Ocean Engineering 239 (2021) 109791

Contents lists available at ScienceDirect

Ocean Engineering

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

Sampling properties and empirical estimates of extreme events

Ed Mackay [a,*], Philip Jonathan [b,c]

[a] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn, UK
[b] Department of Mathematics and Statistics, Lancaster University, UK
[c] Shell Research Ltd., London, UK

ARTICLE INFO

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

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, 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.

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, model tests or field data.

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

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

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

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

We denote the return period of the kth 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 Balakrishnan and Rao, 1998; David and Nagaraja, 2003).

In the following, we make no assumptions about how $X_1, \ldots, X_n$ are sampled, only that observations are independent and identically distributed (iid). If $X_1, \ldots, X_n$ are a sample of annual maxima, then T

* Corresponding author.
E-mail address: e.mackay@exeter.ac.uk (E. Mackay).

https://doi.org/10.1016/j.oceaneng.2021.109791
Received 5 March 2021; Received in revised form 15 July 2021; Accepted 3 September 2021

has units of years. If $X_1, \ldots, X_n$ 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 kth 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 $\hat{p}_k = k/n$ (Beirlant et al., 2004; de Haan and Ferreira, 2006). 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 Coles, 2001; Davison, 2003; Madsen et al., 2006):

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

This definition was originally proposed by Weibull (1939) and was popularised by Gumbel (1958). 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 (Gumbel, 1943; Bernard and Bos-Levenbach, 1953; Blom, 1958; Gringorten, 1963; Cunnane, 1978; Adamowski, 1981; Harter, 1984; Harter and Wiegand, 1985; Arnell et al., 1986; Guo, 1990; Hosking and Wallis, 1995; De, 2000; Yu and Huang, 2001; Erto and Lepore, 2011, 2013; Cook and Harris, 2013; Fuglem et al., 2013; Hong and Li, 2014; Lozano-Aguilera et al., 2014; Hosseini and Takemura, 2016; Lepore, 2017).

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 $\hat{p}_k = \mathrm{median}(P_k)$, rather than $\hat{p}_k = \mathrm{E}(P_k)$. The use of the median in this context has been advocated by various authors in the past (Beard, 1943; Bernard and Bos-Levenbach, 1953; Yu and Huang, 2001; Folland and Anderson, 2002; Erto and Lepore, 2011, 2013; Lozano-Aguilera et al., 2014; Hosseini and Takemura, 2016). 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 $F_X(x) \leq p$:

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

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, (see e.g. Wadsworth, 1960), 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, $P_k \sim \mathrm{beta}(k, n-k+1)$. 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 Fig. 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.

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

$$\begin{aligned} f_{T_k}(t) &= \frac{d}{dt} F_{T_k}(t) = \frac{dp}{dt} \frac{d}{dp} F_{P_k}(p) = \frac{n!}{(n-k)!(k-1)!} p^{k-1}(1-p)^{n-k} t^{-2} \\ &= \frac{n!}{(n-k)!(k-1)!} (t-1)^{k-1} t^{-n-1}. \end{aligned} \tag{9}$$

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

$$\begin{aligned} f_{X_{(k)}}(x) &= \frac{d}{dx} F_{X_{(k)}}(x) = \frac{dF_X}{dx} \frac{d}{dF_X} F_{P_k}(F_X) \\ &= \frac{n!}{(n-k)!(k-1)!} (F_X(x))^{k-1}(1-F_X(x))^{n-k} f_X(x). \end{aligned} \tag{10}$$

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 $\hat{f}_{X_{(k)}}(x)$.

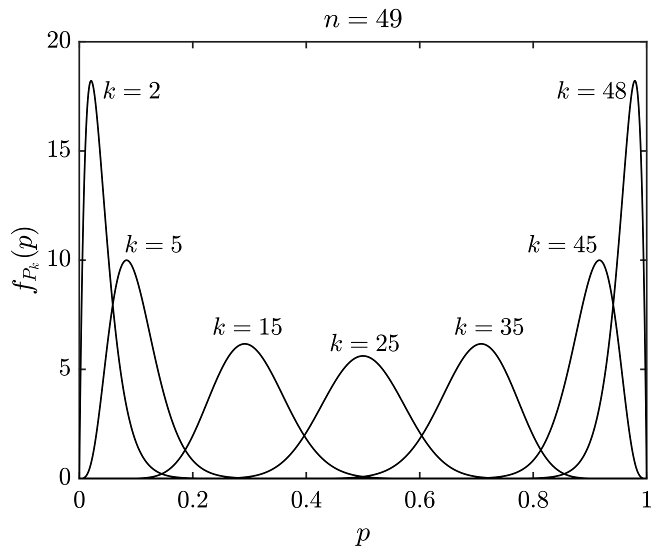

Fig. 1. PDF of non-exceedance probability, P_k, associated with order statistic $X_{(k)}$ for a sample size of $n = 49$ and various values of k.

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

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

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

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

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

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 simple approximations for the median (see Appendix). However, approximations for the quantiles are more complicated (see Abramowitz and Stegun, 1964, §26.5.22 or Temme, 1992 for large sample approximations). 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

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

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

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

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

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

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

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

(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 (Cook, 2012). 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 $\left(1-\mathrm{E}(P_k)\right)^{-1}$, $\mathrm{E}(T_k)$, $\mathrm{mode}(T_k)$ 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 $\mathrm{mode}(T_n) = \left(1-\mathrm{E}(P_{n-1})\right)^{-1}$. Finally, for the median value, we have $\mathrm{median}(T_n) \approx 1.44n$ compared to $\left(1-\mathrm{E}(P_n)\right)^{-1} = n+1$.

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:

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

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

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 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

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

The variation of the different statistics of P_k and T_k with sample size is shown in Fig. 2, plotted against the ranks $j = n-k+1$ (where ranks are defined here in descending order, so that $j = 1$ indicates the largest observation). Logarithmic scales are used for both plots and the direction of the x-axis has been reversed so that the value associated with the largest observation appears on the right. When j is large there is little difference between the various definitions of empirical exceedance probabilities and return periods, and $j \approx n/\hat{t}_k$. So the x-axis values in Fig. 2 can be interpreted as the ratio of sample size to the return period of an observation. From (11b), we see that $1-\mathrm{mode}(P_n) = 0$, so this cannot 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 Fig. 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

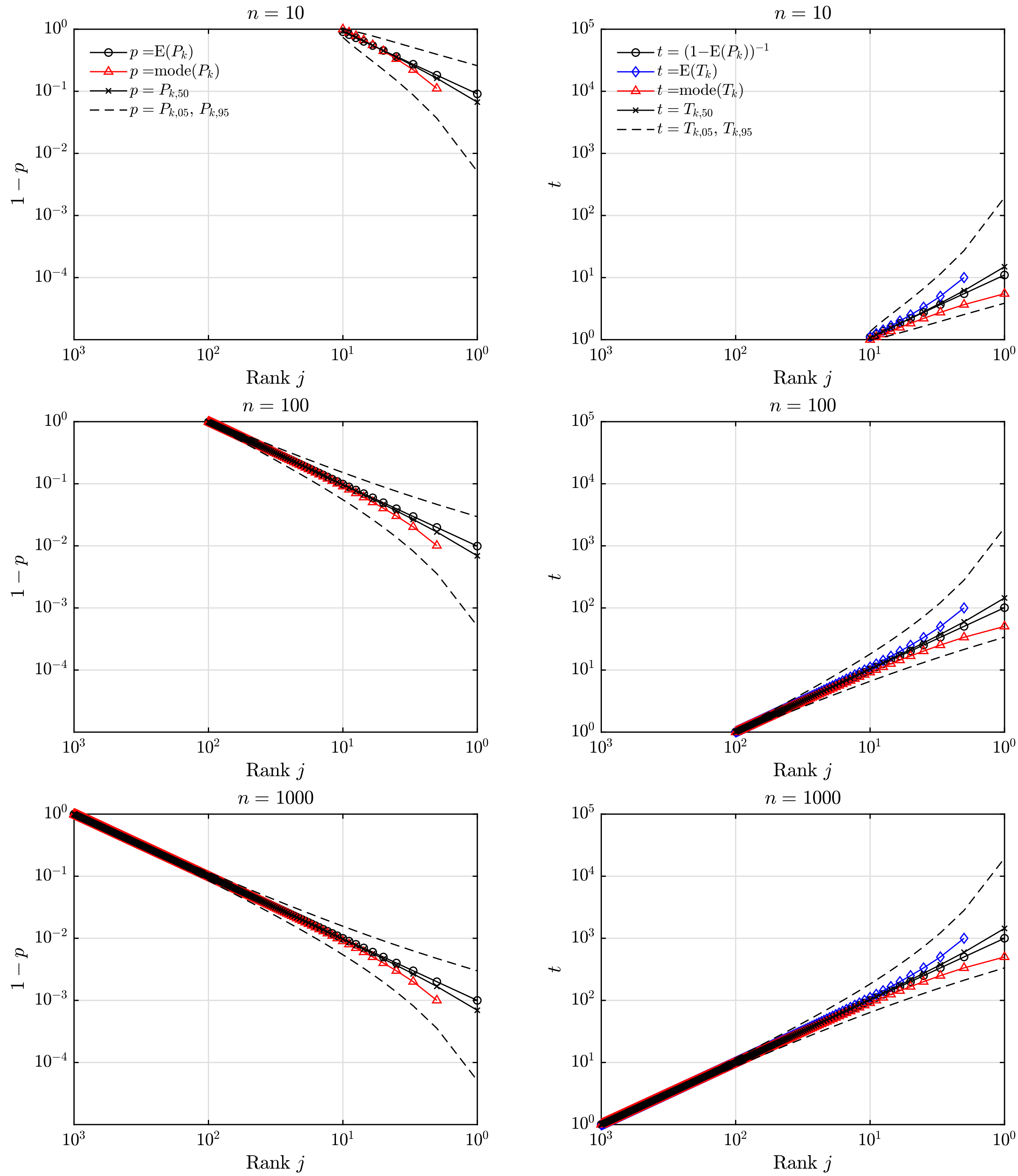

Fig. 2. Values of various statistics of P_k and T_k against rank $j = n - k + 1$ (i.e. $j = 1$ is the largest observation) for various sample sizes, n.

both exceedance probabilities and return periods. Moreover, the width of the confidence interval as a function of rank, appears not to vary with sample size. This is considered further in 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 Fig. 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 is due to the parameter estimation method (see e.g. de Zea Bermudez and Kotz, 2010; Mackay et al., 2011). 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

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, $\hat{t}_k$.

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

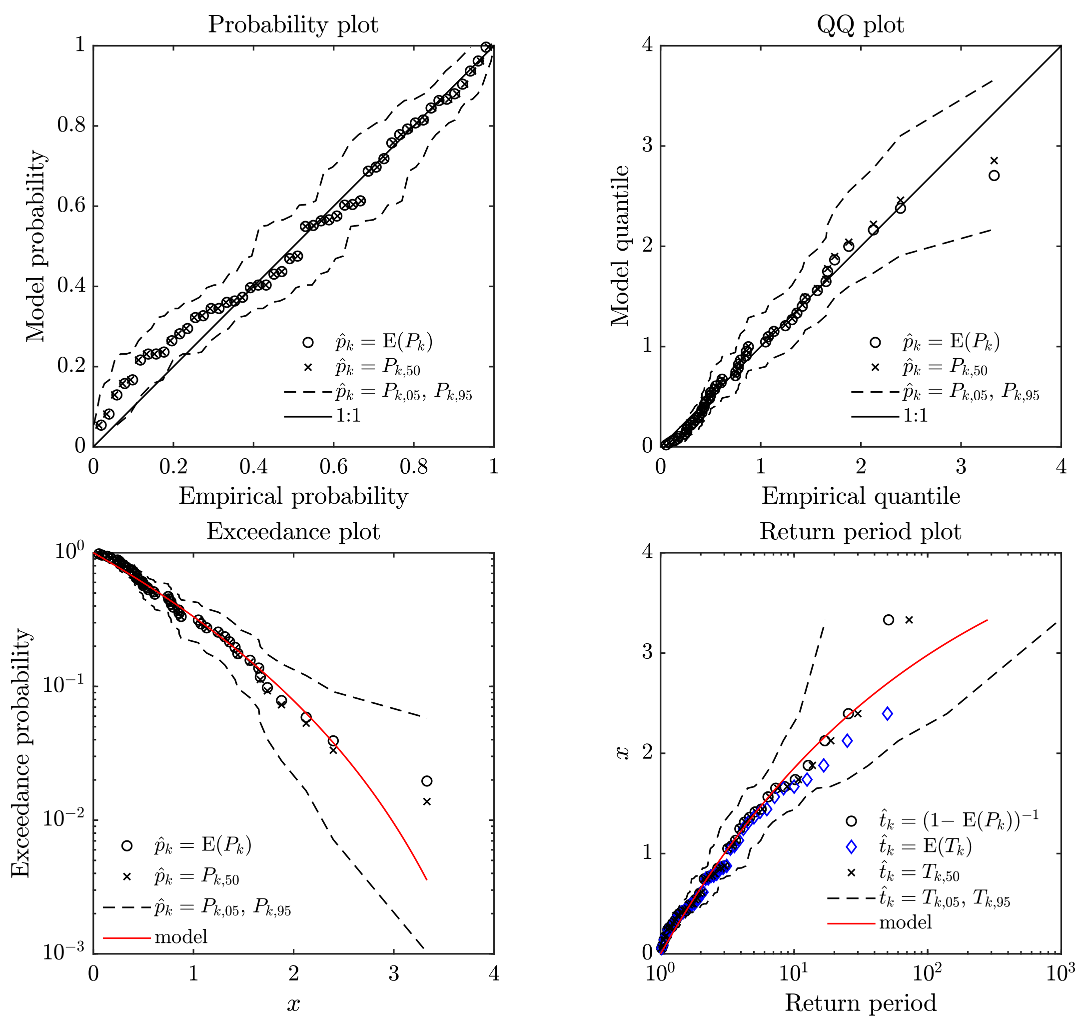

Fig. 3. Examples of model diagnostic plots for GP fit to data generated from GP distribution.

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 $\hat{p}_k = \mathrm{E}(P_k)$ 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 $\{(-\ln(-\ln(\hat{p}_k)), x_{(k)}) : k = 1, \dots, n\}$ with both axes on linear scales. This is similar to a QQ plot if the fitted model is a Gumbel distribution, since $-\ln(-\ln(p)) = (x - \mu)/\sigma$ 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 some 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 from Monte Carlo simulation, model tests or field data.

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

We denote the quantiles of the exceedance probability associated with the kth order statistic as $Q_{k,\alpha} = 1 - P_{k,1-\alpha}$. Using (15a), the width of the $1 - 2\alpha$ CI for Q_k, when plotted on a log scale, is given by

$$\ln(Q_{k,1-\alpha}) - \ln(Q_{k,\alpha}) = -\ln(Q^{-1}_{k,1-\alpha}) + \ln(Q^{-1}_{k,\alpha}) = \ln(T_{k,1-\alpha}) - \ln(T_{k,\alpha}). \quad (18)$$

Therefore, the width of CIs for exceedance probabilities and return periods are equal on a log scale. This was apparent in Figs. 2 and 3. Due to the equivalence between the width of the CIs for exceedance probabilities and return values, we will consider only exceedance probabilities in the remainder of the section.

5.2. Width of CI for exceedance probabilities is asymptotically independent of sample size

This result was also apparent in Fig. 2. For the largest observation, the asymptotic width of the $1 - 2\alpha$ CI, when plotted on a log scale, can be calculated using (14):

$$\begin{aligned}\ln(Q_{n,1-\alpha}) - \ln(Q_{n,\alpha}) &= \ln(1 - \alpha^{1/n}) - \ln(1 - (1 - \alpha)^{1/n}),\\ &\to \ln\left(-\frac{\ln(\alpha)}{n}\right) - \ln\left(-\frac{\ln(1-\alpha)}{n}\right), \quad n \to \infty\\ &= \ln\left(\frac{\ln(\alpha)}{\ln(1-\alpha)}\right). \end{aligned} \quad (19)$$

This limit is independent of sample size, n.

For the general case, consider the approximate form of the inverse incomplete beta function (Abramowitz and Stegun, 1964, §26.5.22):

$$Q_{k,\alpha} \approx \frac{j}{j + ke^{2w(\alpha)}}, \quad (20)$$

where

$$\begin{aligned} w(\alpha) &= \left(\frac{1}{2j-1} - \frac{1}{2k-1}\right)\left(\lambda(\alpha) + \frac{5}{6} - \frac{2}{3h}\right) - \Phi^{-1}(\alpha)\frac{(h + \lambda(\alpha))^{1/2}}{h},\\ \lambda(\alpha) &= \frac{(\Phi^{-1}(\alpha))^2 - 3}{6},\\ h &= 2\left(\frac{1}{2j-1} + \frac{1}{2k-1}\right)^{-1}, \end{aligned}$$

and Φ is the CDF of the standard normal distribution. Suppose that both n and k are large and that $j \ll k$. The width of the CI on a log scale can then be approximated as:

$$\begin{aligned}\ln(Q_{k,1-\alpha}) - \ln(Q_{k,\alpha}) &\approx \ln\left(\frac{j}{j + ke^{2w(1-\alpha)}}\right) - \ln\left(\frac{j}{j + ke^{2w(\alpha)}}\right)\\ &= \ln\left(\frac{j + ke^{2w(\alpha)}}{j + ke^{2w(1-\alpha)}}\right)\\ &\to \ln\left(\frac{e^{2w(\alpha)}}{e^{2w(1-\alpha)}}\right), \quad k \to \infty\\ &= 2\,(w(\alpha) - w(1-\alpha)). \end{aligned} \quad (21)$$

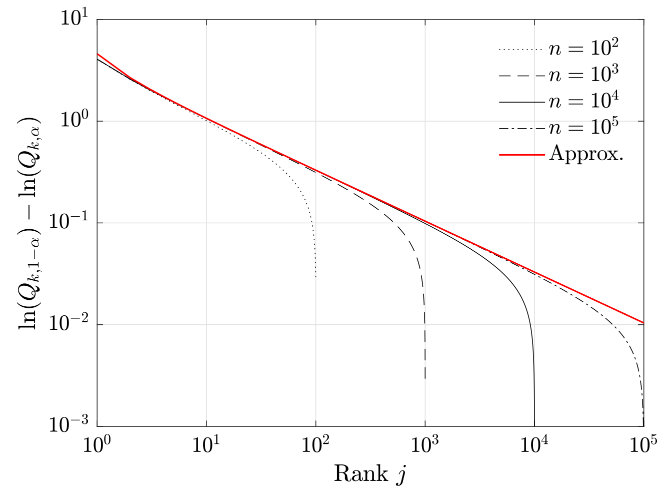

Fig. 4. Width of 90% confidence interval for exceedance probabilities against rank j for various samples sizes and asymptotic approximation (22).

The functions w and h tend to limits that are independent of sample size and dependent only the rank, j. Taking the limits as $k \to \infty$, we can write

$$\ln(Q_{k,1-\alpha}) - \ln(Q_{k,\alpha}) \approx -2\Phi^{-1}(\alpha)\frac{(4j - 2 + \lambda(\alpha))^{1/2}}{2j - 1}. \quad (22)$$

This approximation only depends on the rank, j, and the value of α, and is independent of sample size. Fig. 4 shows the width of the 90% CI for Q_k against rank j for various sample sizes, together with the asymptotic approximation (22). The agreement between the exact results and the asymptotic approximation is good for $2 < j < n/10$. In Fig. 4, the width of the confidence interval using a natural logarithm scale is shown. This scale is used due to the connection with the confidence intervals for quantiles, discussed in the next section. In extreme value plots, it is more common to use a $\log_{10}$ scale. The CI width on the $\log_{10}$ scale can be obtained from Fig. 4 by dividing by $\ln(10) \approx 2.30$. It is apparent that to obtain a 90% CI width of less than 0.1 on a $\log_{10}$ scale, the sample size needs to be approximately 200 times the return period. Of course, the quantity that is probably of greater interest is the CI for the return values rather than the return periods. As discussed in Section 2, this CI is dependent on the data-generating distribution. Various methods for obtaining this CI from the sampling properties of the order statistics are discussed in the following subsection.

5.3. Relation between CI for probabilities and CI for quantiles

Given that the quantiles of $X_{(k)}$ are directly related to the quantiles of Q_k through (15b), an estimate of the CI for the quantiles can be obtained by substituting either the exact or approximate expressions for the quantiles of Q_k into the quantile function for the fitted distribution. In the case that n is large and $1 \ll k \ll n$, the beta distribution of Q_k can be approximated by a normal distribution (Johnson et al., 1995), giving a particularly simple expression for the quantiles of Q_k. From the properties of the beta distribution, the expected value of Q_k is given by

$$\mathrm{E}(Q_k) = 1 - \mathrm{E}(P_k) = \frac{n - k + 1}{n + 1}, \quad (23)$$

and the variance is

$$\mathrm{var}(Q_k) = \frac{k(n + 1 - k)}{(n + 1)^2(n + 2)}. \quad (24)$$

The quantiles of the exceedance probabilities can then be approximated as:

$$Q_{k,\alpha} \approx \mathrm{E}(Q_k) + \Phi^{-1}(\alpha)\sqrt{\mathrm{var}(Q_k)}. \quad (25)$$

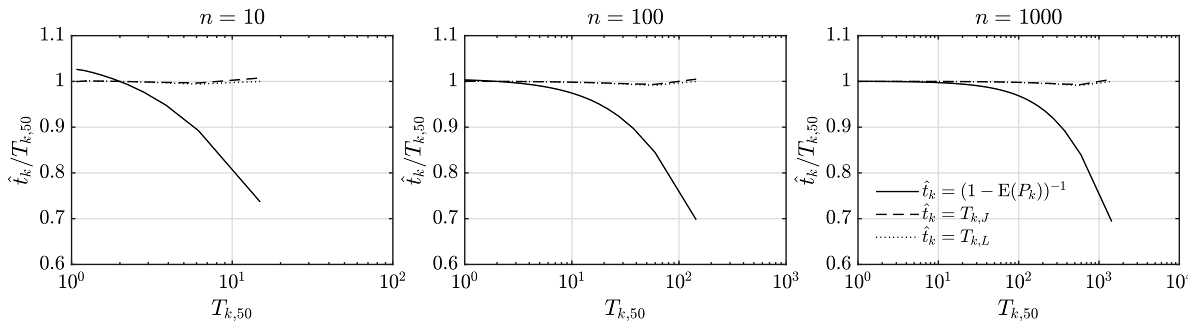

Fig. A.1. Ratio of return periods calculated using various methods to the median return period, $T_{k,50}$, for various sample sizes n.

We can also establish a more direct link between the CI for exceedance probabilities and the CI for quantiles. Suppose we are trying to estimate the T-year return value of a variable X using Monte Carlo simulation from a sample of size $n \geq T$. Monte Carlo simulation might be used when we do not have an explicit model for X, but the values of X can be calculated in terms of other variables, e.g. when X represents the response of a structure to environmental loading. Let k be the value corresponding to $\hat{t}_k = T$, where k is not necessarily an integer. The estimate of the T-year return value is $\hat{x}_T = x_{(k)}$ and the $1-2\alpha$ CI for the estimate is given by $(F_X^{-1}(P_{k,\alpha}), F_X^{-1}(P_{k,1-\alpha}))$. If Monte Carlo simulation is being used, then it is unlikely that F_X is known explicitly. However, if the interest is in the extreme values of X, asymptotic arguments can be used to show that if X are block maxima then X will be well-modelled using a generalised extreme value (GEV) distribution, for sufficiently large block size. Alternatively, if X are peaks-over-threshold, then X will be well-modelled by a generalised Pareto (GP) distribution, for sufficiently high threshold levels (Coles, 2001). In these cases, the CDF of X is given by:

$$F_X(x) = \begin{cases} \exp(-z), & \text{if } X \sim GEV \\ 1 - z, & \text{if } X \sim GP \end{cases} \tag{26}$$

where

$$z = \begin{cases} \exp\left(-\dfrac{x-\mu}{\sigma}\right), & \xi = 0, \\ \left(1 + \xi\dfrac{x-\mu}{\sigma}\right)_+^{-1/\xi}, & \xi \neq 0, \end{cases} \tag{27}$$

and $s_+ = \max\{s, 0\}$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$. For large return periods, z is small and $\exp(z) \approx 1 - z$, so the tail of the GEV distribution tends to a GP distribution with the same parameters. Therefore, we can just consider the GP case, where the quantile function is given by:

$$F_X^{-1}(p) = \begin{cases} \mu - \sigma \ln(1-p), & \xi = 0, \\ \mu + \dfrac{\sigma}{\xi}\left((1-p)^{-\xi} - 1\right), & \xi \neq 0, \end{cases} \tag{28}$$

for $p \in [0, 1]$. The normalised width of the $1 - 2\alpha$ CI for $\hat{x}_T = x_{(k)}$ for large T can therefore be approximated by:

$$\frac{x_{T,1-\alpha} - x_{T,\alpha}}{\sigma} \approx \begin{cases} \ln\left(Q_{k,1-\alpha}\right) - \ln\left(Q_{k,\alpha}\right), & \xi = 0, \\ \frac{1}{\xi}\left(Q_{k,\alpha}^{-\xi} - Q_{k,1-\alpha}^{-\xi}\right), & \xi \neq 0, \end{cases} \tag{29}$$

So, in the special case that $\xi = 0$, the normalised width of the CI for the return value of X is given by the width of the CI for exceedance probabilities, when plotted on a natural logarithmic scale. If $\xi > 0$ then the width of the CI for $\hat{x}_T$ will be greater than the width of the CI for exceedance probabilities, and vice versa when $\xi < 0$. In ocean engineering, many key variables, such as wave heights or wind speeds, are often assumed to follow distributions which have exponential tails (i.e. $\xi = 0$), such as Weibull or lognormal distributions. In these cases, the relationship between the CI for probabilities and quantiles is particularly useful.

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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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

Appendix. Approximations for the beta median

Various authors have derived approximations for the median value of P_k. Bernard and Bos-Levenbach (1953) derived

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

while Jenkinson (1977) derived a very similar approximation (for the derivation see Folland and Anderson (2002)):

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

Yu and Huang (2001) used numerical simulation to derive

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

Finally, Lepore (2010) used a slightly different analytical approach, to derive

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

Fig. 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 Fig. 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

Abramowitz, M., Stegun, I., 1964. Handbook of Mathematical Functions. United States Department of Commerce, National Bureau of Standards, p. 1046.

Adamowski, K., 1981. Plotting formula for flood frequency. J. Am. Water Resour. Assoc. 17 (2), 197–202. http://dx.doi.org/10.1111/j.1752-1688.1981.tb03922.x.

Arnell, N.W., Beran, M., Hosking, J.R., 1986. Unbiased plotting positions for the general extreme value distribution. J. Hydrol. 86 (1–2), 59–69. http://dx.doi.org/10.1016/0022-1694(86)90006-5.

Balakrishnan, N., Rao, C.R., 1998. Order Statistics: Theory and Methods. In: Handbook of Statistics, vol. 16, North-Holland, Amsterdam.

Beard, L., 1943. Statistical analysis in hydrology. Trans. Am. Soc. Civ. Eng. 108, 1110–1160.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Wiley, Chichester, UK.

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

Blom, G., 1958. Statistical Estimates and Transformed Beta-Variables. Wiley, New York, http://dx.doi.org/10.2307/3614144.

Coles, S., 2001. An Introduction to Statistical Modelling of Extreme Values. Springer.

Cook, N.J., 2012. Rebuttal of 'Problems in the extreme value analysis'. Struct. Saf. 34 (1), 418–423. http://dx.doi.org/10.1016/j.strusafe.2011.08.002.

Cook, N.J., Harris, R.I., 2013. The Gringorten estimator revisited. Wind Struct. Int. J. 16 (4), 355–372. http://dx.doi.org/10.12989/was.2013.16.4.355.

Cunnane, C., 1978. Unbiased plotting positions - A review. J. Hydrol. 37, 205–222.

David, H., Nagaraja, H., 2003. Order Statistics, third ed. Wiley Series in Probability and Statistics.

Davison, A.C., 2003. Statistical Models. Cambridge University Press, http://dx.doi.org/10.1017/cbo9780511815850.

De, M., 2000. A new unbiased plotting position formula for Gumbel distribution. Stoch. Environ. Res. Risk Assess. 14 (1), 0001. http://dx.doi.org/10.1007/s004770050048.

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

de Zea Bermudez, P., Kotz, S., 2010. Parameter estimation of the generalized Pareto distribution-Part I. J. Statist. Plann. Inference 140 (6), 1353–1373. http://dx.doi.org/10.1016/j.jspi.2008.11.019.

Erto, P., Lepore, A., 2011. A note on the plotting position controversy and a new distribution-free formula. In: Proceedings of the 45th Scientific Meeting of the Italian Statistical Society. University of Padua, pp. 16–18.

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

Folland, C., Anderson, C., 2002. Estimating changing extremes using empirical ranking methods. J. Clim. 15 (20), 2954–2960. http://dx.doi.org/10.1175/1520-0442(2002)015<2954:ECEUER>2.0.CO;2.

Fuglem, M., Parr, G., Jordaan, I., 2013. Plotting positions for fitting distributions and extreme value analysis. Can. J. Civil Eng. 40, 130–139. http://dx.doi.org/10.1139/cjce-2013-0427.

Gringorten, I.I., 1963. A plotting rule for extreme probability paper. J. Geophys. Res. 68 (3), 813–814. http://dx.doi.org/10.1029/jz068i003p00813.

Gumbel, E., 1943. On the plotting of flood-discharges. Trans. Am. Geophys. Union 24 (2), 699. http://dx.doi.org/10.1029/TR024i002p00699.

Gumbel, E., 1958. Statistics of Extremes. Columbia University Press.

Guo, S.L., 1990. A discussion on unbiased plotting positions for the general extreme value distribution. J. Hydrol. 121 (1–4), 33–44. http://dx.doi.org/10.1016/0022-1694(90)90223-K.

Harter, H.L., 1984. Another look at plotting positions. Comm. Statist. Theory Methods 13 (13), 1613–1633. http://dx.doi.org/10.1080/03610928408828781.

Harter, H.L., Wiegand, R.P., 1985. A Monte Carlo study of plotting positions. Comm. Statist. Simulation Comput. 14 (2), 317–343. http://dx.doi.org/10.1080/03610918508812442.

Hong, H.P., Li, S.H., 2014. Plotting positions and approximating first two moments of order statistics for Gumbel distribution: Estimating quantiles of wind speed. Wind Struct. Int. J. 19 (4), 371–387. http://dx.doi.org/10.12989/was.2014.19.4.371.

Hosking, J.R.M., Wallis, J.R., 1995. A comparison of unbiased and plotting-position estimators of L moments. Water Resour. Res. 31 (8), 2019–2025. http://dx.doi.org/10.1029/95WR01230.

Hosseini, R., Takemura, A., 2016. An objective look at obtaining the plotting positions for QQ-plots. Comm. Statist. Theory Methods 45 (16), 4716–4728. http://dx.doi.org/10.1080/03610926.2014.927498.

Jenkinson, A.F., 1977. The Analysis of Meteorological and Other Geophysical Extremes, Vol. 58. Tech. rep., Met Office Synoptic Climatology Branch Memo, p. 41.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions, Vol. 2, second ed. Wiley-Interscience.

Lepore, A., 2010. An Integrated Approach to Shorten Wind Potential Assessment (Ph.D. thesis). University of Naples Federico II.

Lepore, A., 2017. Nearly unbiased probability plots for extreme value distributions. In: Petrucci, A., Racioppi, F., Verde, R. (Eds.), New Statistical Developments in Data Science. In: Springer Proceedings in Mathematics & Statistics, vol. 288, pp. 457–468.

Lozano-Aguilera, E.D., Estudillo-Martínez, M.D., Castillo-Gutiérrez, S., 2014. A proposal for plotting positions in probability plots. J. Appl. Stat. 41 (1), 118–126. http://dx.doi.org/10.1080/02664763.2013.831814.

Mackay, E., Challenor, P., Bahaj, A., 2011. A comparison of estimators for the generalised Pareto distribution. Ocean Eng. 38 (11–12), http://dx.doi.org/10.1016/j.oceaneng.2011.06.005.

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

Temme, N.M., 1992. Asymptotic inversion of the incomplete beta function. J. Comput. Appl. Math. 41 (1–2), 145–157. http://dx.doi.org/10.1016/0377-0427(92)90244-R.

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

Weibull, W., 1939. A statistical theory of the strength of materials. Ingvetensk Akad. Handl. 151.

Yu, G.H., Huang, C.C., 2001. A distribution free plotting position. Stoch. Environ. Res. Risk Assess. 15 (6), 462–476. http://dx.doi.org/10.1007/s004770100083.