Sampling Properties and Empirical Estimates of Extreme Events

Ed Mackay

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn, UK

Philip Jonathan

Department of Mathematics and Statistics, Lancaster University, UK

Shell Research Ltd., London, UK

Abstract

The statistical characteristics of the largest observations in a sample are highly uncertain. In this work we consider the problem of how to define empirical estimates of exceedance probabilities and return periods associated with an ordered sample of observations. Understanding the sampling properties of these quantities is important for assessing the fit of a statistical model and also for placing confidence bounds on estimates of extreme events from Monte Carlo simulations. The empirical distribution function (EDF) is often defined as the expected non-exceedance probability (NEP) associated with sample order statistics. Yet, due to the non-linearity of the relations between return periods and quantiles and NEP, the return period (or quantile) associated with the expected NEP is not equal to the expected return period (or quantile), leading to ambiguity. However, the sampling distributions of exceedance probabilities, return periods and quantiles are, in fact, linked by a simple relation. From this relation, it follows that defining the EDF in terms of the median NEP of the order statistics gives a consistent framework for defining empirical estimates of all three quantities. We demonstrate that the median value of the return period of the largest observation is 44% larger than the return period calculated using the common definition of the EDF in terms of the expected NEP of the order statistics. We also derive some new results about the size of the confidence intervals for exceedance probabilities and return periods.

Keywords: Sampling variability, Return period, Return value, Plotting position, Model diagnostics, Empirical distribution function, Confidence interval

1. Introduction

Estimating the frequency of occurrence of extreme events is an important topic in offshore and coastal engineering. Many design standards for marine structures require the design to be assessed in sea states associated with specific return periods. The usual approach for estimating return values of metocean variables is to fit a statistical model to observed or hindcast data and extrapolate into the tail of the fitted model. For extreme value analyses, the fit of the model is usually assessed using plots of the observations together with various quantities derived from the empirical distribution function (EDF), such as exceedance probabilities, return periods or quantiles.

In this work we consider the problem of how to define empirical estimates of exceedance probabilities, return periods and quantiles and the related problem of calculating their sampling properties. Understanding the sampling properties of these quantities is important for assessing the fit of a statistical model and also for placing confidence bounds on estimates of extreme events from Monte Carlo simulations.

Suppose we have a sequence of n independent random variables $X_1, ..., X_n$, assumed to have common distribution function F_X. The ordered variables, denoted $X_{(1)} \leq ... \leq X_{(n)}$, are referred to as the order statistics. We denote the non-exceedance probability associated with the k^{th} order statistic as

$$P_k = F_X(X_{(k)}) = \Pr\left\{X \leq X_{(k)}\right\} \in [0, 1]. \tag{1}$$

The return period, T, of level x is defined as the inverse of the exceedance probability

$$T(F_X(x)) = \frac{1}{1 - F_X(x)} \in [1, \infty]. \tag{2}$$

We denote the return period of the k^{th} order statistic as $T_k = T(P_k)$. The quantities $X_{(k)}$, P_k and T_k are random variables. In general, for extreme value analysis, the data-generating distribution F_X is not known. Consequently, for a given sample, the values of P_k and T_k are not known. However, as discussed below, the sampling distribution of P_k is well known and is straightforward to derive (e.g [1, 2]).

In the following, we make no assumptions about how $X_1, ..., X_n$ are sampled, only that observations are independent and identically distributed. If $X_1, ..., X_n$ are a sample of annual maxima, then T has units of years. If $X_1, ..., X_n$ are a sample of peaks-over-threshold, with a mean rate of m peaks per year, then mT has the unit of years. Thus, the difference between samples of annual maxima and peaks-over-threshold only affects the units of T.

The EDF is the non-exceedance probability assigned to the order statistics. Here, we denote the EDF as $\hat{p}_k$, to indicate that it is an estimate of the unknown non-exceedance probability associated with the k^{th} order statistic. One way to define the EDF is in terms of proportion of observations in the sample less than or equal to $X_{(k)}$, so that $\hat{p}_k = k/n$ [3, 4]. However, this gives $\hat{p}_n = 1$, which is undesirable when considering the extremal properties of a sample, since it implies that the largest observation is the upper end point of the distribution. To avoid this issue, the EDF can be defined as the expected value of P_k (e.g [5–7]):

$$\hat{p}_k = \mathrm{E}(P_k) = \frac{k}{n+1}. \tag{3}$$

This definition was originally proposed by Weibull [8] and was popularised by Gumbel [9]. It has the attractive properties that it is simple and has a well-founded theoretical justification. However, since $T(p)$ and $F_X^{-1}(p)$ are nonlinear functions of p (apart from the special case where X is uniformly distributed in which case F_X^{-1} is linear), we have

$$\mathrm{E}(T_k) \neq T(\mathrm{E}(P_k)), \tag{4a}$$

$$\mathrm{E}(X_{(k)}) \neq F_X^{-1}(\mathrm{E}(P_k)). \tag{4b}$$

In fact, since $T(p)$ is a convex function, from Jensen's inequality, we have $\mathrm{E}(T_k) \geq T(\mathrm{E}(P_k))$. However, since the quantile function $F_X^{-1}(p)$ can be either concave or convex, $\mathrm{E}(X_{(k)})$ can be either less than on greater than $F_X^{-1}(\mathrm{E}(P_k))$. The difference between $\mathrm{E}(X_{(k)})$ and $F_X^{-1}(\mathrm{E}(P_k))$ has led to many other definitions of the EDF, so-called plotting positions, $\hat{p}'_k$, for which $F_X^{-1}(\hat{p}'_k) \approx \mathrm{E}(X_{(k)})$. Since $\mathrm{E}(X_{(k)})$ depends on the data-generating distribution F_X, unbiased estimates of $\hat{p}'_k$ are dependent on F_X. The problem of defining plotting positions has been considered in many studies over the years and is still an active topic of research [10–30].

It used to be common practice to fit extreme value models to observations using a least-squares fit to observed quantiles plotted on probability paper (plots of order statistics against the EDF, with various transformations applied to the axes, such that if the data follow a straight line this indicates that it follows a certain distribution). In this case, the choice of plotting position can affect the inferences made from the data. Although modern methods for statistical inference, such as maximum likelihood or Bayesian inference, mean that this type of least-squares fitting is now less common, using graphical means to assess the fit of a model is still commonplace. For extreme value models, we are interested in the fit of the model for the largest observations. The purpose of the present work is to illustrate how the definition of the EDF can have a large impact on the probabilities and return periods associated with the largest observations.

In this work, we argue that a common framework can be used for defining the empirical estimates of either exceedance probabilities, return periods or quantiles, where the EDF is defined as $\hat{p}_k = \text{median}(P_k)$, rather than $\hat{p}_k = \text{E}(P_k)$. The use of the median in this context has been advocated by various authors in the past [11, 22–24, 28, 29, 31, 32]. In the current paper, we present a brief review the theory of the sampling properties of order statistics. We show that the sampling distributions of the exceedance probabilities, return periods and quantiles are all linked by a simple relation, that makes the use of the median value appropriate for all cases. This relation is then used to derive some results about the confidence intervals associated with extreme observations. The results are of interest either when using diagnostic plots for assessing the fit of an extreme value model, or when Monte Carlo simulation is used to estimate extreme events.

The work is organised as follows. The sampling distributions of exceedance probabilities, return periods and quantiles are discussed in Section 2 and the definition of the EDF is discussed in Section 3. The impact of sampling variability and the definition of the EDF on diagnostic plots for extreme value models is discussed in Section 4. Finally, some properties of confidence intervals for empirical estimates of extreme events are derived in Section 5. Conclusions are presented in Section 6.

2. Sampling distributions of exceedance probabilities, return periods and quantiles

To derive the sampling distribution of P_k, first note that the probability that an individual observation, x_k, has $F_X(x_k) \leq p$ is a Bernoulli trial with probability of success p. As the observations are independent, the cumulative distribution function (CDF) of P_k, denoted $F_{P_k}(p)$, is simply the probability that at least k observations have $F_X(x) \leq p$:

$$F_{P_k}(p) = \sum_{j=k}^{n} \begin{pmatrix} n \\ j \end{pmatrix} p^j (1-p)^{n-j}. \tag{5}$$

For the smallest and largest observations we have $F_{P_1}(p) = (1-p)^n$ and $F_{P_n}(p) = p^n$. From the relationship between binomial sums and the regularised incomplete beta function, I, [33], we can write

$$F_{P_k}(p) = I(p, k, n-k+1) = \frac{n!}{(n-k)!(k-1)!} \int_0^p s^{k-1}(1-s)^{n-k} \mathrm{d}s. \tag{6}$$

So P_k follows a beta distribution, $P_k \sim \text{beta}(k, n-k+1)$. Taking the derivative, we obtain the probability density function (PDF) as

$$f_{P_k}(p) = \frac{n!}{(n-k)!(k-1)!} p^{k-1}(1-p)^{n-k}. \tag{7}$$

Examples of the PDF $f_{P_k}(p)$ for a sample size of $n = 49$ and various values of k are shown in Figure 1. For $1 \ll k \ll n$ the PDF is approximately symmetric, with exact symmetry for $k = (n+1)/2$. For k close to 1 the distribution is positively skewed and for k close to n the distribution is negatively skewed.

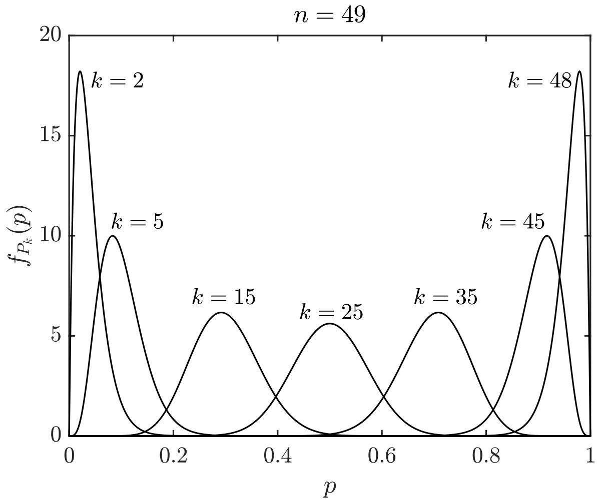

Figure 1: PDF of non-exceedance probability, P_k, associated with order statistic $X_{(k)}$ for a sample size of $n = 49$ and various values of k.

To derive the sampling distributions of the order statistics and their return periods, we note that since $F_X^{-1}(p)$ and $T(p)$ are monotonically increasing functions we have

$$F_{P_k}(p) = F_{T_k}(T(p)) = F_{X_{(k)}}(F_X^{-1}(p)). \tag{8}$$

The PDF of T_k is therefore given by

$$f_{T_k}(t) = \frac{dF_{T_k}(t)}{dt} = \frac{dF_{P_k}(p)}{dp}\frac{dp}{dt} = \frac{n!}{(n-k)!(k-1)!}p^{k-1}(1-p)^{n-k}t^{-2} = \frac{n!}{(n-k)!(k-1)!}(t-1)^{k-1}t^{-n-1}. \tag{9}$$

The PDF of $X_{(k)}$ is given by

$$f_{X_{(k)}}(x) = \frac{dF_{X_{(k)}}(x)}{dx} = \frac{dF_{P_k}(F_X(x))}{dF_X(x)}\frac{dF_X(x)}{dx} = \frac{n!}{(n-k)!(k-1)!}(F_X(x))^{k-1}(1-F_X(x))^{n-k}f_X(x). \tag{10}$$

It is important to note that the PDFs of P_k and T_k do not depend on the data-generating distribution F_X, so these can be calculated without having to estimate F_X. However, obviously, the PDF of $X_{(k)}$ does depend on F_X. In practice, the data-generating distribution F_X is not known. Instead, we can substitute an estimate, $\hat{F}_X$, into (10) to obtain an estimate $\hat{f}_{X_{(k)}}(x)$.

3. Definition of the empirical distribution function and empirical return periods

From the properties of the beta distribution, the expected value, mode, median and quantiles of P_k are given by

$$\mathrm{E}(P_k) = \frac{k}{n+1}, \tag{11a}$$

$$\mathrm{mode}(P_k) = \frac{k-1}{n-1}, \tag{11b}$$

$$\mathrm{median}(P_k) = I^{-1}(0.5, k, n-k+1), \tag{11c}$$

$$P_{k,\alpha} \stackrel{\mathrm{def}}{=} F_{P_k}^{-1}(\alpha) = I^{-1}(\alpha, k, n-k+1), \quad \text{for } \alpha \in [0,1]. \tag{11d}$$

Apart from the special cases of $k = 1$ and $k = n$, it is not possible to write down an explicit expression for the median or quantiles $P_{k,\alpha}$. It is possible to derive various approximations for the median (see Appendix A), although it appears that there are no corresponding approximations for the quantiles. However, efficient algorithms to compute I and I^{-1} are available in most software languages (MATLAB, Python and R all have built-in functions, and various libraries are available for Fortran and C++), so the lack of an explicit formula is not restrictive.

We denote empirical estimates of the return period T_k as $\hat{t}_k$. If the standard definition of the EDF, given in (3), is used to estimate return periods, then we have

$$\hat{t}_k = \frac{1}{1 - \mathrm{E}(P_k)} = \frac{n+1}{n+1-k}. \tag{12}$$

From the sampling distribution of T_k, (9), and the relation to F_{P_k}, (8), we can derive expressions for the expected value, mode and quantiles of T_k as

$$\mathrm{E}(T_k) = \frac{n}{n-k}, \tag{13a}$$

$$\mathrm{mode}(T_k) = \frac{n+1}{n+2-k}, \tag{13b}$$

$$T_{k,\alpha} \overset{\mathrm{def}}{=} F_{T_k}^{-1}(\alpha) = \left(1 - I^{-1}(\alpha, k, n-k+1)\right)^{-1}. \tag{13c}$$

Note that for largest observation we can approximate the quantiles of the distribution of T_k as

$$T_{n,\alpha} = \left(1 - \alpha^{1/n}\right)^{-1} \approx -\frac{n}{\log(\alpha)}. \tag{14}$$

(This approximation can be obtained by expanding $\left(1-\alpha^{1/n}\right)^{-1}$ as a Laurent series). Moreover, from (13a) we can see that the expected value of the return period of the largest observation is $\mathrm{E}(T_n) = \infty$. This has led some authors to dismiss the use of return periods for graphical assessments of extreme value models [34]. However, in Section 5, it will be shown that if a logarithmic scale is used, then return periods are as useful as exceedance probabilities for model diagnostics, despite the infinite expected value for the largest observation.

Nevertheless, it is important to note that for the largest observations there are large differences between $(1 - \mathrm{E}(P_k))^{-1}$, $\mathrm{E}(T_k)$, $\mathrm{mode}(T_k)$ and $\mathrm{median}(T_k)$. Table 1 lists the values of these quantities for the largest two observations. Both the exact expressions and approximate values for large n are listed. If $\mathrm{E}(P_k)$ is used to define an empirical estimate of the return period (or *empirical return period*, ERP), then the ERPs for the two largest observations are approximately $n/2$ and n. In contrast, the expected values of T_k for the two largest observations are n and ∞, so there is a factor of 2 difference for the second largest observation, whereas the expected value of T_n is infinite. For the mode of T_k we have $\mathrm{mode}(T_n) = (1 - \mathrm{E}(P_{n-1}))^{-1}$. Finally, for the median value, we have $\mathrm{median}(T_n) \approx 1.44n$ compared to $(1 - \mathrm{E}(P_n))^{-1} = n + 1$.

So, if empirical estimates of return periods are defined in terms of $\mathrm{E}(P_k)$, then the values obtained do not correspond to the mean, mode or median of T_k. However, from (8), we see that the quantiles of T_k and $X_{(k)}$ are given directly in terms of the quantiles of P_k:

$$T_{k,\alpha} = F_{T_k}^{-1}(\alpha) = (1 - P_{k,\alpha})^{-1}, \tag{15a}$$

$$X_{(k),\alpha} = F_{X_{(k)}}^{-1}(\alpha) = F_{X_{(k)}}^{-1}(P_{k,\alpha}). \tag{15b}$$

Therefore, if we define empirical estimates of probabilities, return periods and quantiles in terms of the medians of the sampling distribution, then this gives a consistent framework that can be used in all types of model diagnostic plots, as

Quantity	$\boldsymbol{k = n-1}$		$\boldsymbol{k = n}$	
	Exact	Approx.	Exact	Approx.
$(1 - \mathrm{E}(P_k))^{-1}$	$\frac{n+1}{2}$	$0.5n$	$n+1$	n
$\mathrm{E}(T_k)$	n	n	∞	∞
$\mathrm{mode}(T_k)$	$\frac{n+1}{3}$	$0.33n$	$\frac{n+1}{2}$	$0.5n$
$\mathrm{median}(T_k)$	$\left(1 - I^{-1}(0.5, n-1, 2)\right)^{-1}$	$0.60n$	$\left(1 - 0.5^{1/n}\right)^{-1}$	$1.44n$

Table 1: Various statistics of the return periods for the largest two observations. Approximate expressions are for large sample size, n.

discussed further in the next section. It is therefore recommended that the EDF and ERP are defined as

$$\hat{p}_k = \mathrm{median}(P_k) = I^{-1}(0.5, k, n-k+1), \tag{16a}$$

$$\hat{t}_k = \mathrm{median}(T_k) = (1 - \hat{p}_k)^{-1}. \tag{16b}$$

Using this definition, we then have

$$\mathrm{median}(X_{(k)}) = F^{-1}_{X_{(k)}}(\hat{p}_k). \tag{17}$$

The variation of the different statistics of P_k and T_k with sample size is shown in Figure 2, plotted against the ratios $T_{k,50}/T_{n,50}$. Logarithmic scales are used for the ordinates of both plots. From (11b), we see that $1 - \mathrm{mode}(P_n) = 0$, so this can not be plotted on a logarithmic scale, since $\log(0) = -\infty$. The 5% and 95% quantiles of P_k and T_k are also shown in Figure 2. It is apparent that, when plotted on a log scale, the width of the 90% confidence interval is similar between sample sizes and similar for both exceedance probabilities and return periods. Moreover, the width of the confidence interval for the largest observation is appears not to vary with sample size. This is considered further in the Section 5.

4. Diagnostic plots for extreme value models

Four types of diagnostic plots that are commonly-used to assess the fit of an extreme value model are listed in Table 2. The plots all involve quantities derived from the observed order statistics, $x_{(k)}$, and quantities derived from either the EDF or ERP. Probability plots and quantile-quantile (QQ) plots both just involve a single set of points, relating the observations and fitted model. In contrast, exceedance probability plots and return period plots both involve two sets of points, one corresponding to the observations and the other corresponding to the fitted model.

Examples of the plots listed in Table 2 are shown in Figure 3. In these plots, 50 independent observations have been simulated from a generalised Pareto (GP) distribution with shape parameter $\xi = 0$, scale parameter $\sigma = 1$. The data have been fitted with a GP model using maximum likelihood. In this case, we know that the fitted model is the same as the data-generating model, so any differences are due to sampling effects and any bias due to the parameter estimation method (see e.g. [35, 36]). The plots are shown using various definitions of the EDF and ERPs.

The various type of plot show the correspondence between the fitted model and observations in different ways. Probability plots give an indication of the agreement over the full range of observations. For extreme value models, we are typically interested in the fit of the model in the tail. This can be difficult to assess from a probability plot, as the tail is compressed in the upper right corner of the plot. For this type of plot, there is little difference in defining the EDF as

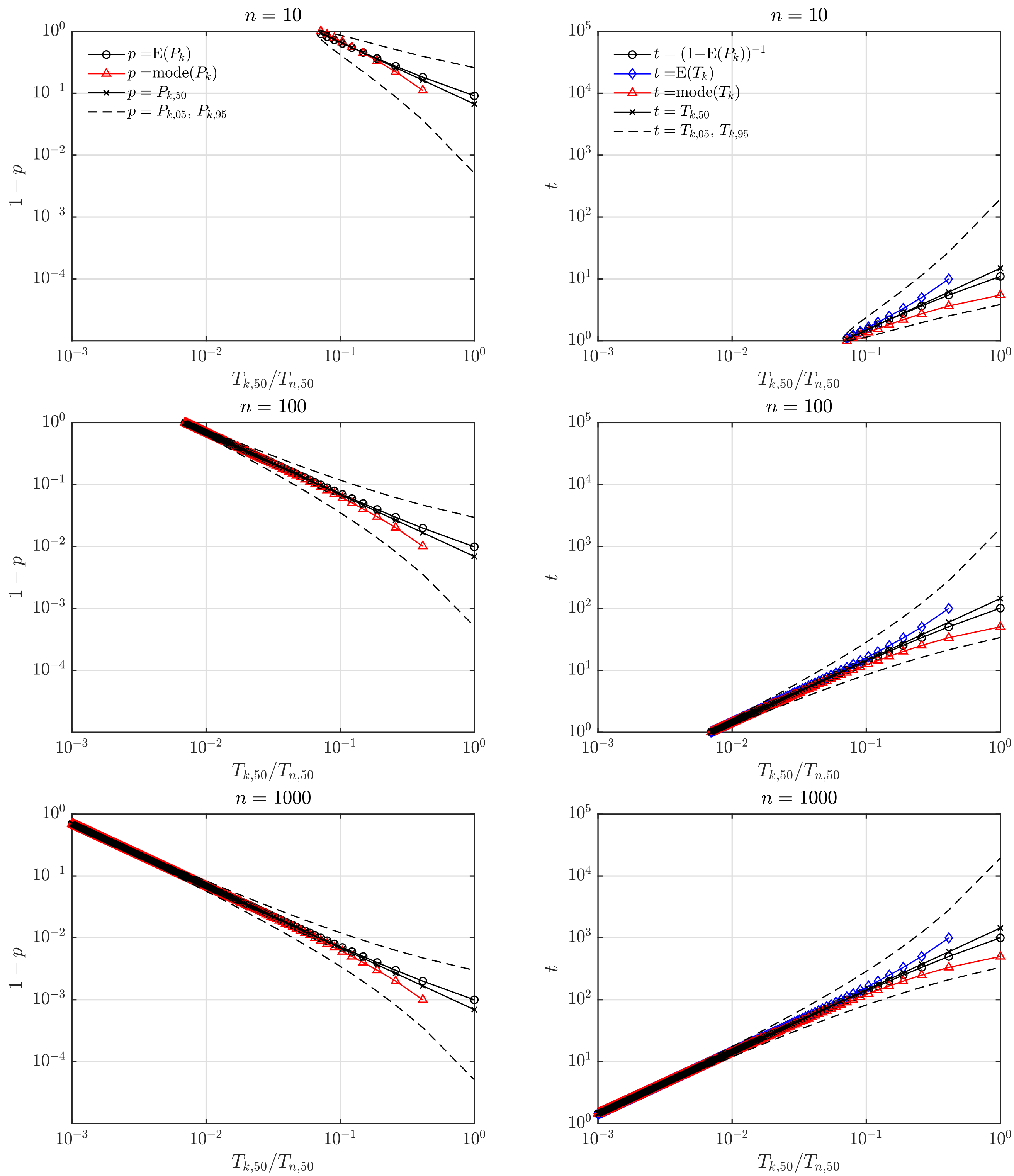

Figure 2: Values of various statistics of P_k and T_k against normalised empirical return period $T_{k,50}/T_{n,50}$ for various sample sizes, n.

Description	Variables	Axes scales
Probability plot	$\left\{\left(\hat{p}_k, \hat{F}_X(x_{(k)})\right) : k = 1, ..., n\right\}$	Linear
Quantile-quantile (QQ) plot	$\left\{\left(x_{(k)}, \hat{F}_X^{-1}(\hat{p}_k)\right) : k = 1, ..., n\right\}$	Linear
Exceedance probability plot	$\left\{\left(x_{(k)}, 1-\hat{p}_k\right) : k = 1, ..., n\right\}$ $\left\{\left(x, 1-\hat{F}_X(x)\right) : x_0 \leq x \leq x_1\right\}$	Ordinate on log-scale
Return period plot	$\left\{\left(\hat{t}_k, x_{(k)}\right) : k = 1, ..., n\right\}$ $\left\{\left((1-\hat{F}_X(x))^{-1}, x\right) : x_0 \leq x \leq x_1\right\}$	Abscissas on log-scale

Table 2: Diagnostic plots used to assess the fit of extreme value models, involving the fitted distribution function, $\hat{F}_X$, order statistics, $x_{(k)}$, empirical distribution function, $\hat{p}_k$, and empirical return period, $\hat{t}_k$.

$\hat{p}_k = \mathrm{E}(P_k)$ or $\hat{p}_k = P_{k,50}$. However, substituting $\hat{p}_k = P_{k,\pm\alpha/2}$ gives a $1-\alpha$ confidence interval for $\hat{p}_k$, which is useful for judging if the fitted model differs from the observations at a given significance level.

QQ plots give an assessment of the fit of the model on the scale of the data, which gives a better indication of the fit in the tail. For QQ plots, the alternative definitions of $\hat{p}_k = \mathrm{E}(P_k)$ or $\hat{p}_k = P_{k,50}$ result in only small differences for most observations, but do lead to a visible difference for the largest value. However, the difference is relatively small compared to the width of the 90% CI.

In a QQ plot there two types of uncertainty in the model quantiles: the sampling uncertainty in the estimate of the empirical non-exceedance probability, $\hat{p}_k$, and the uncertainty in the estimated distribution $\hat{F}_X$. The uncertainty in $\hat{F}_X$ is related to sampling effects, but also includes uncertainty due to model misspecification and any bias introduced by the inference method. In exceedance plots and return period plots, these two types of uncertainty can be considered separately. For exceedance plots, the difference in the definition of $\hat{p}_k$ is, again, relatively small compared to the width of the confidence interval. For the return period plot, three alternative definitions of ERPs are used. In this case, defining $\hat{t}_k = \mathrm{E}(T_k)$ results in slightly larger values than the other definitions. However, since $\mathrm{E}(T_n) = \infty$, this value cannot be plotted.

Gumbel plots are used by practitioners in some fields of extreme value analysis, such as estimation of extreme wind speeds. These plots consist of the points $\left\{\left(-\log(-\log(\hat{p}_k)), x_{(k)}\right) : k = 1, ..., n\right\}$ with both axes on linear scales. This is similar to a QQ plot if the fitted model is a Gumbel distribution, since $-\log(-\log(P)) = (x-\mu)/\sigma$ are normalised quantiles of the Gumbel distribution, where μ and σ are the location and scale parameters. Due to the similarity with QQ plots, the comments on QQ plots above, also apply to Gumbel plots.

5. Confidence intervals

In this section we derive three simple results about the width of confidence intervals (CIs) for the exceedance probabilities and return periods of observations. The results are useful when interpreting either model diagnostic plots or when considering the accuracy of extreme quantities estimated through Monte Carlo simulation.

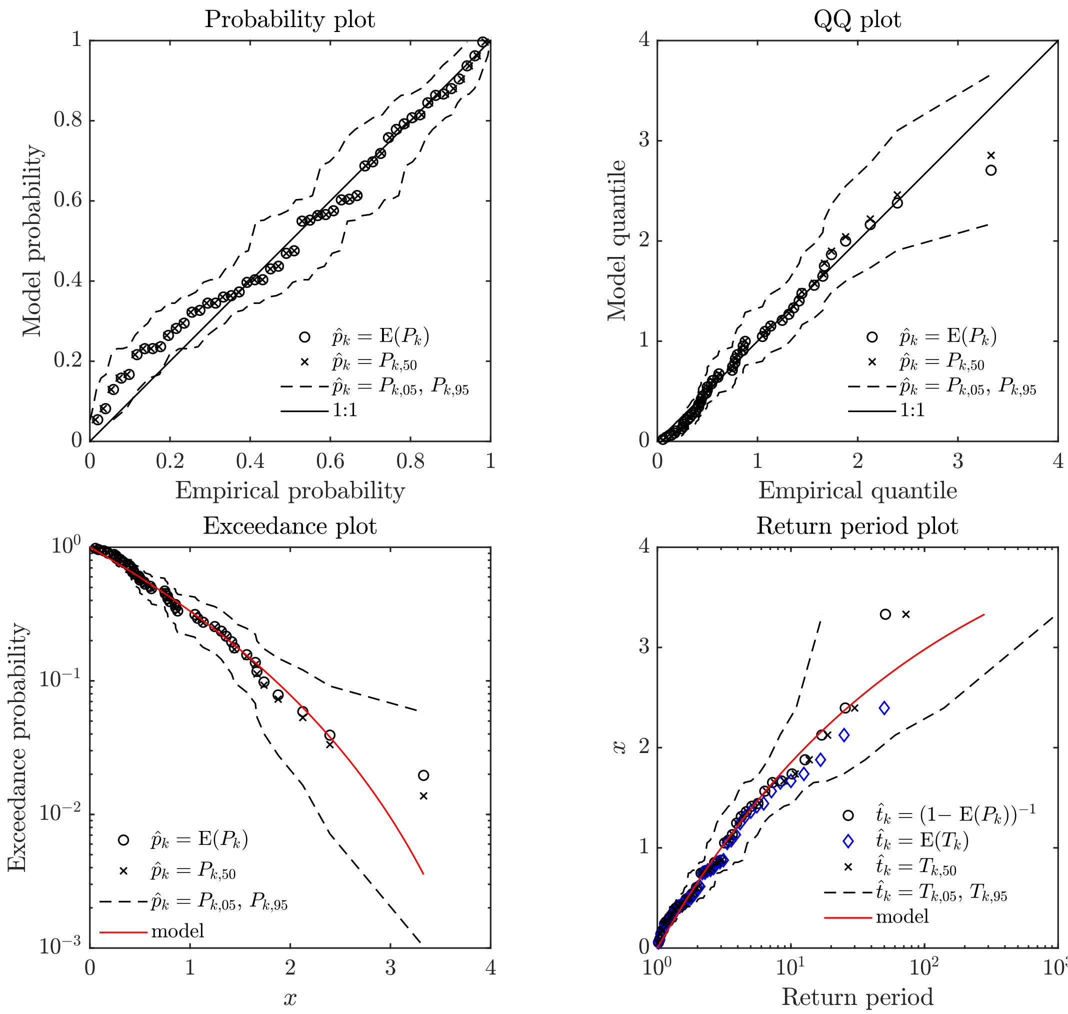

Figure 3: Examples of model diagnostic plots for GP fit to data generated from GP distribution.

5.1. *CIs for exceedance probabilities and return periods have equal widths on a log scale*

It is convenient to denote the quantiles of the exceedance probabilities as

$$Q_{k,\alpha} = 1 - P_{k,1-\alpha}. \tag{18}$$

Then, using (15a), the width of the $1-2\alpha$ CI for Q_k, when plotted on a log scale, is given by

$$\log(Q_{k,1-\alpha}) - \log(Q_{k,\alpha}) = -\log(Q^{-1}_{k,1-\alpha}) + \log(Q^{-1}_{k,\alpha}) = \log(T_{k,1-\alpha}) - \log(T_{k,\alpha}). \tag{19}$$

Therefore, the width of CIs for exceedance probabilities and return periods are equal on a log scale. So, provided that a log scale is used for both exceedance plots and return period plots, then the width of the CI for the empirical estimates is equal. This was apparent in Figures 2 and 3.

5.2. Width of CI for return period of largest observation is asymptotically independent of sample size

Using (14), the width of the $1-2\alpha$ CI for the largest observation, when plotted on a log scale tends to an asymptotic limit for large n:

$$
\begin{aligned}
\log(Q_{n,1-\alpha}) - \log(Q_{n,\alpha}) &= \log(1-\alpha^{1/n}) - \log(1-(1-\alpha)^{1/n}), \\
&\to \log\left(-\frac{\log(\alpha)}{n}\right) - \log\left(-\frac{\log(1-\alpha)}{n}\right), \quad n \to \infty \\
&= \log\left(\frac{\log(\alpha)}{\log(1-\alpha)}\right).
\end{aligned} \tag{20}
$$

This limit is independent of sample size, n. This means that for a large enough sample size, when exceedance plots or return period plots are used to assess the fit of a model, the width of the CI for the largest observation will always be the same on a log scale, regardless of sample size. This was apparent in Figure 2. This provides a quantification of the intuitive notion that there is a large sampling uncertainty associated with the most extreme observation in a sample. The result tells us that no matter how much data we gather, the exceedance probability and return period of the largest observation is highly uncertain. However, the uncertainty in the exceedance probabilities and return periods of less extreme observations does decrease with sample size. This is quantified by the next result.

5.3. Asymptotic approximation for width of CI

To derive an approximation for the width of the CI as a function of the rank, it is useful to start by considering

$$
\mathrm{E}(Q_k) = 1 - \mathrm{E}(P_k) = \frac{n+1-k}{n+1}. \tag{21}
$$

Since Q_k follows a beta distribution, the variance of Q_k can be written explicitly, as

$$
\mathrm{var}(Q_k) = \frac{k(n+1-k)}{(n+1)^2(n+2)} = \frac{\mathrm{E}(Q_k)\mathrm{E}(P_k)}{n+2}. \tag{22}
$$

For large n and low exceedance probabilities we have

$$
\mathrm{var}(Q_k) \approx \frac{\mathrm{E}(Q_k)}{n}. \tag{23}
$$

For large n and $1 \ll k \ll n$, the beta distribution of Q_k can be approximated by a normal distribution [37] with mean $\mathrm{E}(Q_k)$ and variance given above. Using the normal approximation to the beta distribution we can write the quantiles as

$$
Q_{k,\alpha} \approx \mathrm{E}(Q_k) + \Phi^{-1}(\alpha)\sqrt{\mathrm{E}(Q_k)/n}, \tag{24}
$$

where Φ is the CDF of the standard normal distribution. For $1 \ll k \ll n$ the difference between the mean and median is small and we can write $Q_{k,\alpha} \approx Q_{k,50} - \Phi^{-1}(\alpha)\sqrt{Q_{k,50}/n}$. We can use this to derive an approximation for the width of the CI for return periods as follows:

$$
\begin{aligned}
\frac{T_{k,1-\alpha} - T_{k,\alpha}}{T_{k,50}} &\approx Q_{k,50}\left(\frac{1}{Q_{k,50} + \Phi^{-1}(\alpha)\sqrt{Q_{k,50}/n}} - \frac{1}{Q_{k,50} - \Phi^{-1}(\alpha)\sqrt{Q_{k,50}/n}}\right) \\
&= -2\Phi^{-1}(\alpha)\frac{\sqrt{Q_{k,50}/n}}{Q_{k,50} - (\Phi^{-1}(\alpha))^2/n} \\
&\approx -2\Phi^{-1}(\alpha)\sqrt{T_{k,50}/n}.
\end{aligned} \tag{25}
$$

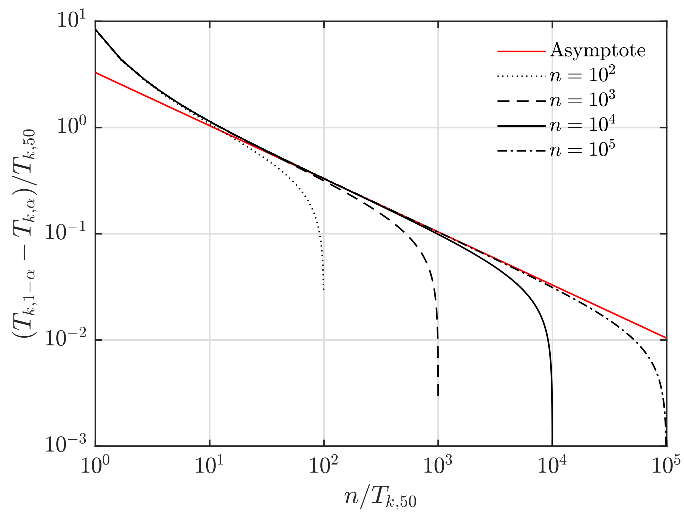

Figure 4: Width of 90% confidence interval for $T_{k,50}$ against $n/T_{k,50}$ for various samples sizes and asymptotic approximation (25).

Figure 4 shows the width of the 90% CI for T_k against $n/T_{k,50}$ for various sample sizes, together with the asymptotic approximation (25). The agreement is good for $10 < T_{k,50} < n/10$.

So, if we want to estimate the T-year return value (RV) based on Monte Carlo simulations using a sample of length $100T$ years, then the 90% CI for the return period of the observation estimated to be the T-year RV is $\approx 0.33T$. And if we generate $1000T$ years, then the width of the CI is $\approx 0.1T$. Of course, the quantity that is probably of greater interest is the CI for the RV rather than the RP of the estimated RV. As discussed in Section 2, this CI is dependent on the data-generating distribution. However, given that the quantiles of $X_{(k)}$ are directly related to the quantiles of Q_k through (15b), an approximation to the width of the CI can easily be generated by substituting the approximation (24) into the quantile function for the fitted distribution.

6. Conclusions

In this work it is argued that defining the empirical distribution function in terms of the median value of the non-exceedance probability of the order statistics, provides a consistent framework for making empirical estimates of exceedance probabilities, return periods and quantiles of extreme events. It is shown that the median value of the return period of the largest observation is 44% larger than the return period calculated using the common definition of the EDF in terms of the expected value of the non-exceedance probability of the order statistics. Although the definition of the EDF influences model diagnostic plots for assessing extreme value models, arguably, a more important consideration is adding confidence bounds to these plots. Some simple results concerning the size of confidence intervals for exceedance probabilities and return periods are derived. These can aid the interpretation of model diagnostic plots and quantify the uncertainty related to Monte Carlo estimates of extreme events.

Acknowledgements

EM was funded by the EPSRC Supergen Offshore Renewable Energy Hub [grant no: EP/S000747/1] Flexible Fund project "Improved Models for Multivariate Metocean Extremes (IMEX)".

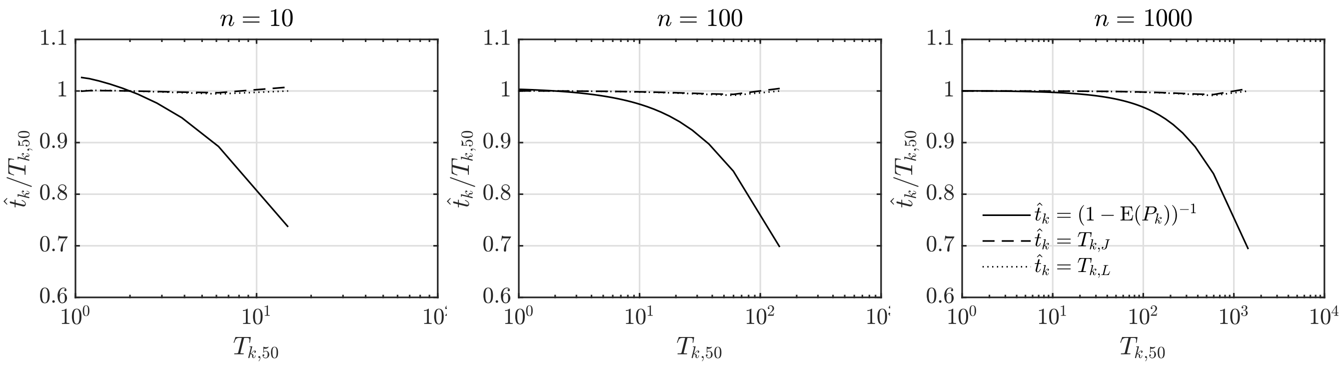

Figure A.1: Ratio of return periods calculated using various methods to the median return period, $T_{k,50}$, for various sample sizes n. .

Appendix A. Approximations for the beta median

Various authors have derived approximations for the median value of P_k. Bernard and Bos-Levenbach [11] derived

$$P_{k,50} \approx \frac{k - 0.3}{n + 0.4}, \tag{A.1}$$

while Jenkinson [38] derived a very similar approximation (for the derivation see [32]):

$$P_{k,50} \approx \frac{k - 0.31}{n + 0.38}. \tag{A.2}$$

Yu and Huang [22] used numerical simulation to derive

$$P_{k,50} \approx \frac{k - 0.326}{n + 0.348}. \tag{A.3}$$

Finally, Lepore [39] used a slightly different analytical approach, to derive

$$P_{k,50} \approx \frac{k - a}{n + 1 - 2a}, \quad a = n + \frac{n-1}{2^{1/n} - 2}. \tag{A.4}$$

Figure A.1 shows the ratio of return periods calculated using various methods to the median return period, $T_{k,50}$, for various sample sizes n. Despite the relatively small differences, it was found that Jenkinson's approximation was more accurate than Bernard and Bos-Levenbach's or Yu and Huang's approximations, so these are not shown in Figure A.1. The figure compares return periods derived using the EDF defined in terms of the expected non-exceedance probability (12), or using Jenkinson's approximation for the median non-exceedance probability, denoted $T_{k,J}$, or using Lepore's approximation, denoted $T_{k,L}$. There is a large difference between return periods calculated using expected non-exceedance probability, with a difference of around 30% for the largest value. Both Jenkinson's and Lepore's approximations give errors less than 1% over all values of k and n considered. Jenkinson's approximation gives a slightly smaller errors for $T_{k,50}/T_{n,50} < 0.7$, whereas Lepore's method is exact for $k = 1$ or n. For practical purposes, both Jenkinson's and Lepore's approximations appear adequate if an exact method for calculating the inverse incomplete beta function is not available.

References

[1] N. Balakrishnan, C. R. Rao, Order Statistics: Theory and Methods. Handbook of statistics, Volume 16, North-Holland, Amsterdam, 1998.

[2] H. David, H. Nagaraja, Order Statistics, 3rd Edition, Wiley Series in Probability and Statistics, 2003.

[3] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of extremes: theory and applications, Wiley, Chichester, UK, 2004.

[4] L. de Haan, A. Ferreira, Extreme Value Theory, An Introduction, Springer, 2006.

[5] S. Coles, An introduction to statistical modelling of extreme values, Springer, 2001.

[6] A. C. Davison, Statistical Models, Cambridge University Press, 2003. doi:10.1017/cbo9780511815850.

[7] H. O. Madsen, S. Krenk, N. Lind, Methods of Structural Safety, Dover Publications, Mineola, New York, 2006.

[8] W. Weibull, A statistical theory of the strength of materials, Ingvetensk Akad. Handl. 151 (1939).

[9] E. Gumbel, Statistics of extremes, Columbia University Press, 1958.

[10] E. Gumbel, On the plotting of flood-discharges, Transactions, American Geophysical Union 24 (2) (1943) 699. doi:10.1029/TR024i002p00699.

[11] A. Bernard, E. J. Bos-Levenbach, The plotting of observations on probability-paper, Statistica (1953) 163–173.

[12] G. Blom, Statistical Estimates and Transformed Beta-Variables, Wiley, New York, 1958. doi:10.2307/3614144.

[13] I. I. Gringorten, A plotting rule for extreme probability paper, Journal of Geophysical Research 68 (3) (1963) 813–814. doi:10.1029/jz068i003p00813.

[14] C. Cunnane, Unbiased plotting positions - A review, Journal of Hydrology 37 (1978) 205–222.

[15] K. Adamowski, Plotting Formula for flood frequency, Journal of the American Water Resources Association 17 (2) (1981) 197–202. doi:10.1111/j.1752-1688.1981.tb03922.x.

[16] H. L. Harter, Another Look at Plotting Positions, Communications in Statistics - Theory and Methods 13 (13) (1984) 1613–1633. doi:10.1080/03610928408828781.

[17] H. L. Harter, R. P. Wiegand, A Monte Carlo Study of Plotting Positions, Communications in Statistics - Simulation and Computation 14 (2) (1985) 317–343. doi:10.1080/03610918508812442.

[18] N. W. Arnell, M. Beran, J. R. Hosking, Unbiased plotting positions for the general extreme value distribution, Journal of Hydrology 86 (1-2) (1986) 59–69. doi:10.1016/0022-1694(86)90006-5.

[19] S. L. Guo, A discussion on unbiased plotting positions for the general extreme value distribution, Journal of Hydrology 121 (1-4) (1990) 33–44. doi:10.1016/0022-1694(90)90223-K.

[20] J. R. M. Hosking, J. R. Wallis, A Comparison of Unbiased and Plotting-Position Estimators of L Moments, Water Resources Research 31 (8) (1995) 2019–2025. doi:10.1029/95WR01230.

[21] M. De, A new unbiased plotting position formula for Gumbel distribution, Stochastic Environmental Research and Risk Assessment 14 (1) (2000) 0001. doi:10.1007/s004770050048.

[22] G. H. Yu, C. C. Huang, A distribution free plotting position, Stochastic Environmental Research and Risk Assessment 15 (6) (2001) 462–476. doi:10.1007/s004770100083.

[23] P. Erto, A. Lepore, A Note on the Plotting Position Controversy and a New Distribution-free Formula, Proceedings of the 45th Scientific Meeting of the Italian Statistical Society, University of Padua (2011) 16–18.

[24] P. Erto, A. Lepore, New distribution-free plotting position through an approximation to the beta median, Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies (2013) 23–27doi:10.1007/978-3-642-35588-2_3.

[25] N. J. Cook, R. I. Harris, The Gringorten estimator revisited, Wind and Structures, An International Journal 16 (4) (2013) 355–372. doi:10.12989/was.2013.16.4.355.

[26] M. Fuglem, G. Parr, I. Jordaan, Plotting positions for fitting distributions and extreme value analysis, Canadian Journal of Civil Engineering 40 (2013) 130–139. doi:10.1139/cjce-2013-0427.

[27] H. P. Hong, S. H. Li, Plotting positions and approximating first two moments of order statistics for Gumbel distribution: Estimating quantiles of wind speed, Wind and Structures, An International Journal 19 (4) (2014) 371–387. doi:10.12989/was.2014.19.4.371.

[28] E. D. Lozano-Aguilera, M. D. Estudillo-Martínez, S. Castillo-Gutiérrez, A proposal for plotting positions in probability plots, Journal of Applied Statistics 41 (1) (2014) 118–126. doi:10.1080/02664763.2013.831814.

[29] R. Hosseini, A. Takemura, An objective look at obtaining the plotting positions for QQ-plots, Communications in Statistics - Theory and Methods 45 (16) (2016) 4716–4728. doi:10.1080/03610926.2014.927498.

[30] A. Lepore, Nearly Unbiased Probability Plots for Extreme Value Distributions, in: A. Petrucci, F. Racioppi, R. Verde (Eds.), New Statistical Developments in Data Science, Springer Proceedings in Mathematics & Statistics, Volume 288, 2017, pp. 457–468.

[31] L. Beard, Statistical Analysis in Hydrology, Transactions of the American Society of Civil Engineers 108 (1943) 1110–1160.

[32] C. Folland, C. Anderson, Estimating changing extremes using empirical ranking methods, Journal of Climate 15 (20) (2002) 2954–2960. doi:10.1175/1520-0442(2002)015<2954:ECEUER>2.0.CO;2.

[33] G. P. Wadsworth, Introduction to Probability and Random Variables, McGraw-Hill, New York, 1960.

[34] N. J. Cook, Rebuttal of 'Problems in the extreme value analysis', Structural Safety 34 (1) (2012) 418–423. doi:10.1016/j.strusafe.2011.08.002.

[35] P. de Zea Bermudez, S. Kotz, Parameter estimation of the generalized Pareto distribution-Part I, Journal of Statistical Planning and Inference 140 (6) (2010) 1353–1373. doi:10.1016/j.jspi.2008.11.019.

[36] E. Mackay, P. Challenor, A. Bahaj, A comparison of estimators for the generalised Pareto distribution, Ocean Engineering 38 (11-12) (2011). doi:10.1016/j.oceaneng.2011.06.005.

[37] N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, Vol 2, 2nd Edition, Wiley-Interscience, 1995.

[38] A. F. Jenkinson, The analysis of meteorological and other geophysical extremes, Tech. rep., Met Office Synoptic Climatology Branch Memo. 58, 41 pp (1977).

[39] A. Lepore, An integrated approach to shorten wind potential assessment, Phd, University of Naples Federico II (2010).