Ocean Engineering 220 (2021) 107725

Contents lists available at ScienceDirect

Ocean Engineering

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

Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution

Philip Jonathan [a,b,*], David Randell [c], Jenny Wadsworth [b], Jonathan Tawn [b]

[a] Shell Research Limited, London, SE1 7NA, United Kingdom
[b] Department of Mathematics and Statistics, Lancaster University, LA1 4YF, United Kingdom
[c] Shell Global Solutions International BV, 1031 HW, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Return value
Extreme
Significant wave height
Predictive distribution

ABSTRACT

We consider the estimation of return values in the presence of uncertain extreme value model parameters, using maximum likelihood and other estimation schemes. Estimators for return value, which yield identical values when parameter uncertainty is ignored, give different values when uncertainty is taken into account. Given uncertain shape ξ and scale parameters of a generalised Pareto (GP) distribution, four sample estimators q for the N-year return value q_0, popular in the engineering community, are considered. These are: q_1, the quantile of the distribution of the annual maximum event with non-exceedance probability $1-1/N$, estimated using mean model parameters; q_2, the mean of different quantile estimates of the annual maximum event with non-exceedance probability $1-1/N$; q_3, the quantile of the predictive distribution of the annual maximum event with non-exceedance probability $1-1/N$; and q_4, the quantile of the predictive distribution of the N-year maximum event with non-exceedance probability $\exp[-1]$. Using theoretical arguments, and simulation of samples of GP-distributed peaks over threshold (with $\xi \in [-0.4, 0.1]$) and different GP parameter estimation schemes, we show that the rank order of estimators q and true value q_0 can be predicted, and that differences between estimators q and q_0 can be large. Judgements concerning the relative performance of estimators depend on the choice of utility function adopted to assess them. We consider bias in return value, bias in exceedance probability and bias in log exceedance probability. None of the four estimators performs well with respect to all three utilities under maximum likelihood estimation, but the mean quantile q_2 is probably the best overall. The estimation scheme of Zhang (2010) provides low bias for q_1.

1. Introduction

Current engineering guidelines (e.g. ISO19901-1, 2015, NORSOK N-003, 2017, DNVGL-RP-C205, 2017) require that extreme ocean environments are characterised marginally and conditionally in terms of return values, often with respect to covariates such as direction. Sometimes the metocean engineer's task is to identify combinations of return values for dominant variables in combination with associated values for other variables to represent extreme conditions for a given environment; these can also be summarised in terms of environmental design contours.

Recent years have seen marked improvements in quality and availability of field measurements, better numerical models for storm environments, and computationally feasible approaches to statistical inference. These allow for realistic quantification of uncertainty in estimation of the tails of distributions of quantities such as significant wave height. Specifically, we can quantify the uncertainty with which the parameters of the distribution of peaks over high threshold, or of block maxima, are estimated from a sample of data. Since return values are defined as functions of tail parameters, we can then quantify the effect of parameter uncertainty on estimates for return values.

When the uncertainty in model parameters is ignored (e.g. when point estimates are used), return values can be estimated using a number of different approaches popular in the ocean engineering literature, yielding identical results. For example, the N-year return value is typically defined as the quantile of the distribution of the annual maximum of a random quantity with non-exceedance probability $1-1/N$. However, some authors use a definition of return value in terms of the quantile of the distribution of the N-year maximum with non-exceedance probability $\exp[-1]$. For large N and given model

* Corresponding author. Shell Research Limited, London, SE1 7NA, United Kingdom.
E-mail address: philip.jonathan@shell.com (P. Jonathan).

https://doi.org/10.1016/j.oceaneng.2020.107725
Received 29 March 2020; Received in revised form 18 June 2020; Accepted 27 June 2020
Available online 4 January 2021

parameters, these two definitions yield effectively identical results as discussed below. However, once parameter uncertainty is incorporated, different approaches (including the two just mentioned) to return value estimation yield different results. These differences are of fundamental concern to the practising ocean engineer, and have been noted at least anecdotally for some time.

Estimation of the N-year return value often relies on an extreme value model fitted to a sample of data, from which an estimate for the distribution of the annual maximum event is made. The characteristics of fitted model parameter estimates depend on (a) the nature of the sample (e.g. block maxima or peaks over threshold, sample size; complications due to lack of independence of observations, covariates, etc.) and (b) the choice of estimation strategy (e.g. generalised extreme value (GEV) distribution for block maxima or generalised Pareto (GP) distribution for threshold exceedances; estimation method including e.g. maximum likelihood or probability weighted moments; accommodations for dependence and covariates, etc.). The characteristics of the fitted parameter estimates in turn influence those of the estimated N-year return value.

Peaks over threshold analysis is often preferred over direct analysis of annual maxima since in typical applications there are multiple occurrences of threshold exceedances per annum. This might be the case for analysis of storm peak significant wave height from extra-tropical storms and typical threshold levels. In this sense, the peaks over threshold analysis can be more efficient since the same period of observation provides more observations for model building and tail. Madsen et al. (1997) provides expressions for the asymptotic variance of return value estimates from GEV and GP estimation. They conclude that for all practical purposes, GP provides the most efficient return value estimator in the case of maximum likelihood estimation. Dombry and Ferreira (2019) further show that peaks over threshold is preferable under maximum likelihood estimation in terms of (a) lower asymptotic bias and (b) lower asymptotic mean square error of the GP shape parameter estimate; however, the block maximum (GEV) approach is preferred in terms of asymptotic variance. Similar findings are given by Ferreira and de Haan (2015) when parameter estimation is performed using the method of probability weighted moments.

From a Bayesian perspective, an optimal value (such as return value) for some random quantity can only be provided with respect to an appropriate loss or utility structure. Christensen and Huffman (1985), building on the work of Geisser (1971), discuss Bayesian point estimation when no loss function is specified, and argue in favour of the posterior mean estimator (analogous to q_2 in the notation of the abstract). Smith (2003) discusses issues related to incorporation of parameter uncertainty in estimates for return values. Serinaldi (2015) emphasises that the preferred approach to estimate environmental risk subject to uncertainty should involve direct estimation of probabilities of exceedance, or of structural failure over a given design life, incorporating all sources of epistemic and aleatory uncertainty in play. Fawcett and Green (2018) assess the performance of (posterior) predictive return levels relative to their estimative counterparts, such as the mean and mode of the posterior distribution of return value. They conclude that, for the cases considered in a simulation study, the predictive return value (analogous to q_3 in the notation of abstract) yields estimates of exceedance probabilities with much higher precision than the corresponding exceedance probabilities obtained from estimative summaries.

Design standards such as DNVGL-RP-C205 (2017) also support the estimation of return values using so-called ‘all sea state’ or ‘global’ models. In this situation, the full distribution of e.g. sea-state significant wave height is estimated from a sample of (typically dependent) values of sea-state significant wave height using some parametric form e.g. related to a Weibull distribution (see Haselsteiner and Thoben, 2020). This approach is not addressed in the current work, which focusses on return value estimation using peaks over threshold analysis.

The purpose of this article is to describe some of the more common estimators of return values used in ocean engineering, and to consider how uncertainties in parameters of extreme value models influence estimation of the bias of (a) return value, (b) probability of exceedance and (c) logarithm of the exceedance probability associated with a return value.

The article is arranged as follows. In Section 2 we provide an outline of different definitions of the N-year return value used by ocean engineers. For a given application, when parameter uncertainty is ignored, all these approaches yield the same estimate of return value. In Section 3, we extend the definition of return value to incorporate parameter uncertainty, suggesting six different estimators; Section 4 presents the corresponding estimators for samples of peaks over threshold following a GP distribution. Section 5 then illustrates the characteristics of the different estimators in a simulation study, assuming that maximum likelihood is used to estimate GP parameters. Section 6 provides simple theoretical arguments to help understand why estimators from Section 4 yield different values. Section 7 provides discussion and recommendations.

2. Return value estimators in the absence of tail parameter uncertainty

Suppose that a random variable A represents the maximum value of some physical quantity X (such as storm peak significant wave height) per annum. The N-year return value q of X is then conventionally defined by the equation

$$F_A(q) = \Pr(A \leq q) = 1 - \frac{1}{N} \tag{1}$$

where F_A is the cumulative distribution function of the annual maximum A. We might also say that q is the quantile of the distribution of A with non-exceedance probability $1 - 1/N$ and write

$$q = Q_A(1 - 1/N) \tag{2}$$

where Q_A is the inverse of F_A, i.e. $Q_A = F_A^{-1}$. To estimate q, we need knowledge of F_A. We might estimate F_A using extreme value analysis on a sample of independent observations of A. For ocean engineering applications however, it is typically more efficient to estimate the distribution $F_{X|X>\psi}$ of threshold exceedances of X above some high threshold ψ using a sample of independent observations of X, and use this in turn to estimate F_A as outlined in the Appendix and discussed in Section 3.

Applying the expression in Equation (1) to a time period of N years, we can also write

$$F_{A_N}(q) = F_A^N(q) = \left(1 - \frac{1}{N}\right)^N$$

where A_N is the N-year maximum event, F_{A_N} its cumulative distribution function, and Q_{A_N} its quantile function. Motivated by the fact that $(1 - 1/N)^N$ converges to $\exp[-1]$ with increasing N (with percentage error of $\approx 1\%$ for $N = 50$ and $\approx 0.1\%$ for $N = 500$), we can write

$$F_{A_N}(q') = \exp[-1], \text{ or} \tag{3}$$

$$q' = Q_{A_N}(\exp[-1]) \tag{4}$$

and solve for q', the quantile of the distribution of the N-year maximum with non-exceedance probability $\exp[-1]$. In the absence of uncertainty, it is clear that with increasing N the estimators q and q' of return value from Equations (1) and (3) are asymptotically equivalent by definition.

The discussion above illustrates how return values might be estimated from the distribution of the annual maximum (Equation (1)) or the N-year maximum (Equation (3)), when we are certain (or have perfect knowledge) about the tail of the distribution of annual maxima. In reality, this is never the case, because we estimate the tail and its parameters from a sample of data. It is natural therefore to seek to understand how this epistemic uncertainty affects estimation for return values, in particular to consider if systematic bias occurs.

3. Return value estimators accommodating uncertain tail parameters

We now assume that the distribution of annual maximum A is known conditional on uncertain extreme value model parameters $\mathbf{Z}$, so that (given $\mathbf{Z}$) we can use any of the approaches in Section 2 to estimate return values. We seek to propagate uncertainty about $\mathbf{Z}$ through to estimates of return values. There are a number of plausible ways to achieve this based on extensions of Equations (1) and (3). These different estimators perform differently given uncertain knowledge of model parameters.

The basic mathematical tools used to incorporate uncertainty are now explained, for some random variable Y (e.g. A or A_N) whose distribution is known conditional on random variables $\mathbf{Z}$ (e.g. uncertain model parameters). If the conditional cumulative distribution function $F_{Y|\mathbf{Z}}$ and the joint density $f_{\mathbf{Z}}$ of $\mathbf{Z}$ are known, the unconditional predictive distribution $\tilde{F}_Y$ can be evaluated using

$$\tilde{F}_Y(y) = \int_{\boldsymbol{\zeta}} F_{Y|\mathbf{Z}}(y|\boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}$$

where $\boldsymbol{\zeta} \in \mathscr{D}_{\mathbf{Z}}$ for range of integration $\mathscr{D}_{\mathbf{Z}}$ is understood throughout. When the joint density $f_{\mathbf{Z}}$ is estimated using Bayesian inference (as joint density $f_{\mathbf{Z}|D}$ from sample D) the unconditional distribution is known as a posterior predictive distribution for Y; alternatively, $f_{\mathbf{Z}}$ might be estimated using a bootstrapping scheme, or some other procedure to generate empirical estimates of the relative probabilities of different values of $\mathbf{Z}$ from data. Analogously, the expected value $E[g(\mathbf{Z})]$ of deterministic function g of uncertain parameters $\mathbf{Z}$ given $f_{\mathbf{Z}}$ is given by

$$E[g(\mathbf{Z})] = \int_{\boldsymbol{\zeta}} g(\boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}.$$

We might consider $g(\mathbf{Z})$ to be the N-year return value as estimated from Equation (1) or Equation (3), a function of uncertain parameters. This motivates considering the following estimators for return values.

3.1. Return value estimators using expected values of parameters: q_1, $q_1{}'$

The first estimator is motivated by the widespread approach of ignoring uncertainty in parameters $\mathbf{Z}$ for estimation of return values. A suitable estimator in this case (see Equation (2)) would be the return value estimated using the expected values of parameters

$$q_1 = Q_{A|\mathbf{Z}}(1 - 1/N \mid E[\mathbf{Z}]) \tag{5}$$

where $E[\mathbf{Z}] = \int_{\boldsymbol{\zeta}} \boldsymbol{\zeta} f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}$, and $Q_{A|Z}(p|\mathbf{Z})$ is the conditional quantile function for the annual maximum evaluated at non-exceedance probability p given model parameters $\mathbf{Z}$. A related estimator q_1' is defined using the distribution of the N-year maximum A_N

$$q_1{}' = Q_{A_N|\mathbf{Z}}(\exp[-1]|E[\mathbf{Z}]).$$

By analogy with Equations (1) and (3), q_1' is asymptotically equivalent to q_1 and will be considered no further.

3.2. Expected quantiles of distribution of A and A_N: q_2, q_2'

The second estimator q_2 is the expected quantile of distribution of A with non-exceedance probability $1 - 1/N$ given by

$$q_2 = E\left[Q_{A|\mathbf{Z}}(1 - 1/N \mid \mathbf{Z})\right] = \int_{\boldsymbol{\zeta}} Q_{A|\mathbf{Z}}(1 - 1/N \mid \boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}. \tag{6}$$

Estimator q_2 involves first solving for the quantile of the distribution of the annual maximum with non-exceedance probability $1 - 1/N$ for each a large number of parameter choices $\boldsymbol{\zeta}$, and then averaging. A related estimator q_2' is the expected quantile of distribution of A_N with non-exceedance probability $\exp[-1]$,

$$q_2' = \int_{\boldsymbol{\zeta}} Q_{A_N|Z}(\exp[-1]|\boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}.$$

That is, q_2' involves first solving for the quantile of the distribution of the N-year maximum with non-exceedance probability $\exp[-1]$ for a large number of parameter choices $\boldsymbol{\zeta}$, and then averaging. Since $Q_{A_N|\mathbf{Z}}(\exp[-1]|\boldsymbol{\zeta})$ is asymptotically equivalent to $Q_{A|\mathbf{Z}}(1 - 1/N|\boldsymbol{\zeta})$, by construction estimator q_2' is asymptotically equivalent to q_2. Hence, q_2' is no longer considered.

3.3. Quantiles of predictive distributions of A and A_N: q_3, q_4

The third estimator is the quantile of predictive distribution of A with non-exceedance probability $1 - 1/N$ defined by

$$q_3 = \tilde{Q}_A(1 - 1/N) \tag{7}$$

where $\tilde{Q}_A$ is the predictive quantile function for A, defined as the inverse of the predictive cumulative distribution function $\tilde{F}_A$ given by

$$\tilde{F}_A(x) = \int_{\boldsymbol{\zeta}} F_{A|\mathbf{Z}}(x \mid \boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}.$$

Estimating q_3 involves first estimating $\tilde{F}_A$, then solving for its $1 - 1/N$ quantile. The related fourth estimator is the quantile of predictive distribution of A_N with non-exceedance probability $\exp[-1]$

$$q_4 = \tilde{Q}_{A_N}(\exp[-1]) \tag{8}$$

where $\tilde{Q}_{A_N}$ is the predictive quantile function for A_N, defined as the inverse of the predictive cumulative distribution function $\tilde{F}_{A_N}$ given by

$$\tilde{F}_{A_N}(x) = \int_{\boldsymbol{\zeta}} F_{A_N|\mathbf{Z}}(x \mid \boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}.$$

Estimating using q_4 involves first estimating $\tilde{F}_{A_N}$, then solving for its $\exp[-1]$ quantile.

In the absence of parameter uncertainty, q_1, q_2 and q_3 are equal to q. Further q_1', q_2' and q_4 are equal to q', which converges to q with increasing N, as discussed in Section 2. That is, when there is no uncertainty in the parameters $\mathbf{Z}$, or when that uncertainty is ignored, all estimators will yield the same return value estimate. In the presence of uncertainty, however, there is no reason to expect equality between estimates using the four estimators q_1, q_2, q_3 and q_4 because they incorporate the effects of parameter uncertainty in different ways.

Numerous questions therefore arise: which estimator is most appropriate to use in metocean design? Is there a universally best choice, or is the choice problem-specific? For typical samples of metocean data, how different are the estimates obtained? How reasonable is it to calculate and quote return values using different estimators in the presence of uncertainty? How should we quantify and communicate the characteristics of extreme environments for metocean design, subject to uncertainty? How should good design decisions be made? Are there approaches to specification of design conditions that are inherently better than return values?

4. Sample estimators for peaks over threshold

Suppose we observe samples of independent threshold exceedances which follow a GP distribution. We now define four sample estimators for return value corresponding to those introduced in Section 3, based on sample estimates of the GP shape and scale parameters.

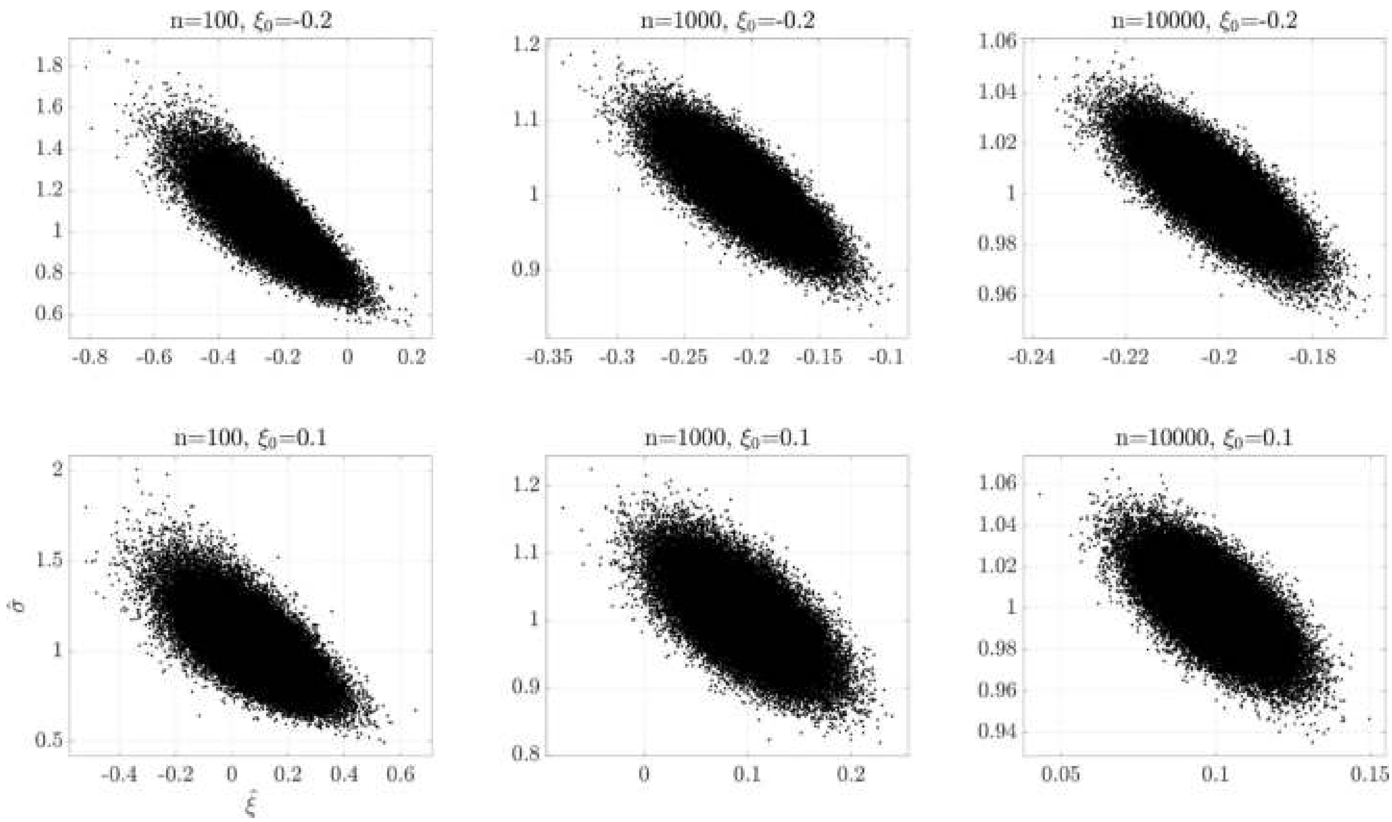

Fig. 1. Scatter plots of 10^5 maximum likelihood GP parameter estimates ξ_i and σ_i for the cases $\xi_0 = -0.2$ (top) and $\xi_0 = 0.1$ (bottom). Columns from left to right correspond to sample sizes n of 100, 1000 and 10,000.

4.1. Generalised Pareto data

Independent observations X of threshold exceedance are assumed to be GP-distributed with shape parameter ξ and scale parameter σ

$$F_{X|X>\psi,Z}(x|\boldsymbol{\zeta}) = 1 - \left(1 + \frac{\xi}{\sigma}(x-\psi)\right)_+^{-1/\xi}, \quad \xi \neq 0 \tag{9}$$

for $x > \psi$, $\psi \in (-\infty, \infty)$, $\xi \in (-\infty, \infty)$, $\sigma \in (0, \infty)$, $\boldsymbol{\zeta} = (\xi, \sigma)$ and $(y)_+ = y$ for $y > 0$ and $= 0$ otherwise. When $\xi = 0$, the conditional distribution takes the form $1 - \exp[-(x-\psi)/\sigma]$. In the absence of parameter uncertainty, the Appendix shows that the value of the N-year return value q solving Equation (1) or (3) for large N is given by

$$q = \frac{\sigma}{\xi}\left(N_*^{\xi} - 1\right) + \psi, \quad \xi \neq 0 \tag{10}$$

where N_* is the expected number of threshold exceedances in N years, and $q = \sigma \log[N_*] + \psi$ when $\xi = 0$.

Equation (10) illustrates that the relationship between q and ξ is non-linear, and that the cases $\xi < 0$ and $\xi \geq 0$ show different behaviours with increasing N_*: for $\xi < 0$, a finite upper limit for q as $N_* \to \infty$ exists given by $\psi + \sigma/(-\xi)$, whereas for $\xi \geq 0$ we see $q \to \infty$ as $N_* \to \infty$. We might expect therefore that uncertain knowledge of ξ influences the value of q differently for different ξ, and that the approach we choose to take to estimate ξ (and σ, ψ in general) and incorporate parameter uncertainty will influence our estimate of q. Section 4.2 below defines four estimators q_1, q_2, q_3 and q_4 of return value subject to uncertainty in ξ and σ, motivated by Section 3. Section 5 then explores differences between estimators q using numerical simulation. It is also possible to demonstrate some differences between return value estimators theoretically, as discussed in Section 6.

4.2. Sample estimators of return values

We have a set $\mathscr{Z} = \{\xi_i, \sigma_i\}_{i=1}^m$ of m independent estimates for the underlying (true known) parameters ξ_0, σ_0 of a GP distribution $F_{X|X>\psi,Z}$ (see Equation (9)), with fixed known extreme value threshold ψ corresponding to non-exceedance probability τ. Without loss of generality, we assume that the pairs of elements of $\mathscr{Z}$ are ordered in terms of decreasing ξ, so that ξ_1 is the largest value of ξ. $\mathscr{Z}$ might be output of a Markov chain Monte Carlo inference from some sample, a set of maximum likelihood parameter estimates from m different bootstrap resamples, or the judgements of m experts. We use $\mathscr{Z}$ to construct return value estimators q_1, q_2, q_3 and q_4 as follows. First the Appendix provides expressions for F_A and F_{A_N} (from Section 2) and solves Equation (2), for $F_{X|X>\psi,Z}$ taking GP form; it also defines the expected number N_* of threshold exceedances in N years, and the expected number λ of events per annum. Then, replacing integration with respect to $f_Z(\boldsymbol{\zeta})d\boldsymbol{\zeta}$ with summation over the set $\mathscr{Z}$, Equations (5) and (10) yield

$$q_1 = \frac{\sum_i \sigma_i}{\sum_i \xi_i}\left(N_*^{\sum_i \xi_i/m} - 1\right) + \psi \tag{11}$$

where summation is for $i = 1, 2, \ldots, m$. That is, q_1 is formed using the mean values of uncertain GP parameter estimates in the standard equation for a return value. To estimate q_2, Equations (6) and (10) yield

$$q_2 = \frac{1}{m}\sum_i \frac{\sigma_i}{\xi_i}\left(N_*^{\xi_i} - 1\right) + \psi. \tag{12}$$

That is, q_2 corresponds to the mean of m estimates, each corresponding to a different pair from $\mathscr{Z}$. To define q_3, we write Equation (7) as

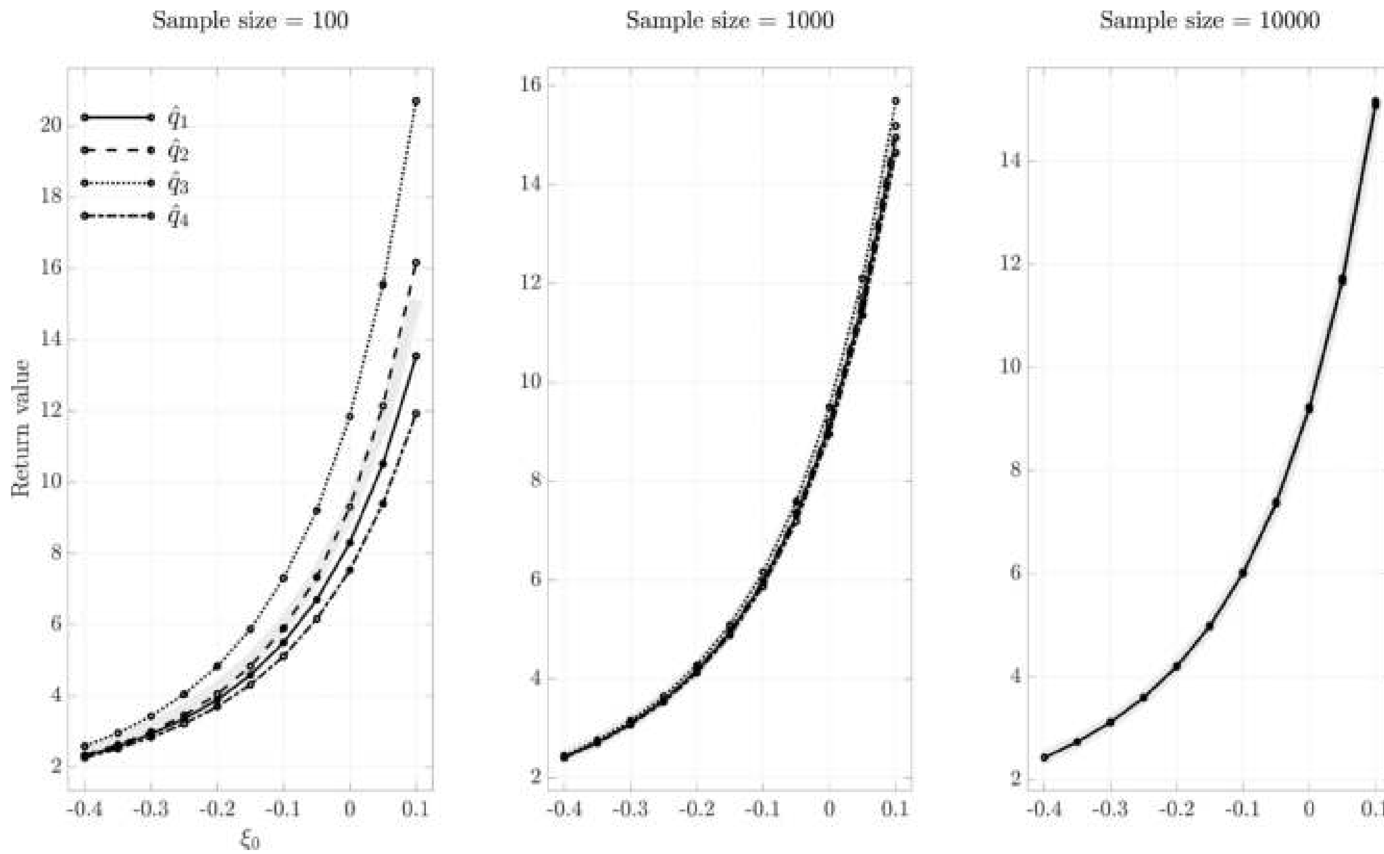

Fig. 2. Return value estimates $\widehat{q}_1$ (solid), $\widehat{q}_2$ (dashed), $\widehat{q}_3$ (dotted), $\widehat{q}_4$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a generalised Pareto distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Return values estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right). Thick grey lines indicate true return value behaviour.

$$\tilde{F}_A(x) = \frac{1}{m}\sum_i \exp\left[-\lambda(1-\tau)\left(1+\frac{\xi_i}{\sigma_i}(x-\psi)\right)^{-1/\xi_i}\right] \tag{13}$$

for $\tilde{F}_A$; that is, we take the mean over m different estimates for the annual distribution (each corresponding to a different pair from $\mathscr{Z}$) to form $\tilde{F}_A(x)$ for each x. Then we set $\tilde{F}_A(q_3) = 1 - 1/N$ such that

$$1 - \frac{1}{N} = \frac{1}{m}\sum_i \exp\left[-\lambda(1-\tau)\left(1+\frac{\xi_i}{\sigma_i}(q_3-\psi)\right)^{-1/\xi_i}\right]. \tag{14}$$

Finally, we estimate $\tilde{F}_{A_N}$ as mean of m different estimates for the distribution of the N-year maximum using GP parameters taken from $\mathscr{Z}$

$$\tilde{F}_{A_N}(q_4) = \frac{1}{m}\sum_i \exp\left[-N_*\left(1+\frac{\xi_i}{\sigma_i}(q_4-\psi)\right)^{-1/\xi_i}\right].$$

Equation (8) then gives q_4 as quantile of the resulting distribution with non-exceedance probability $\exp[-1]$

$$\exp[-1] = \frac{1}{m}\sum_i \exp\left[-N_*\left(1+\frac{\xi_i}{\sigma_i}(q_4-\psi)\right)^{-1/\xi_i}\right]. \tag{15}$$

5. Differences between return value estimators explored by simulation and maximum likelihood estimation

We use simulation to explore the characteristics of the four return value estimators q_1, q_2, q_3 and q_4. We assume that observations of threshold exceedances follow a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$ (and threshold non-exceedance probability $\tau = 0$) for sample sizes are $n = 100$, 1000 and 10,000. We estimate return values q using maximum likelihood estimates for ξ and σ from GP fits to $m = 10^5$ realisations of samples, for true ξ_0 values of -0.4, -0.35, …, 0.1. This selection of values of ξ_0 corresponds to those typically found in application to storm peak significant wave height from different ocean basins; for example, in the northern North Sea, values of $\xi \in [-0.4, -0.1]$ have been reported (e.g. Elsinghorst et al., 1998; Ewans and Jonathan, 2008; Randell et al., 2016); in the South China Sea, values of $\xi \in [-0.1, +0.1]$ (e.g. Randell et al., 2015); and in the Gulf of Mexico, $\xi \in [-0.1, +0.05]$ (e.g. Raghupathi et al., 2016).

Fig. 1 gives scatter plots of the set $\mathscr{Z} = \{\xi_i, \sigma_i\}$ of maximum likelihood parameter estimates ξ_i, σ_i for the cases $\xi_0 = -0.2$ and $\xi_0 = 0.1$. For larger sample sizes, the characteristics of the sample resemble those of the known Gaussian asymptotic form (e.g. Hosking and Wallis, 1987), with ξ_i and σ_i approximately unbiased, but negatively correlated. For $n = 100$, the scatter plots appear non-Gaussian, with more occurrences of large negative values of $\xi_i - \xi_0$ than of large positive ones. We might anticipate that the properties of return value estimators q relative to q_0 depend on the characteristics of sample $\mathscr{Z}$, and hence on the method of parameter estimation.

Differences between return value estimators and the known population return values are assessed as a function of ξ_0 in Figs. 2–5 below. For the purposes of return value calculations, we assume that the rate of occurrence $\lambda = 100$ per annum. First we consider return period $N = 100$ years. Fig. 2 illustrates the variation of true return value q_0 as a function of ξ_0, for the three sample sizes under consideration, as a thick grey line. Differences in estimates $\widehat{q}$ for return values q are greatest for sample size $n = 100$; for larger sample sizes the differences between return value estimators is relatively small relative to the range of return values.

Fig. 3 provides a better illustration of the differences between estimators q in terms of fractional bias

$$\text{Fractional bias} = \frac{q_j}{q_0} - 1 \qquad \text{for } j = 1, 2, 3, 4.$$

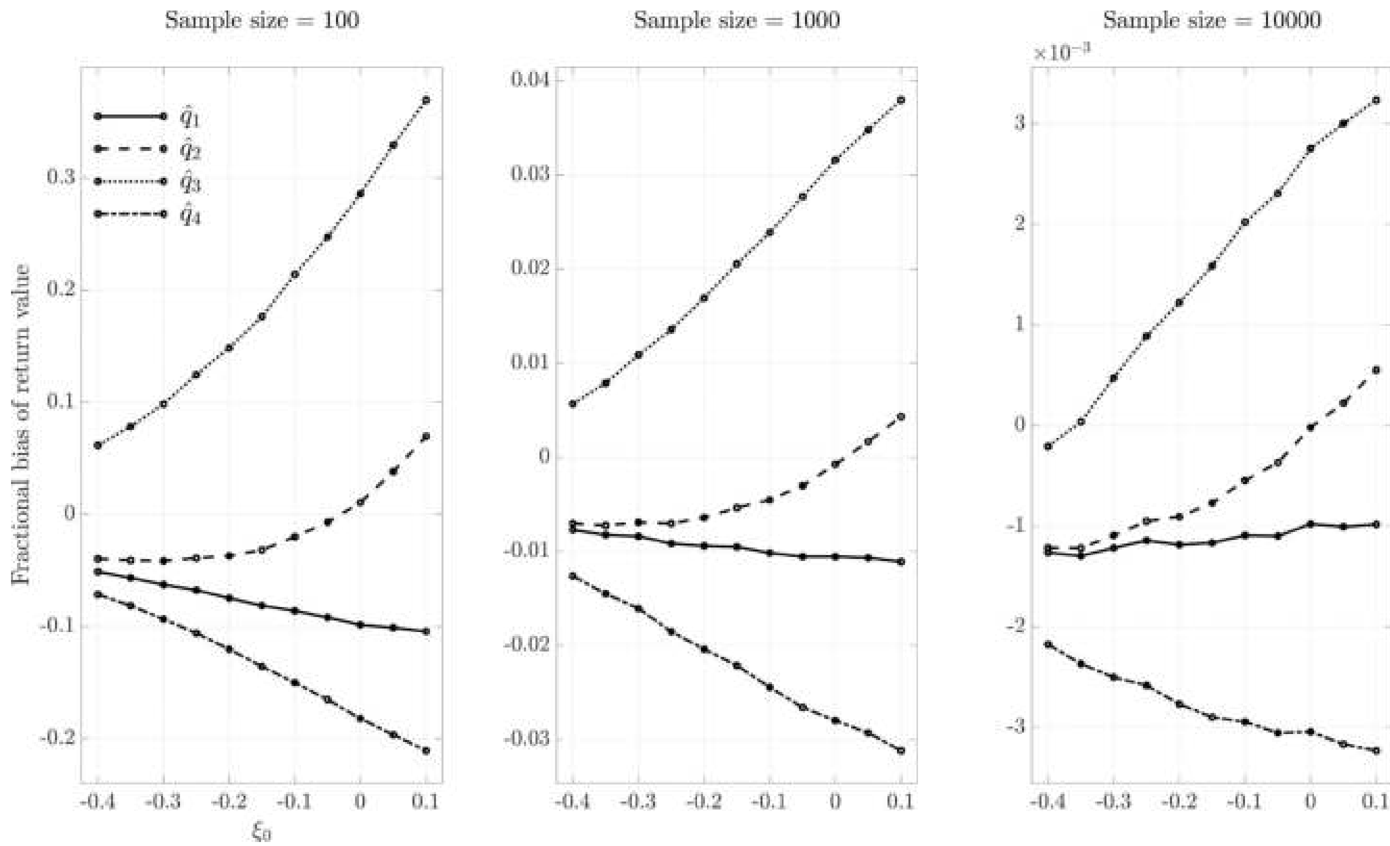

Fig. 3. Fractional bias for return value estimates $\widehat{q}_1$ (solid), $\widehat{q}_2$ (dashed), $\widehat{q}_3$ (dotted), $\widehat{q}_4$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right). Unbiased estimates would have fractional bias of zero.

Independent of sample size n, estimator q_2 provides the smallest fractional bias. Recall, in the notation of Section 3, that q_2 takes the form $E[Q_{A|Z}(1 - 1/N \mid \boldsymbol{Z})]$; that is, q_2 is the average of m return value estimates, each corresponding to a pair of maximum likelihood parameter estimates. Return value estimator q_1 provides consistent underestimation of return value. Recall that q_1 is obtained by first estimating the mean of maximum likelihood parameters, then plugging these into the return value equation; q_1 takes the form $Q_{A|Z}(1 - 1/N \mid E[\boldsymbol{Z}])$ in the notation of Section 3. q_1 can therefore be viewed as the return value estimated when we ignore variability in maximum likelihood parameter estimates, and simply use the average values of ξ and σ for return value calculation. Return value estimator q_4 provides the largest negative bias, underestimating the return value. Referring to Section 3, q_4 corresponds to the quantile of the predictive distribution for the N-year maximum A_N with non-exceedance probability $\exp[-1]$. Return value estimator q_3 provides the largest positive bias, consistently overestimating the return value. Referring to Section 3, q_3 corresponds to the quantile of the predictive distribution for the annual maximum A with non-exceedance probability $1 - 1/N$.

Fig. 3 shows that the magnitude of fractional bias increases with ξ_0 for all return value estimators except q_2, and for all sample sizes n, for the interval $\xi_0 \in [-0.4, 0.1]$ (but see theoretical results in Section 6 which show that $q_2 > q_3$ in the very unlikely case that $\xi_0 > 1$). The figure also shows that, as n increases, fractional bias for all q reduces suggesting that all four approaches provide consistent estimation. However, for finite sample sizes, the bias in q_3 in particular is very large, corresponding to around 20% for $\xi_0 \approx -0.1$. This is somewhat alarming given that q_3 would be the default choice of many with a statistical background, defined in terms of the predictive distribution for the annual maximum. q_2 clearly provides the least biased results, with fractional bias $\leq 5\%$ for $\xi_0 \in [-0.4, 0.05)$, increasing for largest ξ_0.

Fig. 4 assesses q_1, q_2, q_3 and q_4 in terms of the bias in their exceedance probabilities

$$\text{Bias} = \Pr(A > q_j) - \frac{1}{N} = 1 - F_A(q_j) - \frac{1}{N} \qquad \text{for } j = 1, 2, 3, 4.$$

Estimator q_3 based on the annual predictive distribution provides the overall best performance for the smallest sample size $n = 100$, certainly for $\xi_0 \in [-0.4, -0.1]$. Unsurprisingly given the evidence of Fig. 3, q_3 nevertheless underestimates exceedance probability for all n. In fact, closer inspection of results shows that $\Pr(A > q_3) = 0$ for $n = 100$ and $\xi_0 < -0.2$, as discussed further in relation to Fig. 5 below. That is, estimator q_3 lies beyond the upper end point of the distribution of the annual maximum A in these cases. Estimator q_4 based on the predictive distribution for the N-year maximum provides the worst performance, and large overestimation of exceedance probability. For $n = 1,000$, q_2 provides best performance for $\xi_0 \geq -0.3$. For $n = 10,000$, q_3 provides best performance for $\xi_0 \leq -0.3$, and q_2 for $\xi_0 \geq -0.2$.

A third useful metric for comparison of return value estimators is bias in log exceedance probabilities

$$\text{Bias} = \log_{10}(\Pr(A > q_j)) - \log_{10}\left(\frac{1}{N}\right) = \log_{10}(1 - F_A(q_j)) + \log_{10}(N) \qquad \text{for } j = 1, 2, 3, 4.$$

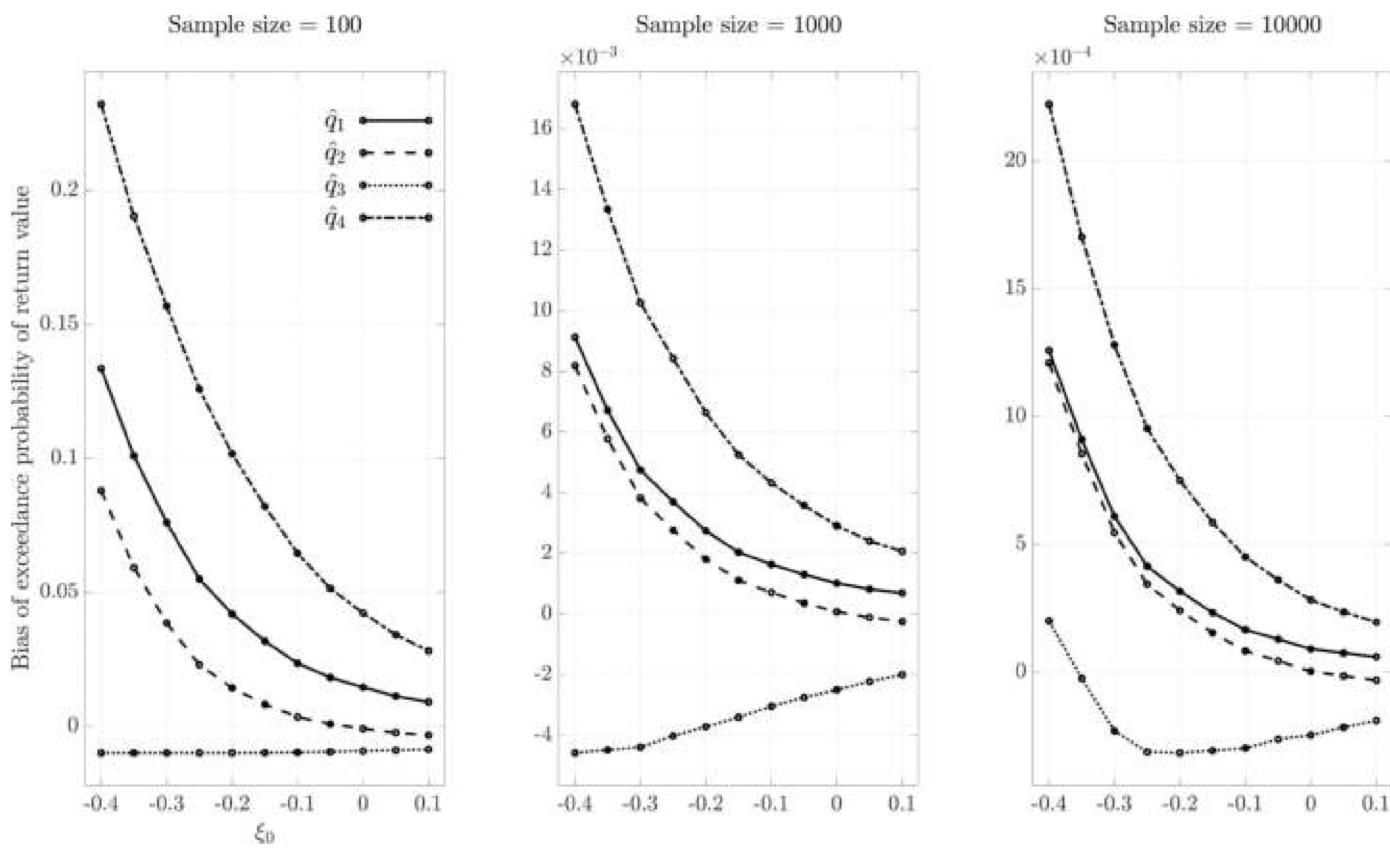

Fig. 4. Bias in exceedance probabilities corresponding to return value estimates $\widehat{q}_1$ (solid), $\widehat{q}_2$ (dashed), $\widehat{q}_3$ (dotted), $\widehat{q}_4$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right).

This scale of comparison is useful since it is often more intuitive to assess errors in probabilities multiplicatively rather than additively. Results for $N = 100$ years are shown in Fig. 5. The predictive estimator q_3 performs poorly for sample size $n = 100$, with infinite bias in log exceedance probability for $\xi_0 < -0.2$. For $n = 10,000$, q_3 performs better. Overall, q_2 provides best performance.

Fig. 6 considers return value estimator behaviour for return period $N = 1,000$ years. The general features of Figs. 6 and 3 (for $N = 100$) are very similar: the magnitude of fractional bias for $N = 1,000$ is greater than for $N = 100$ since we are extrapolating farther. Fig. 7 considers the corresponding result for $N = 10,000$ years. We infer that return period (within the range considered) has little effect on the relative performance of return value estimators.

Fig. 8 shows the bias in estimated exceedance probability for the case $N = 10,000$ years, for comparison with the results in Fig. 4 (for $N = 100$). q_3 provides the least biased estimate of exceedance probability for sample size $n = 100$. For larger n, $\Pr(A > q_3) < 1/N$ for the interval of ξ_0 considered. Conversely, all other estimators q produce overestimates of the exceedance probability. For $n \geq 1,000$ and $\xi_0 \geq -0.2$ there is little to choose between the estimators.

In terms of bias in log exceedance probability, Fig. 9 once again illustrates the relatively poor performance of the predictive estimator q_3, which yields infinite bias in a large proportion of the cases considered for return period $N = 10,000$ years since the estimate for q_3 is beyond the upper end point of the distribution of the annual maximum A. Overall, q_2 provides best performance.

Motivation for differences between return value estimators q and q_0 is provided by Fig. 10, which shows different annual maximum and N-year maximum cumulative distribution functions, and differences of cumulative distribution functions, for the case $\xi_0 = -0.2$, sample size $n = 1,000$ and return period $N = 100$ years.

Comparison of the true N-year distribution F_{A_N} and its predictive estimate $\tilde{F}_{A_N}$ in the left hand panel suggests that the latter is more spread; this is reasonable given evidence of parameter estimation variation in Fig. 1. Consequently we expect predictive distributions $\tilde{F}_A$ and $\tilde{F}_{A_N}$ to have more mass in their tails compared to the true distributions F_A and F_{A_N}: this is shown to be the case in the top right hand panel. With increasing x, differences $\tilde{F}_A(x) - F_A(x)$ and $\tilde{F}_{A_N}(x) - F_{A_N}(x)$ start at zero, are then positive and then negative and ultimately zero again in both cases, except that the intervals corresponding to positive and negative differences is not the same. Clearly the locations of positive and negative differences are determined by the characteristics of set $\mathscr{Z}$. We are specifically interested in the value of the difference at the (true) N-year return value, indicated by a vertical line in the top right panel. Here, $\tilde{F}_{A_N}$ overestimates the true distribution, and hence a return value estimator based on $\tilde{F}_{A_N}$ will be biased low (and therefore $q_4 < q_0$). The opposite is true for $\tilde{F}_A$ (and therefore $q_3 > q_0$).

Referring to Equation (15), taking the N^{th} root of each side, and applying the approximation $\exp[-1/N] \approx 1 - 1/N$ for large N suggests that solving $1 - 1/N = \tilde{F}_{A_N}^{1/N}(q_4)$ provides a good approximation to q_4. The top right panel illustrates that the difference $\tilde{F}_{A_N}^{1/N} - \tilde{F}_A$ is never negative; this suggests that in particular $q_3 \geq q_4$. This ordering is clear from inspection of the bottom right panel, where the $1 - 1/N = 0.99$ level is shown as a horizontal line.

6. Differences between return value estimators demonstrated theoretically

We show theoretically that, in certain situations, there are systematic

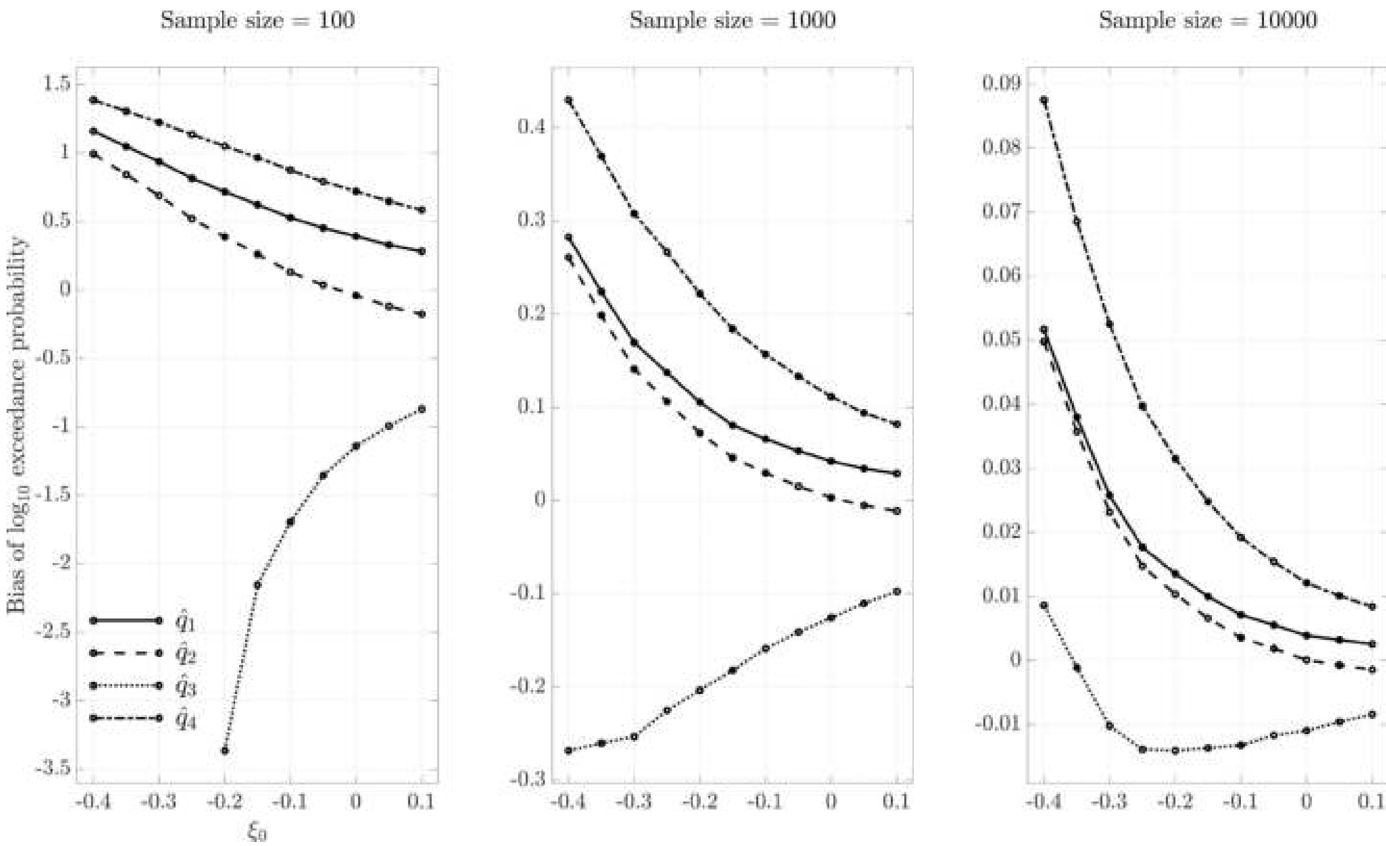

Fig. 5. Bias in logarithm (base 10) of exceedance probabilities corresponding to return value estimates $\hat{q}_1$ (solid), $\hat{q}_2$ (dashed), $\hat{q}_3$ (dotted), $\hat{q}_4$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right). For $\xi_0 < -0.2$, $\Pr(A > \hat{q}_3)$ is estimated to be zero.

differences between different point estimators of return values. We first present a tabular summary (see Table 1) of differences between return value estimators q_1, q_2, q_3 and q_4 and the true known value q_0 given certain assumptions about the set $\mathscr{Z} = \{\xi_i, \sigma_i\}$ of GP parameter estimates, and the true parameter values ξ_0, σ_0, with $\psi_0 = 0$ assumed, and sufficiently large N_*. We then provide outline proofs of the inequalities. These inequalities are not specific to maximum likelihood estimates for GP parameter estimates.

Inequality I1 occurs since x^N $(x \geq 0, N = 1, 2, \ldots)$ is a convex function. Jensen's inequality states that for m weights $\{\omega_i\}_{i=1}^m$ such that $\omega_i \geq 0$ and $\Sigma_i \omega_i = 1$, and a convex function $\phi(x)$ on some domain (such that $\partial^2\phi/\partial x^2 \geq 0$), then $\phi(\Sigma_i \omega_i x_i) \leq \Sigma_i \omega_i \phi(x_i)$. Setting $\phi(x) = x^N$, $\omega_i = 1/m$ and $x_i = \exp[-\lambda(1-\tau)(1+(\xi_i/\sigma_i)(x-\psi))^{-1/\xi_i}]$, Jensen's inequality gives

$$\left(\frac{1}{m}\sum_i \exp\left[-\lambda(1-\tau)(1+(\xi_i/\sigma_i)(x-\psi))^{-1/\xi_i}\right]\right)^N \leq \frac{1}{m}\sum_i \exp\left[-N\lambda(1-\tau)(1+(\xi_i/\sigma_i)(x-\psi))^{-1/\xi_i}\right] \text{ or}$$

$$\tilde{F}_A^N(x) \leq \tilde{F}_{A_N}(x)$$

for all x, referring to Equations (14) and (15). Therefore, with increasing x, $\tilde{F}_{A_N}(x)$ will rise to a value of $\exp[-1]$ (at q_4) before $\tilde{F}_A^N(x)$ does (at q_3). For large N, the latter is equivalent to $\tilde{F}_A(x)$ reaching a value of $\exp[-1/N] \approx (1 - 1/N)$. We deduce that $q_3 \geq q_4$ for all choices of $\mathscr{Z}$. Inspection of the top right panel of Fig. 10 illustrates this behaviour for specific conditions. Inspection of Figs. 3, 5 and 6 shows that this inequality holds in our simulation study.

Inequalities I2 and I3 occur when the maximum value ξ_1 of ξ from $\mathscr{Z}$ is greater than the maximum of the true underlying value ξ_0, and zero. Recall we assume pairs of elements in $\mathscr{Z}$ to be ordered in terms of decreasing ξ, so that ξ_1 is the largest value of ξ in $\mathscr{Z}$. $\xi_1 > \max(\xi_0, 0)$ is a relatively likely scenario in practice for a variable such as storm peak H_S. Fundamental physical understanding suggests a finite upper value to storm peak H_S (and hence $\xi < 0$); there is considerable empirical evidence to support this also. Nevertheless, for a Bayesian inference we might specify a prior for ξ which includes 0 so as not to be too restrictive; estimates of $\xi > 0$ would therefore be expected. For maximum likelihood inference with small samples, estimates for GP shape parameter exceeding zero are typical even for storm peak H_S. For other oceanographic variables such as wind speed, $\xi > 0$ is typically expected.

To motivate I2, Equations (12) and (10) give

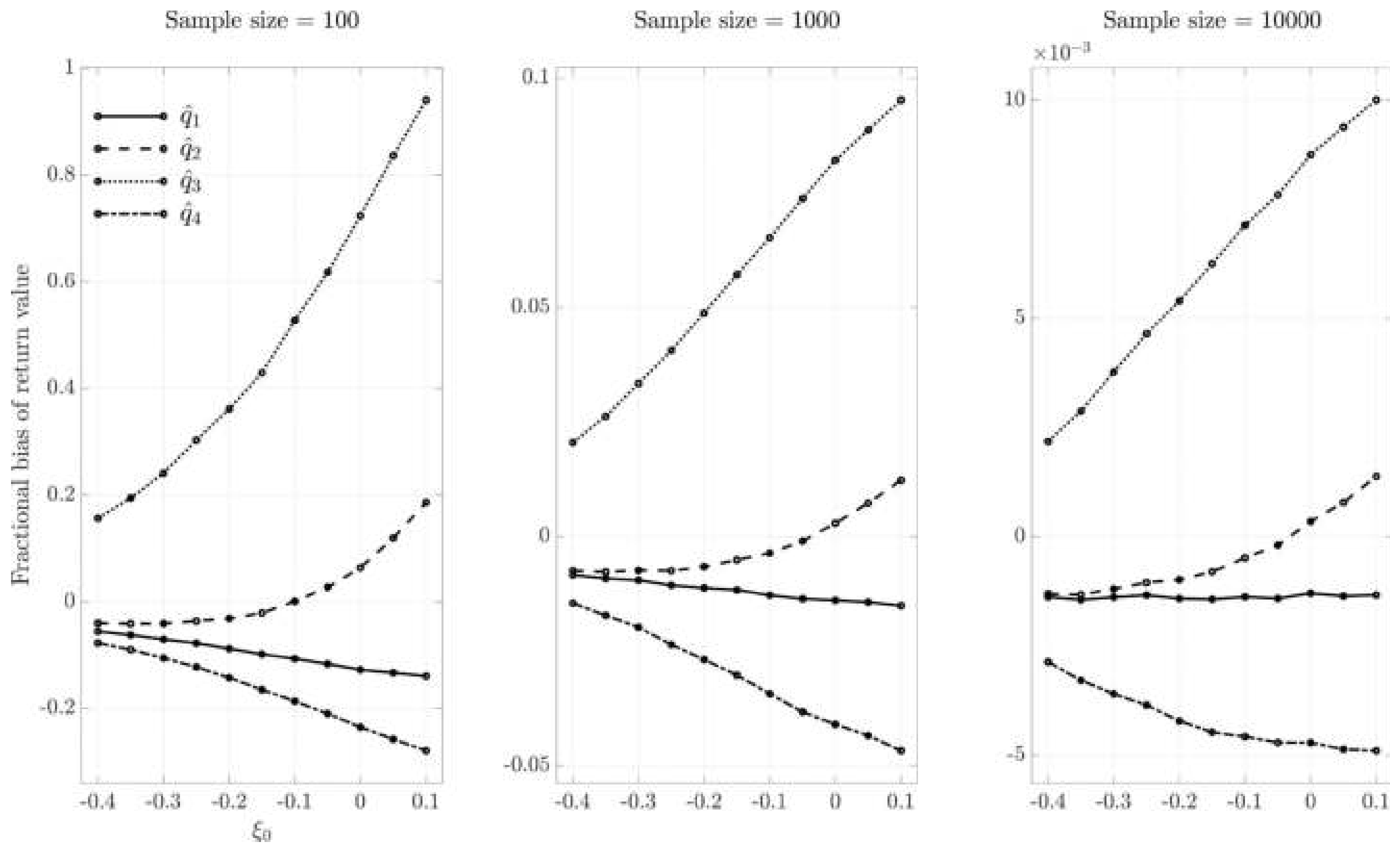

Fig. 6. Fractional bias for return value estimates $5\hat{q}_1$ (solid), $\hat{q}_2$ (dashed), $\hat{q}_3$ (dotted), $\hat{q}_4$ (dot-dashed) for a return period of 1000 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right).

$$\frac{q_2}{q_0} = \frac{\frac{1}{m}\sum_i (\sigma_i/\xi_i)\left[N_*^{\xi_i} - 1\right]}{(\sigma_0/\xi_0)\left[N_*^{\xi_0} - 1\right]}$$

for $\xi \neq 0$ and the analogous expression when $\xi = 0$. For $\xi_1 > \max(\xi_0, 0)$, we can see that

$$\frac{q_2}{q_0} \approx \begin{cases} m^{-1}[(\sigma_1\xi_0)/(\sigma_0\xi_1)]N_*^{\xi_1-\xi_0}, & \xi_0 > 0 \\ m^{-1}[\sigma_1/(\sigma_0\xi_1)]N_*^{\xi_1}/\log N_*, & \xi_0 = 0 \\ m^{-1}[(-\sigma_1\xi_0)/(\sigma_0\xi_1)]N_*^{\xi_1}, & \xi_0 < 0, \end{cases}$$

all of which diverge to infinity with increasing N_*. That is, q_2 always overestimates the true return value for large N_* when the largest value of ξ in $\mathscr{Z}$ exceeds the maximum of the true value ξ_0, and 0. Inspection of Figs. 3, 6 and 7 illustrates this for $N = 1,000$ and $N = 10,000$.

To demonstrate I3, we first approximate Equation (14) of Section 4.2 when q_3 is large using the approximation $\exp[x] \approx 1 + x$ for small x, so that

$$1 - \frac{1}{N} \approx \frac{1}{m}\sum_i \left(1 - \lambda(1-\tau)\left(1 + \frac{\xi_i}{\sigma_i}(q_3 - \psi)\right)^{-1/\xi_i}\right)$$
$$= 1 - \frac{1}{m}\sum_i \left(\lambda(1-\tau)\left(1 + \frac{\xi_i}{\sigma_i}(q_3 - \psi)\right)^{-1/\xi_i}\right)$$

from which we obtain

$$\frac{1}{N_*} \approx \frac{1}{m}\sum_i \left(1 + \frac{\xi_i}{\sigma_i}(q_3 - \psi)\right)^{-1/\xi_i}.$$

Substituting this expression for N_* into Equation (12) gives

$$q_2 = \frac{1}{m}\sum_{i=1}^{m}\frac{\sigma_i}{\xi_i}\left[\left\{\frac{1}{m}\sum_{j=1}^{m}\left[1 + \frac{\xi_j}{\sigma_j}q_3\right]^{-1/\xi_j}\right\}^{-\xi_i} - 1\right].$$

With $\xi_1 > \max(\xi_0, 0)$, the predictive distribution $\tilde{F}_A$ has an infinite upper end point, and hence $q_3 \to \infty$ as $N_* \to \infty$. Hence we can expand the above expression to give

$$q_2 = \frac{1}{m}\left[\frac{\sigma_1}{\xi_1}m^{\xi_1}\left[1 + \frac{\xi_1}{\sigma_1}q_3\right]\left(1 + \sum_{j=2}^{m}\left[1 + \frac{\xi_j}{\sigma_j}q_3\right]^{-1/\xi_j}\left[1 + \frac{\xi_1}{\sigma_1}q_3\right]^{1/\xi_1}\right)^{-\xi_1}\right]$$
$$+ \left[\sum_{i=2}^{m}\frac{\sigma_i}{\xi_i}m^{\xi_i}\left[1 + \frac{\xi_1}{\sigma_1}q_3\right]^{\xi_i/\xi_1}\left(1 + \sum_{j=2}^{m}\left[1 + \frac{\xi_j}{\sigma_j}q_3\right]^{-1/\xi_j}\left[1 + \frac{\xi_1}{\sigma_1}q_3\right]^{1/\xi_1}\right)^{-\xi_i}\right]$$
$$- \frac{1}{m}\sum_{i=1}^{m}\frac{\sigma_i}{\xi_i} \approx m^{\xi_1 - 1}q_3 \text{ for large } N_*.$$

Hence as $N_* \to \infty$, $q_2, q_3 \to \infty$ and $q_2/q_3 \to m^{\xi_1 - 1}$. For $\xi_1 < 1$, highly likely for environmental variables like storm peak H_S, we conclude that $q_3 > q_2$. That is, q_3 exceeds q_2 for large N_* when $1 > \xi_1 > \max(\xi_0, 0)$. Combining this result with that from paragraph above, we deduce that $q_3 > q_2 > q_0$ for large N_* when $1 > \xi_1 > \max(\xi_0, 0)$. Figs. 3, 6 and 7 illustrate this ordering. For $\xi_1 = 1$ and $\xi_1 > \max(\xi_0, 0)$, we observe that q_2 and q_3 are asymptotically equivalent. For $\xi_1 > 1$, we have $q_2 > q_3$ for large N_* and $\xi_1 > \max(\xi_0, 0)$.

Unfortunately useful approximation of Equation (15) for large q_4 and N_*, mimicking the approximation of Equation (14) for large q_3 and large N here, is not possible, since the arguments of the exponent summands in Equation (15) do not become small with increasing q_4 and N_*.

We can also demonstrate orderings of return value estimators when $\xi_1 < \max(\xi_0, 0)$ to support the simulation study results. Inequalities I4, I5 and I6 occur when $\xi_0 < 0$, typically of relevance for extremes of storm

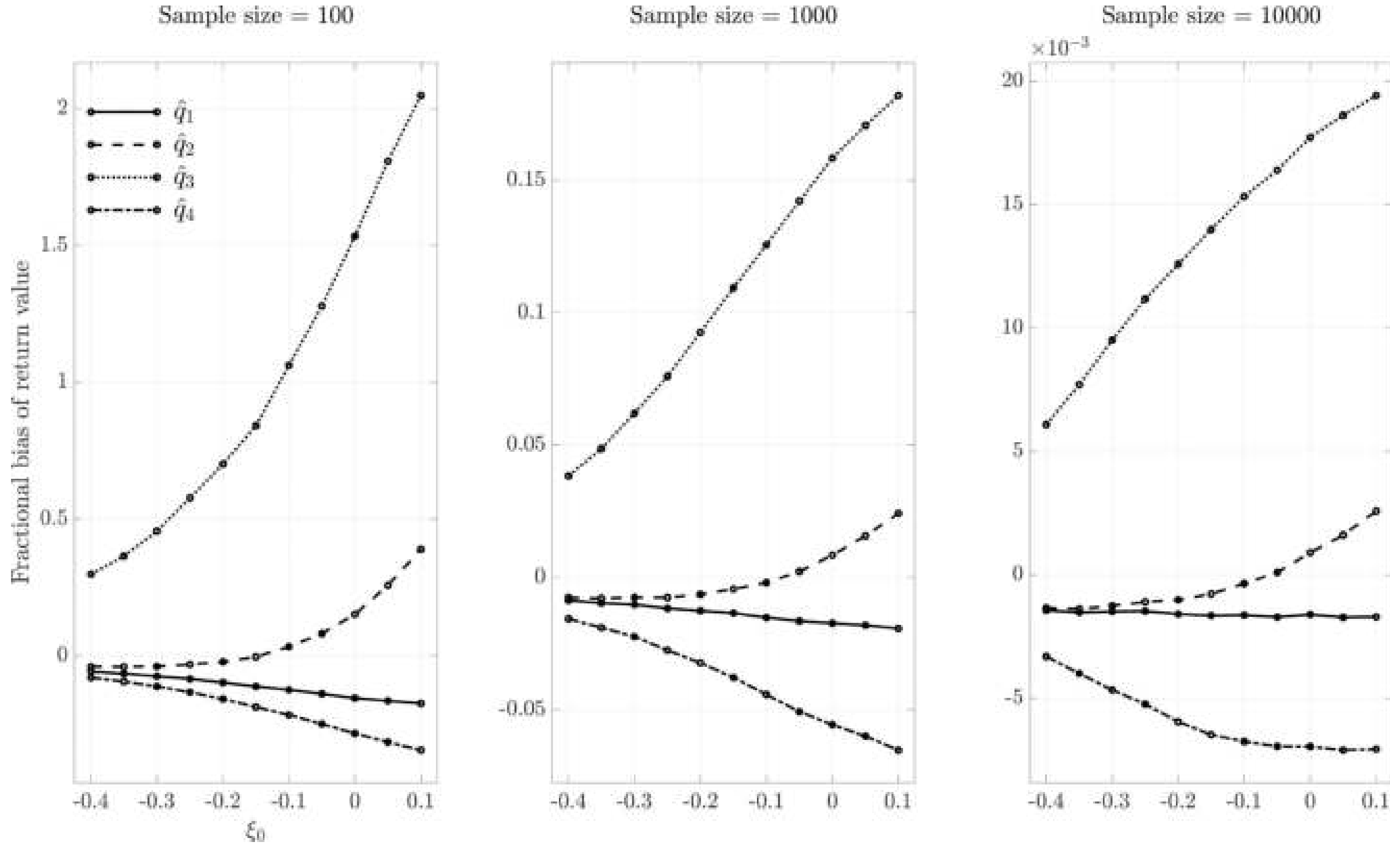

Fig. 7. Fractional bias for return value estimates $\hat{q}_1$ (solid), $\hat{q}_2$ (dashed), $\hat{q}_3$ (dotted), $\hat{q}_4$ (dot-dashed) for a return period of 10,000 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right).

peak H_S. In this situation, a true upper end point $\psi + \sigma_0/(-\xi_0)$ exists. We can restrict discussion to $\xi_1 < 0$, since the case $\xi_1 > 0$ has already been considered in I2 and I3. The relative ordering of q_0, q_1, q_2 and q_3 can be shown to be related to the relative sizes of different estimates for ratios of σ/ξ from $\mathscr{Z}$ as outlined below, relative to q_0.

To motivate I4, we see from Equation (11) that the maximum value for q_1 (at $N_* = \infty$) is $\psi + (1/m\sum_i \sigma_i)/(1/m\sum_i(-\xi_i))$. Therefore, if $(1/m\sum_i \sigma_i)/(1/m\sum_i(-\xi_i)) > \sigma_0/(-\xi_0)$, the maximum value of q_1 will exceed the true upper end point of the distribution F_A; q_1 must show positive bias in this case for sufficiently large N_*. More generally, the bias in q_1 for large N_* is determined by the relative sizes of $(1/m\sum_i \sigma_i)/(1/m\sum_i(-\xi_i))$ and $\sigma_0/(-\xi_0)$, and is not known prior to analysis. For I5, a similar argument, using Equation (12), shows that if $(1/m)\sum_i(\sigma_i/(-\xi_i)) > \sigma_0/(-\xi_0)$, the maximum value of q_2 will exceed the true upper end point of the distribution F_A; q_2 must show positive bias in this case for sufficiently large N_*. More generally, the bias in q_2 for large N_* is determined by the relative sizes of $(1/m)\sum_i(\sigma_i/(-\xi_i))$ and $\sigma_0/(-\xi_0)$, and is not known prior to analysis. For I6, referring to Equation (13) for $\xi_0 < 0$ and $\xi_1 \in (\xi_0, 0)$, suppose that the pair $(\xi_{k^*}, \sigma_{k^*})$ provides the maximum value of $\sigma/(-\xi)$ in $\mathscr{Z}$. Then $\tilde{F}_A$ could have upper end point $\psi + \sigma_{k^*}/(-\xi_{k^*}) > \psi + \sigma_0/(-\xi_0)$. In this situation, q_3 would exhibit positive bias for sufficiently large N_*. More generally, the bias in q_3 for large N_* is determined by the relative sizes of $\sigma_{k^*}/(-\xi_{k^*})$ and $\sigma_0/(-\xi_0)$, and is not known prior to analysis.

For illustration, in the simulation study in Section 5, the values of $\sum_i \sigma_i/\sum_i(-\xi_i)$, $\sum_i(\sigma_i/(-\xi_i))/m$ and $\sigma_{k^*}/(-\xi_{k^*})$ produce an increasing sequence for every choice of n and ξ_0 considered, when $\xi_1 = \max_{k\in(1,2,\ldots,m)}(\xi_k) < 0$; that is, we expect $q_1 < q_2 < q_3$. Further, for small ξ_0 the value of $\sigma_0/(-\xi_0)$ falls between the second and third terms in the sequence, indicating $q_1 < q_2 < q_0 < q_3$ as observed in the simulation results. For large sample size $n = 10,000$ (for which $\xi_1 < 0$) and $\xi_0 \geq -0.15$, the value of $\sigma_0/(-\xi_0)$ falls between the first and second terms so that $q_1 < q_0 < q_2 < q_3$ for sufficiently large N. In fact, for simulations of the $N = 10^8$ year return value, we observe that the function $q_2 - q_0$ (of ξ_0) crosses zero for $n = 10,000$ at approximately $\xi_0 = -0.13$.

7. Discussion and conclusions

This paper discusses the estimation of return values in the presence of uncertainty in extreme value model parameters. We show that different estimators for return value, which yield identical estimates when parameter uncertainty is ignored, yield different values when uncertainty is taken into account. Given uncertain shape and scale parameters $\mathbf{Z}$ of a GP distribution, four sample estimators for the N-year return value are considered: q_1 ($= Q_{A|\mathbf{Z}}(1 - 1/N \mid E[\mathbf{Z}])$, where $Q_{A|\mathbf{Z}}$ is the quantile of the annual maximum event A given parameters $\mathbf{Z}$), q_2 ($= E(Q_{A|\mathbf{Z}}(1 - 1/N \mid \mathbf{Z}))$, q_3 ($= \tilde{Q}_A(1 - 1/N)$, where $\tilde{Q}$ is a predictive quantile) and q_4 ($= \tilde{Q}_{A_N}(\exp[-1])$). We show by simulation and theory that, for given circumstances, the order of estimators q and true value q_0 can be predicted, and that differences between estimators q and q_0 can be very large. The ordering $q_4 \leq q_1 \leq q_2 \leq q_3$ is observed in our simulation study, when $\mathbf{Z}$ corresponds to maximum likelihood estimates for a sample of peaks over threshold of sizes $n = 100$, 1000 and 10,000 for true parameters $\xi_0 \in [-0.4, 0.1]$ and $\sigma_0 = 1$. The true value q_0 lies between q_1 and q_3.

In general for maximum likelihood estimation and values of parameters typical for extreme value modelling of storm peak significant wave height, estimator q_2 yields lowest bias, and the predictive estimator q_3 is obviously the worst performer in terms of bias. In terms of exceedance probability $\Pr(A > q)$, q_3 is in general a relatively good

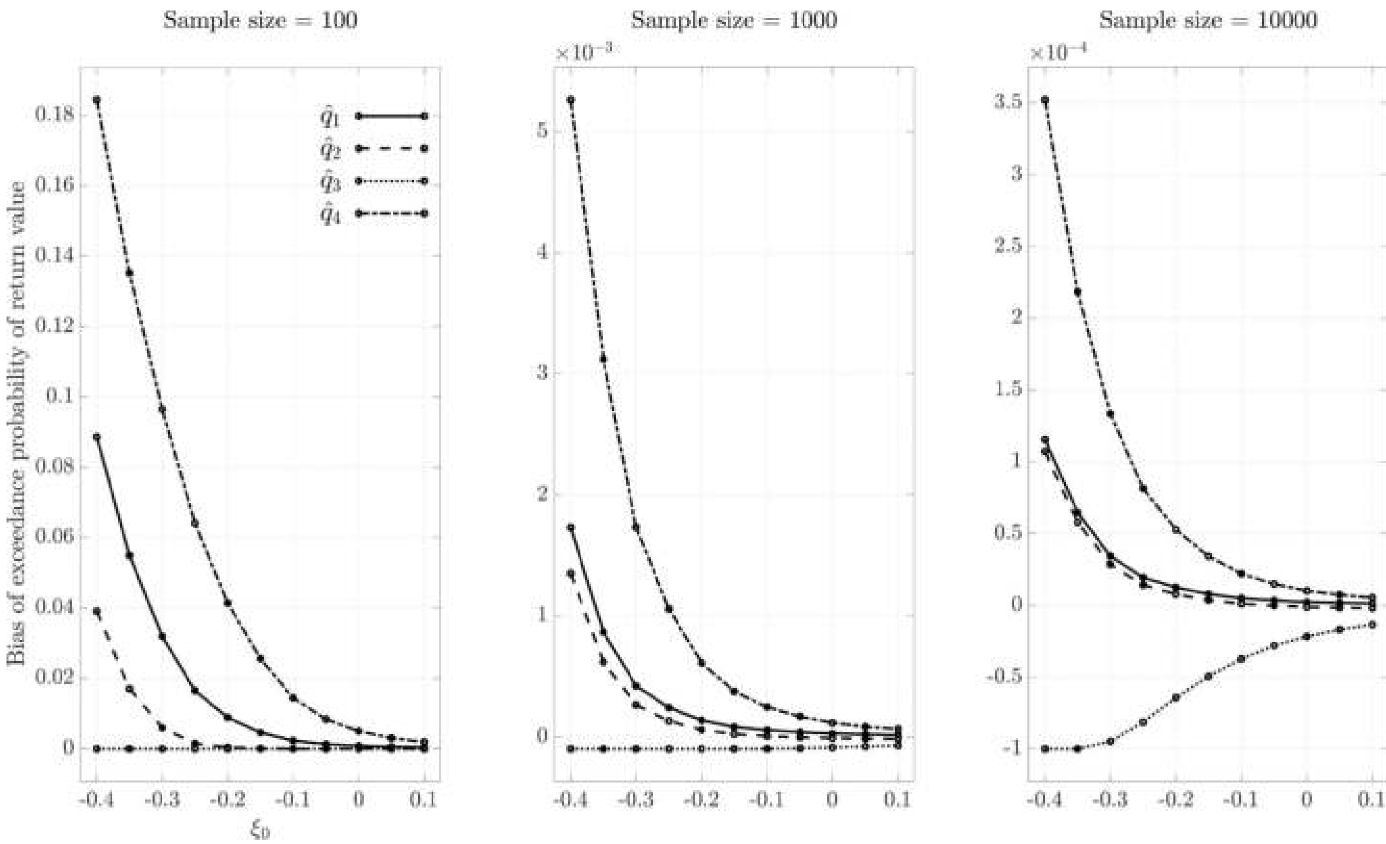

Fig. 8. Bias in exceedance probabilities corresponding to return value estimates $\hat{q}_1$ (solid), $\hat{q}_2$ (dashed), $\hat{q}_3$ (dotted), $\hat{q}_4$ (dot-dashed) for a return period of 10,000 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right).

performer for $\xi_0 < -0.2$. In our simulations, this tends to be because q_3 provides an estimate which is beyond the (known) upper end point of the distribution of annual maximum A. For larger ξ_0, q_2 and q_3 are competitive. In terms of log exceedance probability, q_3 performs spectacularly poorly for small sample size n; q_2 is overall the best performer.

7.1. The role of parameter estimation scheme

Many approaches exist to fit a GP distribution to a sample of independent threshold exceedances, e.g. as reviewed by Hosking and Wallis (1987); Madsen et al. (1997); Ashkar and Nwentsa Tatsambon (2007); de Zea Bermudez and Kotz (2010a, b); Mackay et al. (2011). Maximum likelihood estimation is the most popular approach (de Zea Bermudez and Kotz, 2010a), but the method of moments (MOM) and probability weighted moments (PWM) are also popular in the environmental and engineering literature. The procedures of Zhang and Stephens (2009) and Zhang (2010), referred to here as empirical Bayesian (EB), also attractive due to their low parameter bias characteristics. The simulation study of Section 5 assumes that maximum likelihood estimation is used for model parameter inference and construction of set $\mathscr{Z}$; the findings of the simulation study are therefore only directly relevant for this method of inference. Maximum likelihood estimation is a natural choice from the statistical perspectives of consistency, asymptotic Gaussianity and asymptotic efficiency (for $\xi > -0.5$, Davison and Smith, 1990). However, other estimation schemes are known to perform particularly well e.g. for small samples. For large samples, it is also known (Hosking and Wallis, 1987) that shape and scale parameters estimates for each of maximum likelihood, MOM and PWM inference are asymptotically Gaussian-distributed with variance-covariance matrix which converges to zero with increasing sample size n, and that shape and scale parameter estimates exhibit negative correlation. This suggests that the results of the maximum-likelihood-based simulation study are generally indicative of the influence of model parameter uncertainty on return value estimators for large sample size n.

Moreover, the theoretical arguments presented in Section 6 to explain the rank ordering of estimators q and truth q_0 for given conditions are applicable using parameter estimates $\mathscr{Z}$ from any estimation scheme of choice, including maximum likelihood, MOM, PWM and other approaches such as the method of Zhang (2010).

The values of estimates for q_1, q_2, q_3 and q_4 are deterministic functions of the set $\mathscr{Z}$ of parameter estimates. Different parameter estimation schemes provide different sets $\mathscr{Z}$, as illustrated in Fig. 11. Visual inspection of the figure suggests that, for given sample size n, differences between sets $\mathscr{Z}$ are relatively small.

Nevertheless, for finite sample sizes n, it is interesting to estimate the relative characteristics of estimators q using this set of relatively popular estimation schemes. Fig. 12 shows estimates $\hat{q}_1$, $\hat{q}_2$, $\hat{q}_3$ and $\hat{q}_4$ for sample size $n = 100$ from each of empirical Bayesian (EB), method of moments (MOM) and probability weighted moments (PWM) estimation schemes. Details of the methods used are given by e.g. Mackay et al. (2011), and software is available from Jonathan (2020).

The general characteristics of Fig. 12 are similar to those of Fig. 3 for maximum likelihood, in particular regarding the ordering of estimators by bias. The EB estimation scheme of Zhang (2010) is designed to provide low bias in GP parameter estimates, and hence provides estimates for q_1 with lowest bias. However, the performance of the EB scheme is qualitatively no better to that of the other estimation schemes for estimates of q_2, q_3 and q_4. Results for other sample sizes reflect those for $n = 100$, and are reported in Jonathan (2020), together with simulation code for estimation of bias in return value, exceedance probability and log exceedance probability for any combination of estimator and estimation scheme. For larger sample sizes, the EB again

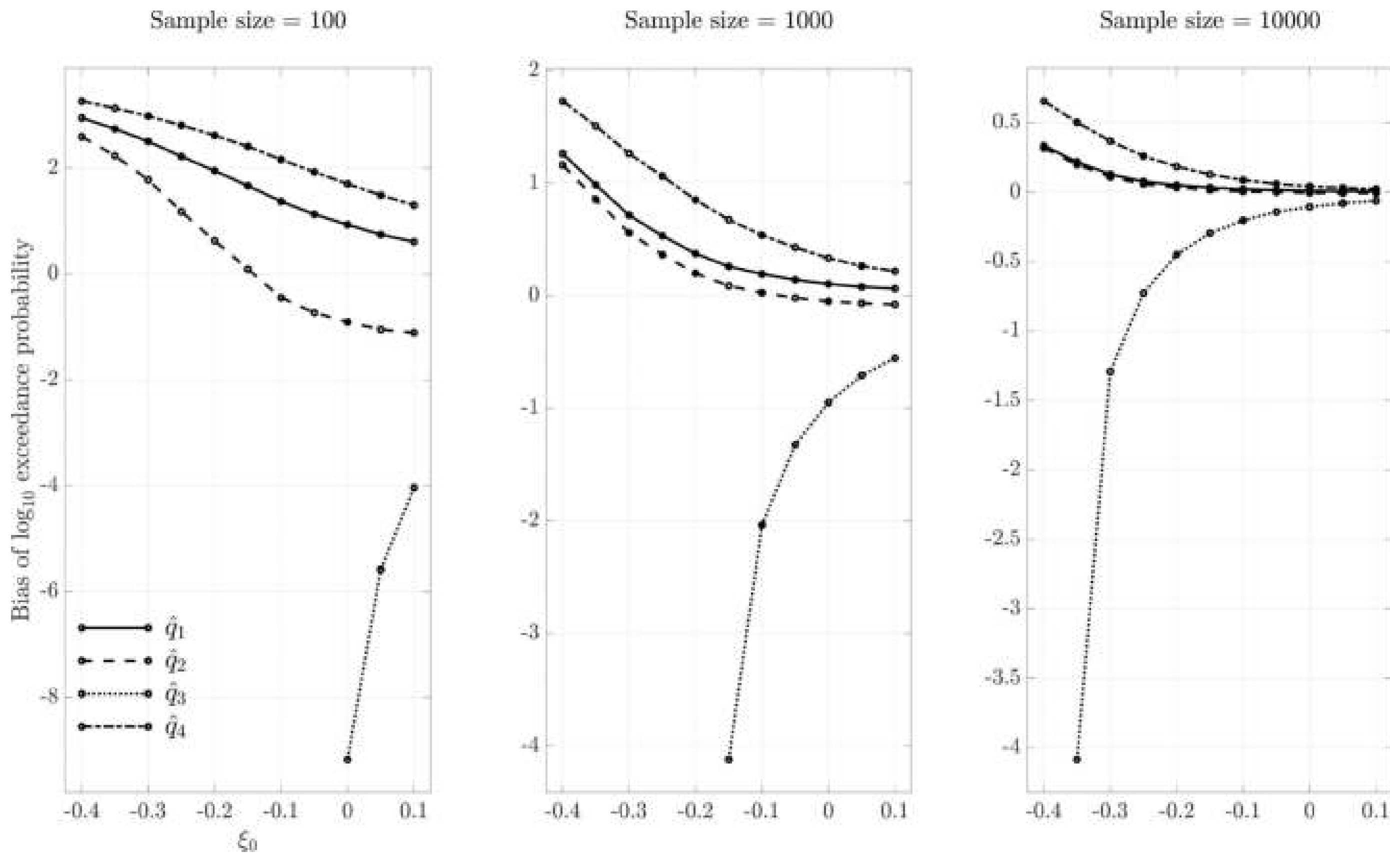

Fig. 9. Bias in logarithm (base 10) of exceedance probabilities corresponding to return value estimates $\widehat{q}_1$ (solid), $\widehat{q}_2$ (dashed), $\widehat{q}_3$ (dotted), $\widehat{q}_4$ (dot-dashed) for a return period of 10,000 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right).

provides q_1 estimates with low bias, and hence relatively low bias with respect to (log-) exceedance probability also.

More generally, characteristics of posterior Bayesian estimates for GP shape and scale can of course be strongly influenced by prior specification. Zhang (2007) proposes a likelihood moment estimator with high asymptotic efficiency. Further, non-parametric approaches including the moment (M) estimator of Dekkers et al. (1989) are available.

7.2. Wave loading

The wave loading on a structure, or its response to wave loading, might typically vary as a monotonically increasing function $\mathscr{R}$ of variable X. In this situation, we might be interested in quantities such as $\Pr(A^* > s)$, where A^* is the annual maximum event for loading (or response) $\mathscr{R}(X)$, and s might represent structural strength (or critical response). Following the arguments of Section 2, in the absence of uncertainty, the N-year structural strength q^* for $\mathscr{R}(X)$ (analogous to the N-year return value q for X in Equation (10)) is given by $q^* = \mathscr{R}(q)$ where $q = \frac{\sigma}{\xi}(N_*^{\xi} - 1) + \psi$, for $\xi \neq 0$ with the corresponding limiting expression when $\xi = 0$. Note that the * superscript q indicates a quantity related to a load or response variable. Sample estimators q_1^*, q_2^*, q_3^* and q_4^* for q^* are easily derived from those in Section 4. Fig. 13 shows fractional bias for estimators of the 100-year structural strength using the simulation procedure described in Section 5, for the case $\mathscr{R}(x) = x^2$.

Comparing Figs. 13 and 3, gross bias characteristics of q and q^* estimators are the same for this choice of $\mathscr{R}$, except that the extent of bias is greater for q^* that for q. By definition, the exceedance probabilities associated with q^* (in the distribution of the annual maximum A^* of $\mathscr{R}(X)$) and q (in the distribution of the annual maximum A of X) must be equal. Therefore Figs. 4 and 5 provide the relevant assessment. The bias of estimators q relative to q_0 will be influenced by the specification of $\mathscr{R}$.

More generally, loading R given environmental variable X will not be a deterministic function. However, the conditional distribution $F_{R|X}$ for a single loading event is usually accessible using numerical modelling, wave tank experiments or Morison-type fluid loading approximations (e.g. Tromans and Vanderschuren, 1995; Ross et al., 2020). The conditional distribution $F_{R|\mathbf{Z}}$ for load given uncertain extreme value parameters $\mathbf{Z}$ for X can be evaluated using

$$F_{R|\mathbf{Z}}(r|\boldsymbol{\zeta}) = \int_x F_{R|X}(r|x) f_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})\, dx$$

for $x \in \mathscr{D}_X$ for domain $\mathscr{D}_X$, where $f_{X|\mathbf{Z}}$ is the conditional density of X given $\mathbf{Z}$. Following the derivation $F_{A_N|\mathbf{Z}}$ in the Appendix, the conditional cumulative distribution function $F_{A_N^*|\mathbf{Z}}$ of the maximum N-year loading given $\mathbf{Z}$ is then

$$F_{A_N^*|\mathbf{Z}}(r|\boldsymbol{\zeta}) = \exp\Big[-N\lambda\big(1 - F_{R|\mathbf{Z}}(x|\boldsymbol{\zeta})\big)\Big].$$

The probability of structural failure p_F during a design life of N years given fixed structural strength s due to loading R can then be estimated using the predictive distribution $\tilde{F}_{A_N^*}$ given by

$$\tilde{F}_{A_N^*}(r) = \int_{\zeta} F_{A_N^*|\mathbf{Z}}(r|\boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\zeta}$$

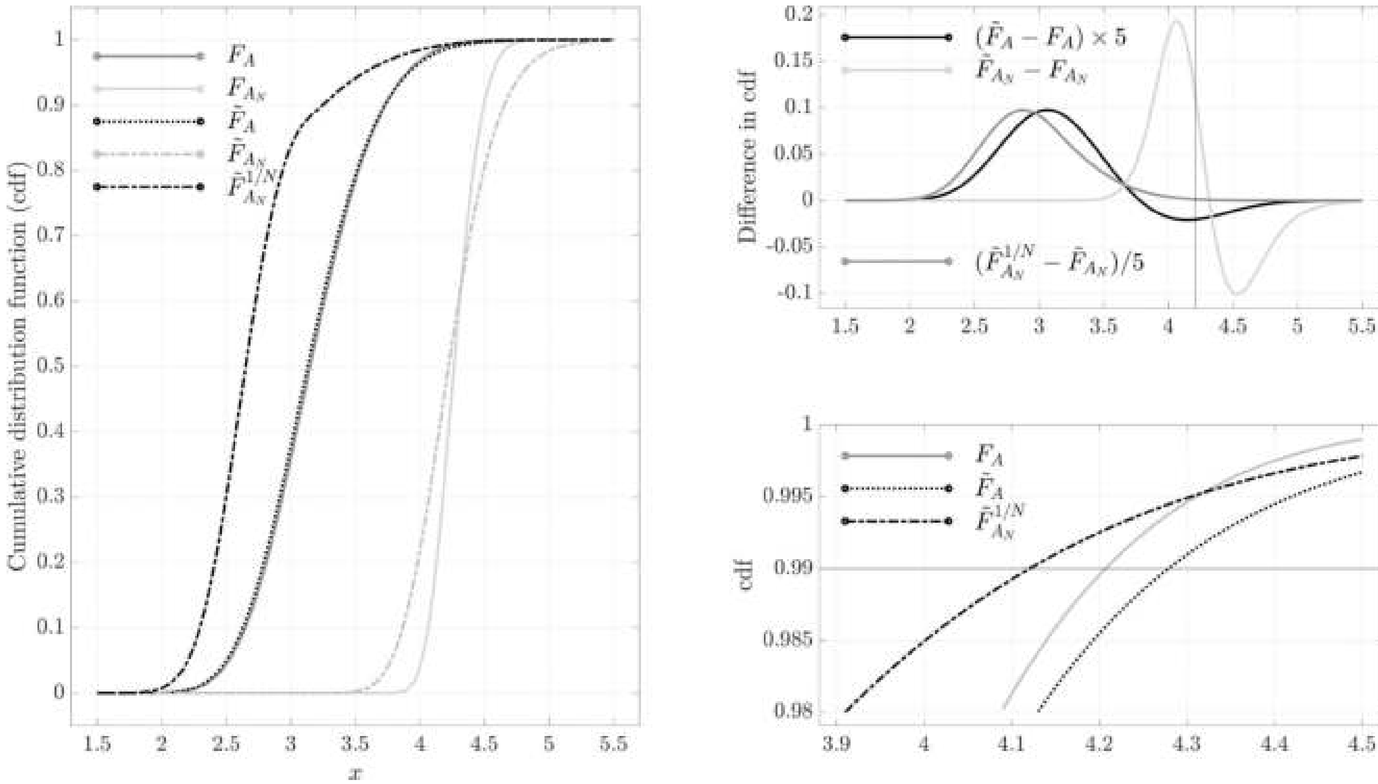

Fig. 10. Motivation for differences between return value estimators for the case $\xi_0 = -0.2$, sample size $n = 1,000$ and return period $N = 100$ years. The left hand panel shows the following cumulative distribution functions (cdfs): F_A, the (true) distribution of the annual maximum; F_{A_N}, the (true) distribution of the N-year maximum; $\tilde{F}_A$, the predictive distribution for the annual maximum; $\tilde{F}_{A_N}$, the predictive distribution for the N-year maximum; $\tilde{F}_{A_N}^{1/N}$, the N^{th} root of the predictive distribution for the N-year maximum. The top right panel gives differences between cdfs, with the vertical line indicating the true N-year return value. The bottom right panel illustrates annual maximum cdfs F_A, $\tilde{F}_A$ together with $\tilde{F}_{A_N}^{1/N}$ for non-exceedance probabilities near to $1 - 1/N$; the horizontal line is drawn at this level.

Table 1
Inequalities involving sample estimators for N-year return values, and conditions necessary for their validity. Note that the condition $N \to \infty$ applies to all these cases. Inequalities are not specific to maximum likelihood estimation of GP parameters.

	Inequality	Condition
I1	$q_3 \geq q_4$	Always
I2	$q_2 > q_0$	$\xi_1 > \max(\xi_0, 0)$
I3	$q_3 > q_2$	$1 > \xi_1 > \max(\xi_0, 0)$
I4	$q_1 > q_0$	$\xi_0, \xi_1 < 0$, $\sum_i \sigma_i / \sum_i (-\xi_i) > \sigma_0 / (-\xi_0)$
I5	$q_2 > q_0$	$\xi_0, \xi_1 < 0$, $(1/m) \sum_i (\sigma_i / (-\xi_i)) > \sigma_0 / (-\xi_0)$
I6	$q_3 > q_0$	$\xi_0, \xi_1 < 0, \max_{k \in (1,2,\ldots,m)} (\sigma_k / (-\xi_k)) > \sigma_0 / (-\xi_0)$

and setting $p_F = \Pr(A_N^* > s) = 1 - \tilde{F}_{A_N^*}(s)$. This derivation is analogous to that underlying return value estimator q_4, except that the value of N would typically be smaller, and p_F would be considerably smaller than $\exp[-1]$. However, as the discussion around the characteristics of q_4 and Fig. 13 shows, this estimator will also typically suffer from bias.

7.3. Conclusions

Theoretical results in Section 6 demonstrate that a systematic ordering of return value estimators q occurs for any parameter estimation scheme, including maximum likelihood. Theoretical results supported by simulations in Sections 5 suggest that none of the estimators q performs well with respect to all three performance measures considered here for maximum likelihood estimation; the predictive mean estimator q_2 performs best overall.

The work of Zhang and Stephens (2009) and Zhang (2010) supported by simulation results in Section 7.1 suggests that their empirical Bayesian estimation scheme provides low bias in GP parameter estimates and hence good performance for estimation of q_1. In relatively simple applications, where the effects of other sources of random and systematic variability can be ignored, estimator q_1 estimated using the method of Zhang (2010) offers lowest bias in return values over all estimators and estimation schemes.

As sample size n increases, biases in the four return value estimators q reduce in magnitude. Given a sufficiently large n, differences between estimators can be reduced to acceptable levels using any parameter estimation scheme; but the required n might well be very large. Moreover, a typical environmental extreme value analysis requires modelling of other sources of variation including extreme value threshold, rate of occurrence, potential extremal dependence and the effects of covariates. Even with a large n, the effective sample size for GP parameter estimation given specific covariate values might be < 100 in many applications. We might therefore anticipate large differences between estimators q. From a statistical modelling perspective, maximum likelihood estimation provides a flexible framework for incorporation of competing sources of variation; it also offers asymptotic efficiency.

From the perspective of predictive inference, estimators q_3 and q_4 might be considered preferable since they are estimated from predictive distributions which incorporate modelling uncertainty explicitly. In this case, inequality I1 shows that $q_3 \geq q_4$ always. Simulation suggests further that $q_3 \geq q_0 \geq q_4$ always for the situations considered. An estimator of the form $\alpha q_3 + (1 - \alpha) q_4$ (for $\alpha \in [0, 1]$) might provide a pragmatic compromise with better performance than either q_3 and q_4. Preliminary evaluation of $(q_3 + q_4)/2$ using maximum likelihood

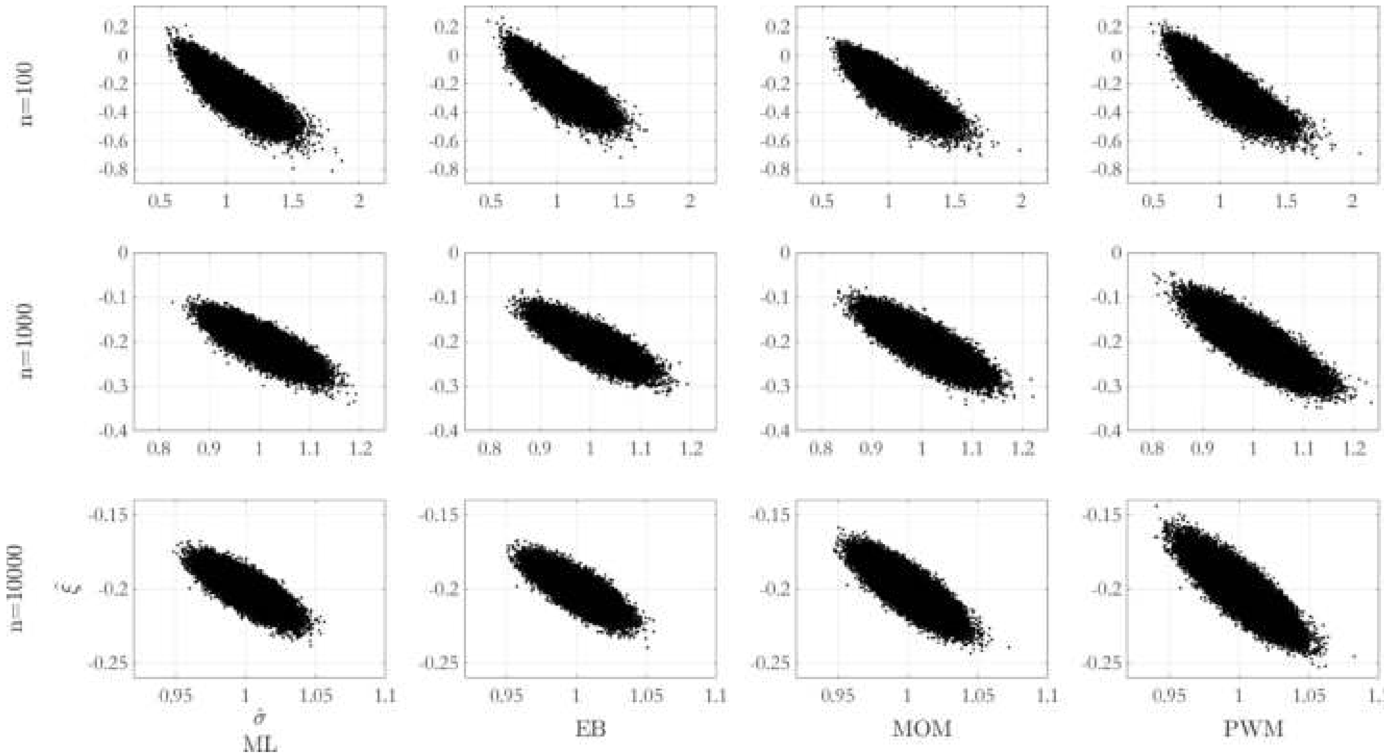

Fig. 11. Scatter plots of sets $\mathscr{Z}$ consisting of 10^5 GP parameter estimates for different sample sizes and estimation schemes for the case $\xi_0 = -0.2, \sigma = 1$. Rows from top correspond to sample sizes n of 100, 1000 and 10,000. Columns correspond to maximum likelihood (ML), the empirical Bayesian method (EB) of Zhang (2010), the method of moments (MOM) and probability weighted moments (PWM).

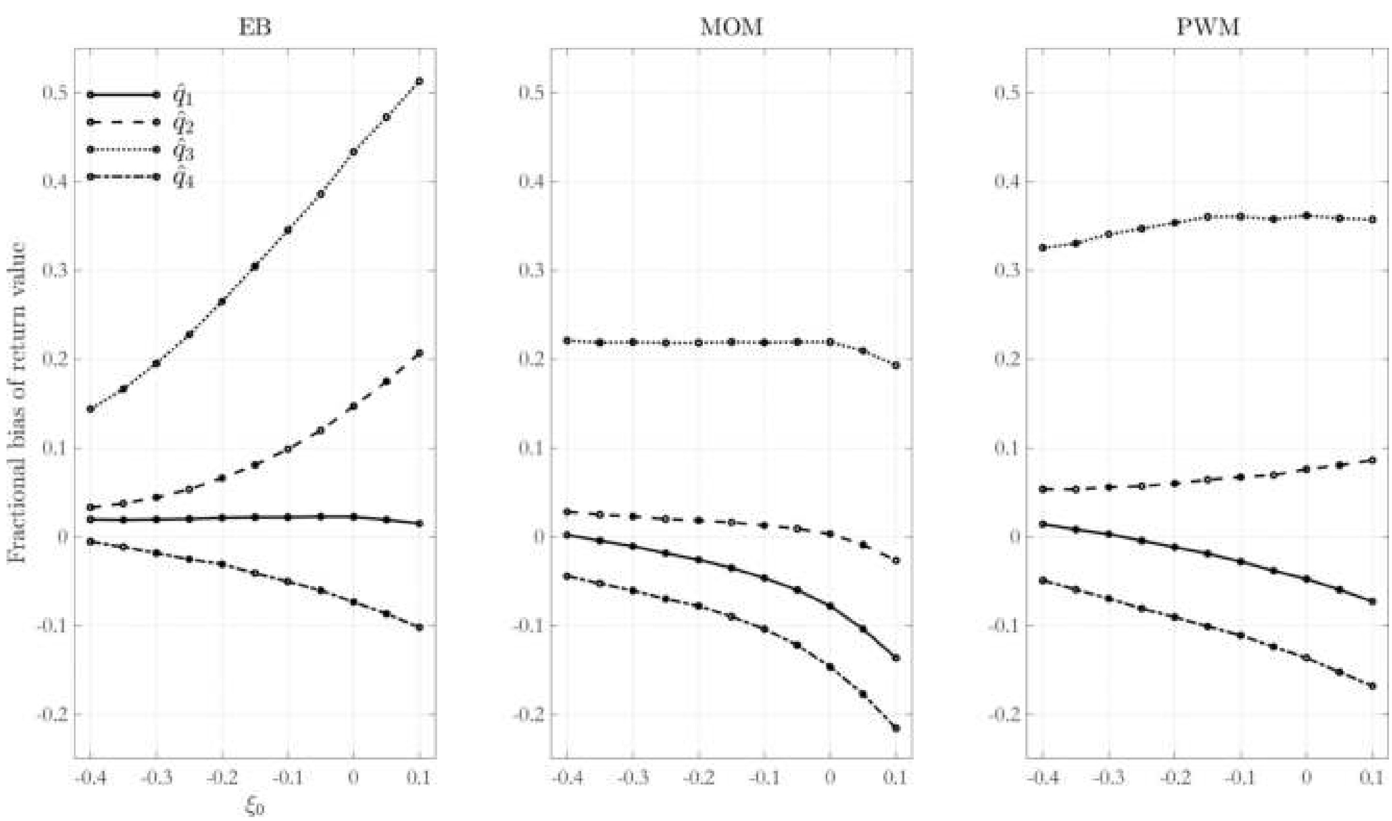

Fig. 12. Fractional bias for return value estimates $\hat{q}_1$ (solid), $\hat{q}_2$ (dashed), $\hat{q}_3$ (dotted), $\hat{q}_4$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$. Parameter estimates estimated from 10^5 realisations of samples of size 100 using empirical Bayesian (EB), method of moments (MOM) and probability weighted moments (PWM) estimation schemes. Unbiased estimates would have fractional bias of zero. Corresponding results for maximum likelihood estimation given in Fig. 3.

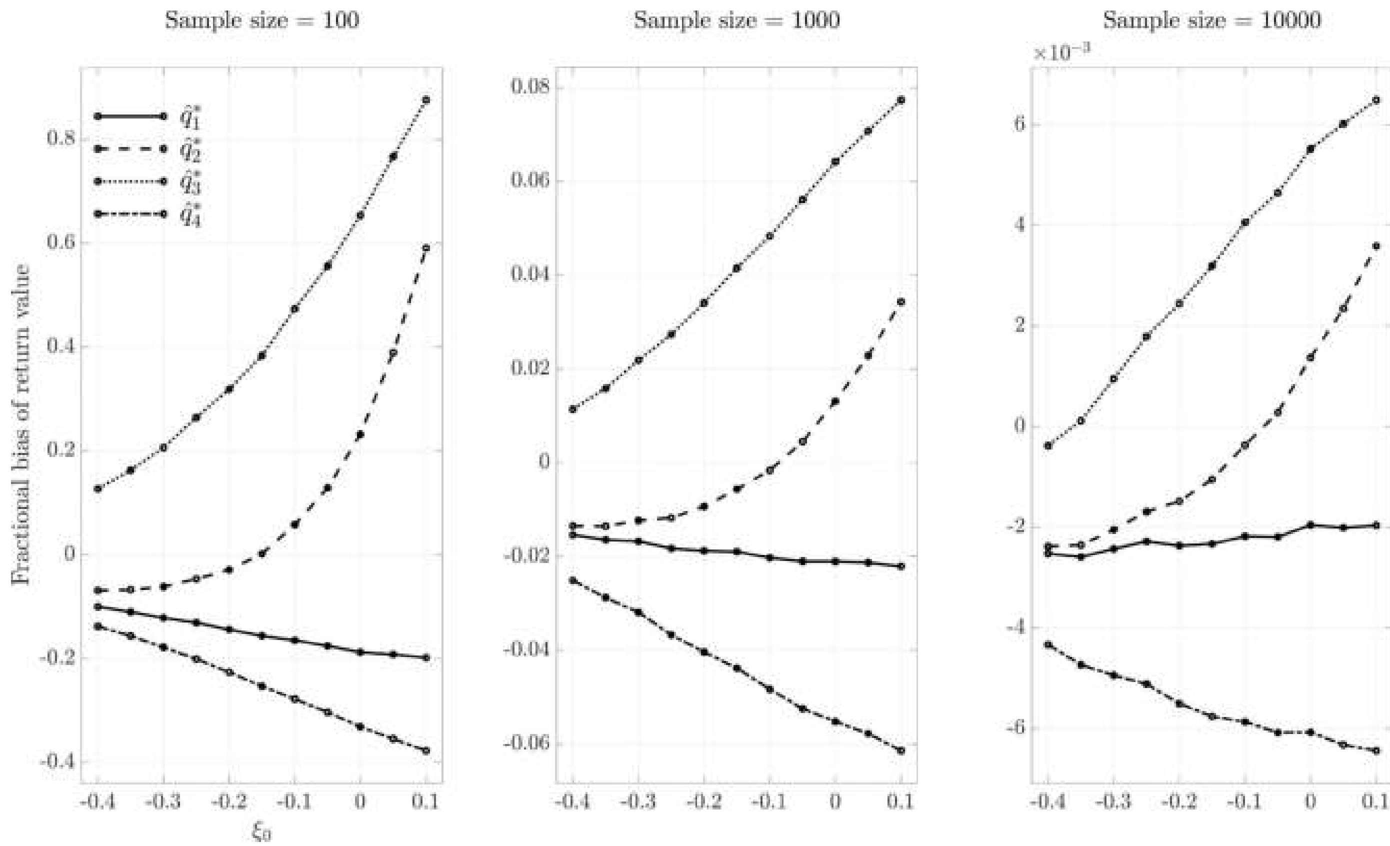

Fig. 13. Fractional bias for N-year structural strength estimates $\hat{q}_1^*$ (solid), $\hat{q}_2^*$ (dashed), $\hat{q}_3^*$ (dotted), $\hat{q}_4^*$ (dot-dashed) for a return period of 100 years, assuming 100 events per annum drawn from a GP distribution with shape ξ_0, scale $\sigma_0 = 1$ and threshold $\psi_0 = 0$ and wave loading function $\mathscr{R}(x) = x^2$. Bias estimated using maximum likelihood parameter estimates from 10^5 realisations of samples of size 100 (left), 1000 (centre) and 10,000 (right). Unbiased estimates would have fractional bias of zero.

estimation suggests this is a promising candidate. In terms of bias, it is competitive with q_2. The bias of its exceedance probability is similar to that of q_3 and generally better than that of q_2. The bias of its log probability is lower in magnitude than that of q_3, but not as low as that of q_2.

Summarising multivariate distributions for metocean variables in terms of return values has obvious advantages in terms of conciseness of description of an extreme ocean environment, for communication between different parties involved in offshore structural design. However, in reality, the accurate estimation of probability of structural failure should be the clear focus of analysis. From a predictive perspective, the effects of all sources of modelling uncertainty should be propagated carefully through the entire sequence of design calculations, expressed probabilistically, so that (a) the estimation of failure probability reflects these uncertainties as fully and fairly as possible, and (b) resources can hence be devoted to reducing the largest sources of uncertainty on failure probability in a rational and systematic manner. In this light, the use of summary statistics such as metocean return values at intermediate design stages in place of full distributions of variables (when available) should be avoided.

Judgements regarding the performance of estimators depend on the choice of utility or loss function adopted to assess it. It is likely that many discussions of return value estimates by environmental and engineering practitioners occur with insufficient awareness of systematic differences between estimators and estimation schemes highlighted by the current work, and of the importance of appropriate utilities to assess return value estimator characteristics appropriately.

CRediT authorship contribution statement

Philip Jonathan: Simulation, Theory, Writing. **David Randell:** Simulation. **Jenny Wadsworth:** Theory. **Jonathan Tawn:** Theory.

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

The authors gratefully acknowledge useful discussions with Matthew Jones (Shell, Amsterdam) and Graham Feld (Shell, Aberdeen). We further thank three reviewers for comments on an earlier version of the manuscript.

Appendix. Estimating the distribution of annual maximum A, the N-year maximum A_N and return values

According to asymptotic statistical theory, under the assumption that the random variable X belongs to the maximum domain of attraction of a non-degenerate distribution, the distribution of X exceeding threshold ψ converges (Pickands, 1975) to the generalised Pareto (GP) distribution as the threshold increases. We therefore typically assume, for sufficiently large threshold ψ that the conditional distribution function $F_{X|X>\psi,\mathbf{Z}}$ for parameters $\mathbf{Z}$ follows GP form as defined in Section 3. Usually, the value of threshold is set prior to estimation of ξ and σ from the sample of values for X (e.g. by examining a mean residual life plot, Coles, 2001), although this is not always the case (Scarrott and MacDonald, 2012). Various approaches to estimation of ξ and σ as used, including maximum likelihood, the method of moments and probability weighted moments. We emphasise the conditioning of distributions with respect to $\mathbf{Z}$ explicitly, since different approaches to incorporating the uncertainty in $\mathbf{Z}$ lead to the differences between return value estimators discussed in this work.

$F_{A|\mathbf{Z}}$ and $F_{A_N|\mathbf{Z}}$

Given $F_{X|X>\psi,\mathbf{Z}}$, the distribution $F_{A|\mathbf{Z}}$ of annual maxima can be derived using

$$F_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \Pr(A \le x|\boldsymbol{\zeta}) = \sum_{k=0}^{\infty} f_C(k) F^k_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})$$

where C is the number of occurrences of X per annum, with probability mass function f_C and

$$F_{X|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \tau + (1-\tau) F_{X|X>\psi,\mathbf{Z}}(x|\boldsymbol{\zeta})$$

where $\tau = \Pr(X < \psi)$. In practice, f_C is unknown and must also be estimated from data. Density f_C is often described by a Poisson distribution such that $f_C(k) = \exp[-\lambda]\lambda^k/k!$, $k = 0, 1, 2, \ldots$, for annual rate $\lambda > 0$ to be estimated. Assuming for simplicity that λ is known, the expression for $F_{A|\mathbf{Z}}$ simplifies (e.g. Ross et al., 2017) to

$$F_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \exp\left[-\lambda\left(1 - F_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})\right)\right].$$

Applying the above expression to a time period of N years, we evaluate the distribution of the N-year maximum A_N to be

$$F_{A_N|\mathbf{Z}}(x|\boldsymbol{\zeta}) = F^N_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \exp\left[-N\lambda\left(1 - F_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})\right)\right].$$

Writing $\overline{F}_{X|X>\psi,\mathbf{Z}}$ for the conditional tail distribution $1 - F_{X|X>\psi,\mathbf{Z}}$, the expressions for $F_{A|\mathbf{Z}}$ and $F_{A_N|\mathbf{Z}}$ become

$$F_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \exp\left[-\lambda(1-\tau)\overline{F}_{X|X>\psi,\mathbf{Z}}(x|\boldsymbol{\zeta})\right] \text{ and}$$
$$F_{A_N|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \exp\left[-N\lambda(1-\tau)\overline{F}_{X|X>\psi,\mathbf{Z}}(x|\boldsymbol{\zeta})\right]$$

where for the GP distribution

$$\overline{F}_{X|X>\psi,\mathbf{Z}}(x|\boldsymbol{\zeta}) = \left(1 + \frac{\xi}{\sigma}(x-\psi)\right)_+^{-1/\xi}$$

and $= \exp[-(x-\psi)/\sigma]$ when $\xi = 0$. That is, the annual maximum A and N-year maximum A_N follow generalised extreme value distributions. An alternative approach to the derivation above (e.g. Smith, 1987; Northrop et al., 2017) starts by assuming that the number λ of events X per annum is given (i.e. estimated independently), but that the number of events exceeding threshold ψ is random and follows a binomial distribution $\text{Bin}(\lambda, (1-\tau))$, which can be approximated by $\text{Poiss}(\lambda(1-\tau))$ when λ is large and $(1-\tau)$ is near zero. Then for $F_{X|\mathbf{Z}} \approx 1$, $\overline{F}_{X|\mathbf{Z}} << 1$ and $\log(1 - \overline{F}_{X|\mathbf{Z}}) \approx -\overline{F}_{X|\mathbf{Z}}$. Hence $F_{A|\mathbf{Z}}(x)$ can be approximated using

$$F_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = F_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})^{\lambda} = \exp\left[\lambda\log\left(F_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})\right)\right] = \exp\left[\lambda\log\left(1 - \overline{F}_{X|\mathbf{Z}}(x|\boldsymbol{\zeta})\right)\right] \approx \exp\left[-\lambda\Pr(X>x|X>\psi,\mathbf{Z}=\boldsymbol{\zeta})(1-\tau)\right] \text{ when } F_{X|\mathbf{Z}} \approx 1 = \exp\left[-\lambda(1-\tau)\overline{F}_{X|X>\psi,\mathbf{Z}}(x|\boldsymbol{\zeta})\right]$$

as before.

Return values

The expressions for $F_{A|\mathbf{Z}}$ however derived and $F_{A_N|\mathbf{Z}}$ facilitate estimation of different return value estimators given $F_{X|X>\psi,\mathbf{Z}}$ (with parameters ξ and σ) for any return period N, as described in Section 4. For brevity, we write the expected number N_* of occurrences of threshold exceedances in N years as $N_* = N\lambda(1-\tau)$.

Given ξ and σ, the N-year return value q (conditional on $\mathbf{Z} = \boldsymbol{\zeta}$) can be found by solving Equation (1)

$$1 - 1/N = \exp\left[-\lambda(1-\tau)\overline{F}_{X|X>\psi,\mathbf{Z}}(q|\boldsymbol{\zeta})\right] \text{ so that}$$

$$\log\left[\frac{1}{1-1/N}\right] = \lambda(1-\tau)\overline{F}_{X|X>\psi,\mathbf{Z}}(q|\boldsymbol{\zeta})$$

With $\log[1/(1-1/N)] = 1/N + 1/(2N^2) + 1/(3N^3) + \ldots \approx 1/N$ for large N, this yields

$$\frac{1}{N_*} = \overline{F}_{X|X>\psi,\mathbf{Z}}(q|\boldsymbol{\zeta}) = \left(1 + \frac{\xi}{\sigma}(q-\psi)\right)_+^{-1/\xi}$$

and hence, conditional on $\mathbf{Z} = \boldsymbol{\zeta}$, the return value is given by $q = \frac{\sigma}{\xi}(N_*^{\xi} - 1) + \psi$ as given in the main text. When $\xi = 0$, the last two equations read $1/N_* = \exp[-(q-\psi)/\sigma]$ and $q = \sigma\log[N_*] + \psi$.

Dependence

When observations of X exhibit dependence, the equations for $F_{A|\mathbf{Z}}$ and $F_{A_N|\mathbf{Z}}$ can be adjusted by incorporating an extremal index $\in [0,1]$ (e.g. Davison et al., 2012) such that

$$F_{A|\mathbf{Z}}(x) = \sum_{k=0}^{\infty} f_C(k) F_{X|\mathbf{Z}}^{k\theta}(x|\boldsymbol{\zeta})$$

where θ must also be estimated. This approach might be advantageous when estimating e.g. $F_{A|Z}$ from dependent sea-state H_S instead of independent storm peak H_S.

Covariates

Further, the equations above also assume that X does not exhibit systematic variation with covariates (such as storm direction or season in the case of significant wave height). In the presence of a single covariate Φ, distributions above are conditional on Φ, and parameters functions of Φ. The inference becomes non-stationary and computationally more demanding, but we can still evaluate unconditional distributions e.g. $F_{A|Z}$ by marginalisation using

$$F_{A|\mathbf{Z}}(x|\boldsymbol{\zeta}) = \int_{\varphi} F_{A|\mathbf{Z},\Phi}(x|\boldsymbol{\zeta}, \varphi) f_{\Phi}(\varphi)\, d\varphi$$

where f_{Φ} is the density of Φ which must itself also be estimated.

References

Ashkar, F., Nwentsa Tatsambon, C., 2007. Revisiting some estimation methods for the generalized Pareto distribution. J. Hydrol. 346, 136–143.
Christensen, R., Huffman, M.D., 1985. Bayesian point estimation using the predictive distribution. Am. Statistician 39, 319–321.
Coles, S., 2001. An Introduction to Statistical Modelling of Extreme Values. Springer.
Davison, A., Smith, R.L., 1990. Models for exceedances over high thresholds. J. Roy. Stat. Soc. B 52, 393.
Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modelling of spatial extremes. Stat. Sci. 27, 161–186.
de Zea Bermudez, P., Kotz, S., 2010a. Parameter estimation of the generalized Pareto distribution-Part I. J. Stat. Plann. Inference 140, 1353–1373.
de Zea Bermudez, P., Kotz, S., 2010b. Parameter estimation of the generalized Pareto distribution-Part II. J. Stat. Plann. Inference 140, 1374–1388.
Dekkers, A.L.M., Einmahl, J.H.J., Haan, L.D., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Stat. 17, 1833–1855.
DNVGL-RP-C205, 2017. Environmental Conditions and Environmental Loads. Det Norske Veritas group, Norway.
Dombry, C., Ferreira, A., 2019. Maximum likelihood estimators based on the block maxima method. Bernoulli 25, 1690–1723.
Elsinghorst, C., Groeneboom, P., Jonathan, P., Smulders, L., Taylor, P., 1998. Extreme value analysis of North Sea storm severity. J. Offshore. Arct. Eng. 120, 177–183.
Ewans, K.C., Jonathan, P., 2008. The effect of directionality on northern North Sea extreme wave design criteria. J. Offshore. Arct. Eng. 130, 041604:1–041604:8.
Fawcett, L., Green, A., 2018. Bayesian posterior predictive return levels for environmental extremes. Stoch. Environ. Res. Risk Assess. 32, 2233–2252.
Ferreira, A., de Haan, L., 2015. On the block maxima method in extreme value theory: pwm estimators. Ann. Stat. 43, 276–298.
Geisser, S., 1971. In: Godambe, V.P., Sprott, D.A. (Eds.), The Inferential Use of Predictive Distributions. Holt, Rinehart and Winston, Toronto, pp. 456–469.
Haselsteiner, A.F., Thoben, K.-D., 2020. Predicting wave heights for marine design by prioritizing extreme events in a global model. Renew. Energy 156, 1146–1157.
Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339–349.
ISO19901-1, 2015. Petroleum and Natural Gas Industries. Specific Requirements for Offshore Structures. Part 1: Metocean Design and Operating Considerations, first ed. International Standards Organisation.
Jonathan, P., 2020. Uncertainties in return values. https://github.com/ygraigarw/UncertaintiesReturnValues.
Mackay, E.B.L., Challenor, P.G., Bahaj, A.S., 2011. A comparison of estimators for the generalised Pareto distribution. Ocean Eng. 38, 1338–1346.
Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. at-site modeling. Water Resour. Res. 33, 747–757.
NORSOK N-003, 2017. NORSOK Standard N-003:2017: Actions and Action Effects. NORSOK, Norway.
Northrop, P., Attalides, N., Jonathan, P., 2017. Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. J. Roy. Stat. Soc. C 66, 93–120.
Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Stat. 3, 119–131.
Raghupathi, L., Randell, D., Ewans, K., Jonathan, P., 2016. Fast computation of large scale marginal extremes with multi-dimensional covariates. Comput. Stat. Data Anal. 95, 243–258.
Randell, D., Feld, G., Ewans, K., Jonathan, P., 2015. Distributions of return values for ocean wave characteristics in the South China Sea using directional-seasonal extreme value analysis. Environmetrics 26, 442–450.
Randell, D., Turnbull, K., Ewans, K., Jonathan, P., 2016. Bayesian inference for non-stationary marginal extremes. Environmetrics 27, 439–450.
Ross, E., Astrup, O.C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., Jonathan, P., 2020. On environmental contours for marine and coastal design. Ocean Engineering 195, 106194.
Ross, E., Randell, D., Ewans, K., Feld, G., Jonathan, P., 2017. Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. Ocean Eng. 142, 315–328.
Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Rev. Stat. 10, 33–60.
Serinaldi, F., 2015. Dismissing return periods! Stoch. Environ. Res. Risk Assess. 29, 1179–1189.
Smith, R.L., 1987. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. Stat. Sci. 2, 45–52.
Smith, R.L., 2003. Statistics of extremes, with applications in environment, insurance, and finance. In: Finkenstadt, B., Rootzén, H. (Eds.), Extreme Values in Finance, Telecommunications, and the Environment. Chapman and Hall/CRC, pp. 20–97.
Tromans, P.S., Vanderschuren, L., 1995. Variable Based Design Conditions in the North Sea: Application of a New Method. Offshore Technology Confernence, Houston (OTC–7683).
Zhang, J., 2007. Likelihood moment estimation for the generalized Pareto distribution. Aust. N. Z. J. Stat. 49, 69–77.
Zhang, J., 2010. Improving on estimation for the generalized pareto distribution. Technometrics 52, 335–339.
Zhang, J., Stephens, M.A., 2009. A new and efficient estimation method for the generalized pareto distribution. Technometrics 51, 316–325.