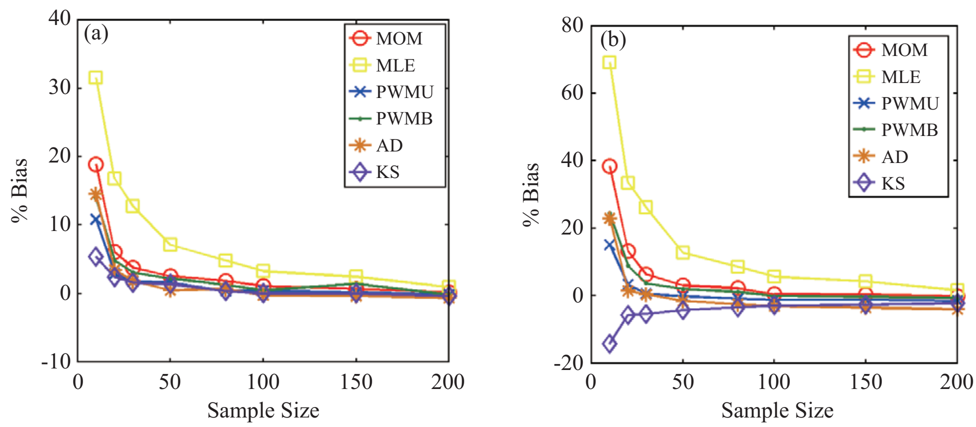
GPD: $u = 1 + N(0,0.5^2)$, $\xi = -0.5$, $\sigma = 2$

Fig. 4. Bias of estimated (a) scale parameter and (b) shape parameter with noise effect in location parameter in GPD model.



GPD: $u = 1$, $\xi = -0.5$, $\sigma = 2 + N(0,1.0^2)$

Fig. 5. Bias of estimated (a) scale parameter and (b) shape parameter with noise effect in scale parameter in GPD model.



GPD: $u = 1$, $\xi = -0.5 + N(0,0.25^2)$, $\sigma = 2$

Fig. 6. Bias of estimated (a) scale parameter and (b) shape parameter with noise effect in shape parameter in GPD model.

important when the sampling frequency is high or when extreme events cause subsequent significant events.

To analyze the effects of serial dependency, an autoregressive model (AR) of order one is utilized to simulate a weakly stationary time series for this investigation,

$$X_t = c + \varphi X_{t-1} + \varepsilon_t \,, \tag{13}$$

where φ is the parameter of the model, c is a constant and ε_t

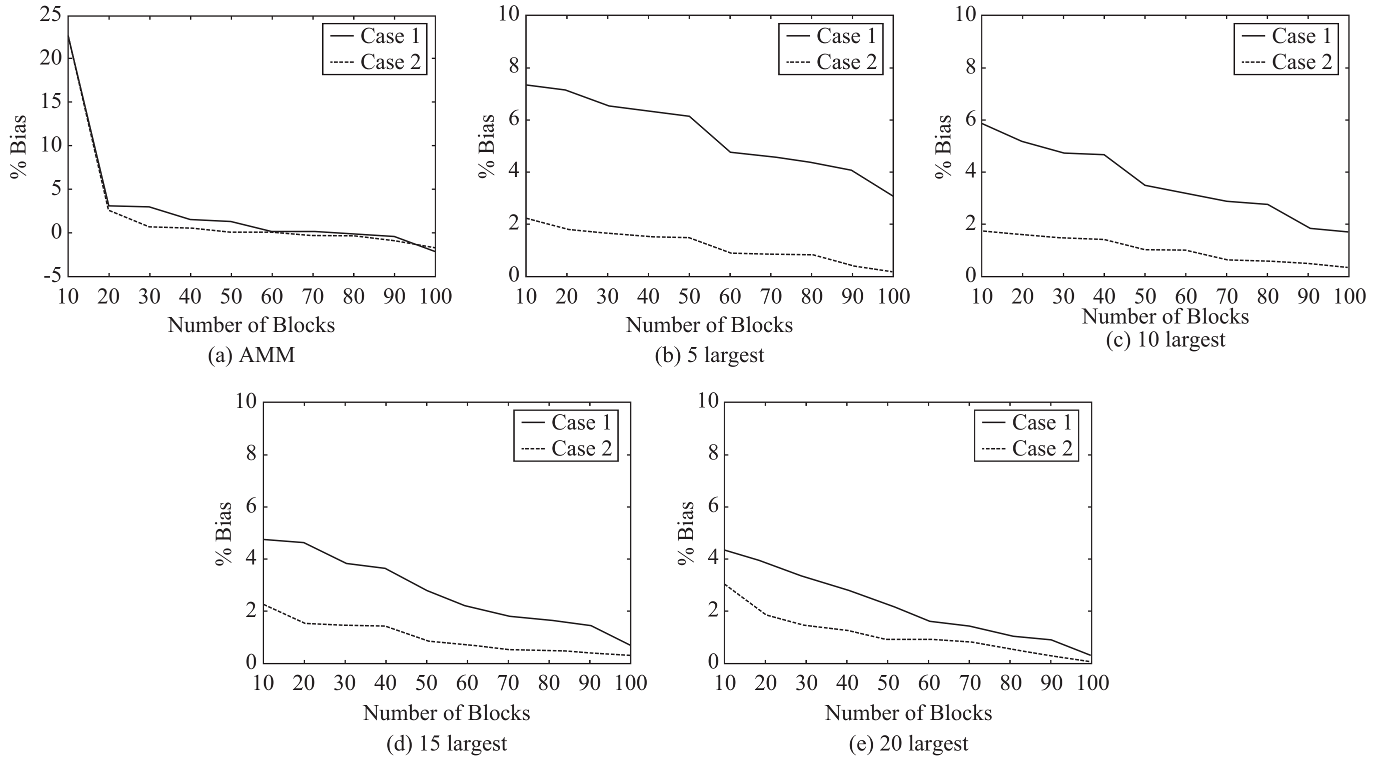


Fig. 7. Biases of estimated 99th percentile in AMM and r largest order statistic method for two time series. (I) Case 1: φ = 0.95 (II) Case 2: φ = 0.

is the noise term. For comparison purpose, the values of φ are set to 0.95 and 0 which correspond to a highly-correlated and an uncorrelated time series, respectively.

In this study, a value of 0 is given to c and ε_t is assumed to follow an exponential distribution which has a rate parameter equal to 1 (that is, $\varepsilon_t \sim \text{Exp}(1)$). Theoretically, the parameters of the extreme value model for these simulated time series have the corresponding values $u = 4.605$, $\sigma = 1$ and $\xi = 0$ (in GPD model). For each simulated realization (time series), a group consisting of 100 continuous time series data is defined as a block (e.g. one block represents one year in the annual maximum method). The block is used here to represent a reference time unit (for example, the AMM will only utilize the maximum value within each block). In order to test the estimates with the effects of different lengths of time series, the data simulated will have lengths of 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 blocks for comparison purpose. The results of interest in this work is to estimate the 99th percentile from each simulated time series. Each estimate is calculated based on an average of the results from 100 simulations (e.g. number of time series data set).

For the purpose of demonstrating the practical advantage of using POT, two commonly used approaches in establishing an extreme value model are also employed here to estimate the 99th percentile. These are annual maximum method (AMM) and r largest order statistic method (Guedes Soares and Scotto, 2004). Within the r largest order statistic method, four values of r are considered, namely, r = 5, 10, 15 and 20. Within the POT method, four different values of threshold and time span are used, denoted as U3T0, U3T10, U5T0 and U5T10, where the notation UiTj refers to a threshold value of i in identifying the excess values, and j represents the value of time span (number of continuous time series data, e.g. excesses separated by less than this period of time are considered as one "event", and the highest value is identified as a peak value) used in de-clustering the peak values. The results in terms of bias in estimating the 99th percentile for each cases are plotted in Figs. 7-8. The findings based on the calculated results are summarized as followings:

- Compared to POT and r largest order statistic methods, AMM is least affected by serial dependencies within the time series with regards to the 99th percentile estimates, provided the sample size is larger than 20 blocks. For example, for time series that only have 10 blocks, the bias of the estimated 99th percentile is quite large (> 20%). This is because the AMM filters out only a small amount of data (only the maximum value within each block is filtered) in the time series. When the number of blocks is limited, the statistical uncertainty resulting from small sample size is high.
- r largest order statistic method filters out more data per block than AMM and hence the statistical uncertainty is smaller. However, it is more sensitive to the serial dependency. This is particular obvious when r is small where fewer data are filtered. For example, when r = 5 and only

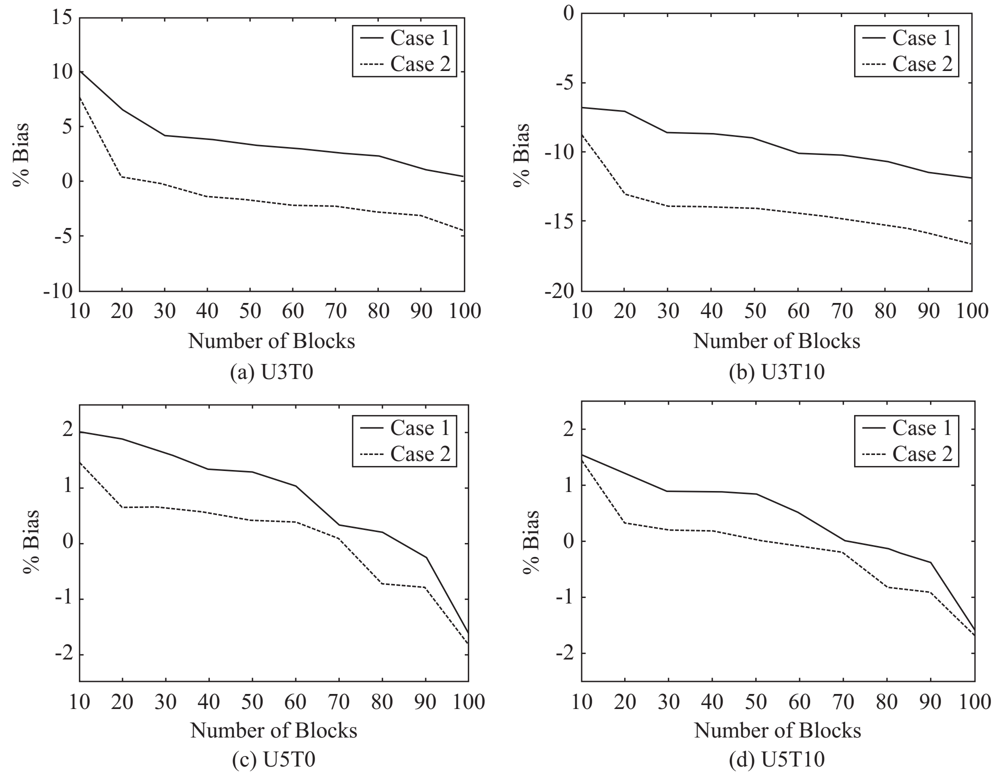


Fig. 8. Biases of estimated 99th percentile in POT method for two time series (a) Case 1: $\varphi = 0.95$ (b) Case 2: $\varphi = 0$.

10 blocks of time series data are available, the estimated biases in case 1 is much higher (7.3%) compared to case 2 (2.2%). These biases can be reduced by utilizing more data within the block. For instance, for $r = 20$, the bias of the estimate for case 1 is much less than for $r = 5$. But it does not imply utilizing more data within the block is always helpful because the basic assumption of asymptotic property in order statistic theory is violated for large r. For this reason, the results utilizing 20 largest values is less accurate compared to 10 largest values in the case of uncorrelated time series.

- The performance of POT is dependent on the given values of threshold and time span. As shown in Fig. 8, U3T0 gives a large positive bias, while U3T10 gives a large negative bias. However, if the threshold changes to 5, the error associated with the estimations in U5T0 and U5T10 cases are very small (-2%~2%). This implies that the threshold value of 3 is too small and is not a suitable value for use in POT method. It is noted that sample size has lesser influence on the results in POT method, as it filters more data compared to AMM and r largest order statistic method. It can be seen from the comparison between (c) and (d) in Fig. 7, the use of time span in U5T10 leads to a smaller bias compared to U5T0. However, the serial dependency in the time series has very little influence to the accuracy in POT method (the difference between case 1 and case 2 in Fig. 8(c) and (d) are quite small) and only leads to a small positive bias in the estimates.

In conclusion, the model selection is a compromise again. AMM has very good performance when there is a large amount of data and it is not affected by the serial dependency effect in the time series. The r largest order statistic method does not need a large amount of data compared to AMM, but it is not suitable for highly correlated time series. POT method gives the most suitable results even for time series that has high serial correlations. However, the accuracy of performing POT method is quite sensitive to the selected values of threshold and time span.

VI. CONCLUSION

In this paper, several issues regarding the establishing of an extreme value model from ocean data have been investigated, focusing primarily on the peak over threshold method. Simulation studies are conducted to test the robustness of the established extreme value model from various methods. It was found that MOM, PWMB and PWMU are the better parameter estimation methods. Besides the sample size effect, the tail behavior can influence the accuracy of the estimated parameter values significantly, especially for light tail in the extreme data. The presence of random noise in the collected data increases the uncertainty in parameter estimations. Noise in location parameter has the most significant influence and the bias of the estimate arising from this may not be reduced much with more data provided. When limited time series data are available, POT method may be the most appropriate approach compared to annual maximum method and *r* largest order statistic method. In addition, serial correlations have little impact on the results from POT method. However, the performance of POT method is largely dependent on the appropriate use of time span and threshold. The current study is limited to the considered uncertainties highlighted in the cases studies. Future work could be focusing on finding the most appropriate parameter values in approaching the extreme values. More efforts could be put on the real case study on different real observed time series data of ocean parameters. The time dependent uncertainties would be an interesting direction for a further development.

REFERENCES

ABS (2013). Accidental load analysis and design for offshore structures. Guidance Notes.

Bitner-Gregersen, E. M., K. C. Ewans and M. C. Johnson (2014). Some uncertainties associated with wind and wave description and their importance for engineering applications. Ocean Engineering 86, 11-25.

Castillo, E. and A. S. Hadi (1997). Fitting the generalized Pareto distribution to data. Journal of American Statistics Association 92, 1609-1620.

Cheng, P. W., G. J. W. van Bussel, G. A. M. van Kuik and J. H. Vugts (2003). Reliability-based design methods to determine the extreme response distribution of offshore wind turbines. Wind Energy 6, 1-22.

Coles, S. G. (2001). An introduction to statistical modeling of extreme values. Springer, London.

Deidda, R. and M. Puliga (2009). Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples. Physics and Chemistry of the Earth 34, 626-634.

DNV (2012). Environmental conditions and environmental loads. Recommended Practice, DNV-RP-C205.

Ferreira, J. A. and C. Guedes Soares (1998). An application of the peaks over threshold method to predict extremes of significant wave height. Journal of Offshore Mechanics and Arctic Engineering 120, 165-176.

Guedes Soares, C. and M. G. Scotto (2004). Application of the *r* largest-order statistics for long-term predictions of significant wave height. Coastal Engineering 51, 387-394.

Gumbel, E. J. (1958). Statistics of Extremes. Columbia University Press, New York.

Jonathan, P. and K. Ewans (2011). Modeling the seasonality of extreme waves in the Gulf of Mexico. Journal of Offshore Mechanics and Arctic Engineering 133, 0211041-49.

Jonathan, P. and K. Ewans (2013). Statistical modeling of extreme ocean environments for marine design: A review. Ocean Engineering 62, 91-109.

Kyselý, J., J. Picek and R. Beranová (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. Global and Planetary Change 72, 55-68.

Mackay, E. B. L., P. G. Challenor and A. S. Bahaj (2011). A comparison of estimators for the generalised Pareto distribution. Ocean Engineering 38, 1338-1346.

Méndez, F. J., M. Menéndez, A. Luceño and I. J. Losada (2006). Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. Journal of Geophysical Research 111, 8-13.

Moharram, S. H., A. K. Gosain and P. N. Kapoor (1993). A comparative study for the estimators of the generalized Pareto distribution. Journal of Hydrology 150, 169-185.

Morton, I. D., J. Bowers and G. Mould (1997). Estimating return period wave heights and wind speeds using a seasonal point process model. Coastal Engineering 31, 305-326.

Muraleedharan, G., A. D. Rao, P. G. Kurup, N. Unnikrishnan Nair and M. Sinha (2007). Modified Weibull distribution for maximum and significant wave height simulation and prediction. Coastal Engineering 54, 630-638.

Naess, A. and P. H. Clausen (2001). Combination of the peaks-over-threshold and bootstrapping methods for extreme value prediction. Structural Safety 23, 315-330.

Petrov, V., C. Guedes Soares and H. Gotovac (2013). Prediction of extreme significant wave heights using maximum entropy. Coastal Engineering 74, 1-10.

Ribereau, P., P. Naveau and A. Guillou (2011). A note of caution when interpreting parameters of the distribution of excesses. Advances in Water Resources 34, 1215-1221.

Smith, R. L. (2001). Environmental statistics, technical report. Dep. of Stat., Univ. of North Carolina, Chapel Hill.

Winterstein, S. R., G. Kleiven and Ø. Hagen (2001). Comparing extreme wave estimates from hourly and annual data, Proceedings ISOPE. Conference.

Zhang, J. and M. A. Stephens (2007). A new and efficient estimation method for the generalized Pareto distribution. Technometrics 51, 316-325.

Zhang, Y. (2013). Modeling time varying and multivariate environmental conditions for extreme load prediction on offshore structures in a reliability perspective. Ph. D Thesis, National University of Singapore, Singapore.

Zhang, Y. (2015a). On the climatic uncertainty to the environment extremes: a Singapore case and statistical approach. Polish Journal of Environmental Studies 24, 1413-1422.

Zhang, Y. (2015b). Comparing the robustness of offshore structures with marine deteriorations — a fuzzy approach. Advances in Structural Engineering 18(8), 1159-1172.

Zhang, Y. and Y. Y. Cao (2015). A fuzzy quantification approach of uncertainties in an extreme wave height modeling. Acta Oceanologica Sinica 34, 90-98.