Ocean Engineering 35 (2008) 1211–1225

Contents lists available at ScienceDirect

Ocean Engineering

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

# Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects

Philip Jonathan [a,*], Kevin Ewans [b], George Forristall [c]

[a] Shell Technology Centre Thornton, P.O. Box 1, Chester, UK
[b] Shell International Exploration and Production, P.O. Box 60, 2280 AB Rijswijk, The Netherlands
[c] Forristall Ocean Engineering, Camden, ME 04843, USA

ARTICLE INFO

*Article history:*
Received 16 December 2007
Accepted 5 April 2008
Available online 12 April 2008

*Keywords:*
Extremes
Directional models
Design criteria
Generalised Pareto

ABSTRACT

With increasing availability of good directional data, provision of directional estimates of extreme significant wave heights, in addition to the omni-directional estimates, is more common. However, interpretation of directional together with omni-directional design criteria is subject to inconsistency, even in design guidelines. In particular, omni-directional criteria are usually estimated ignoring directional effects. In this article, for data which exhibit directional effects, we show that a directional extreme value model generally explains the observed variation significantly better than a model which ignores directionality, and that omni-directional criteria developed from a directional model are different from those generated when directionality is not accounted for. We also show that omni-directional criteria derived from a directional model are more accurate and should be preferred in general over those based on models which ignore directional effects. We recommend use of directional extreme value models for estimation of both directional and omni-directional design criteria in future, when good directional data are available. If effects of other covariates (e.g. time or space) are suspected, we similarly recommend use of extreme value models which adequately capture sources of covariate variability for all design analysis.



## 1. Introduction

Historically, design wave criteria have generally been based on omni-directional estimates—that is, sea state criteria that are relevant for any direction of approach. However, to take advantage of directional effects in optimising cost of offshore facilities, and with increasing availability of good wave direction data, it is common to provide directional estimates of extreme significant wave heights, in addition to the omni-directional estimate. The practice has been to associate an appropriate wave direction, such as the mean wave direction or the direction of spectral peak, with the significant wave height of a given sea state. The significant wave height values selected for extreme value analysis are then binned into directional sectors, with sector limits chosen arbitrarily either on a fixed number (typically eight) centred on the cardinal directions or on the basis of a perceived directionality in the data. Both sectoring approaches make the assumption that statistics are homogeneous within each sector. Extremal analyses are then performed on the significant wave height values in each sector, and also on the complete set of values ignoring direction. Design values with a given return period are then specified for each directional sector and for all directions, the latter being the omni-directional estimate.

Unfortunately, this approach has often led to inconsistencies in design criteria, such as that the probability of exceedance of a given significant wave height when calculated from the directional criteria is different from that obtained from omni-directional criteria. A discussion of this problem is given by Forristall (2004). Guidelines such as API (2005) and ISO (2005) provide recommendations on treating directional criteria, but even when these are followed, either inconsistency remains (in the case of API) or insufficient detail is given on how to make the criteria consistent (in the case of ISO).

Numerous authors have reported the essential features of extreme value analysis (e.g. Davison and Smith, 1990) and the importance of considering different aspects of covariate effects. Coles and Walshaw (1994) describe directional modelling of extreme wind speeds. Robinson and Tawn (1997) describe directional modelling of extreme sea currents. Anderson et al. (2001) report that estimates for 100-year maximum $H_S$ from an extreme value model ignoring seasonality are considerably

* Corresponding author. Tel.: +44 151 373 5421.
*E-mail addresses:* philip.jonathan@shell.com (P. Jonathan), kevin.ewans@shell.com (K. Ewans), george@forocean.com (G. Forristall).

0029-8018/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oceaneng.2008.04.002

smaller than those obtained using a number of different seasonal extreme value models. Chavez-Demoulin and Davison (2005) and Coles (2001) provide straight-forward descriptions of a non-homogeneous Poisson model in which occurrence rates and extremal properties are modelled as functions of covariates. Carter and Challenor (1981) consider estimation of annual maxima from monthly data, when the distribution functions of monthly extremes are known. They prove in this case that the distribution function of the maximum from a random sampling of the annual population is greater than or equal to the distribution function of the *proportionally sampled* monthly sub-populations, with equality only when monthly distributions are equal.

Recently, Jonathan and Ewans (2007) and Ewans and Jonathan (2008) have borrowed elements of developments from the statistics literature to model directional effects smoothly in a way that better reflects nature and observed data in application to Gulf of Mexico and North Sea hindcast data. Consistency between directional and omni-directional criteria follows directly in a rigourous way in these analyses, but the analyses also demonstrate a perhaps more significant aspect to establishing omni-directional criteria that we believe is not widely recognised by practitioners. Specifically, (i) a directional model generally explains the observed variation of extremes significantly better than a model which ignores directionality, (ii) omni-directional criteria developed from a directional model are different from those generated when directionality is not accounted for, and thus (iii) omni-directional criteria derived from a directional model are more realistic and should be preferred in general. These facts challenge traditional practice and warrant special attention. Accordingly, we have undertaken specific simulation studies to demonstrate unambiguously that omni-directional criteria should be derived using a directional model in cases where good directional data exist.

Covariate effects are not restricted to direction, but are equally relevant for other covariates, such as time and space. We focus here on directionality due to widespread current practice of providing directional criteria. Nevertheless, findings are transferable to covariate effects generally.

The outline of the article is as follows. In Section 2 we give some theoretical background to the problem at hand. In Section 3 we introduce a series of case studies based on an idealised model used thereafter to demonstrate the importance of directional modelling on estimation of omni-directional criteria. Sections 4 and 5 give results and discussion for various scenarios corresponding to different assumed extremal characteristics for the toy model. Conclusions are drawn and recommendations made in Section 6. In the appendix, following a query from a discussant, we demonstrate informally that the distribution of the $p$-year maximum can be obtained equivalently from either the conditional distribution of peaks over threshold or from the unconditional distribution.

## 2. Theory

Extreme value analysis of ocean storms is performed as part of design assessment of offshore and coastal structures. In essence, for data in the form of block maxima (e.g. monthly or annual maxima), this analysis involves fitting a generalised extreme value (GEV) distribution. Then extreme quantiles of the fitted distribution are used as design values. The choice of GEV distribution for block maxima is justified asymptotically, since the distribution of the maxima of independent random variables converges to GEV (see, e.g., Coles, 2001). When data take the form of peaks over threshold, we fit using a generalised Pareto (GP) distribution, since again the distribution of peaks over high threshold for independent random observations converges in a certain sense to GP (see, e.g., Embrechts et al., 2003). The GEV and GP distributions are intimately related; if the number of exceedances of a high threshold is assumed Poisson-distributed, and exceedances over threshold are characterised by GP, then the distribution of the maximum over threshold is GEV.

Extreme value behaviour of ocean storms is often dependent on covariates such as storm direction, geographic location and season. Extremal behaviour may also show long-term trends in time throughout the period under consideration. Covariate effects are often ignored in practice because of small sample size and complexity of extreme value modelling with covariates. However, it is important that effects of covariates are adequately modelled. Improvements in structural reliability for no additional cost are achievable in principle if covariate effects are incorporated in the specification of design criteria in general, e.g. by (a) designing structures with different strengths in different directions, (b) locating structures in the most benign environments, and (c) performing activities of limited duration during the more benign periods, all consistent with other operational requirements for the offshore structure. Moreover, estimates for extreme quantiles obtained from modelling that ignores covariate effects will not be the same as those obtained from models incorporating covariate effects, in general. If models incorporating covariate effects provide statistically better fits to data than those ignoring covariate effects, there is good reason to believe that models incorporating covariate effects will provide better estimates for extreme quantiles.

In engineering practice, when covariate effects are suspected, the data sample is often partitioned by covariate and independent analyses performed per subsample. For instance, directional modelling may be performed for a set of directional sectors which are assumed homogeneous with respect to the covariate (see, e.g., Forristall, 2004); omni-directional quantiles are then estimates on the basis of all sector models. However, in the statistics literature, extreme value models with parameters varying smoothly with one or more covariates are used routinely. This has the advantage of varying extremal properties with covariates in a manner consistent with underlying physics. The current authors have reported the application of directional modelling to Gulf of Mexico and Northern North Sea hindcast data (Jonathan and Ewans, 2007; Ewans and Jonathan, 2008). These articles modelled extremes of storm peak $H_S$ over threshold, each storm contributing exactly one storm peak observation, itself associated with a unique storm peak direction. Storm periods were selected to be sufficiently far apart that storm peak events could be reasonably assumed to be independent at a given location. Data from multiple locations were included, each location contributing exactly one observation for each storm, all locations assumed to have identical (marginal) extreme value behaviour. Since marginal distributions per location are certainly dependent in general, and since modelling both dependency and marginal behaviour together is technically challenging, we proceed to model assuming marginal distributions independent, but then carefully correct estimates for parameter uncertainties using a block bootstrapping approach (Chavez-Demoulin and Davison, 2005).

To estimate extreme quantiles, it is necessary to characterise both the distribution of peaks over threshold and the rate of occurrence of extremes. For peaks $x$ over threshold $u$, the form of the GP distribution $F_c(x;\theta)$ with shape parameter $\gamma$ and scale parameter $\sigma$ is

$$F_c(x;\theta) = 1 - \left(1 + \frac{\gamma(\theta)}{\sigma(\theta)}(x - u(\theta))\right)^{-1/\gamma(\theta)}$$

for all $x$ such that $1+(\gamma(\theta)/\sigma(\theta))(x-u(\theta))>0$. The form of the unconditional distribution of peaks over threshold is

$$F(x;\theta)=(1-p(\theta))+p(\theta)F_c(x;\theta)$$

where $p$ is the probability of threshold exceedance. Here, each of $\gamma$, $\sigma$ and $p$ are assumed to be functions of covariates $\theta=\{\theta_i\}$ which are estimated directly from the data using maximum likelihood. Informally, the distribution of the storm peak maximum in some time period $P$ for covariate values in interval $I$ is

$$F_P(x;\theta\in I)=\prod_{i=1}^{n}\sum_{k=0}^{\infty}F^k(x;I_i)\Pr(N_n(I_i,P)=k)$$

where $\bigcup_{i=1}^{n}I_i$ is a partition of the domain $I$ of covariate(s) of interest into small homogeneous intervals, and $N_n(I_i,P)$ is the number of occurrences of storm peak events corresponding to interval $I_i$ in period $P$, assumed to be independently Poisson-distributed. Thus, in the case of storm direction as a covariate, we can evaluate extreme quantiles for arbitrary directional sectors, including the so-called omni-directional extreme quantile.

## 3. The toy model case studies

We illustrate the importance of directional extreme value modelling using a series of case studies based on a simple conceptual model, henceforth referred to as the 'toy model'.

In the toy model, we observe storm peak $H_S$ events associated with one of two directional sectors (S1 and S2). Storm peak events are independent, and events in a given directional sector are identically GP distributed. Numbers of events occurring per annum per directional sector are exactly equal and constant in time. We use the toy model to generate a sample consisting of $n$ data corresponding to a period $P_0$ (years). Our objective is to estimate the distribution of the 100-year maximum storm peak event (or characteristics thereof).

For convenience, we refer to the set of GP shape $\gamma$, scale $\sigma$ and threshold $u$ parameters by $E=(\gamma,\sigma,u)$. The marginal distributions of sectors S1 and S2 are characterised by $E_1$ and $E_2$, respectively.

Using the toy model we can calculate the characteristics of extremes from theory. Since the number of occurrences in 100 years, $m$ is known to be $m=100n/P_0$ (and is not a random quantity), the distribution of the 100-year maximum event is

$$F_{100\mathrm{Theory}}(x)=F_1^{m/2}(x)\times F_2^{m/2}(x)$$

where $F_1$ and $F_2$ are GP sector distributions for S1 and S2.

The distribution of the 100-year maximum can also be estimated by simulating multiple realisations of periods of 100 years, recording the maximum observed in each realisation and constructing an empirical distribution $F_{100\mathrm{Sim}}$.

Further, we can generate a sample of data from the toy model, fit various extreme value models and estimate the distribution of extreme quantiles of interest. We compare sample-based estimates of extreme quantiles with those obtained from theory or directly from simulations. Specifically, we compare the bias and precision of sample-based estimates for omni-directional criteria based on either a directional extreme value model or an extreme value model which ignores directionality (henceforth the 'constant model', since extremal characteristics are assumed constant with direction) with reality (i.e. theory or direct simulation).

Using the directional model, the distribution of the 100-year maximum event is

$$F_{100\mathrm{Dir}}(x)=((1-\hat{p}_1)+\hat{p}_1F_c(x\,|\,\hat{E}_1))^{m/2}((1-\hat{p}_2)+\hat{p}_2F_c(x\,|\,\hat{E}_2))^{m/2}$$

where $\hat{p}_1$, $\hat{p}_2$ are estimated threshold exceedance probabilities for the respective sectors, and $\hat{E}_1$, $\hat{E}_2$ are estimated GP parameters. Using the constant model, the distribution of the 100-year maximum is

$$F_{100\mathrm{Con}}(x)=((1-\hat{p})+\hat{p}F_c(x\,|\,\hat{E}))^m$$

where $\hat{p}$ is the estimated (omni-directional) exceedance probability and $\hat{E}$ the estimated GP parameters.

We estimate bias and precision of $F_{100\mathrm{Dir}}$ and $F_{100\mathrm{Con}}$ with respect to $F_{100\mathrm{Theory}}$ or $F_{100\mathrm{Sim}}$ by simulation of 1000 realisations of the complete sample-based estimation. For definiteness and clarity of discussion, we focus on estimation of the median, 25th and 75th percentiles of the 100-year maximum event. Inferences from these case studies are generally applicable to estimation of omni-directional criteria for directional data, and more widely to sample-based estimation of extreme quantiles in the presence of covariate effects.

## 4. Results

This section presents results for three case studies, and summarises findings for supporting case studies not reported fully here. The modelling procedure is explained in more detail for Case Study 1.

### 4.1. Case Study 1

We observe a sample of $n=2500$ storm peak $H_S$ events in a period $P_0=25$ years. Exactly 1250 events occur in each of directional sectors S1 and S2. The sector GP parameters are taken to be $E_1=(-0.1,2,4)$ and $E_2=(-0.3,4,4)$. Therefore, extreme value distributions for sectors S1 and S2 are bounded on the right-hand side at 24 and 17.33 m, respectively (obtained using the expression $u-\sigma/\gamma$). These parameter values might correspond, e.g. to a Northern North Sea environment. The marginal densities of events per directional sector are illustrated in Fig. 1.

The distribution of the 100-year maximum event can be calculated from theory and simulation as discussed in Section 3, using 1000 realisations of samples of 100 years. Resulting cumulative distribution functions, illustrated in Fig. 2, are in good agreement. Behaviour of the probability of threshold exceedance with threshold was monitored and is illustrated in Fig. 3.

We need to assess the quality of fit of the directional GP model to the data, and estimate an appropriate value of threshold to adopt. In this simple situation, fitting the directional model corresponds to fitting sectors 1 and 2 independently using the GP distribution. We select an appropriate threshold value by examining the variation of estimated parameters $\gamma$ and $\sigma$ with threshold $u$ (see, e.g., Embrechts et al., 2003). In the case of $\gamma$, we expect the estimate to remain steady as a function of threshold over a reasonable interval, since theory tells us that varying threshold does not affect the shape parameter estimate, if the GP form is appropriate. This is indeed the case from inspection of Fig. 4. For $\sigma$, we expect from theory that the estimate is a linear function of threshold (with gradient $\gamma$) over a reasonable interval, if the GP form is appropriate. Again, this appears to be the case from inspection of Fig. 5. The expected behaviour with threshold also suggests that the GP model appears reasonable for fitting to data from individual sectors (but see Case Study 3) and that estimation of extreme quantile characteristics will be relatively insensitive to choice of thresholds between 4 and 8 m.

Next we adopt the constant model, fitting the GP distribution to all data regardless of sector. The variation of parameter estimates with threshold (Fig. 6) is now inconsistent with expected behaviour for GP-distributed data. The estimate for $\gamma$ is

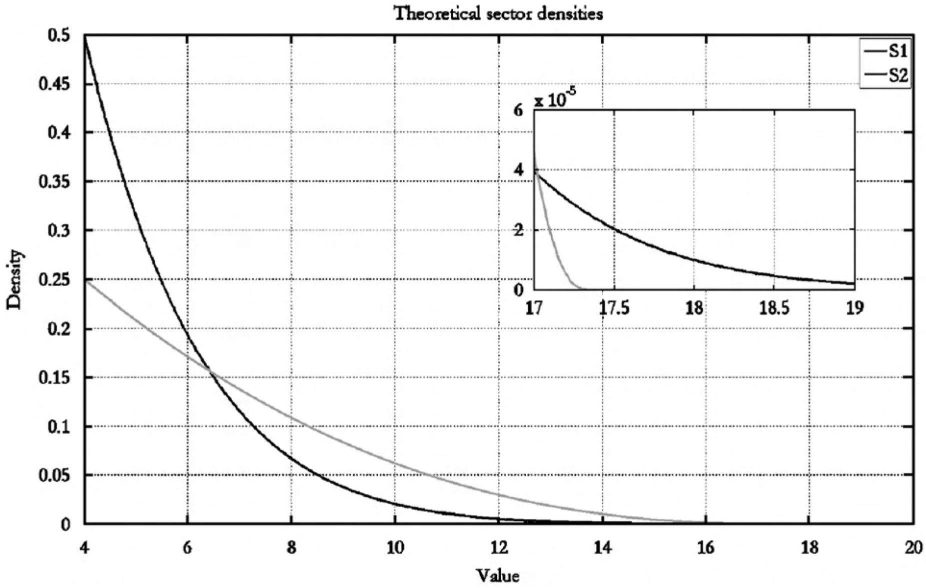


**Fig. 1.** Theoretical sector densities for Case Study 1, for sectors S1 (black) and S2 (grey).

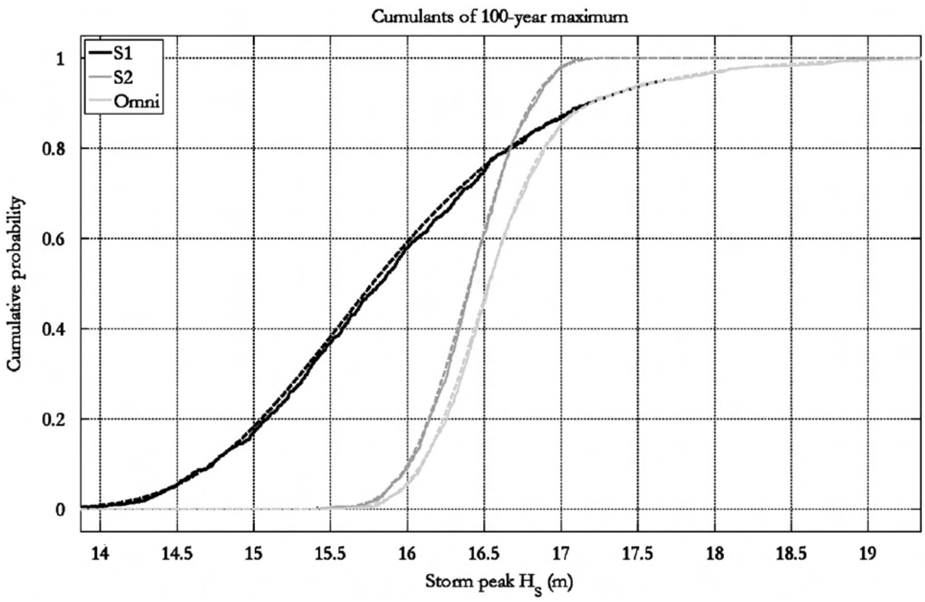


**Fig. 2.** Distribution of 100-year maximum from theory and simulation for Case Study 1, for sectors S1 (black), S2 (dark grey) and omni-directionally (light grey).

not constant with threshold, and the estimate for $\sigma$ varies non-linearly with threshold. We conclude that the GP distribution does not provide a reliable fit, and that choice of threshold will affect estimates for extreme quantiles.

To assess which of the constant or directional models is to be preferred for these data, we further visualise the quality of fit of the two models in various ways. For example, for a typical sample realisation of Case Study 1, Fig. 7 shows empirical cumulative distributions, obtained by simply ordering the sample data, allocating equal probability mass to each observation to obtain a distribution. The figure also shows corresponding fitted GP distributions, for individual sector fits (l.h.s., assuming the smallest threshold of 4 m, since this is consistent with a good GP fit from Figs. 4 and 5) and the fit to all data (r.h.s., with a threshold of 6m, chosen as a typical value for which we might consider fitting the constant model). The quality of fit of the two models cannot be easily distinguished. However, formal statistical testing is more informative and is the preferred approach to model selection. We use the likelihood ratio test (Kalbfleisch, 1985) to compare the adequacy of directional and constant models for Case Study 1. This test compares the goodness of fit of the respective models, whilst accounting for the added flexibility of the directional model. Specifically, the difference of the log likelihoods of the data, fitted with either model in turn and corrected for model complexity, is calculated. If the constant model is adequate, this difference should follow a $\chi^2$ distribution with degrees of freedom equal to the difference in numbers of parameters between the directional and constant models (since in essence the directional model contributes nothing new). If the constant model is inadequate, the difference of log likelihoods will be extreme in the $\chi^2$ distribution. Fig. 8 shows the estimated probability of rejecting the constant model in favour of the directional model, as a function of threshold. It is apparent that the directional model is a significantly better fit to the data than

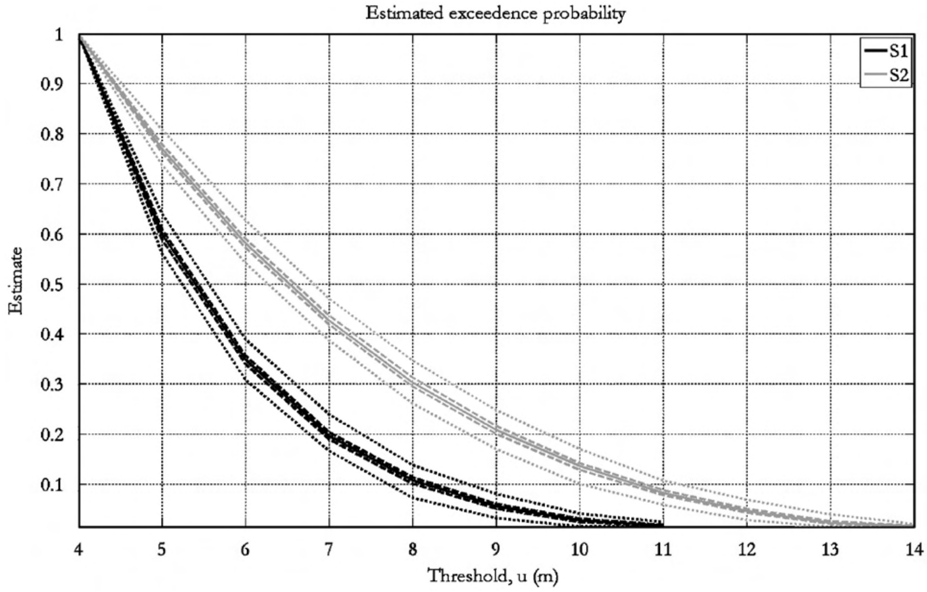


**Fig. 3.** Probability of threshold exceedance for each sector, as a function of threshold, in Case Study 1, for sectors S1 (black) and S2 (grey).

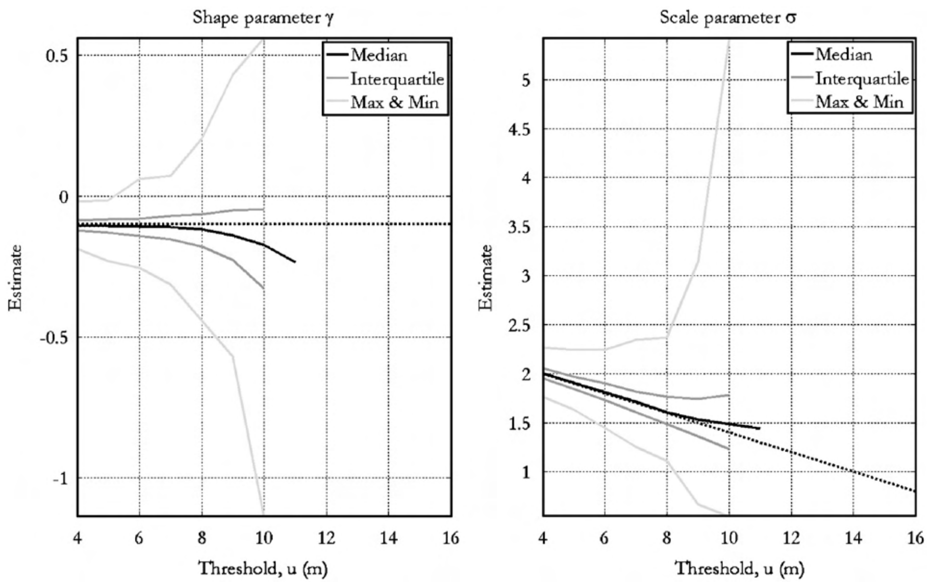


**Fig. 4.** GP parameter estimates for sector 1, as a function of threshold, in Case Study 1. The dotted line indicates expected behaviour for GP data. Interquartile uncertainty (dark grey) and maximum and minimum (light grey) also shown.

the constant model for all but the highest thresholds, for which the sample size is small. We conclude from the statistical diagnostic tests conducted that the directional model is preferable to the constant model for these data.

Finally, we use the directional and constant models to estimate statistical properties of the 100-year maximum event, and compare estimates obtained with theory (and direct simulation). For estimation using the directional model, we recommend using the smallest threshold consistent with a good GP fit (from Figs. 4 and 5) since this corresponds to the largest sample size for model fitting; however, for clarity in this discussion and fairness of comparison, we present estimates as a function of threshold. For the constant model, since the choice of threshold is unclear, we must examine estimates as a function of threshold. Fig. 9 shows estimates for the quartiles of the distribution of the 100-year maximum. The directional model is in excellent agreement with theory over a range of thresholds, whereas the constant model estimates vary with threshold. A fortunate choice of threshold around 9 m would give relatively good results using the constant model for this return period, but no objective a priori basis exists for threshold selection. It is clear that the constant model is less reliable that the directional model for these data, and that the directional model provides realistic parameter and quantile estimates over a range of thresholds.

Fig. 10 illustrates the uncertainty in the estimate of the median 100-year maximum in terms of the interquartile range for that estimate. The widths of uncertainties bands from directional and constant model are similar.

In this case study, for smaller thresholds, the constant model overestimates return values for a return period of 100 years. This bias is due to the fact that the sample data, known to come from a mixture of sector GP distributions, are fitted assuming them to be drawn from a GP distribution; this is not generally the case. Overestimation at 100-year level does not imply overestimation at

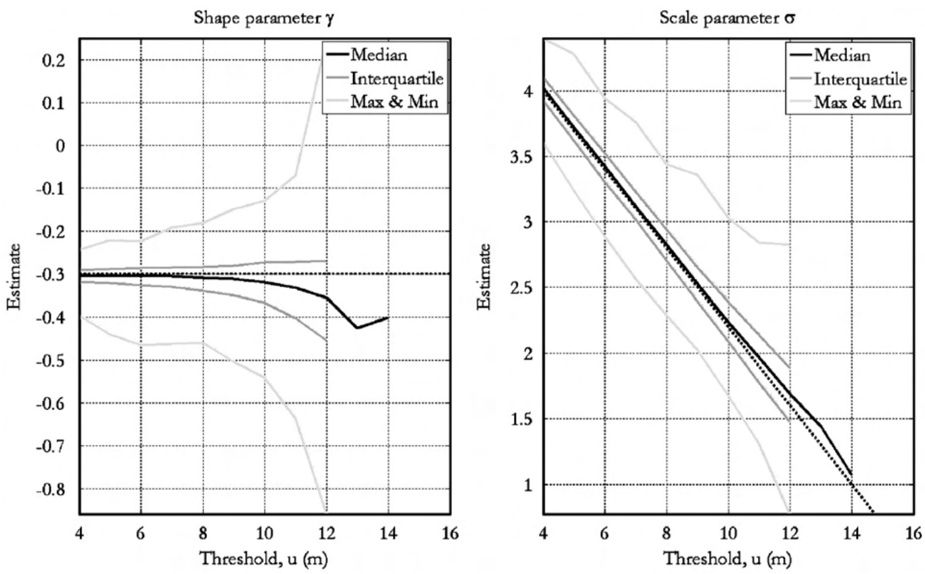


**Fig. 5.** GP parameter estimates for sector 2, as a function of threshold, in Case Study 1. The dotted line indicates expected behaviour for GP data. Interquartile uncertainty (dark grey) and maximum and minimum (light grey) also shown.

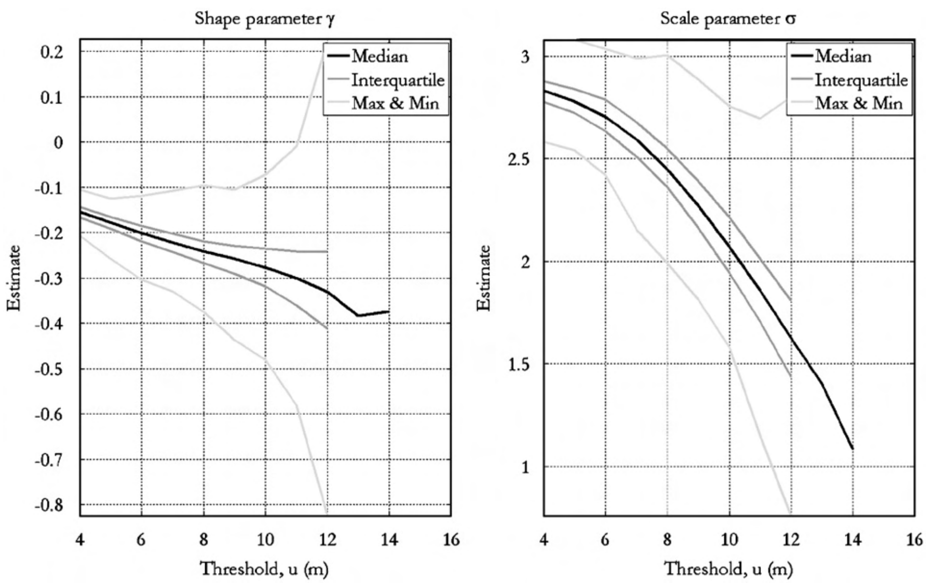


**Fig. 6.** GP parameter estimates for constant model, as a function of threshold, in Case Study 1.

other return periods as discussed in Section 6. As threshold varies, sample proportions from underlying sector distributions vary, as would return values, consistent with the findings of Carter and Challenor (1981). For the largest values of threshold, the sample is composed overwhelmingly of observations from the more extreme sector, which is GP distributed. We would therefore expect GP parameter estimates and return values from the constant model to be relatively unbiased, but with high variance due to the relatively small sample size available for fitting. In this case, since shape parameter $\gamma$ is negative in each sector, upper bounds for sectors S1 and S2 of $4 + 2/0.1 = 24$ m and $4 + 4/0.3 \simeq 17.3$ m exist. Thus, if we were able to set the extreme value threshold at approximately 17.3 m, we would be sure of having homogeneous GP data. However, we would need a period of approximately 1200 years to obtain a single observation in this case. By setting a reduced threshold of approximately 9.6 m, we obtain a sample of size 250 with approximately 20% of all observations from sector S1 on average in 25 years (as can be seen from Fig. 3).

### 4.2. Case Study 2

In the first case study, we were blessed with a large sample. Case Study 2 is identical to Case Study 1 except that the number of occurrences in the period $P_0 = 25$ years is reduced from $n = 2500$ to 250, with exactly 125 from each sector. Smaller samples increase modelling uncertainty generally, as seen in Fig. 11, which shows (in comparison with Fig. 8) that only for the smallest thresholds (less than 6 m) can the inadequacy of the constant model be reliably detected. Nevertheless, for small thresholds the directional model is preferred. Plots of parameter estimates with threshold (not shown) also lead us to doubt the adequacy of the constant model.

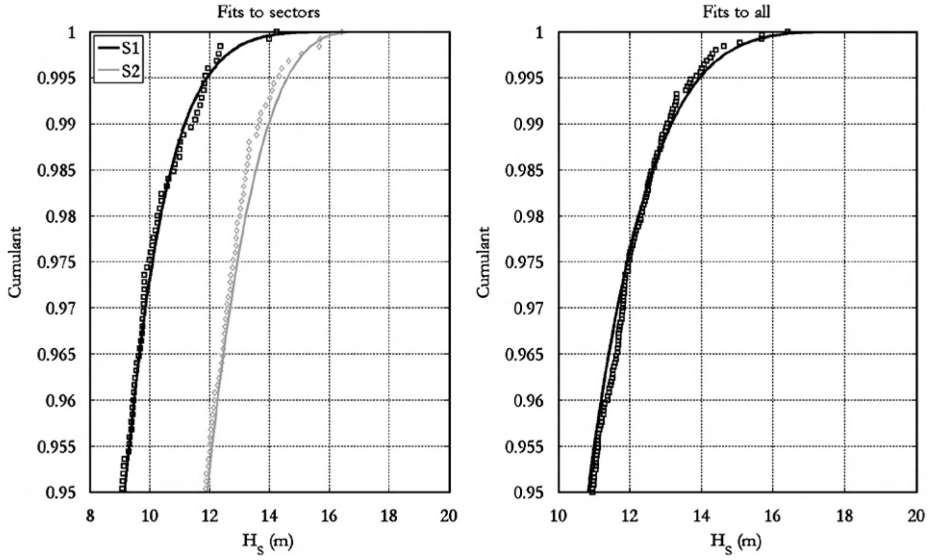


**Fig. 7.** Comparison of tails of empirical and fitted cumulative distribution functions, for a typical sample realisation in Case Study 1. Fits to sectors S1 and S2 (black and grey, respectively, l.h.s.) and to all data (r.h.s.).

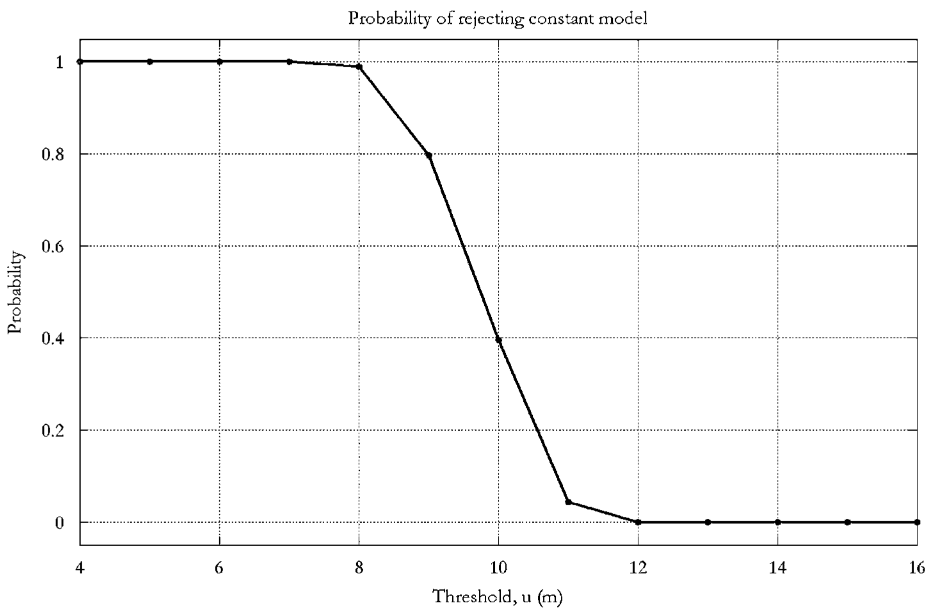


**Fig. 8.** Probability of rejecting constant model, as a function of threshold, in Case Study 1.

Estimates for quartiles of the 100-year maximum are illustrated in Fig. 12 (following the format as Fig. 9). Estimates from the directional model are more stable with threshold, and generally more accurate than those obtained using the constant model. But for the 100-year return period, the difference in the value of estimates obtained is rather small. The uncertainty of the median 100-year maximum is shown in Fig. 13.

### 4.3. *Case Study 3*

We observe a sample of $n = 2500$ storm peak $H_S$ events in a period $P_0 = 25$ years, 1250 events in each of directional sectors S1 and S2. The sector GP parameters are taken to be $E_1 = (-0.1, 3, 4)$ and $E_2 = (-0.3, 4, 6)$. Extreme values for sectors S1 and S2 are bounded on the right-hand side at 34 and 18.33 m, respectively. The marginal densities of events per directional sector are plotted in Fig. 14. Sector exceedance probabilities (from 1000 realisations of the sample) are given in Fig. 15. The parameter values for $E_1$ and $E_2$ were chosen so that domains of sample data for each sector be similar.

To assess model fit, we examine the behaviour of GP parameter estimates with threshold for fitting to individual sectors and overall. From Figs. 16–18 it is not at all clear that the directional model is preferable even at lower thresholds. Parameter estimates vary with threshold as would be expected for GP distributed data, over a range of threshold values, for both directional and constant models.

However, the likelihood ratio test results (Fig. 19) provide clear evidence that the directional model is preferable to the constant model at thresholds below 8 m.

Estimates for the quartiles of the 100-year maximum event are illustrated in Fig. 20, which follows the format as Figs. 9 and 12. The uncertainty of the median 100-year maximum is illustrated in Fig. 21.

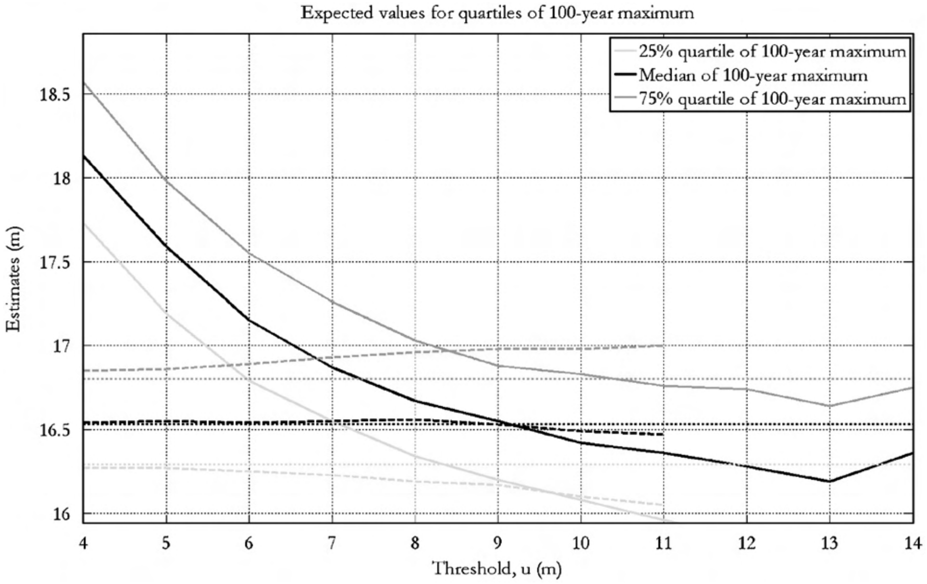


**Fig. 9.** Estimates for quartiles of 100-year maximum, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates with threshold, for Case Study 1.

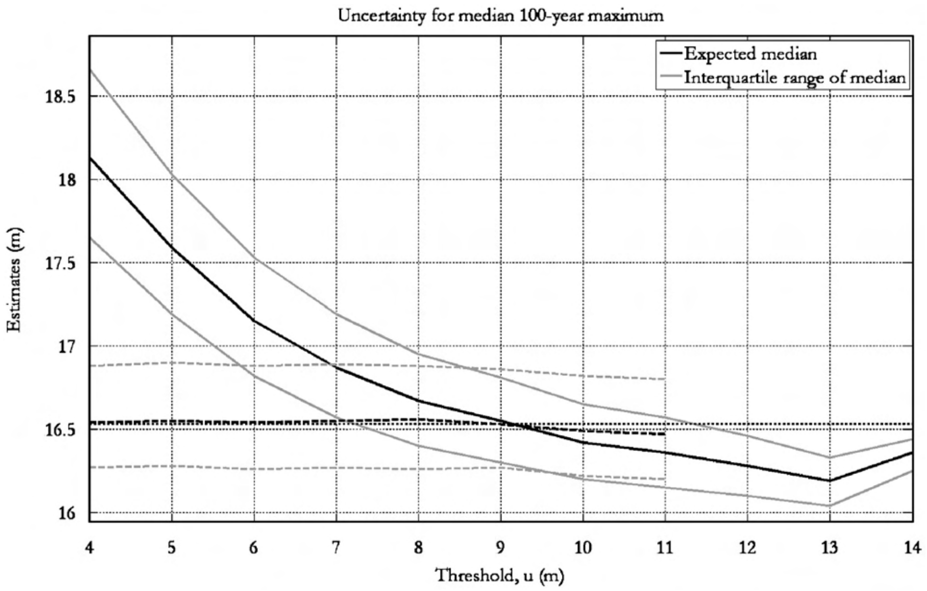


**Fig. 10.** Interquartile uncertainty for median 100-year maximum, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates with threshold, for Case Study 1.

A number of characteristics of Figs. 20 and 21 is consistent with observations from Case Studies 1 and 2. Notably, for lower thresholds in Fig. 20 corresponding to largest sample sizes for fitting, the directional model is considerably more accurate than the constant model. At a threshold of 6 m, the directional model underestimates the true median 100-year value by less than 0.2 m, whereas the constant model underestimates by more than 1.5 m. For all thresholds less than 10 m, the directional model is within 1 m of the truth. The upper (75%) quartile of the 100-year maximum distribution is also well-estimated by the directional model. A cautionary note in this case study is the behaviour of parameter estimates with threshold for fitting the constant in Fig. 18. The behaviour observed is that which would be expected if the data fit the GP model well, yet we know this not to be the case. Statistical testing, the likelihood ratio test, here, should be used for model selection. Parameter behaviour with threshold is used primarily to select an appropriate value for threshold.

The poor performance of the constant model in this case can be explained as follows. Sector S1 ($\gamma = -0.1$) is more extreme than S2 ($\gamma = -0.3$). However, by construction, the proportion of observations in the data sample from sector S2 is higher than from S1 even for the highest thresholds. We know asymptotically that the proportion of observations from S1 would increase with threshold, eventually to dominate. In this sense, S1 is always under-represented in the sample. Since the maximum values in the sample are also approximately equal (again by construction), the constant model tends to compromise between S1 and S2 at all thresholds. If data corresponding to a longer period were

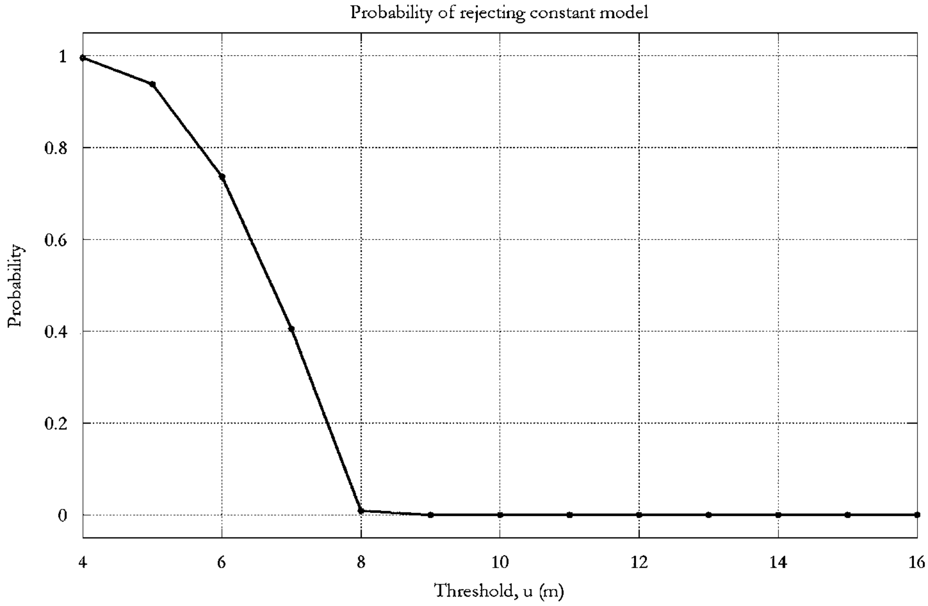


**Fig. 11.** Probability of rejecting constant model, as a function of threshold, in Case Study 2.

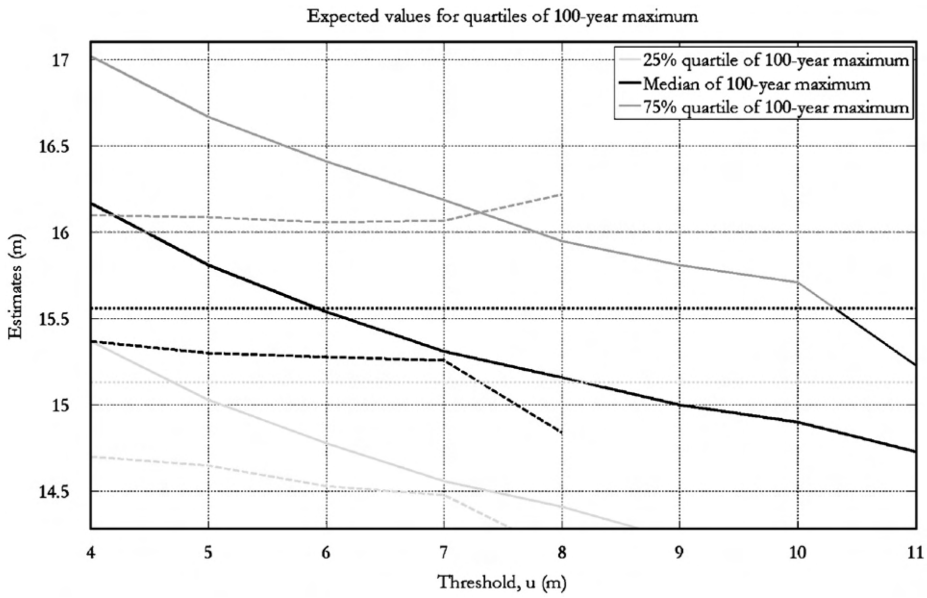


**Fig. 12.** Estimates for quartiles of 100-year maximum event, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates are given as a function of threshold, for Case Study 2.

available, the two sectors would be more easily distinguishable at high thresholds by the constant model. Nevertheless, the directional model is clearly preferable in this case study.

### 4.4. Other case studies

In addition to the case studies reported here, a number of other cases have been examined. Details are not reproduced here, but are available from the authors. Noteworthy are the following. When the sectors have identical extreme value parameters, the likelihood ratio test indicates that the constant model is adequate for all thresholds. Parameter estimates vary with threshold as would be expected for GP data for both directional and constant models. Estimates for distributional characteristics of the 100-year maximum from the directional and constant models are very similar, but the constant model should be preferred since it is more parsimonious. When sector distributions are assumed to be Weibull rather than GP, and fitting is performed using the Weibull distribution, effects analagous to those reported here are observed. When Weibull-distributed data are modelled using GP, bias in extreme quantile estimates due to both model mis-specification and neglect of directional effects are observed. Only asymptotically can reasonable estimates be anticipated (see, e.g., Elsinghorst et al., 1998).

## 5. Discussion

The period of data available for extreme value estimation is usually shorter than the return period of interest. Therefore the analysis entails extrapolation beyond the domain of measurements.

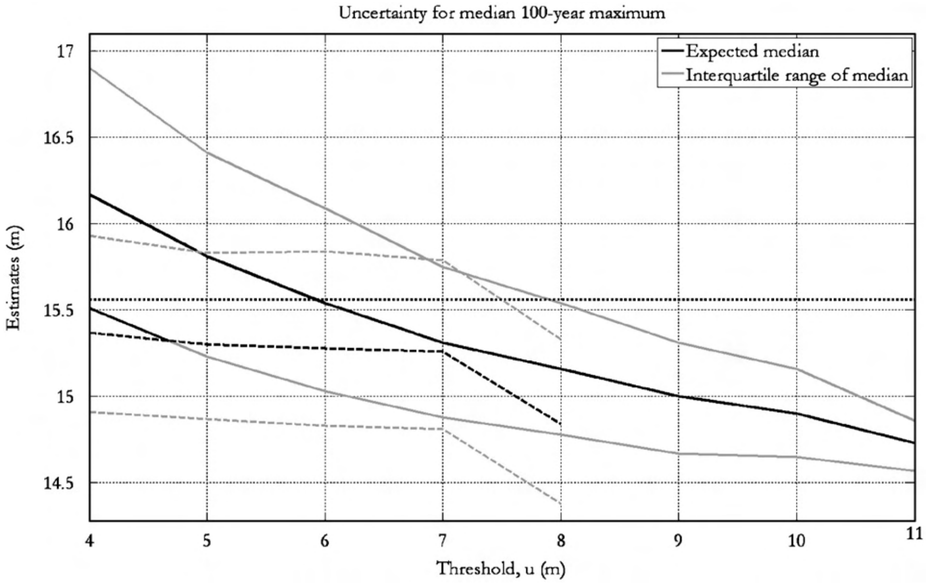


**Fig. 13.** Interquartile uncertainty for median 100-year maximum, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates given as a function of threshold, for Case Study 2.

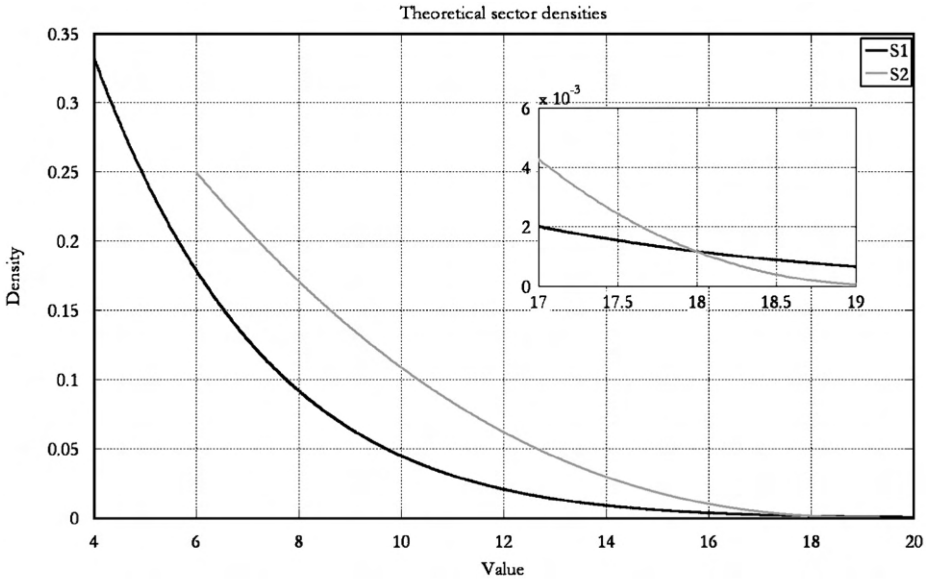


**Fig. 14.** Theoretical sector densities for Case Study 3, for sectors S1 (black) and S2 (grey).

Even after assuming homogeneity and regularity, uncertainties associated with estimates will be high compared with the usual and preferred case in statistical modelling corresponding to interpolation between data within the domain of the sample (Anderson, 1990). In practice, extreme events with low rates of occurrence, not seen during the period of observation, may have an important influence on return values. If 1000 years of data were available with which to estimate the distribution of 100-year maximum at a given location over the past millennium, progress would be relatively straight forward, even in the presence of directional, seasonal and possible long-term variation with time. This regrettably is not usually the case in offshore design.

Extreme value analysis involves modelling the most unusual events in the sample, rather that the typical. Uncertainties of extreme values (from measurement or hindcast) are likely to be different to those of the bulk of the data. The few largest values in the sample are typically highly influential for estimates (Davison and Smith, 1990). The model is most sensitive to the most informative observations. Compounded with limited sample size, selection of the optimal model form that characterises the observed extreme values is therefore difficult. Minor differences in model form or parameter estimates, which would be of little consequence for interpolation within the domain of the data, can produce material discrepancies in return values, especially at long return periods. To alleviate sample size problems, it is appealing to combine data from different sources, e.g. from different locations in the same neighbourhood, anticipated to have similar extreme value behaviour, including covariates as necessary to explain differences in behaviour. This increases sample size at the possible expense of introducing dependency which must be treated carefully, especially with respect to estimation of parameter and extreme

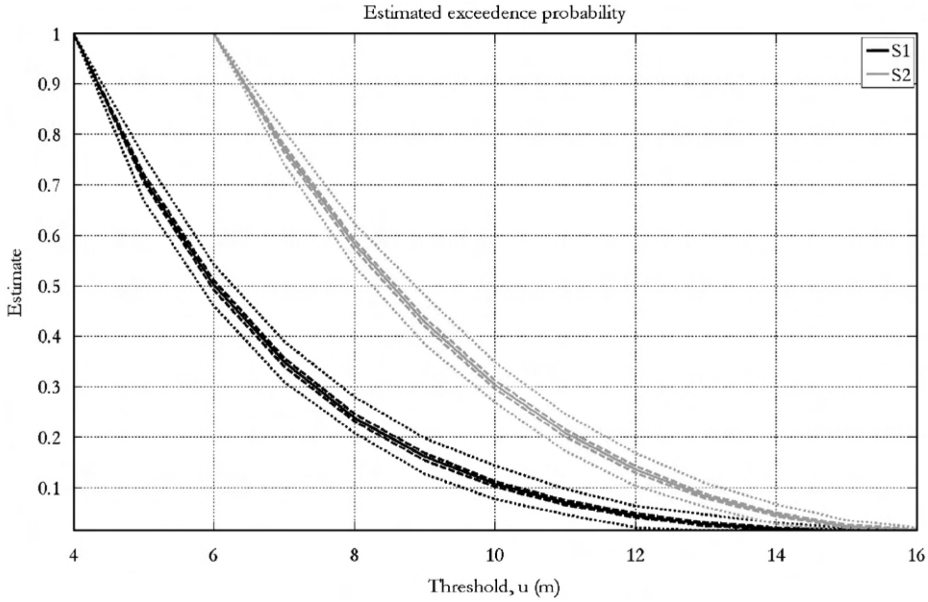


**Fig. 15.** Probability of threshold exceedance for each sector, as a function of threshold, in Case Study 3, for sectors S1 (black) and S2 (grey).

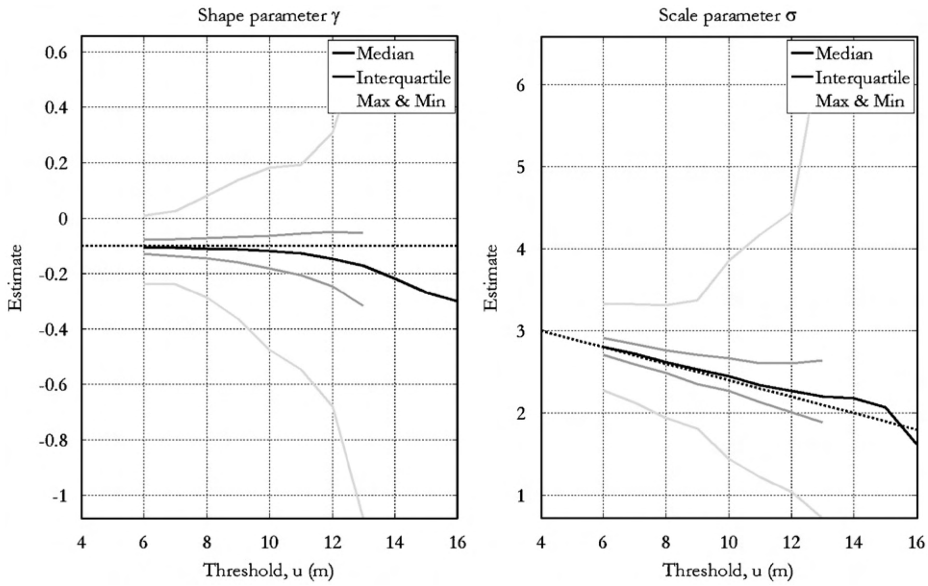


**Fig. 16.** GP parameter estimates for sector 1, as a function of threshold, in Case Study 3. The dotted line indicates expected behaviour for GP data.

quantile uncertainty. Block bootstrapping in an important tool in this respect.

Extreme value theory shows that, asymptotically, distributions of threshold exceedances converge to the GP distribution given a sufficiently high threshold, provided the limit exists. Fitting samples of extreme values using the GP distribution is therefore only defensible given that the threshold used is large enough. Using diagnostic tests, such as the mean residual life plot (e.g. Embrechts et al., 2003), comparing the variability of parameter estimates with expected behaviour as threshold is varied and confirming the stability of extreme quantile estimates with threshold, the plausibility of the GP model form can be ascertained (e.g. Coles, 2001). However, there is no direct evidence that the chosen threshold is sufficiently high that the GP distribution is an adequate model. In general, it would be advantageous to have sufficient understanding of the underlying physics to suggest an appropriate model selection, so that as much of the sample as possible (not just the largest values) would aid parameter estimation.

The GP distribution is the limiting form for threshold exceedances for all distributions. The shape parameter $\gamma$ determines tail behaviour. When $\gamma$ is negative, the domain of extreme values is bounded on the right-hand side. Otherwise, the distribution is unbounded on the right. Many practitioners in engineering disciplines adopt specific model forms to fit extreme values, notably variants of the Weibull distribution. There is often little evidence in the data to distinguish between quality of fit for different model forms. Yet asymptotically, the Weibull distribution is of the Gumbel type (with $\gamma = 0$); that is, fitting using the Weibull distribution implicitly constrains the solution to be unbounded on the right-hand side. Modelling using GP avoids this constraint. For this reason, fitting using Weibull and GP are expected to given different estimates for return values especially at long return periods.

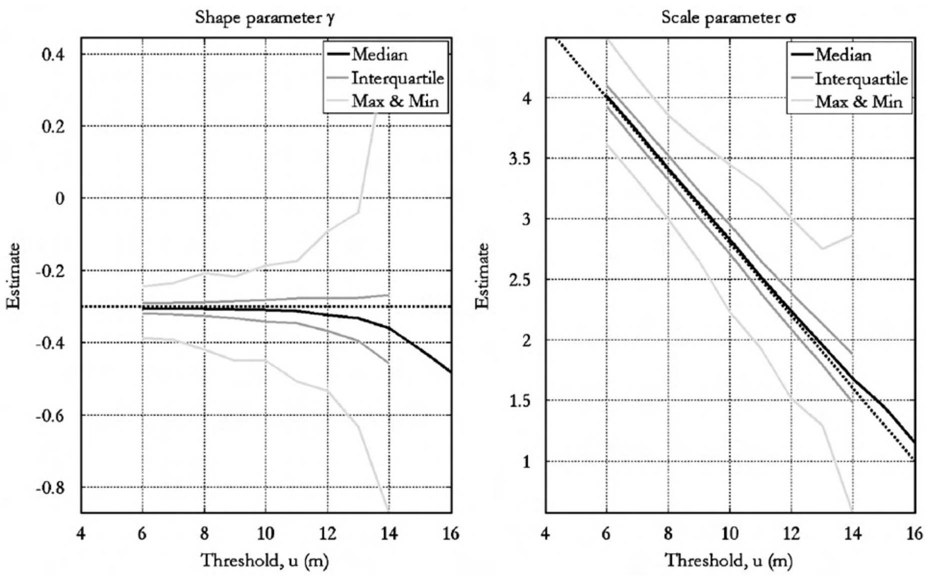


**Fig. 17.** GP parameter estimates for sector 2, as a function of threshold, in Case Study 3. The dotted line indicates expected behaviour for GP data.

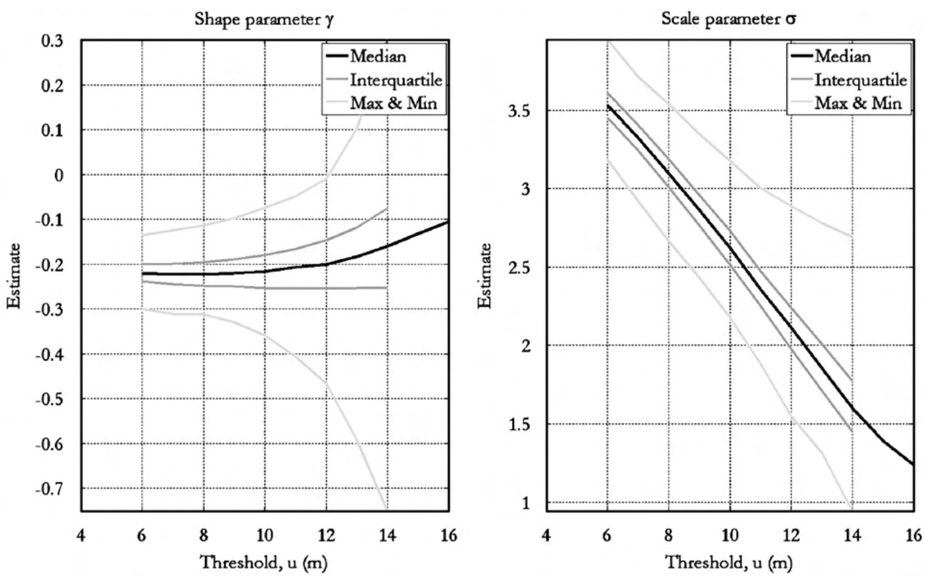


**Fig. 18.** GP parameter estimates for constant model, as a function of threshold, in Case Study 3.

An important aim in extreme value analysis is the estimation of an extreme quantile corresponding to a particular return period, *P*, or the maximum value observed in *P* years. Typically, a single value is reported, such as the value of $H_S$ which is exceeded with probability $1/P$ per annum. Yet the *P*-year maximum is itself a random variable. Given perfect model specification and parameter estimation, the *P*-year maximum event will follow a skewed distribution with long right-hand tail. Model and parameter uncertainty broaden this distribution. Typically, the value reported is the mode of this distribution, which because of skewness, is less than both the expected and median values. Further, the ratio of the 95th percentile of the distribution of the *P*-year maximum to its mode also varies systematically with the shape parameter of the GP used. This ratio is greater in Gulf of Mexico conditions ($\gamma$ near $-0.1$) than in the Northern North Sea conditions ($\gamma = -0.3$). That is, for the same value of most probable 100-year $H_S$, we expect more exceptionally large values in the Gulf than in the Northern North Sea (Jonathan and Ewans, 2008).

A typical storm is not uni-directional, generating peak over threshold events in a number of directions, not just the storm peak direction. In estimating directional design criteria, we must also account for the influence of storms over their full range of sea states and wave directions, e.g. using directional dissipation (Jonathan and Ewans, 2007).

The requirement for modelling of covariate effects must be considered in light of all these issues. When physical considerations such as fetch variability with direction, seasonal cycles of storm severity, variation of water depth or climate change is suspected to be present, inclusion of the corresponding covariates is likely to be important for good model fit. Conversely, it is important also not to ‘overfit’ by including covariate effects which are not supported by the data. Statistical tests can be used to select formally between models with or without covariate terms based on their goodness of fit to the sample. Techniques such as cross-validation offer one approach to ensuring parsimony in terms of predictive performance. The behaviour of parameter and

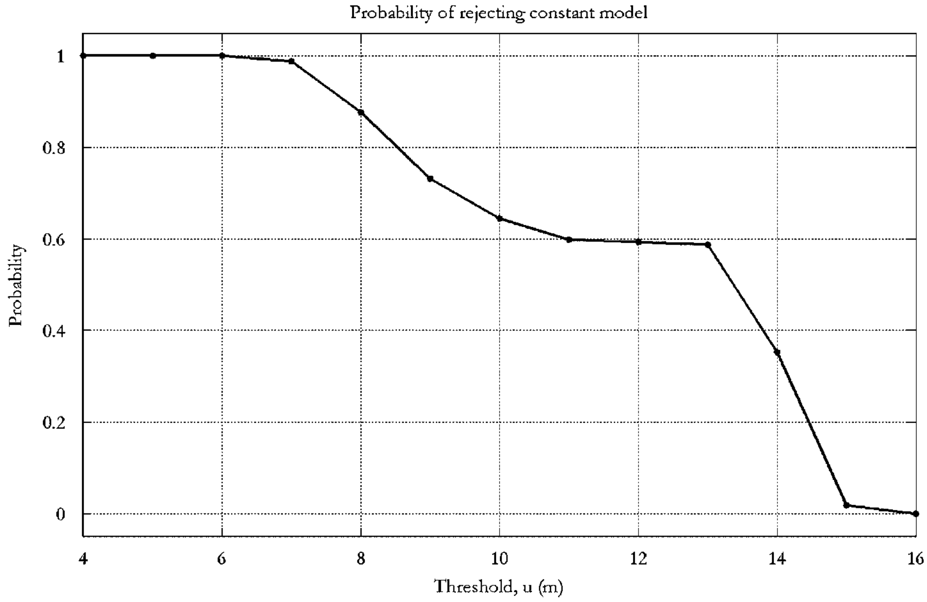


**Fig. 19.** Probability of rejecting constant model, as a function of threshold, in Case Study 3.

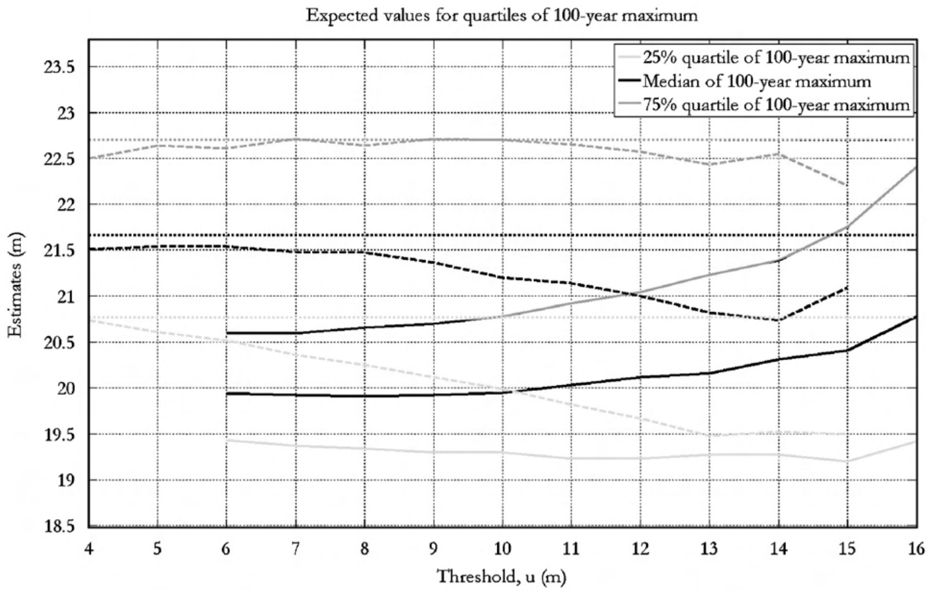


**Fig. 20.** Estimates for quartiles of 100-year maximum, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates are given as a function of threshold, for Case Study 3.

quantile estimates with threshold should be examined and compared with expectation from theory and previous application. The stability of quantile estimates for directional sectors with respect to threshold and small changes to sector boundary specifications should also be examined. As illustrated in this work, neglect of covariates can lead to biased estimates of extreme quantiles. The extent of any bias cannot be anticipated a priori in general. Moreover, the relative size and direction of biases due to neglect of covariate(s) compared to other sources (e.g. mis-specification of threshold, of model form, outliers, sample size) is also difficult to estimate before the analysis is performed. Nevertheless, failure to model covariate effects adequately when there is evidence for their presence will likely compound errors.

In the current work, with extremes GP-distributed in each of sectors S1 and S2, an arbitrary extreme event is distributed as a mixture of two GP distributions. For a sufficiently high threshold, threshold exceedances will follow (at least approximately) the GP distribution corresponding to the sector with highest value for $\gamma$. When the sample is sufficiently large (in Case Study 1, e.g.), this threshold is prohibitively high, corresponding to a very long return period. For smaller threshold values, only the directional model can adequately describe the sample and provide realistic quantile estimates. We have concentrated on estimation of the properties of the 100-year maximum, but longer return periods are sometimes also of interest. Fig. 22 gives estimates for the median of the $P$-year maximum for values of $P$ between a hundred and a million years, for Case Study 1. The figures illustrate that directional model estimates (using the smallest value of threshold available consistent with GP) agree reasonably will theory, whereas estimates from the constant model (for thresholds of $6, 7, \ldots, 12\,\mathrm{m}$) are generally inaccurate.

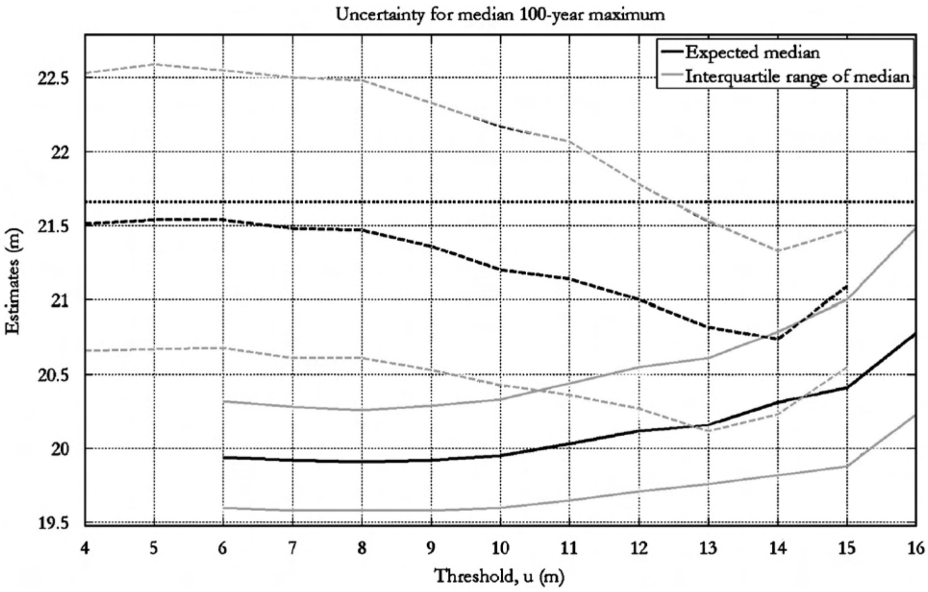


**Fig. 21.** Interquartile uncertainty for median 100-year maximum, from theory (dotted), directional (dashed) and constant (solid) models. Directional and constant model estimates given as a function of threshold, for Case Study 3.

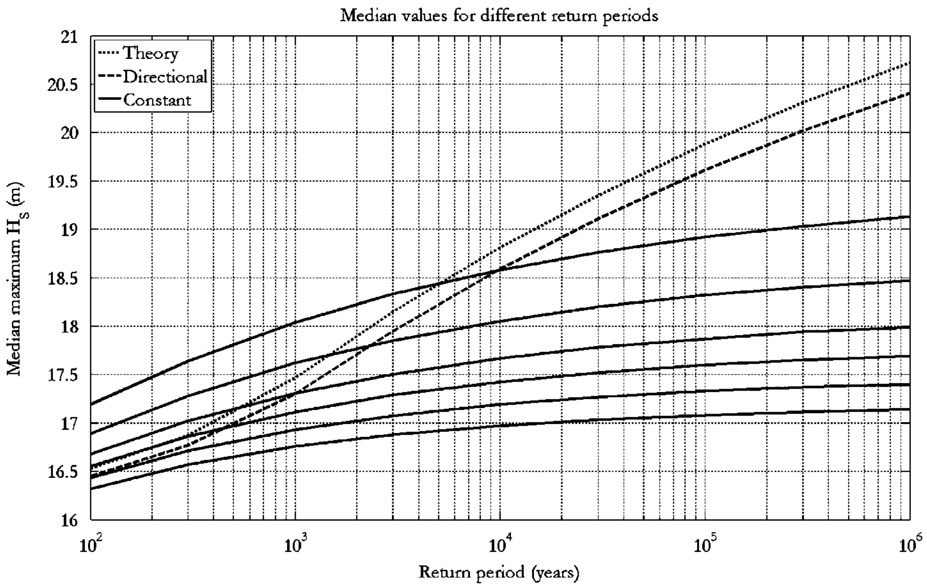


**Fig. 22.** Median return values for different return periods, from theory (solid black), directional model fitting (dashed black) and constant model fitting for different thresholds (grey), for Case Study 1. For constant model, curves (from top to bottom) correspond to thresholds of 6, 7, ..., 11 m.

In the case studies, we have chosen to fit either the directional or constant models, using the likelihood ratio test for model selection. We might enhance the set of candidate models further, since there may be models of intermediate complexity which are more efficient. For example, we might consider models for which the extreme value shape is assumed constant with direction with scale allowed to vary (or vice versa).

## 6. Conclusions and recommendations

We have demonstrated that model-fitting to a heterogenous sample ignoring covariate effects leads to biased estimates of return values, whereas models incorporating covariate effects provide relatively unbiased estimates. For data that exhibit directional effects, we show that a directional extreme value model generally explains observed variation significantly better than a model which ignores directionality, and that omni-directional criteria developed from a directional model are different from those generated when directionality is not accounted for. Finally we show that omni-directional criteria derived from a directional model are more accurate and should be preferred in general over those based on models which ignore directional effects.

We recommend the use of directional extreme value models for estimation of both directional and omni-directional design criteria in future, when good directional data are available. When the effects of other covariates (e.g. time or space) are suspected, we similarly recommend the use of extreme value models which adequately capture sources of covariate variability for all design analysis.


## Acknowledgements

The authors acknowledge useful discussions with Clive Anderson, Idris Eckley, Jonathan Tawn, Peter Tromans and Michael Vogel. The authors further acknowledge the support of Shell International Exploration and Production and Shell Research Ltd.


## Appendix A. Conditional and unconditional distribution to estimate extreme quantiles

Assume $n$ independent storm events occurring in period $P_0$. Extremal characteristics vary with covariate $\theta$. Partition the domain of $\theta$ into $K$ intervals $\{I_k\}_{k=1}^K$ such that the number of occurrences in $I_k$ is $n_k$ (so that $\sum_{k=1}^K n_k = n$). We are interested in the distribution of the maximum storm peak event $X_{\max}$ for period $P$, $P > P_0$, given the distribution of storm peaks $X$ over threshold $u_k$ in each of the $K$ covariate intervals. We show informally that the distribution of $X_{\max}$ is obtained equivalently from either the conditional or unconditional distribution of storm peak.

For any interval, the unconditional distribution function of storm peak $X$ can be written

$$\begin{aligned}
&\Pr(X \leqslant x \mid \theta \in I_k) \\
&\quad = \Pr(X \leqslant u \mid \theta \in I_k) \\
&\qquad + \Pr(X > u \mid \theta \in I_k) \Pr(X \leqslant x \mid X > u \mid \theta \in I_k) \\
&\quad = (1 - \alpha_k) + \alpha_k F_{Ck}(x) \\
&\quad = 1 - \alpha_k (1 - F_{Ck}(x))
\end{aligned}$$

where $\alpha_k$ is the probability of exceedance of threshold $u_k$ for interval $I_k$, and $F_{Ck}(x)$ is the conditional distribution of $X$ given $X > u_k$ in $I_k$.

In period $P$, the distribution $F_{X_{\max}}$ of the maximum storm peak $X_{\max}$ is given by

$$F_{X_{\max}}(x) = \prod_{k=1}^K \sum_{j=0}^{\infty} \Pr(X \leqslant x \mid M_k = j) \Pr(M_k = j)$$

where $M_k$ is the number of storms in interval $I_k$ in period $P$. In interval $I_k$, we expect to observe $m_k = (P/P_0) n_k$ peaks over threshold $u_k$ with probability $\alpha_k$ in $P$ years. Therefore we expect to observe (unconditionally) $m_k/\alpha_k$ storm peaks in the same period (so that the expected number of exceedances of the threshold is $(m_k/\alpha_k)\alpha_k = m_k$). Assuming the number of storms $M_k$ to be Poisson-distributed with expectation $m_k/\alpha_k$ per interval $I_k$, we have

$$\begin{aligned}
F_{X_{\max}}(x) &= \prod_{k=1}^K \sum_{j=0}^{\infty} (1 - \alpha_k (1 - F_{Ck}(x)))^j \left(\frac{m_k}{\alpha_k}\right)^j \frac{1}{j!} \mathrm{e}^{-m_k/\alpha_k} \\
&= \prod_{k=1}^K \mathrm{e}^{-m_k (1 - F_{Ck}(x))}
\end{aligned}$$

Conversely, we calculate the distribution $F^*_{X_{\max}}$ of the maximum storm peak in period $P$ using the conditional distribution $F_{Ck}$ of storm peaks over threshold directly

$$F^*_{X_{\max}}(x) = \prod_{k=1}^K \sum_{j=0}^{\infty} \Pr(X \leqslant x \mid X > u_k \mid N_k = j) \Pr(N_k = j)$$

where $N_k$ is the number of storm peaks over threshold $u_k$ in interval $I_k$ in period $P$. Assuming $N_k$ to be Poisson-distributed with expectation $m_k$, we have

$$\begin{aligned}
F^*_{X_{\max}}(x) &= \prod_{k=1}^K \sum_{j=0}^{\infty} (F_{Ck})^j (x) (m_k)^j \frac{1}{j!} \mathrm{e}^{-m_k} \\
&= \prod_{k=1}^K \mathrm{e}^{-m_k (1 - F_{Ck}(x))}
\end{aligned}$$

The expressions for $F_{X_{\max}}$ and $F^*_{X_{\max}}$ are identical regardless of functional form $F_{Ck}$.

## References


Anderson, C.W., 1990. Discussant of Davison and Smith (1990), referenced herein (Royal Statistical Society).

Anderson, C.W., Carter, D.J.T., Cotton, P.D., 2001. Wave climate variability and impact on offshore design extremes. Report Commissioned from the University of Sheffield and Satellite Observing Systems for Shell International.

API, 2005. API Recommended Practice 2A-WSD (RP 2A-WSD), Recommended Practice for Planning, Designing and Constructing Fixed Offshore Platforms—Working Stress Design. API.

Carter, D.J.T., Challenor, P.G., 1981. Estimating return values of environmental parameters. Quarterly Journal of the Royal Meteorological Society 107, 259.

Chavez-Demoulin, V., Davison, A.C., 2005. Generalized additive modelling of sample extremes. Journal of the Royal Statistical Society Series C: Applied Statistics 54, 207.

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

Coles, S.G., Walshaw, D., 1994. Directional modelling of extreme wind speeds. Applied Statistics 43, 139.

Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. Journal of the Royal Statistical Society B 52, 393.

Elsinghorst, C., Groeneboom, P., Jonathan, P., Smulders, L., Taylor, P.H., 1998. Extreme value analysis of North Sea storm severity. Journal of Offshore Mechanics and Arctic Engineering 120, 177.

Embrechts, P., Klueppelberg, C., Mikosch, T., 2003. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Ewans, K.C., Jonathan, P., 2008. The effect of directionality on Northern North Sea extreme wave design criteria. Journal of Offshore Mechanics and Arctic Engineering, accepted October 2007.

Forristall, G.Z., 2004. On the use of directional wave criteria. Journal of Waterway, Port, Coastal and Ocean Engineering 130, 272.

ISO, 2005. BS EN ISO 19901-1:2005. Petroleum and natural gas industries—specific requirements for offshore structures—part 1: metocean design and operating conditions. British Standards Institute.

Jonathan, P., Ewans, K.C., 2007. The effect of directionality on extreme wave design criteria. Ocean Engineering 34, 1977–1994.

Jonathan, P., Ewans, K.C., 2008. Uncertainties in extreme wave height estimates for hurricane dominated regions. Journal of Offshore Mechanics and Arctic Engineering 129, 300.

Kalbfleisch, J.G., 1985. Probability and Statistical Inference. Statistical Inference, vol. 2. Springer, Berlin.

Robinson, M.E., Tawn, J.A., 1997. Statistics for extreme sea currents. Applied Statistics 46, 183.