C. Elsinghorst

P. Groeneboom

Delft University of Technology,
The Netherlands

P. Jonathan
Shell Research Thornton,
P.O. Box 1,
Chester CH1 3SH, U.K.

L. Smulders
Delft University of Technology,
The Netherlands

P. H. Taylor
Shell Research Rijswijk,
The Netherlands

Extreme Value Analysis of North Sea Storm Severity

In this paper we consider the estimation of North Sea storm severity, for storms with return periods in the interval 100 to 500 yr. The analysis consists of: modeling the tail-distribution for a set of data for storm severity (using, e.g., storm hindcast data); estimating extreme storm severity; estimating confidence intervals for extreme storm severity; validating the bias and variance of estimates using simulation studies, for known underlying model forms; and estimating the robustness of extreme quantile estimates with respect to misspecification of the underlying model for the tail-distribution of storm severity. Applications to NESS (Northern European Hindcast Study) hindcast data at clusters of locations in the northern, central and southern North Sea are considered. Results suggest, in particular, the existence of a physical upper limit for storm severity in the North Sea and a close to constant value for the extreme value index, $\gamma \approx -0.2$.

1 Introduction

Design and re-assessment of offshore structures requires the modeling of extreme ocean environments and structural loads, which occur on average as seldom as once every 100–10,000 yr. In the North Sea, meteorological and oceanographic records extend over a few decades. Therefore, we need to extrapolate to environmental conditions far beyond the domain of the set of measurements.

As a basis for extreme value estimation, we have used data sets of significant wave height from the NESS hindcast (Peters et al., 1993). NESS uses 25 yr of meteorological data (atmospheric pressure, wind speed, etc., at various locations) for the period 1964–1989 as input to a wave-field model from which met-ocean conditions are estimated once every 3 h, on a grid of locations throughout the North Sea. The NESS hindcast contains 3-h sequential estimates for the significant wave height H_S (equivalent to average energy density per unit area) within the wave field. We identify each storm as an interval within which H_S exceeds a certain threshold. The severity of each storm is then characterized by the most probable maximum wave height (H_{MP}) within the storm (Tromans and Vanderschuren, 1995).

We have examined hindcast data for five adjacent NESS grid locations (forming a cross configuration with a central point) in each of the northern, central, and southern North Seas (denoted henceforth as NNS1-NNS5, CNS1-CNS5, SNS1-SNS5, respectively), and used various extreme value models to make predictions of extreme storm severity. In the NNS, adjacent grid locations are 30 km apart, but in the CNS and SNS, a spacing of 10 km was employed (Peters et al., 1993).

An outline of the analysis is given in the forthcoming. The first step is to fit a parametric Pareto model to the tail of the distribution of storm severity (estimated from the hindcast data). This is described in detail in Section 2. Particular attention needs to be paid to estimating the variability of model parameters; we use a studentized bootstrap resampling scheme. The next step is to estimate extreme (e.g. 10^n-yr) storm severities and appropriate confidence intervals using the fitted model. This is described in Section 3.

In calculating confidence intervals for model parameter and extreme quantile estimates, a number of theoretical assumptions are necessary. However, we can use brute-force simulation studies to test how good our confidence intervals actually are. In Section 4, we report the results of extensive simulations used to estimate coverage probabilities for the interval estimates for model parameters and extreme quantiles, assuming always that the true underlying distribution is GPD. In fitting a particular model to the hindcast data, we naturally assume that the model is appropriate for the hindcast data, but there is no way to prove this in general. We would, therefore, like to estimate the robustness of our model parameter and extreme quantile estimates with respect to (small) misspecifications of the underlying model. We do this, again by means of simulation, in Section 4. Finally, conclusions are drawn in Section 5.

This paper documents an approach to the estimation of extreme storm conditions for locations in the North Sea. The results quoted should be viewed as examples of the methodology: estimates not polished values for use in design. We have made no allowances for the severe storms which occurred after the NESS period, bias and noise in the data set and/or in comparable measurements and other modifications commonly made on the basis of the intuition of the oceanographic community (see Maes and Gu, 1995).

2 Fitting a Model for the Tail Distribution

For each of the 15 NESS grid locations used in the present study, we have a sample of at least 500 storm severities. For each sample, the empirical distribution function is shown in Fig. 1. From the figure, a number of features are evident. In particular, in the CNS, the distributions appear to be very similar, whereas in the SNS there are marked differences between the distributions for adjacent grid locations.

It can be shown, for a wide class of distributions, that the excesses over some high threshold tend to be distributed according to a generalized Pareto distribution (GPD), provided that the distribution is in the domain of attraction of one of three extreme value distributions (Leadbetter et al.,

Contributed by the OMAE Division for publication in the JOURNAL OF OFFSHORE MECHANICS AND ARCTIC ENGINEERING. Manuscript received by the OMAE Division, September 1997; revised manuscript received March 1998. Associate Technical Editor: C. Guedes Soares.

1983; Pickands, 1975). In this study, we fit the GPD with distribution function

$$F_{\mathrm{GPD}}(x;\gamma,\sigma) = \begin{cases} 1 - \left(1 + \dfrac{\gamma(x-u)}{\sigma}\right)^{-1/\gamma} \\ \text{for } \gamma \neq 0,\ x > u,\ \sigma + \gamma(x-u) > 0, \\ 1 - \exp\left(-\dfrac{(x-u)}{\sigma}\right) \\ \text{for } \gamma = 0,\ x > u,\ \sigma > 0 \end{cases}$$

and density f_{GPD} for some threshold u, to the tail of an empirical distribution function generated using the NESS data shown in Fig. 1. More colloquially, we assume that there is a power law scaling between the number of storms above some threshold and the value of that threshold. Mandelbrot (1983) gives an interesting discussion of similar scaling behavior in other application areas such as economics and statistical physics, where $\gamma > 0$.

Fitting the Model. We use a maximum likelihood procedure to estimate the scale parameter σ and the so-called "extreme value index," γ. Hereafter, estimates are denoted by the superscript ($\hat{}$). The actual fitting is achieved by minimizing the negative of the log likelihood, $L(X;\gamma,\sigma)$, given by

$$L(X;\gamma,\sigma) = \log_e\left(\prod_i f_{\mathrm{GPD}}(X_i;\gamma,\sigma)\right)$$

for observations X_i, using a conjugate gradient numerical solver. For the NESS data considered here, results for fits to the 400 largest storm severities at each location are given in Table 1.

One important feature of the GPD is that, for $\gamma < 0$, a finite upper limit exists for the largest possible value of x, seen from the foregoing equation to be $(u - \sigma/\gamma)$. In contrast, for $\gamma > 0$, there is no upper limit; larger values just become increasingly rarer instead. From Table 1 it is evident that, regardless of geographical location, the estimates for the extreme value index for the present data are negative, with values near $\gamma = -0.2$. That is, for each location, there is a maximum storm severity which cannot be exceeded. Recall that we have defined storm severity in terms of the most probable maximum wave height, related directly to the average energy density within the wave field. It is then not unreasonable that this could have a physical upper limit (arising from limited fetch, limits on wind speed, limited duration of storm winds in the same direction, energy dissipation due to wave breaking, etc.). However, we have no physical explanation for a common value for the extreme value index γ, although this has been seen elsewhere (e.g., Coles and Tawn, 1994).

We have also examined the effect of the number of storms used for the fit, equivalent to the threshold u, on the parameter estimates $\hat{\gamma}$. Results are shown in Fig. 2. For the smallest values of threshold, the assumption of being in the tail of a distribution is probably not valid. In the central and southern North Sea, the estimates are relatively insensitive to the number of data (>100) used to define the empirical-distribution. In the NNS, however, we believe that physical processes such as wave shielding by island groups like the Shetlands make storm severity very sensitive to wave direction. In Fig. 3, we show scatter diagrams of storm severity as a function of mean wave direction

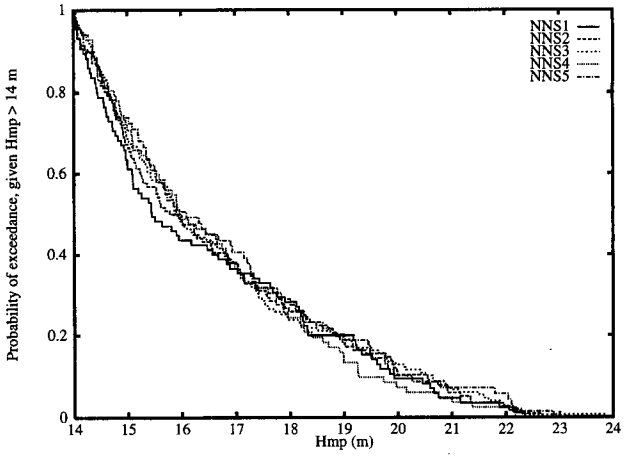

Fig. 1(*a*) Tail function: NNS

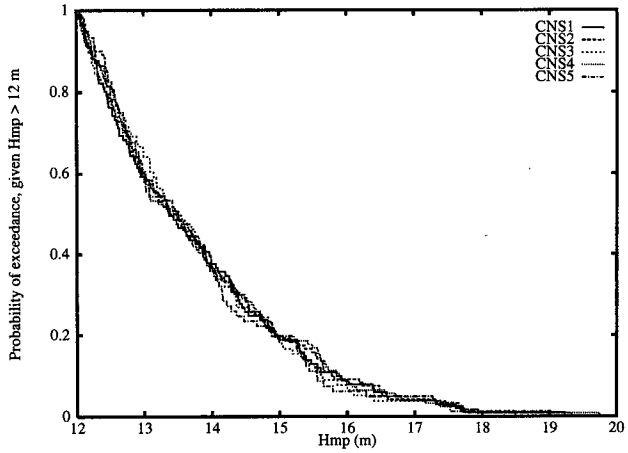

Fig. 1(*b*) Tail function: CNS

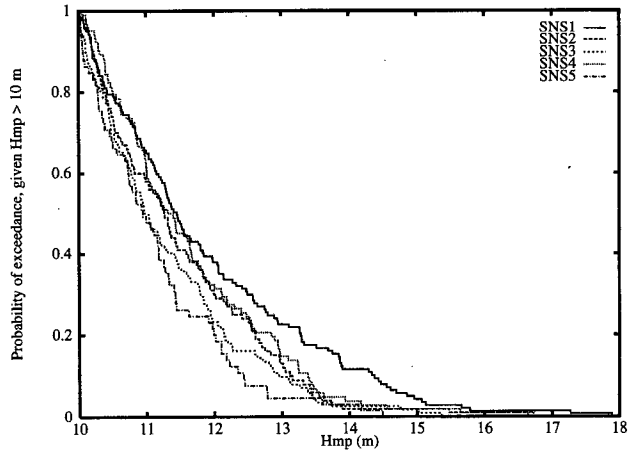

Fig. 1(*c*) Tail function: SNS

to illustrate this point. One approach to overcome directional effects would be to perform analyses based on directionally partitioned data; but this is hampered by the sparsity of observations.

Estimating Confidence Intervals for Model Parameters. The results in Table 1 show not only estimates $\hat{\gamma}$, $\hat{\sigma}$ for γ, σ,

Table 1 Maximum likelihood estimates for γ and σ, with bootstrap 95-percent confidence intervals

Data set	Threshold u	$\hat{\gamma}$	95% CI for $\hat{\gamma}$	$\hat{\sigma}$	95% CI for $\hat{\sigma}$
NNS1	9.3	-0.20	-0.26 , -0.10	3.8	3.4 , 4.2
NNS2	10.2	-0.19	-0.26 , -0.08	3.7	3.2 , 4.1
NNS3	10.4	-0.24	-0.31 , -0.14	4.2	3.7 , 4.7
NNS4	8.6	-0.26	-0.33 , -0.17	4.4	4.0 , 5.0
NNS5	8.6	-0.22	-0.29 , -0.11	4.3	3.7 , 4.9
CNS1	9.2	-0.24	-0.31 , -0.12	2.9	2.5 , 3.3
CNS2	9.7	-0.26	-0.32 , -0.14	2.9	2.5 , 3.3
CNS3	8.9	-0.23	-0.30 , -0.10	2.8	2.4 , 3.2
CNS4	9.6	-0.25	-0.31 , -0.14	3.0	2.6 , 3.3
CNS5	8.8	-0.22	-0.29 , -0.09	2.8	2.4 , 3.2
SNS1	7.4	-0.20	-0.26 , -0.11	2.8	2.5 , 3.1
SNS2	7.1	-0.20	-0.26 , -0.12	2.4	2.2 , 2.7
SNS3	7.4	-0.22	-0.28 , -0.12	2.3	2.0 , 2.5
SNS4	7.1	-0.22	-0.27 , -0.14	2.5	2.3 , 2.8
SNS5	6.7	-0.25	-0.31 , -0.17	2.3	2.1 , 2.6

but also 95-percent confidence intervals for these estimates. The width of the confidence intervals gives essential information about the precision of the estimates. Taking NNS1 as an example, we calculate an estimate of -0.20 for $\hat{\gamma}$; further, we expect the true but unknown value for γ to lie in the interval (-0.26 to -0.10) with a probability of 0.95.

We estimate these confidence intervals using a bootstrapping approach (see Efron, 1982), which consists of three important stages:

1 For the actual NESS data sample of size n at some location, we fit a GPD and estimate $\hat{\gamma}$, $\hat{\sigma}$ as described in the foregoing.

2 Next, we generate a large number P of random samples from this model GPD($\hat{\gamma}$, $\hat{\sigma}$). For each random sample, we refit the GPD model and estimate new parameters $\hat{\gamma}_j^*$, $\hat{\sigma}_j^*$, $j = 1, \ldots, P$. This stage is called "parametric resampling" because we are attempting to generate more samples from the distribution from which the actual NESS data was assumed to be drawn, using a parametric model.

3 Finally, we use the sample statistics of the $\hat{\gamma}_j^*$, $\hat{\sigma}_j^*$ to estimate the statistics of $\hat{\gamma}$, $\hat{\sigma}$, and hence calculate confidence intervals for the latter. More precisely, we use $(\hat{\gamma}_j^* - \hat{\gamma})/s_{\hat{\gamma}_j^*}$, $j = 1, \ldots, P$, to estimate the distributional properties of $(\hat{\gamma} - \gamma)/\sigma_{\hat{\gamma}}$, where $s^2_{\hat{\gamma}_j^*}$ is an estimate for the variance of $\hat{\gamma}_j^*$ using just the information from the jth resample, and $\sigma^2_{\hat{\gamma}}$ is the variance of $\hat{\gamma}$. We perform a similar calculation to obtain confidence intervals for $\hat{\sigma}$. The process of standardising with respect to scale is known as "studentizing" (Hall, 1988).

To achieve this final stage, we need to obtain estimates $s^2_{\hat{\gamma}_j^*}$ and $s^2_{\hat{\sigma}_j^*}$. These are obtained by making the assumption that the estimates $\hat{\gamma}_j^*$, $\hat{\sigma}_j^*$ for $\hat{\gamma}$, $\hat{\sigma}$ are reasonably good. That is, we can neglect $O(\hat{\gamma}_j^* - \hat{\gamma})^3$ and $O(\hat{\sigma}_j^* - \hat{\sigma})^3$, and hence assume (Kalbfleisch, 1979) that $(\hat{\gamma}_j^* - \hat{\gamma})/s_{\hat{\gamma}_j^*}$ and $(\hat{\sigma}_j^* - \hat{\sigma})/s_{\hat{\sigma}_j^*}$ are both normally distributed with zero mean and unit variance. Then, making use of the limiting behavior $(\hat{\gamma}_j^* - \hat{\gamma})/s_{\hat{\gamma}_j^*} \rightarrow (\hat{\gamma} - \gamma)/\sigma_{\hat{\gamma}}$ (and likewise for σ), we can trivially calculate the appropriate 95-percent confidence limits for $\hat{\gamma}$, $\hat{\sigma}$. The assumption of normality means that we can use the inverse of the information matrix $I(\hat{\gamma}_j^*, \hat{\sigma}_j^*)$ to estimate the variances of $\hat{\gamma}_j^*$, $\hat{\sigma}_j^*$ (Kalbfleisch, 1979)

$$\begin{pmatrix} \text{var}\,(\hat{\gamma}_j^*) & \text{cov}\,(\hat{\gamma}_j^*, \hat{\sigma}_j^*) \\ \text{cov}\,(\hat{\gamma}_j^*, \hat{\sigma}_j^*) & \text{var}\,(\hat{\sigma}_j^*) \end{pmatrix} = I^{-1}(\hat{\gamma}_j^*, \hat{\sigma}_j^*)$$

$$= -\begin{pmatrix} \dfrac{\partial^2 L}{\partial \gamma^2} & \dfrac{\partial^2 L}{\partial \gamma \partial \sigma} \\ \dfrac{\partial^2 L}{\partial \gamma \partial \sigma} & \dfrac{\partial^2 L}{\partial \sigma^2} \end{pmatrix}^{-1} \text{ evaluated at } \hat{\gamma}_j^*, \hat{\sigma}_j^*$$

We obtain the estimates (Tiago de Oliveira, 1983) $s^2_{\hat{\gamma}_j^*} = (1 + \hat{\gamma}_j^*)^2$ and $s^2_{\hat{\sigma}_j^*} = 2\hat{\sigma}_j^{*2}(1 + \hat{\gamma}_j^*)$. Using the parametric studentized bootstrap approach, we obtained the 95-percent confidence limits for $\hat{\gamma}$, $\hat{\sigma}$ shown in Table 1. In the central and southern North Sea, the estimates are somewhat more precise that in the northern North Sea, but the confidence intervals are still quite large, reflecting the general difficulty of fitting the tails of distributions.

3 Estimating Extreme Storm Severity

The GPD model established can now be used to estimate extreme storm severity corresponding to a return period of T years (e.g., 100-yr). The estimate $\hat{q}_T$ of the extreme quantile q_T can be written in terms of the GPD model parameter estimates $\hat{\gamma}$, $\hat{\sigma}$ as

$$\hat{q}_T = \frac{\hat{\sigma}}{\hat{\gamma}}(p^{-\hat{\gamma}} - 1) + u$$

$$\text{where} \quad p = \frac{T_0}{TN_0} \quad \text{and } u \text{ is the threshold severity}$$

Here, T_0 is the period of the NESS data (25 yr), and N_0 is the size of the NESS sample (taken as 400). Results obtained are reported in Table 2. It is apparent that storms in the NNS are more severe than in the central or southern regions, due to longer fetches. For the shallow waters of the SNS, seabed topography causes wave refraction and promotes wave breaking, which limits storm severity. For the CNS, quantile estimates at adjacent locations are very similar. In contrast, in the SNS, estimates vary considerably with location, due primarily to local changes in water depth.

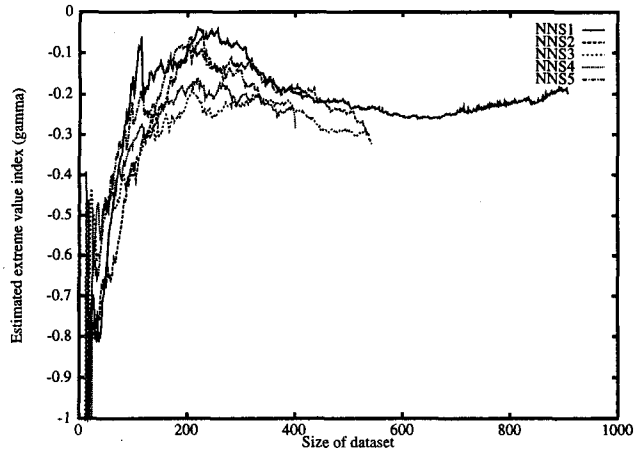

Fig. 2(*a*) Estimated extreme value index: NNS

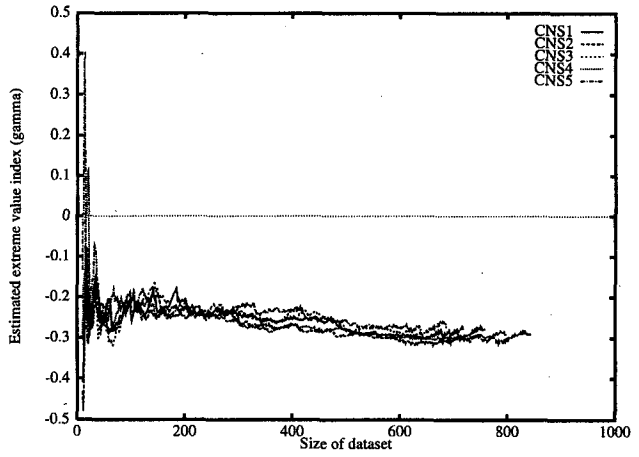

Fig. 2(*b*) Estimated extreme value index: CNS

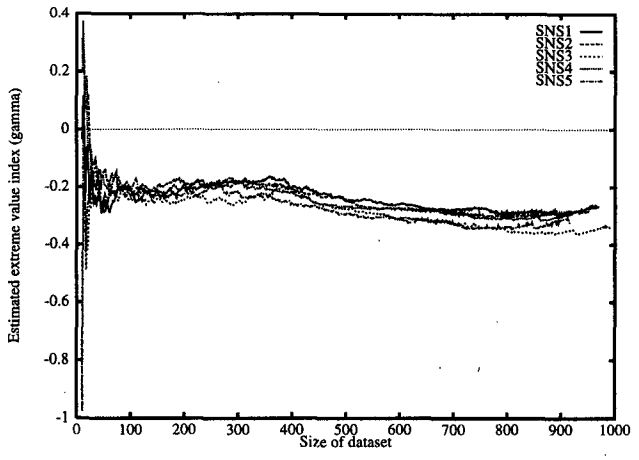

Fig. 2(*c*) Estimated extreme value index: SNS

Estimating extreme storm severity alone is not sufficient for reliable applications of extreme value statistics in offshore engineering; confidence intervals for those estimates should be estimated also. The confidence intervals for q_T in Table 2 were again estimated using a parametric studentized bootstrap, following the approach outlined in Section 2. This time, however, the studentization stage requires that we specify an estimate $s^2_{\hat{q}_T j}$ for the variance of $\hat{q}_T$ (obtained from resample j only). This estimate can also be obtained (Kalbfleisch, 1979) from the inverse of the information matrix, suitably transformed

$$\text{var}\,(\hat{q}^*_{T_j}) = \underline{\alpha}\,\underline{\underline{I}}^{-1}\underline{\alpha}' \quad \text{evaluated at } \hat{\gamma}^*_j,\ \hat{\sigma}^*_j,$$

$$\text{where} \quad \underline{\alpha} = \left(\frac{\partial q_T}{\partial \gamma}, \frac{\partial q_T}{\partial \sigma}\right)$$

which eventually leads to the result

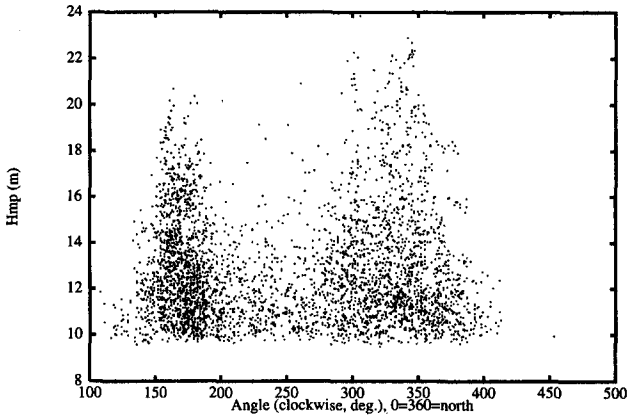

Fig. 3(*a*) Storm direction: NNS

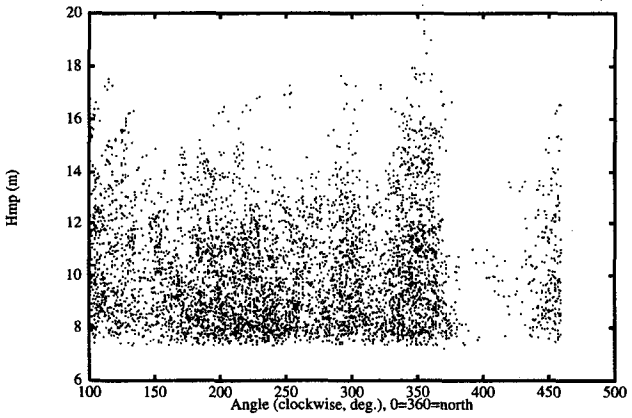

Fig. 3(*b*) Storm direction: CNS

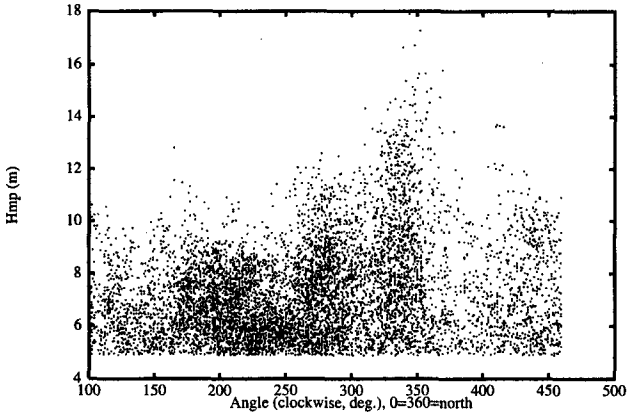

Fig. 3(*c*) Storm direction: SNS

Table 2 Extreme quantile estimates with bootstrap 95-percent confidence intervals

Data set	$\hat{q}_{25}$	95% CI for $\hat{q}_{25}$	$\hat{q}_{100}$	95% CI for $\hat{q}_{100}$	$\hat{q}_{500}$	95% CI for $\hat{q}_{500}$
NNS1	22.6	21.3 , 25.3	24.0	22.3 , 27.7	25.2	23.6 , 30.3
NNS2	23.4	22.0 , 26.4	24.9	22.9 , 29.1	26.1	24.3 , 31.8
NNS3	23.6	22.5 , 26.1	24.8	23.3 , 28.3	25.7	24.3 , 30.4
NNS4	22.0	21.0 , 24.1	23.0	21.8 , 25.9	23.9	22.7 , 27.6
NNS5	22.8	21.5 , 25.6	24.3	22.5 , 28.1	25.5	23.8 , 30.5
CNS1	18.3	17.5 , 20.1	19.1	18.2, 21.7	19.8	18.9 , 23.3
CNS2	18.4	17.8 , 20.0	19.2	18.4 , 21.5	19.8	19.0 , 22.8
CNS3	17.8	17.0 , 19.7	18.7	17.6 , 21.4	19.4	18.4 , 23.2
CNS4	18.6	17.9 , 20.2	19.4	18.6 , 21.7	20.1	19.3 , 23.2
CNS5	18.1	17.2 , 20.0	19.0	17.9 , 21.9	19.8	18.7 , 23.9
SNS1	17.1	16.2 , 18.9	18.1	17.0 , 20.6	19.0	17.9 , 22.4
SNS2	15.4	14.8 , 16.8	16.3	15.5 , 18.2	17.0	16.2 , 19.6
SNS3	15.0	14.3 , 16.4	15.7	14.9 , 17.7	16.3	15.5 , 19.0
SNS4	15.6	15.0 , 17.0	16.4	15.6 , 18.3	17.1	16.3 , 19.5
SNS5	13.9	13.4 , 14.9	14.4	13.9 , 15.8	14.9	14.4 , 16.7

$$s^2_{\hat{q}_T^*} = K^2(1 + \hat{\gamma}_j^*)^2 + 2K\hat{\sigma}_j^* \frac{1 + \hat{\gamma}_j^*}{\hat{\gamma}_j^*}(p^{-\hat{\gamma}_j^*} - 1) + 2\hat{\sigma}_j^{*2} \frac{1 + \hat{\gamma}_j^*}{\hat{\gamma}_j^{*2}}(p^{-\hat{\gamma}_j^*} - 1)^2$$

$$\text{where} \quad K = \frac{\hat{\sigma}_j^*}{\hat{\gamma}_j^{*2}}(p^{-\hat{\gamma}_j^*} - 1) + \frac{\hat{\sigma}_j^*}{\hat{\gamma}_j^*} p^{-\hat{\gamma}_j^*} \log_e p$$

from which confidence intervals for $\hat{q}_T$ can be estimated. As would be expected, the precision of $\hat{q}_T$ degrades as T increases (i.e., as we extrapolate further). Estimates for 25 and 100-yr quantiles are reasonably precise, but for longer return periods, confidence intervals are very wide.

We have also calculated the extreme quantile estimates $\hat{q}_{100}$ as a function of the number of data points used (equivalent to the storm severity threshold u). Results are given in Fig. 4. In the NNS, the estimates are relatively stable when a high threshold is used, i.e., we only keep the largest storm severities to estimate the 100-yr quantile. When we decrease the threshold, the extreme quantile estimate becomes quite unstable (suggesting mixed behavior—more than one family of storms?). In the CNS and SNS, the 100-yr quantile estimate is much less sensitive to the threshold.

Estimates for the maximum (upper limit) value for storm severity ($u - \sigma/\gamma$) have also been examined. In the NNS, these are ~20 percent higher than the largest values in the hindcast data (see Table 1 and Fig. 1), whereas in the CNS and SNS the difference is only ~10 percent. Thus, in some sense, we appear to be extrapolating "further," the further north we go. This has important implications for structural reliability assessment. Although, the extrapolation of storm severity to very long return periods requires a great extrapolation in time (from the 25 yr of the hindcast to 10,000 yr, say), it does not require an equivalently large extrapolation in storm severity: the 10,000-yr storm will be almost as severe as the upper limit on storm severity, which is only 10–20 percent larger than the most severe storms in the hindcast. Then, if the numerical models used to create the hindcast adequately capture the physics for those storms within it, there is little room for new physics to enter for rarer, more severe storms. As a note of caution, even if an upper limit on storm severity exists, this does not necessarily imply that there is a corresponding limit on load and structural response due to short-term wave variability within storms (e.g., Tromans and Vanderschuren, 1995), which is not accounted for in our extrapolation. This would be incorporated in a full reliability analysis.

Confidence intervals for the maximum value of storm severity are of course wider that confidence intervals for q_{25}, q_{100} and q_{500}, since we are again extrapolating further. Typical values for the confidence limits in the NNS are (−3 m, +10 m), although there is variation from location to location (with one interval as large as (−4 m, +18 m)). In the CNS and SNS, the intervals are generally somewhat narrower.

4 On the Quality of the Estimates Obtained

In the analysis presented in Sections 2 and 3, we have made a number of assumptions concerning the properties of the NESS data and the various parameter and quantile estimates. Using numerical simulation, we can test the validity of a number of these assumptions.

In particular, in this section we examine the coverage probabilities for the interval estimates of $\hat{\gamma}$, $\hat{\sigma}$, and $\hat{q}_T$. That is, we seek to confirm that the 95-percent confidence intervals for each of $\hat{\gamma}$, $\hat{\sigma}$, and $\hat{q}_T$, derived using the bootstrap procedure, actually give 95 percent coverage for simulated data drawn from a known GPD distribution.

We also examine the effect of misspecifying the underlying probability distribution from which the NESS data is assumed to be drawn. Specifically, we use the GPD fitting and quantile estimation procedure for data drawn from a known Weibull distribution (for which $\gamma = 0$, as opposed to GPD for which $\gamma \neq 0$ in general). In this situation, what is the bias and variance of extreme quantile estimates?

Bias and Coverage Probabilities for Model Parameter and Extreme Quantile Estimates. Using simulation studies, we can check the coverage of the bootstrap 95-percent confidence intervals for return periods up to 500 yr. We proceed as follows: For 1000 simulated data sets drawn from a known GPD representative of the NESS data, we estimate 95-percent confidence intervals for T-yr storm severity using the bootstrap method, and check whether the true (known) T-yr storm severity lies inside the confidence interval. Results are given in Table 3 for γ, σ, and the 25, 100, and 500-yr quantiles, for sample sizes of 100, 400, and 1000, corresponding to a known underlying GPD distribution with $\gamma = -0.2$ and $\sigma = 1$. For perfect agreement, we require that the left-hand and right-hand tail probabilities (P(lhs) and P(rhs), respectively) are both 0.025 (so that the estimates are unbiased and the coverage probability is 1 −

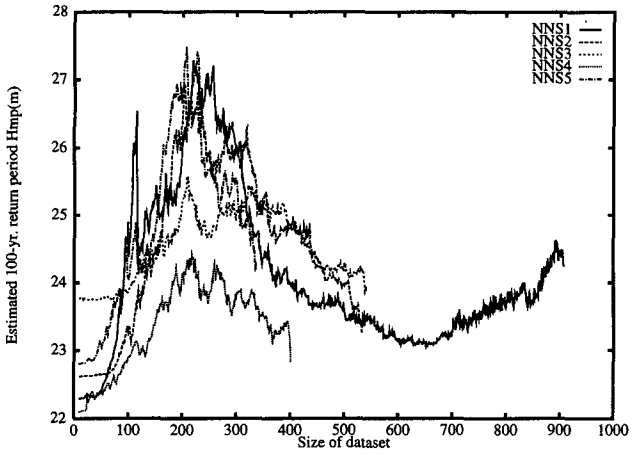

Fig. 4(*a*) Estimated 100-yr Hmp: NNS

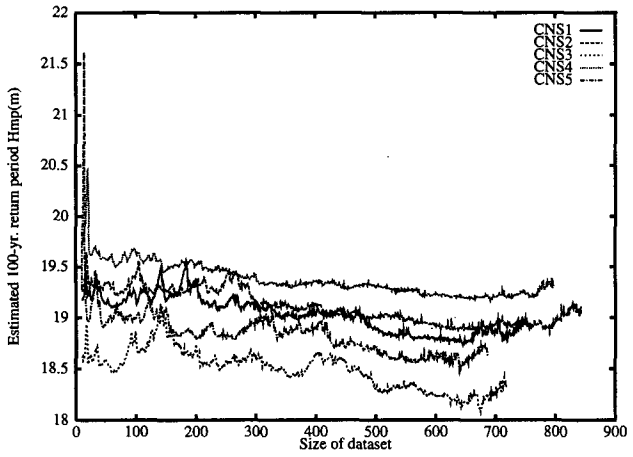

Fig. 4(*b*) Estimated 100-yr Hmp: CNS

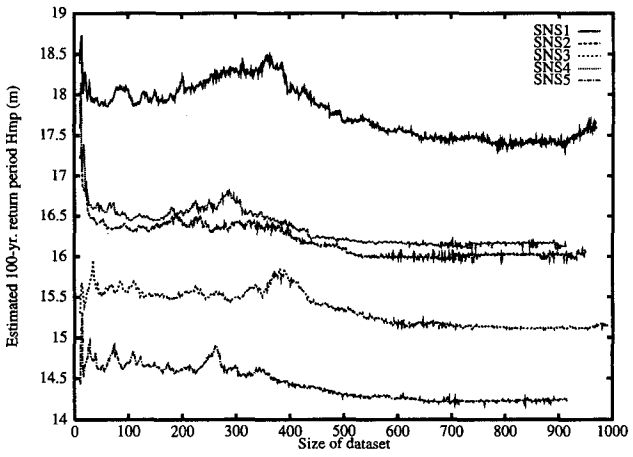

Fig. 4(c) Estimated 100-yr Hmp: SNS

$2*0.025 \approx 0.95$). The table also gives the relative bias for the 25, 100, and 500-yr quantiles, defined in the key. As the magnitude of the bias is small ($\ll 1$), we conclude that the estimation method presented in Sections 2 and 3 is relatively unbiased.

From Table 3, bootstrap estimates of 95-percent confidence intervals for 25 and 100-yr storm severity include the estimated 100-yr storm severity for approximately 95 percent of the simulations performed, confirming that the bootstrap procedure and studentization of estimates works well in these cases.

Robustness of Estimates With Respect to Model Misspecification. There is no guarantee that the data sets of storm severity from NESS correspond to a tail distribution which is exactly GPD. For example, the adoption of a GPD distribution as opposed to a Weibull distribution cannot be justified in general, based on data alone. Thus, we need to test the robustness of our extreme value analysis with respect to misspecification of the underlying tail distribution. The simulation strategy is similar to that described in the foregoing, except that we now produce 1000 simulated samples from a Weibull distribution (not a GPD). As before, we then estimate 25, 100, and 500-yr extreme storm severities using the GPD model. The quality of the agreement obtained in reported in Table 4 in terms of coverage probabilities for the bootstrap 95-percent confidence intervals.

It is somewhat surprising that the coverages in Table 4 are still very good, even when the (synthetic) data is drawn from a Weibull tail distribution, especially for the 25 and 100-yr quantiles. This is of some importance, since Weibull-fitting (in particular) has been prevalent in the oceanographic community. For events with relatively low-return periods (25–100 yr, say), fits to Weibull and generalized Pareto distributions yield similar results. Table 4 also suggests that the bias introduced by the estimation procedure is larger than in Table 3, as might be expected. Note that positive bias here indicates that the GPD modeling and extrapolation overestimates the extreme quantiles considered. However, for extrapolation to very long return periods, this will inevitably be reversed: a GPD can, and for the NESS data does, have an upper limit, whereas a Weibull fit can never have any such limit.

5 Conclusions and Recommendations

In this paper, we present a method for estimation of extreme quantiles using a generalized Pareto model. The precision of extreme quantiles estimated is calculated using a parametric studentized bootstrap procedure.

Using simulation studies, we have demonstrated that the estimation procedure is effectively unbiased for known underlying GPD or Weibull-distributed data. We have also demonstrated that the coverage of bootstrap-estimated 95-percent confidence intervals for model parameters and extreme quantiles is ~0.95, provided that we do not extrapolate too far.

In application to 15 samples from the NESS hindcast database, we have found that the estimated extreme value index is consistently negative, with a value of ~ -0.2. This suggests the existence of a finite upper bound for storm severity in the North Sea. In the central and south North Sea, the value of the extreme

Table 3 Bias and coverage of estimates for synthetic GPD data

Sample size N_0	γ		σ		q_{25}		q_{100}		q_{500}		Bias[1]
	P(lhs)	P(rhs)	P(lhs)	P(rhs)	P(lhs)	P(rhs)	P(lhs)	P(rhs)	P(lhs)	P(rhs)	in q_{100}
100	0.025	0.099	0.078	0.029	0.031	0.041	0.031	0.040	0.036	0.032	-0.26
400	0.022	0.018	0.031	0.021	0.032	0.011	0.029	0.011	0.027	0.011	-0.20
1000	0.029	0.017	0.019	0.027	0.032	0.014	0.033	0.016	0.037	0.016	-0.19

[1] Bias estimated as the difference between the mean $\hat{q}_{100}$ over all simulations and the true value of q_{100}, divided by the standard deviation of $\hat{q}_{100}$ estimated from simulation. A negative bias indicates underestimation of the true q_{100}.

Table 4 Bias and coverage of extreme quantile estimates using a GPD model for Weibull data

Sample size N_0	q_{25}		q_{100}		q_{500}		Bias in q_{100}
	P(lhs)	P(rhs)	P(lhs)	P(rhs)	P(lhs)	P(rhs)	
100	0.019	0.031	0.017	0.033	0.024	0.150	+1.16
400	0.022	0.034	0.027	0.028	0.032	0.097	+0.87

[Definition of bias as in Table 3]

value index is relatively insensitive to the choice of threshold selected for GPD modeling. For the northern North Sea, the Pareto fits are less robust but the extreme value index is still approximately -0.2.

Results also indicate that storm severities at adjacent grid locations in each of the northern, central, and southern North Seas are correlated. However, we have not attempted to use this correlation structure to improve our estimates. Note further that the correlation structure appears weakest in the southern North Sea, where storm severity varies rapidly between grid locations due to varying water depth, among other things. This suggests that modeling the water depth dependence of extreme storm severity in the southern North Sea might be worthwhile.

References

Coles, S. G., and Tawn, J. A., 1994, "Statistical Methods for Multivariate Extremes: An Application to Structural Design," *Applied Statistics,* Vol. 43, pp. 1–48.

Efron, B., 1982, *The Jack-Knife, The Bootstrap and Other Resampling Plans,* SIAM, Philadelphia, PA.

Hall, P., 1988, *The Bootstrap and Edgeworth Expansion,* Springer-Verlag, New York, NY.

Kalbfleisch, J. G., 1979, *Probability and Statistical Inference.* Vol. 2, Springer-Verlag, New York, NY.

Leadbetter, M. R., Lindgren, G., and Rootzén, 1983, *Extremes and Related Properties of Random Sequences and Processes,* Springer-Verlag, New York, NY.

Mandelbrot, B. B., 1983, *The Fractal Geometry of Nature,* W. H. Freeman, New York, NY.

Maes, M. A., and Gu, G. Z., 1994, "Techniques Used to Determine Extreme Wave Heights From the NESS Dataset," *Journal of Research of the National Institute of Standards and Technology,* Vol. 99, No. 4, pp. 435–444.

Peters, D. J., Shaw, C. J., Grant, C. K., Heideman, J. C., and Szabo, D., 1993, "Modelling the North Sea through the North European Storm Study," OTC 7130, 25th Offshore Technology Conference, Houston, TX, pp. 479–493.

Pickands, J., 1975, "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics,* Vol. 3, pp. 119–131.

Tiago de Oliveira, J., ed., 1983, *Statistical Extremes and Applications,* NATO ASI Series C Vol. 31, Reidel, Netherlands.

Tromans, P. S., and Vanderschuren, L., 1995, "Response-Based Design Conditions in the North Sea: Application of a New Method," OTC 7683, 27th Offshore Technology Conference, Houston, TX, pp. 387–397.