Uncertainties in Extreme Wave Height Estimates for Hurricane-Dominated Regions

Philip Jonathan
Shell Research Limited,
P.O. Box 1,
Chester, United Kingdom
e-mail: philip.jonathan@shell.com

Kevin Ewans
Shell International Exploration and Production,
P.O. Box 60,
2280 AB Rijswijk, The Netherlands
e-mail: kevin.ewans@shell.com

Inherent uncertainties in estimation of extreme wave heights in hurricane-dominated regions are explored using data from the GOMOS Gulf of Mexico hindcast for 1900–2005. In particular, the effect of combining correlated values from a neighborhood of 72 grid locations on extreme wave height estimation is quantified. We show that, based on small data samples, extreme wave heights are underestimated and site averaging usually improves estimates. We present a bootstrapping approach to evaluate uncertainty in extreme wave height estimates. We also argue in favor of modeling supplementary indicators for extreme wave characteristics, such as a high percentile (95%) of the distribution of 100-year significant wave height, in addition to its most probable value, especially for environments where the distribution of 100-year significant wave height is strongly skewed. [DOI: 10.1115/1.2746401]

1 Introduction

Environmental design criteria for offshore facilities have inherent uncertainties. These uncertainties are a function of both the climate variability at a location and the amount of data available for modeling and estimating design criteria (e.g., [1]). In regions such as the Gulf of Mexico and off the North West Shelf of Australia, extreme sea states are associated with hurricanes, but these occur relatively infrequently (by comparison to extra-tropical storms in the North Sea, for example). Moreover, hurricanes are relatively small scale by comparison to extra-tropical storms and, as a result, hurricane track has an important influence on severity of sea state at a particular location.

In the Gulf of Mexico, site averaging is used to increase the sample size for modeling and to account for randomness of storm track [2]. However, hurricane data from even quite largely separated locations are highly correlated. As a result, it is not straightforward to determine the reliability (or, equivalently, the degree of uncertainty) associated with design criteria derived from the site-averaging approach.

In this paper, we consider uncertainties of extreme value estimates for hurricane regions, particularly the Gulf of Mexico, motivated by the site-averaging approach. Analysis of illustrative hindcast data is reported, together with results of simulations using realistic extreme value models, to examine the effect of sample size and intersite correlation on estimates for extremes and uncertainties associated with those estimates. For this purpose, it is sufficient to consider only the significant wave height of a sea state, but we acknowledge that an appropriate treatment of the short-term variability is needed in practice to specify, for example, the extreme crest for design (e.g., [3]).

In a previous study [4], application of generalized Pareto modeling to estimation of North Sea storm severity was reported for storms with return periods of 100–500 years based on NESS hindcast data. The study consisted of the following elements. The tail distribution of storm severity was modeled, and magnitudes of extreme events with long return periods were estimated. Uncertainty of estimates was quantified using a bootstrapping approach. Finally, bias and coverage for estimates of uncertainty were quantified by simulation study. The present investigation follows similar lines but focuses on extreme value modeling using data from a neighborhood of locations.

Practical applications of multivariate extreme value modelling are usually limited to two or three dimensions (e.g., [5–7], and references therein). Coles and Simiu [8] explore the use of bootstrapping to obtain reasonable measures of uncertainty when using hindcast data for extreme value analysis. Bootstrapping is a standard approach in statistics and involves estimating parameter uncertainty by resampling the original data sample at random (e.g., [9–13]). Coles and Simiu [8] discuss a number of difficulties in applying bootstrapping for estimation of uncertainties in extreme value analysis, including the tendency to underestimate extreme quantiles. Nevertheless, they conclude that bootstrapping, carefully applied, can be used reliably to give realistic estimates for parameter uncertainties. Heffernan and Tawn [12] report a conditional approach for extreme value analysis applicable to higher-dimensional problems, also incorporating bootstrapping, in which dependence structure is characterised using rank correlation.

The paper is arranged as follows. In Sec. 2, we introduce the GOMOS hindcast data motivating the investigation, illustrating some key features. In Sec. 3, an outline of the generalized Pareto model used to characterise the extreme value behavior of the data is given. We also apply the model to the hindcast data and motivate the analysis in Secs. 4 and 5. In Sec. 4, a simulation study is performed to explore the effects of site averaging on extreme value estimation. In Sec. 5, we introduce a simple method based on bootstrapping to estimate uncertainty in extreme value estimates (such as the magnitude of the most probable 100-year event) and evaluate the performance of the method. In Sec. 6, we discuss characteristics of the distribution of 100-year significant wave height and the need for careful interpretation of most probable extreme values, especially when the extreme value index is thought to be small and negative. In Sec. 7, we summarise findings and make suggestions and recommendations for future studies.

2 The Data

The data examined are significant wave height (H_S) values from the proprietary GOMOS Gulf of Mexico hindcast study [13], for the period from September 1900 to September 2005 inclusive, at 30 min intervals. For a typical Gulf of Mexico location, we selected 72 grid points arranged on a 6×12 rectangular lattice with

Contributed by the Ocean Offshore and Article Engineering Division of ASME for publication in the JOURNAL OF OFFSHORE MECHANICS AND ARCTIC ENGINEERING. Manuscript received September 1, 2006; final manuscript received January 30, 2007. Review conducted by Eddie H. H. Shih. Paper presented at the 25th International Conference on Offshore Mechanics and Arctic Engineering (OMAE2006), June 4–9, 2006, Hamburg, Germany.

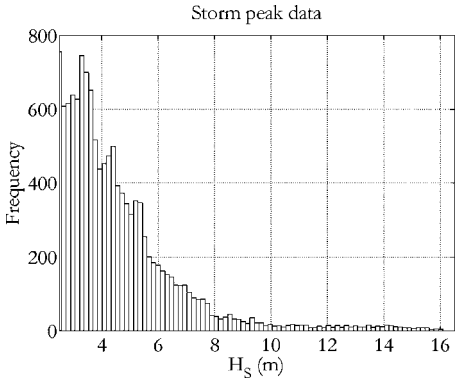

Fig. 1 Histogram of storm peak data, for all values for H_S^{sp} in excess of 2.5 m. Constant histogram bin width.

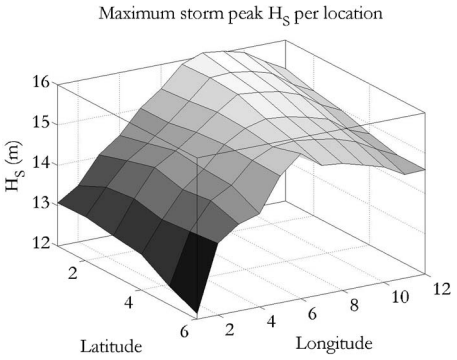

Fig. 3 Maximum value of H_S^{sp} per location. Gray scale indicates value.

spacing with 0.125 deg (≈14 km). For each storm period for each grid point, we isolated storm peak significant wave height H_S^{sp} for modeling purposes.

Figures 1–3 illustrate the key features of the current data, namely that they correspond to a sample of extreme values from 72 locations that are spatially dependent with extreme values of similar magnitudes. A histogram of storm peak data (H_S^{sp}) in excess of 2.5 m is given in Fig. 1. Figure 2(*a*) illustrates the high intercorrelation in H_S^{sp} between one pair of diagonally opposite corner points of our lattice. Figure 2(*b*) shows the Spearman rank correlation matrix for H_S^{sp} values above 2.5 m. Figure 3 gives the maximum value of H_S^{sp} observed at each location for the full period of the hindcast. We note that the spatial dependence evident in Figs. 2 and 3 is a function of the wave environment in the region under consideration. The extent of spatial dependence will, of course, depend on location in general, although we believe the current location to be typical for the Gulf of Mexico.

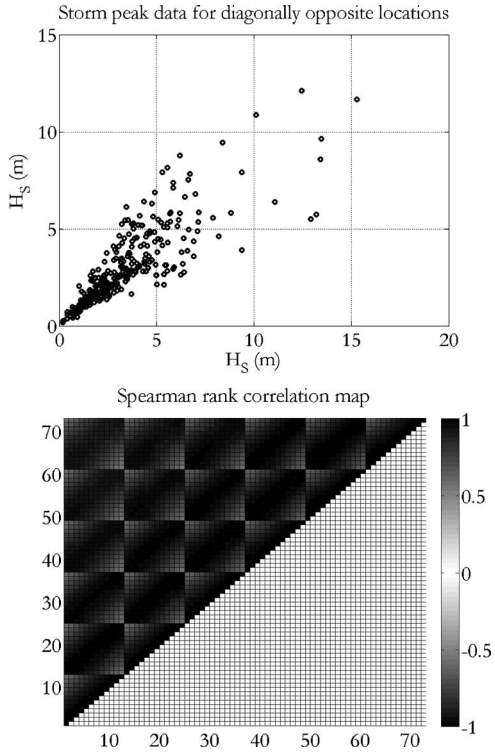

Fig. 2 (*a*) Scatter plot of H_S^{sp} for diagonally opposite grid corners. The two sets of data are highly correlated. (*b*) Spearman rank correlation map showing rank correlation coefficients for all pairs of locations, using data with H_S^{sp} in excess of 2.5 m. Locations are numbered numerically from 1 to 72, such that successive groups of 12 locations correspond to different longitudes at a given latitude. Top left shows positive rank correlations. Bottom right (empty) shows negative rank correlations. Gray scale indicates value of rank correlation. Mean rank correlation between locations is 0.886, with minimum of 0.535 and maximum of 0.997.

3 The Model

Generalized Pareto Modeling. We follow the approach outlined in Elsinghorst et al. [4] for extreme value modeling using the generalized Pareto distribution (GPD) for extremes above a prespecified threshold. Motivation for this choice of distribution is given by Reiss and Thomas [5]. The cumulative density function for the GPD is given by

$$F(x;\gamma,\sigma) = 1 - \left[1 + \frac{\gamma}{\sigma}(x-u)\right]^{-(1/\gamma)} \qquad x > u, \quad \sigma > 0$$

where γ is the extreme value index (also known as the shape parameter), σ is the scale, and u is a prespecified threshold. We use maximum likelihood estimation to estimate the values of parameters $(\hat{\gamma},\hat{\sigma})$ appropriate for a given sample of data. Care must be taken to ensure that values of $(\hat{\gamma},\hat{\sigma})$ are not materially affected by choice of u. Asymptotic variances for $(\hat{\gamma},\hat{\sigma})$ can be derived analytically.

They are

$$\sigma_{a\gamma}^2 = \frac{(1+\hat{\gamma})^2}{n}$$

$$\sigma_{a\hat{\sigma}}^2 = \frac{2\hat{\sigma}^2(1+\hat{\gamma})}{n}$$

Extreme quantiles can also be estimated as follows. The most probable 100-year significant wave height, $H_{S100\text{yrMP}}$, is estimated in terms of the 100-year return level, using

$$H_{S100\text{yrMP}} = \frac{\hat{\sigma}}{\hat{\gamma}}(p^{-\hat{\gamma}} - 1) + u$$

Here, p is the given by $p = P/100n$, where P is the period of the data in years and n is the number of data used to achieve the fit. p^{-1} corresponds to the expected number of occurrences above the GPD threshold in 100 years. An expression for the asymptotic variance of $H_{S100\text{yrMP}}$ is also available

Table 1 Extreme value estimates and uncertainties. Note: q used as shorthand for $H_{S100yrMP}$.

Method	$\hat{q}$(m)	$\sigma_{a\hat{q}}$(m)	$\hat{\gamma}$	$\sigma_{a\hat{\gamma}}$	$\hat{\sigma}$	$\sigma_{a\hat{\sigma}}$
Mean of individual estimates per grid location	13.16	1.42	–0.024	0.08	2.22	0.24
Single estimate based on all grid locations	13.19	0.17	–0.022	0.01	2.22	0.03

$$\sigma^2_{H_{100yrMP}} = \frac{K^2(1+\hat{\gamma})^2 + 2K\hat{\sigma}\left(\frac{1+\hat{\gamma}}{\hat{\gamma}}\right)(p^{-\hat{\gamma}}-1) + 2\hat{\sigma}^2\left(\frac{1+\hat{\gamma}}{\hat{\gamma}^2}\right)(p^{-\hat{\gamma}}-1)^2}{n}$$

where $K=(\hat{\sigma}/\hat{\gamma}^2)(p^{-\hat{\gamma}}-1)+(\hat{\sigma}/\hat{\gamma})p^{-\hat{\gamma}}\log_e p$. γ and σ are reserved throughout the paper to refer to extreme value tail index and scale, respectively.

Estimation for Individual and Combined Locations. Using the GPD model, we have estimated the extreme value parameters and $H_{S100yrMP}$ for grid locations, individually. Results are given in Table 1, in terms of the mean value for extreme value index and scale over the 72 locations. Table 1 also gives the mean value (over the 72 locations) of the asymptotic standard error for each of extreme value index and scale. When limited by the number of measurements of H_S^{sp} available at any one location, it is attractive to combine data from neighboring locations to achieve a bigger sample for extreme value estimation; this is the basis of the site-averaging approach. Our motivation for doing this is that neighboring locations are assumed to have common (or very similar) extreme value behavior. Combining locations gives us more data, with the same underlying extreme value distribution, to model. We would expect to obtain better estimates of extreme value parameters $(\hat{\gamma},\hat{\sigma})$ and extreme quantiles. Table 1 gives values for extreme value index, scale and $H_{S100yrMP}$ estimated using combined data for all locations. 12,226 data were used to provide the single estimate based on all locations. Per location, an average of 170 data were used for the location-specific estimation.

Agreement between estimates for (γ,σ) and $H_{S100yrMP}$ is good. However, there is a clear discrepancy in values for standard errors. Estimation based on a single location uses data from that location only; thus, standard errors are larger. But the estimate above, based on all data, has assumed that different locations represent independent samples from the common extreme value distribution. However, neighboring locations are interdependent. Given this dependence structure, it is likely that we are underestimating the uncertainty associated with the extreme value estimates $(\hat{\gamma},\hat{\sigma})$ and $\hat{H}_{S100yrMP}$. Suppose we have data for n values of H_S^{sp} at each of p locations in a neighborhood. Furthermore, if data for all locations are identical, corresponding to perfect dependence (so that all locations return the same H_S^{sp} for any given storm), combining locations cannot improve estimation. The number of independent data for modeling is still n. Conversely, suppose that H_S^{sp} for each pair of locations is independent; combining locations will improve estimation. The number of independent data for modeling is $n\times p$. In intermediate situations, with some level of dependence between locations, it is difficult to quantify the extent to which combining data from different locations alters the uncertainty in extreme value parameters $(\hat{\gamma},\hat{\sigma})$ and extreme quantiles $\hat{H}_{S100yrMP}$. Furthermore, we know that uncertainty estimates in Table 1 are based on asymptotic results and are unlikely to be reliable in practice, especially for small samples.

The discussion above motivates the current investigation. Using a simulation study, we explore the possible magnitudes of effects due to site averaging on the estimation of extreme value parameters and quantiles. We further investigate a procedure to estimate the uncertainties of extreme value parameters and quantiles directly, given any set of n measurements for H_S^{sp} at each of p locations in a neighborhood, for arbitrary dependence structure.

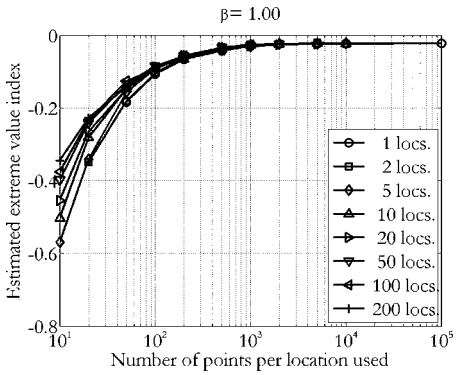

Fig. 4 Variation of estimated extreme value index, for combinations of locations of different sizes, as a function of sample size per location. Gaussian perturbation standard deviation, β =1.

4 Effect of Sample Size and Site Averaging on Extreme Value Estimates

We consider a simple model to construct dependent storm peak data to examine the effects of sample size and site averaging on extreme value estimates. The essence of the model is to construct data based on random samples from the generalized Pareto distribution, perturbed by Gaussian noise. We select the parameters of the GPD to mimic the GoM (Gulf of Mexico) data above and adjust the strength of dependence by changing the standard deviation of the Gaussian noise perturbation. The scheme is explained as follows:

1. We draw a random sample $\{Z_i\}_{i=1}^n$ from GPD with (γ,σ) $=(-0.02,2.4)$ corresponding to n storm peaks at each of p locations.
2. We add Gaussian random noise with standard deviation β to represent data for p locations: $x_{ij}=z_{ij}+\varepsilon_{ij}\beta$, $i=1,2,\ldots,n$, j $=1,2,\ldots,p$, where ε are independent drawings from the standard Gaussian distribution.
3. By construction, each pair of locations has common dependence structure; each pair of locations will be dependent to the same extent. The magnitude of the rank correlation is a function of (γ,σ) and β.
4. By construction, the marginal distributions at all locations are identical, but the overall extreme value characteristics will be modified by the additive noise term.
5. Values exceeding the threshold $u=2.5$ m are used for extreme value modeling.

One-thousand realizations of this sample are created for different combinations of sample size n ranging from 10 to 10^5, number of locations from 1 to 200, and perturbation standard deviations from 0 to 5. Results are presented in terms of variation of $(\hat{\gamma},\hat{\sigma})$ and $\hat{H}_{S100yrMP}$, for combinations of different numbers of locations, with respect to sample size at each location.

Figures 4–6 present results for the variation of $(\hat{\gamma},\hat{\sigma})$ and $\hat{H}_{S100yrMP}$ in the case $\beta=1$. The mean Spearman rank correlation coefficient is 0.73 for these data. We note that all estimates are biased for small sample sizes per location. As sample size per location increases, relative bias reduces. We also note that curves for different numbers of locations combined are distinguishable for $(\hat{\gamma},\hat{\sigma})$; this trend is not evident for $\hat{H}_{S100yrMP}$. We conclude that combining data from different locations reduces the bias of $(\hat{\gamma},\hat{\sigma})$

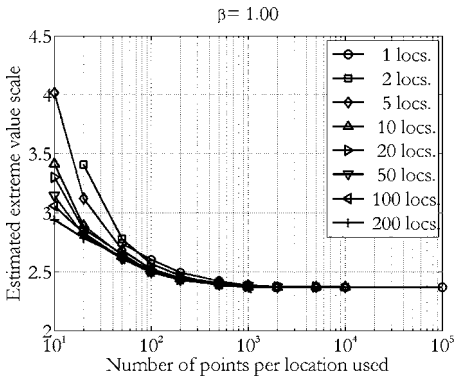

Fig. 5 Variation of estimated extreme value scale, for combinations of locations of different sizes, as a function of sample size per location. Gaussian perturbation standard deviation, β =1.

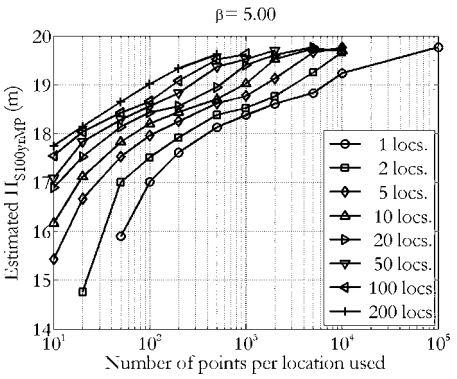

Fig. 7 Variation of $\hat{H}_{S100yrMP}$, for combinations of locations of different sizes, as a function of sample size per location. Gaussian perturbation standard deviation, β=5.

but not that of $\hat{H}_{S100yrMP}$. Corresponding curves for case $\beta=0$ (perfect dependence, not shown) exhibit the same trends with sample size, but no effect due to number of locations combined as would be expected. In passing, we note that these results are consistent with those for the GOMOS hindcast sample in Table 1, for which relatively large samples are available per location, and locations are highly dependent. Note that a minimum sample size of 100 for GPD modeling was imposed in these simulation studies. Therefore, cases corresponding to smaller sample sizes were not evaluated and thus not shown in Figs. 4–6.

As we increase Gaussian noise standard deviation to $\beta=2$ and $\beta=5$, corresponding to weaker dependence between locations, the effect of combining locations becomes clearer, even for $\hat{H}_{S100yrMP}$. Figure 7 illustrates this for $\beta=5$, although it should be noted that extreme value estimates change considerably since the Gaussian perturbation is dominating. We have confirmed that a Weibull fit to this GOMOS data set behaves in a similar way, and we expect other distributions used to model extremes to behave similarly. The mean Spearman rank correlation coefficient for the cases β =2 and $\beta=5$ is 0.47 and 0.15, respectively.

5 Bootstrap Approach to Quantifying Uncertainty in Extreme Value Estimates

In this section, we present a method to estimate uncertainties of extreme value parameters $(\hat{\gamma},\hat{\sigma})$ and quantiles $\hat{H}_{S100yrMP}$ directly, given any set of n measurements for H_S^{sp} at each of $p(\geq 1)$ locations in a neighborhood, for arbitrary dependence structure. The approach uses bootstrapping, a well-established technique for estimation of parameter uncertainty (e.g., [10]) based on data resampling.

Suppose we have a set of observations $\{x_{ij}\}_{i=1\,j=1}^{n\quad p}$ for H_S^{sp} measured at p locations corresponding to n storms. We assume that each location has the same generalized Pareto marginal distribution with parameters (γ,σ), GPD(γ,σ). The spatial dependence between locations is unknown but potentially considerable. Given the original data sample $D=\{x_{ij}\}_{i=1\,j=1}^{n\quad p}$, we proceed as follows.

1. Estimate $(\hat{\gamma},\hat{\sigma})$ and $\hat{H}_{S100yrMP}$ using the whole of the original data sample D.
2. Create a sample $D^{(b)}=\{x_{ij}^{(b)}\}_{i=1\,j=1}^{n\quad p}$ from D by resampling from the n storms in D (using a common stormwise resampling for all locations), at random, with replacement.
3. Obtain estimates $(\hat{\hat{\gamma}}^{(b)},\hat{\hat{\sigma}}^{(b)})$ and $\hat{\hat{H}}_{S100yrMP}^{(b)}$ for the parameters and extreme quantile of the common marginal GPD distribution fitted to all of the data in $D^{(b)}$.
4. Estimate the values of $d_\theta^{(b)}=(\hat{\hat{\theta}}^{(b)}-\hat{\theta})/\sigma_{a\hat{\theta}^{(b)}}$ for $\theta \in \{\gamma,\sigma,H_{S100yrMP}\}$, where $\sigma_{a\hat{\theta}}^2$ is the asymptotic variance of parameter estimate $\hat{\theta}$, introduced above.
5. Repeat steps 2 to 4 a large number of times B (certainly $\geq$200, usually of the order of 1000).
6. Estimate the critical values (c_θ^-,c_θ^+), such that 2.5% of the values of $\{d_\theta^{(b)}\}_{b=1}^B$ are smaller than c_θ^- and 2.5% are larger than c_θ^+.
7. A 95% confidence interval for parameter $\theta \in \{\gamma,\sigma,H_{S100yrMP}\}$ is then given by $(\hat{\theta}-c_\theta^+\sigma_{a\hat{\theta}},\hat{\theta}-c_\theta^-\sigma_{a\hat{\theta}})$

The rationale for the specific nonparametric studentized resampling approach adopted here is as follows. In step 1, we estimate the common marginal distribution of extremes using the GPD. In step 2, based on the assumption that the original sample D provides a valid estimate for the distribution from which D is drawn, we assume that $D^{(b)}$ will also provide a valid estimate. Note that resampling in step 2 retains the dependence structure of D; data for all locations for any given storm are treated as a single multivariate observation for resampling. In steps 3 and 4, we estimate the difference between estimate $\hat{\hat{\theta}}$ and $\hat{\theta}$ on a standardized (studentized) scale. Studentizing improves performance of bootstrap estimation (e.g., [10]). By repeating steps 2–4 a large number of times, we estimate critical values (c_θ^-,c_θ^+) for $\theta \in \{\gamma,\sigma,H_{S100yrMP}\}$ that characterize parameter uncertainty of $\hat{\hat{\theta}}$ (with respect to known $\hat{\theta}$) and, therefore, the parameter uncertainty of $\hat{\theta}$ (with respect to the true unknown θ). Note, in particu-

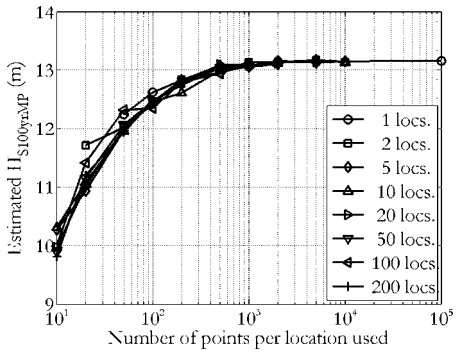

Fig. 6 Variation of $\hat{H}_{S100yrMP}$, for combinations of locations of different sizes, as a function of sample size per location. Gaussian perturbation standard deviation, β=1.

Table 2 Extreme value estimates and uncertainties estimated using H_S^{sp} values above 2.5 m at all 72 grid locations. Note: q used as shorthand for $H_{S100yrMP}$.

Parameter	$\hat{q}$ (m)	$\hat{\gamma}$	$\hat{\sigma}$
Estimate	13.0	−0.098	2.68
Bootstrap 95% interval	(11.5,16.3)	(−0.164,0.015)	(2.28,3.07)

lar, the signs and placement of c_θ^- and c_θ^+ in the expression for the confidence interval for θ at step 7, since $\sigma_{a_{\hat{\theta}}}^{-1}(\hat{\theta}-\theta) \in (c_\theta^-, c_\theta^+\theta)$ is equivalent to $\theta \in (\hat{\theta}-c_\theta^+\sigma_{a\hat{\theta}}, \hat{\theta}-c_\theta^-\sigma_{a\hat{\theta}})$.

We apply the method to a subset of the GoM hindcast data for which all 72 locations provide storm peak H_S^{sp} values in excess of 2.5 m, so that we can assume that data for each location are reasonably modeled by a common extreme value distribution. A total of 148 storms were used. Results of extreme value modeling using $B=500$ bootstrap realizations are given in Table 2. We observe, in particular, that the 95% interval for each parameter is asymmetric. For $\hat{H}_{S100yrMP}$ in particular, asymmetry is consistent with theory and previous findings for single locations [4].

It is interesting to compare estimates in Table 2 to those in Table 1. Recalling that for Table 1, asymptotic 95% confidence bands can be constructed using $(\hat{\theta}-1.96\sigma_{a_{\hat{\theta}}}, \hat{\theta}+1.96\sigma_{a_{\hat{\theta}}})$, we see that confidence intervals for $\hat{\gamma}$ and $\hat{\sigma}$ in Table 2 are narrower than those based on an individual location, but considerably wider than those based on combining data for all locations without regard for dependence structure. That is, site averaging reduces uncertainty of $(\hat{\gamma}, \hat{\sigma})$ to some extent. Conversely, for $\hat{H}_{S100yrMP}$, the confidence band in Table 2 is almost as wide as that estimated using a single location. These inferences concur with those from the simulations in Sec. 4 for case $\beta=1$.

The performance of the bootstrapping method was quantified in terms of the coverage for interval estimates $(\hat{\theta}-c_\theta^+\sigma_{a\hat{\theta}}, \hat{\theta}-c_\theta^-\sigma_{a\hat{\theta}})$ for $\theta \in \{\gamma, \sigma, H_{S100yrMP}\}$, when the true data model is known. We performed the following simulation study. Data samples of size 150 storms for 72 locations were generated from GDP(−0.1,2.7) for three different situations. In the first case (no dependence), independent data samples were generated for each location. In the second case, (perfect dependence), identical data were used for each location for any given storm. In the third case, (GOMOS resample), a resample (stormwise across all locations) of the actual GOMOS data was used. The no-dependence and perfect-dependence cases correspond to limiting dependence structures that we would expect to encounter.

For each of 500 realizations of the data, $B=500$ bootstrap samples were used to estimate extreme value parameter uncertainty. Results are given in Table 3, in terms of the number of exceedances of the bootstrap interval on the left- and right-hand sides.

We expect total exceedance to be 5%, since we are using a 95% interval, with an average 12.5 of the 500 realizations to provide exceedances on each of the left- and right-hand sides. Values for number of exceedances in Table 3 confirm that the bootstrap confidence interval estimate is performing adequately in all three cases; numbers of exceedances are generally consistent with expectation. We also note that intervals $(-c_\theta^+, -c_\theta^-)$ for $\hat{\sigma}$, in particular, correspond reasonably with intuition. For the no-dependence case, $(-c_\theta^+, -c_\theta^-)$ is approximately equal to the asymptotic value (−1.96,1.96). For the perfect-dependence case, the interval estimate is approximately $\sqrt{72} \cong 8.5$ times as wide as the no-dependence case. The GOMOS resample case is intermediate, as would be expected. We note that interval estimates are skewed, skewness being considerable for γ and especially $\hat{H}_{S100yrMP}$.

We note at this point that a brief study of the Heffernan and Tawn [12] approach was undertaken, but found not to provide that same quality of coverage performance especially for the no-dependence and perfect-dependence cases. This is perhaps not surprising since rank correlation is an inadequate characterisation of the dependency structure in these cases.

Table 3 Performance of interval estimates $(\hat{\theta}-c_\theta^+\sigma_{a\hat{\theta}}, \hat{\theta}-c_\theta^-\sigma_{a\hat{\theta}})$ for $\theta \in \{\gamma, \sigma, H_{S100yrMP}\}$ for different known dependence structures. Note: q used as shorthand for $H_{S100yrMP}$.

Critical values of 95% interval estimates

Data	$\theta=\hat{q}$		$\theta=\hat{\gamma}$		$\theta=\hat{\sigma}$	
	$-c_\theta^+$	$-c_\theta^-$	$-c_\theta^+$	$-c_\theta^-$	$-c_\theta^+$	$-c_\theta^-$
No dependence	−1.8	2.1	−1.8	2.1	−1.9	1.9
Perfect dependence	−10.6	39.9	−13.8	28.5	−17.9	18.7
GOMOS resample	−10.3	25.2	−8.2	14.0	−11.0	11.6

Coverage for 95% interval estimates (numbers of left- and right-hand exceedences)

Data	$\theta=\hat{q}$		$\theta=\hat{\gamma}$		$\theta=\hat{\sigma}$	
	LH	RH	LH	RH	LH	RH
No dependence	14/500	14/500	17/500	13/500	18/500	17/500
Perfect dependence	27/500	21/500	8/500	9/500	16/500	7/500
GOMOS resample	21/500	9/500	24/500	12/500	6/500	28/500

6 Discussion

The value of extreme value index for the current data is small and negative. This requires careful interpretation of estimates of $H_{S100yrMP}$ used for structural design purposes. For values of γ of approximately −0.3 (typical for northern North Sea environments), the distribution of the 100-year significant wave height H_{S100yr} is relatively symmetric about $H_{S100yrMP}$. However, as the value of γ increases to zero from below, the distribution of H_{S100yr} becomes increasingly skewed to the right. $H_{S100yrMP}$ remains its most probable value, but there is increasing probability of considerably larger values than $H_{S100yrMP}$. In particular, from a structural design perspective, broad-brush application of a design safety factor may be inappropriate. Indeed, it might be more appropriate to estimate the value of some high percentile (say 95%) of the H_{S100yr} distribution, rather than its most probable value. Specifically, $H_{S100yr0.95}$ is the value of H_{S100yr} exceeded once in every 20 independent locations studied. From above, we have $H_{S100yrMP} = \hat{\sigma}/\hat{\gamma}(p^{-\hat{\gamma}}-1)+u$ where $p=1/n_{100}$ and n_{100} is the expected number of storms in a 100-year period.

A similar expression can be derived for the value of $H_{S100yr(1-q)}$, the value of the 100-year significant wave height exceeded with probability q in any 100-year period, assuming that the number of storms in 100 years is Poisson distributed,

$$H_{S100yr(1-q)} = \frac{\hat{\sigma}}{\hat{\gamma}}\left(\left\{\log_e\left[\frac{(1-q)^{-1}}{n_{100}}\right]\right\}^{-\hat{\gamma}} - 1\right) + u$$

For $H_{S100yr0.95}$, $q=0.05$. Figure 8 shows the behavior of $H_{S100yrMP}$ and H_{S100yr} as a function of $\hat{\gamma}$ for the standard case $\hat{\sigma}=1$, $u=0$. Values for other $\hat{\sigma}$ and u can be read from Fig. 8 by multiplying the value of the ordinate from the figure by $\hat{\sigma}$ then adding u. $H_{S100yr0.95}$ grows more quickly than $H_{S100yrMP}$ as $\hat{\gamma}$ approaches 0 from below. Thus, for the current GOMOS data, the most probable value for H_{S100yr} is ~13 m, whereas the value of $H_{S100yr0.95}$ is above 17 m. It is illustrative to consider the implications of Fig. 8 for northern North Sea conditions compared to Gulf of Mexico. For the northern North Sea, the distribution of H_{S100yr} is approximately symmetric (index is approximately −0.3). However, for the Gulf, the distribution is asymmetric (index is ~0).

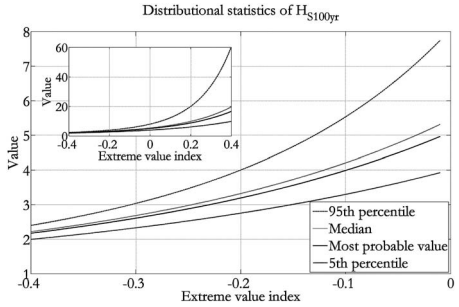

Fig. 8 $H_{S100yrMP}$, $H_{S100yr0.95}$, $H_{S100yr0.50}$, and $H_{S100yr0.05}$ as a function of extreme value index $\hat{\gamma}$ for the standard case $\hat{\sigma}=1$, $u=0$. The distribution becomes skewed to the right as $\hat{\gamma}$ approaches zero from below.

Thus, even for the same value of $H_{S100yrMP}$ in both environments, we expect a greater proportion of larger values of H_{S100yr} in the Gulf than in the northern North Sea.

For comparison, Fig. 8 also shows values of the median $H_{S100yr0.50}$ and fifth percentile $H_{S100yr0.05}$. The inset in Fig. 8 extends the illustration to small positive extreme value index. The current findings clearly demonstrate the need to accommodate wave height uncertainty explicitly (for example in the form of the distribution for H_{S100yr} rather than its most probable value $H_{S100yrMP}$) for reliable structural design.

It might be noted that the current analysis assumes the distribution of extremes of storm peak H_S to be homogeneous in space, time, and direction. The current data does not contradict the assumption of homogeneity in space for the spatial scale under consideration. However, there is clear evidence in the data suggesting temporal variability (e.g., from comparison of data for the periods before and after 1950) and directional variability for hurricanes in the Gulf of Mexico. The intention of the current investigation is to develop a pragmatic but reliable approach to the estimation of the uncertainty of extreme storm conditions for the simple model presented. In a future analysis, it would be worthwhile to extend the model formulation in Sec. 3 to include both directional and temporal variation of the extreme value distribution; an extension incorporating directional variability has recently been presented by Ewans and Jonathan [14]. Over wider spatial domains, spatial variability would also need to be accommodated in a full spatiotemporal directional extreme value model.

7 Conclusions

The main conclusions of the investigation are as follows:

1. Modeling using the generalized Pareto distribution for data typical of the current hindcast example produces biased estimates for extreme value index, scale, and quantiles from small samples. Simulation studies for single and multiple locations show that, for data typical of the GOMOS hindcast examined here, using <100 independent points for modeling leads to underestimation of events, such as $H_{S100yrMP}$. Magnitudes of extreme value index γ and scale σ are overestimated, but these effects partly compensate for $H_{S100yrMP}$ quantile estimates. Similar effects are observed for Weibull fitting of the same data.
2. Combining dependent data from different locations reduces bias of estimates. Given that all locations can be modeled using a common extreme value distribution, estimates for γ, σ, and $H_{S100yrMP}$ become less biased as effective sample size increases, regardless of whether that sample is drawn from multiple dependent locations. Nevertheless, site averaging can have (at worst) no effect and can also be of little benefit when strong location dependence exists. Site averaging also provides less benefit for estimates of extreme quantiles than for index and scale.
3. Reliable estimates for parameter uncertainties for data from multiple dependent locations are not trivial to obtain. Asymptotic forms for standard errors of maximum likelihood estimators for γ, σ, and $H_{S100yrMP}$ are available but cannot be used reliably to quantify uncertainty. We present a nonparameteric studentized bootstrapping approach and demonstrate its performance in providing $\sim 95\%$ confidence intervals using GOMOS hindcast and simulated data corresponding to a range of dependence structures.
4. Estimates of $H_{S100yrMP}$ used for structural design purposes require careful interpretation, especially when the extreme value index is small and negative. As γ increases to zero (and beyond) from below, the distribution of H_{S100yr} becomes increasingly skewed to the right and there is increasing probability of considerably larger values of H_{S100yr} for given $H_{S100yrMP}$.

Acknowledgment

The authors acknowledge the support of Shell International Exploration and Production and Shell Research Limited for this work. The authors further acknowledge supporting discussions with Michael Vogel.

References

[1] Rutten, J. G., Ewans, K. C., van Gelder, P. H. A. J. M., and Efthymiou, M., 2004, *Uncertainties in Extreme Value Analysis and Their Effect on Load Factors*," *Proc., of 23rd International Conference on Offshore Mechanics and Arctic Engineering*, June 20–25, Vancouver.
[2] Forristall, G. Z., Larrabee, R. D., and Mercier, R. S., 1991, "Combined Oceanographic Criteria for Deepwater Structures in the Gulf of Mexico," *Offshore Technology Conference Proceedings*, Houston, Offshore Technology Conference, Houston, paper No. OTC 6541.
[3] Tromans, P. S., and Vanderschuren, L., 1995, "Response Based Design Conditions in the North Sea: Application of a New Method," Offshore Tech. Conf., Paper no. OTC 7683, pp. 387–397.
[4] Elsignhorst, C., Groeneboom, P., Jonathan, P., Smulders, L., and Taylor, P. H., 1998, "Extreme Value Analysis of North Sea Storm Severity," ASME J. Offshore Mech. Arct. Eng., **120**, pp. 177–183.
[5] Reiss, R.-D., and Thomas, M., 2001, *Statistical Analysis of Extreme Values*, Birkhauser Verlag, Basel, Switzerland.
[6] Coles, S. G., and Tawn, J. A., 1994, "Statistical Methods for Multivariate Extremes: An Application to Structural Design," Appl. Stat., **43**, pp. 1–48.
[7] Kotz, S., and Nadarajah, S., 2000, *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.
[8] Coles, S., and Simiu, E., 2003, "Estimating Uncertainty in the Extreme Value Analysis of Data Generated by a Hurricane Simulation Model" J. Eng. Mech., **129**, pp. 1288–1294.
[9] Hall, P., 1988, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, Berlin.
[10] Efron, B., and Tibshirani, R. J., 1993, *An Introduction to the Bootstrap*, Chapman and Hall, New York.
[11] Davison, A. C., and Hinkley, D. A., 1997, *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, England.
[12] Heffernan, J. E., and Tawn, J. A., 2004, "A Conditional Approach for Multivariate Extreme Values," J. R. Stat. Soc. Ser. B (Stat. Methodol.), **66**, pp. 497–546.
[13] Oceanweather Inc., 2005, "GOMOS—USA Gulf of Mexico Oceanographic Study, Northern Gulf of Mexico Archive, Oct.
[14] Ewans, K., and Jonathan, P., 2006, "Estimating Extreme Wave Design Criteria Incorporating Directionality," *9th International Workshop on Wave Hindcasting & Forecasting*, Victoria, BC, Canada, Sept. 24–29.