Omnidirectional return values for storm severity from directional extreme value models: the effect of physical environment and sample size

David Randell
Shell Projects & Technology,
Manchester, M22 0RR, UK.
david.randell@shell.com

Elena Zanini
Department of Mathematics and Statistics,
Lancaster University, LA1 4YF, UK.
ezanini1@lancs.ac.uk

Michael Vogel
Shell Projects & Technology,
Houston, TX 77079-1197, USA.
michael.vogel@shell.com

Kevin Ewans
Sarawak Shell Bhd.,
50450 Kuala Lumpur, Malaysia.
kevin.ewans@shell.com

Philip Jonathan[*]
Shell Projects & Technology,
Manchester, M22 0RR, UK.
philip.jonathan@shell.com

ABSTRACT

Ewans and Jonathan [2008] shows that characteristics of extreme storm severity in the northern North Sea vary with storm direction. Jonathan et al. [2008] demonstrates, when directional effects are present, that omnidirectional return values should be estimated using a directional extreme value model. Omnidirectional return values so calculated are different in general to those estimated using a model which incorrectly assumes stationarity (or inhomogeneity) with respect to direction. The extent of directional variability of extreme storm severity depends on a number of physical factors, including fetch variability. Our ability to assess directional variability of extreme value parameters and return values also improves with increasing sample size in general. In this work, we estimate directional extreme value models for samples of hindcast storm peak significant wave height from locations in ocean basins worldwide, for a range of physical environments, sample sizes and periods of observation. At each location, we compare distributions of omnidirectional 100-year return values estimated using a directional model, to those (incorrectly) estimated assuming stationarity. The directional model for peaks over threshold of storm peak significant wave height is estimated using a non-homogeneous point process model as outlined in Randell et al. [2013]. Directional models for extreme value threshold (using quantile regression), rate of occurrence of threshold exceedances (using a Poisson model), and size of exceedances (using a generalised Pareto model) are estimated. Model parameters are described as smooth functions of direction using periodic B-splines. Parameter estimation is performed using maximum likelihood estimation penalised for parameter roughness. A bootstrap re-sampling procedure, encompassing all inference steps, quantifies uncertainties in, and dependence structure of, parameter estimates and omnidirectional return values. Physical environment has a large effect on estimated distributions of 100-year return values; the most severe environments of those considered are the Gulf of Mexico and northern North Sea. However, when return value distributions are normalised relative to their median values, the (normalised) return value distributions for all locations considered are remarkably similar. Moreover, once the effect of sample size is accounted for, the widths of return value distributions (quantified in terms of the inter-quartile range) are also remarkably consistent. The effect on estimated return value distributions of neglecting the influence of non-stationarity at different stages of the extreme value modelling procedure is unpredictable; a fully non-stationary model is recommended. In general, accommodating non-stationarity in extreme value threshold and rate of occurrence of threshold exceedance appears most critical.

---

[*]Address all correspondence to this author.

1 Introduction

Coastal and marine structures must be designed for extreme environmental conditions. Design codes stipulate that off-shore structures should be designed to exceed specific levels of reliability, expressed in terms of an annual probability of failure or return-period. To define the environmental loading, metocean criteria therefore need to be specified to an appropriate return period, typically 100 years, but sometimes to 10000 years for the required failure probabilities. Extreme value analysis of data from measurements and/or hindcasts are undertaken to derive these criteria, but the resulting uncertainties associated with long return period criteria are usually large, as the number of data available for analysis is usually small by comparison. In addition, there is no standard approach to analysis within the metocean community, and it remains a subject of continuous debate and active research.

There is a large literature on applied extreme value analysis relevant to ocean engineering. Threshold methods in extreme value analysis are reviewed by Scarrott and MacDonald [2012]. Tancredi et al. [2006] considers accounting for threshold uncertainty in extreme value anlaysis. Wadsworth and Tawn [2012] presents likelihood-based procedures for threshold diagnostics and uncertainty. Thompson et al. [2009] proposes automatic threshold selection for extreme value analysis. Thompson et al. [2010] reports Bayesian non-parametric regression using splines. Muraleedharan et al. [2012] and Cai and Reeve [2013] model significant wave height distributions with quantile functions for estimation of extreme wave heights. Scotto and Guedes-Soares [2000] and Scotto and Guedes-Soares [2007] discuss the long-term prediction of significant wave height. Methods for analysis of time-series extremes are reviewed by Chavez-Demoulin and Davison [2012]. Ferro and Segers [2003] and Fawcett and Walshaw [2007] discuss modelling of clustered extremes. Mendez et al. [2006] considers long-term variability of extreme significant wave height using a time–dependent POT model. Ruggiero et al. [2010] reports increasing wave heights and extreme value projections for the US Pacific Northwest. Calderon-Vega et al. [2013] models seasonal variation of extremes in the Gulf of Mexico using a time-dependent GEV model. Mendez et al. [2008] considers the seasonality and duration in extreme value distributions of significant wave height. Mackay et al. [2010] discusses discrete seasonal and directional models for the estimation of extreme wave conditions. Eastoe and Tawn [2012] models non-stationary extremes with application to surface level ozone. Chavez-Demoulin and Davison [2005] provides a nice introduction to modelling non-stationary extremes using splines, and Davison et al. [2012] is a good introduction to spatial extremes. Jonathan and Ewans [2013] overviews extreme value analysis from a met–ocean perspective.

The requirement to consider directional effects in developing extreme criteria has been well demonstrated. Although directional effects have long been known (see, for example, Graham 1981), and rigourous techniques for dealing with covariates like direction in estimates have also been available for some time (see, for example, Davison and Smith 1990), it is only recently that such methods have been adopted for establishing metocean design criteria. Graham [1981] notes inconsistencies associated with estimating directional wave extremes with the omnidirectional extreme, and much debate has ensued. Forristall [2004] notes the importance of ensuring that directional criteria are consistent with omnidirectional criteria in the sense that they yield consistent failure probabilities in design. In particular, he points out that directional criteria that are simply scaled to omnidirectional criteria estimated independently give inconsistent probabilities. In this work, we present a rigorous and consistent approach (see Section 3) to estimation of cumulative distribution functions of omnidirectional design criteria, using a non-stationary directional extreme value model, summarised in Randell et al. [2013] and Jonathan et al. [2014]. The approach requires the estimation of extreme value threshold, rate of threshold exceedance and size of threshold exceedance - all of which are functions of direction. In Section 4 we use the model to estimate omnidirectional design values for samples of storm peak significant wave height from 8 locations worldwide. In Section 5 we quantify the difference between estimates for 100-year return values estimated by simulation from extreme value models with directional effects incorporated fully (in threshold, rate and size), partly (in threshold and rate only) and not at all. Note that, as is common in the statistics literature, we refer to "(non-) stationarity" with respect to direction in this work; some readers may prefer the term "inhomogeneity" in this context.

Uncertainty in return value estimates for storm severity at a location is determined by a number of factors, but predominantly by (a) the physical characteristics of the ocean environment at that location, and (b) the size of the sample available for extreme value modelling. For example, fluctuations in atmospheric pressure over the ocean produces storms whose severities are inherently random; this natural (or aleatory) uncertainty cannot be reduced (other than by intervening directly and changing the physics of the environment!). However, return value uncertainty is also strongly influenced by sample size: all other things being equal, the larger the sample, the smaller the sample (or epistemic) uncertainty. In Section 5 we explore how the uncertainty in the 100-year return value for storm peak significant wave height varies as a function of environment and sample size.

First, however, we introduce the 8 samples of storm severities.

2 Data and general wave climates

Eight samples of storm peak significant wave height H_S^{sp} and corresponding storm peak directions θ^{sp} for different time periods and locations worldwide are considered in this work. Each of the eight samples corresponds to hindcast data for a single geographic location within the ocean basin of interest. Storm peak data are isolated from hindcast time-series of significant wave height H_S and (dominant) wave direction θ in time as described in Ewans and Jonathan [2008]. Briefly,
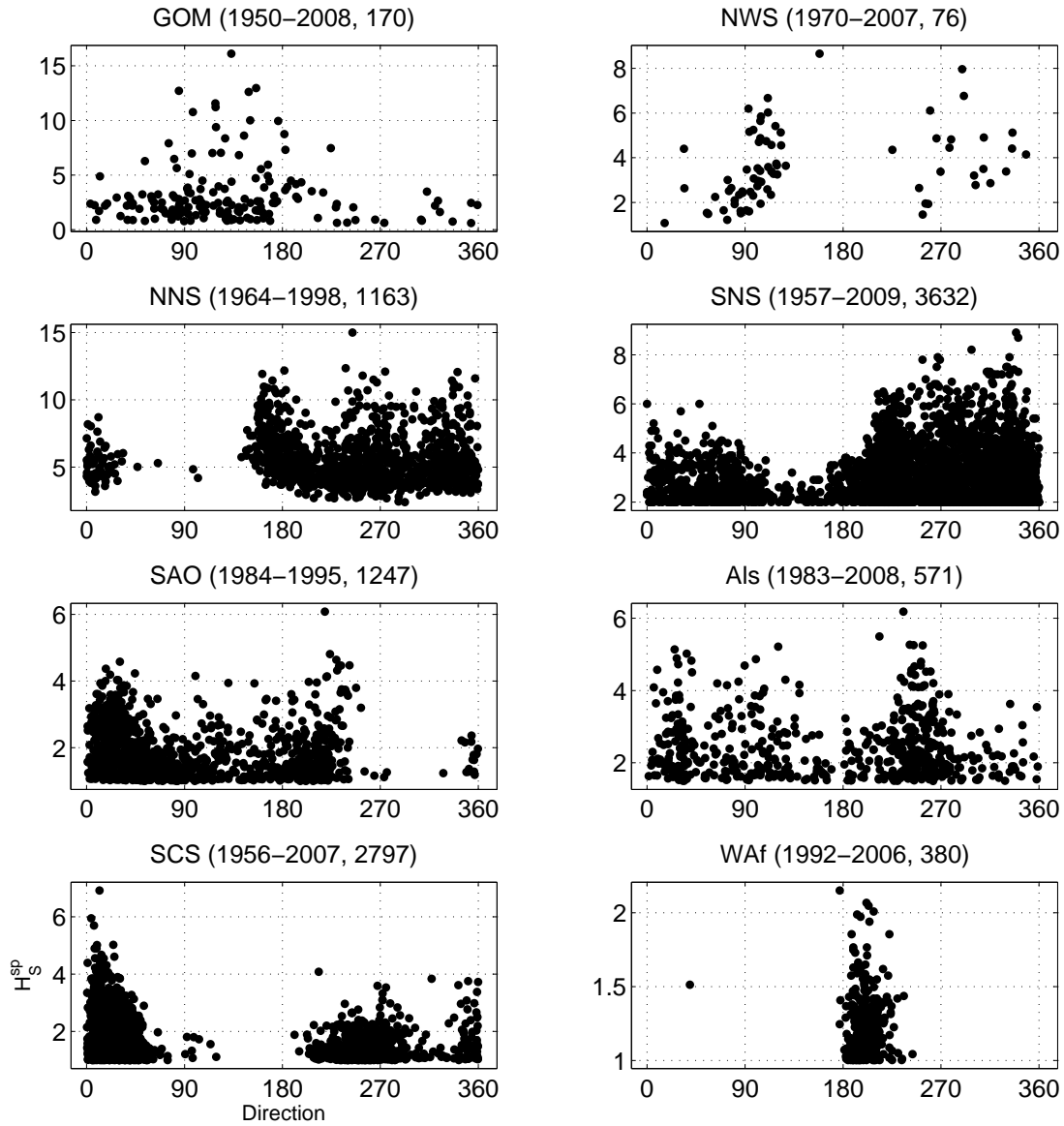
Fig. 1. Storm peak significant wave height H_S^{sp} on storm peak direction θ^{sp} for the 8 locations under consideration. From right to left, top to bottom: Gulf of Mexico (GOM), North-West Shelf of Australia (NWS), Northern North Sea (NNS), Southern North Sea (SNS), South Atlantic Ocean (SAO), Alaska (Als), South China Sea (SCS) and West Africa (WAf). Panel titles give the location, the sample period and storm peak sample size.

contiguous intervals of H_S above a low *peak-picking* threshold are identified, each interval corresponding to a storm event. In this work, the peak-picking threshold for each sample is constant with respect to θ. The maximum of significant wave height during the storm event is taken as the storm peak significant wave height for the storm. The value of wave direction at the time of the storm peak is referred to as storm peak direction. Below, when referring to storm direction, we mean explicitly the dominant wave direction for the storm peak sea state.

The samples were chosen to reflect typical ocean environments of interest to the ocean engineer. The prevailing climates of the 8 locations impose different directional characteristics on H_S^{sp} as illustrated in Figure 1, which shows H_S^{sp} (in metres, corresponding to the storm peak sea-state of three hours duration throughout) on storm peak direction θ^{sp} for all samples. Direction (from which waves emanate) is measured clockwise from North, so that $90°$ corresponds to a storm peak from the east. Hindcast data providers are acknowledged in the descriptions of the 8 environments below. Confidentiality prevents

precise specification of the locations in each case.

In the Gulf of Mexico, hurricanes produce the most severe sea states. Most storms originate in the Atlantic Ocean between June and November and propagate west to northwest into the Gulf producing the largest sea states with dominant directions from the southeast to east directions. Whereas more than 100 years of hurricane data are available for hindcast modelling, data prior to the commencement of overflights in 1944 are regarded as less reliable. The sample used here corresponds to the period 1950-2008, consisting of 170 hurricane events. The sample was drawn from the GOMOS [2008] hindcast study.

Extreme sea states off the northwest shelf of Australia are also dominated by tropical storms, occurring from October to April. Reliable storm data are available from around 1970, with between three to five storms occurring at a given location per annum. Some of these storms do not produce significant sea states, so that the number of data available for a given location is less than 100. The general direction of propagation of the storms over the region is south to southwest, with some re-curving eastward at lower latitudes. Accordingly, the extreme sea states tend to have dominant directions from west through north to east. The sample was drawn from the NAMOS hindcast study.

Extreme sea states in the North Sea are dominated by winter storms originating in the Atlantic Ocean and propagating eastwards across the northern part of the North Sea. Due to their proximity to the storms, sea states at northern North Sea locations are usually more intense than in the southern North Sea. Occasionally, the storms travel southeast-ward and intrude into the southern North Sea producing large sea states (see, for example, Magnusson and Donelan 2012). The directionality of the extreme seas varies considerably with location, depending on land shadows of the UK, Scandinavia, and the coast of mainland Europe, and the fetches associated with the Atlantic Ocean, Norwegian Sea, and the North Sea itself. In the northern North Sea the main fetches are the Norwegian Sea to the North, the Atlantic Ocean to the west, and the North Sea to the south. Extreme sea states from the directions of Scandinavia to the east and the British Isles to the southwest are not possible. The shielding by these land masses is more effective for southern North Sea locations, resulting in a similar directional distribution but reduced wave heights by comparison with northern North Sea locations. With up to several tens of storms impacting the North Sea each winter, the number of events for analysis is several times that of locations in tropical cyclone dominated regions of the Gulf of Mexico and the North-West Shelf of Australia. The samples were drawn from the WAM (Reistad et al. [2009]) hindcast study.

For the South Atlantic Ocean sample, observations correspond to storm peaks for the period June 1984 to July 1995. Three directional regimes are identified by consideration of fetch conditions, corresponding to the land shadow of Brazil (directional sector $[240, 350)$), the northern Atlantic (directional sector $[350, 90)$) and the southern Atlantic (directional sector $[90, 350)$). The largest winds in this region are generated by smaller-scale intensified cyclonic extra-tropical lows. These events cause large winds from the southeast through the southwest and, due to the long fetch from this sector, large waves with particularly long periods can be generated during these events. In general, the open exposure to the southern South Atlantic leads to sea states that have a significant swell component, most prevalent during the austral winter (June, July, August). The sample was drawn from the BOMOS hindcast study.

The Alaskan sample corresponds to a location in the eastern Chukchi Sea off the northwest coast of Alaska. Weinzapfel et al. [2011] report three primary storm tracks affecting coastal waters of northern Alaska in summer and early fall. Wave conditions generated by these systems may be categorized into several broad groups (Francis and Atkinson 2012). The most prevalent track has lows that move from the Bering Sea into the Chukchi Sea from the south. These lows are usually strongest when they first enter the Chukchi Sea. The associated pressure gradient will tighten across the entire North Slope and winds will increase out of the east and east-northeast, increasing most in the northern Chukchi Sea and in the western Beaufort Sea. Lows on this track typically begin to fill and weaken as they reach the northern Chukchi Sea area and then fade to the northwest. Another storm track generates lows that move north from the Gulf of Alaska into the interior of Alaska, continue north or north-northeast, and then emerge in the Canadian portion of the Beaufort Sea. As the low moves through the interior, the gradient tightens along the North Slope. This is apparent more so in the eastern Beaufort Sea than the western Beaufort Sea or the Chukchi Sea. Once these lows reach the eastern Beaufort Sea, they usually begin to weaken and fill like their counterparts in the Chukchi Sea. A third storm track that affects these areas presents lows that move in from the East Siberian Sea and move west to east across the Chukchi and Beaufort Seas. These storm systems induce westerly winds across the region, opposite from the prevailing easterlies. This third storm track often produces the strongest storm systems seen across the region during summer and early autumn. Wind waves are mitigated by the annual development of shore-fast ice. The sample was drawn from the GROW-FINE hindcast study.

The main climatic features of the South China Sea are monsoons. The southwest monsoon occurs in the northern summer and the northeast monsoon in the northern winter. Monsoonal surges cause increased sea states and are the source of the extreme sea states in the southern areas off the west coast of Borneo. Typhoons dominate the extremes in the northern South China Sea, but the extreme sea state data we are examining in this paper are off Borneo and are dominated by monsoonal surges with southeast and northeast directions. The number of monsoonal surges each season is similar to the number of winter storms experienced in the North Sea, and so the number of data available for extreme value analysis is rather similar to that for the North Sea. The sample was drawn from the SEAFINE hindcast study.

In the absence of strong wind forcing in low latitudes off west Africa, extreme sea states correspond to swell events and

are relatively mild. Almost all swell events originate from mid-latitude storms progressing eastward in the south Atlantic and Southern Ocean. Thus, the dominant directions associated with these events at low latitude locations off west Africa occur within a narrow south to southwest sector. As the number of events more or less corresponds to the number of southern mid-latitude storms, the number of data available for extremal analysis is similar to a North Sea or South China Sea location. Although squall events also occur, they are not present in the hindcast sample, drawn from the WANE hindcast study.

3 Model

We seek to estimate a directional extreme value model for storm peak significant wave height H_S^{sp} at each location, the parameters of which vary smoothly with respect to storm peak direction θ^{sp}.

3.1 Model components

Following Randell et al. [2013], for each sample $\{\dot{z}_i\}_{i=1}^{\dot{n}}$ of $\dot{n}$ storm peak significant wave heights observed with storm peak directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$ (with the "sp" superscript suppressed for brevity), we proceed using the peaks over threshold approach as follows.

Threshold: We first estimate a threshold function ψ above which observations $\dot{z}$ are assumed to be extreme. The threshold varies smoothly as a function of direction ($\psi \stackrel{\triangle}{=} \psi(\theta)$) and is estimated using quantile regression. We retain the set of n threshold exceedances $\{z_i\}_{i=1}^{n}$ observed with storm peak directions $\{\theta_i\}_{i=1}^{n}$ for further modelling. The quantile regression lack of fit criterion is given in section 3.2.

Rate of occurrence of threshold exceedance: We next estimate rate of occurrence ρ of threshold exceedance using a Poisson process model with Poisson rate $\rho(\stackrel{\triangle}{=} \rho(\theta))$. The Poisson density is:

$$f_\rho = \exp(-\int \rho(\theta)d\theta)\prod_{i=1}^{n}\rho(\theta_i)$$

Magnitude of occurrence of threshold exceedance: We estimate size of occurrence of threshold exceedance using a generalised Pareto (henceforth GP) model. Briefly, for a peak X over threshold ψ, the form of the generalised Pareto distribution with shape parameter ξ and scale parameter σ is

$$F_\psi(x) = \frac{F(x) - F(\psi)}{1 - F(\psi)} = \Pr(X \leq x | X > \psi) = 1 - (1 + \frac{\xi}{\sigma}(x - \psi))^{-1/\xi} \quad \text{for } \xi \neq 0$$

for $\psi \in (-\infty, \infty)$, $\sigma \in (0, \infty)$ and $\xi \in (-\infty, \infty)$ with $x > \psi$, $1 + \xi(x - \psi)/\sigma > 0$, with the right hand side taken to be $1 - \exp(-(x - \psi)/\sigma)$ when $\xi = 0$. Note that ξ and σ (and ψ) are assumed to vary smoothly with direction.

This approach to extreme value modelling follows that of Chavez-Demoulin and Davison [2005] and is equivalent to direct estimation of a non-homogeneous Poisson point process model (see, for example, Dixon et al. 1998, Jonathan and Ewans [2013]).

3.2 Parameter estimation

For quantile regression, we seek a smooth function ψ of direction corresponding to non-exceedance probability τ of H_S^{sp} for any θ. We estimate ψ by minimising the quantile regression lack of fit criterion

$$\ell_\psi = \{\tau \sum_{i,r_i \geq 0}^{n} |r_i| + (1 - \tau) \sum_{i,r_i < 0}^{n} |r_i|\}$$

for residuals $r_i = z_i - \psi(\theta_i; \tau)$. We regulate the smoothness of the quantile function by penalising lack of fit for parameter roughness R_ψ (with direction), by minimising the penalised criterion

$$\ell_\psi^* = \ell_\psi + \lambda_\psi R_\psi$$

where the value of roughness coefficient λ_ψ is selected using cross-validation to provide good predictive performance.

For Poisson modelling, we use penalised likelihood estimation. The rate ρ of threshold exceedance is estimated by minimising the roughness-penalised (negative log) likelihood

$$\ell_\rho^* = \ell_\rho + \lambda_\rho R_\rho$$

where R_ρ is parameter roughness with respect to direction, λ_ρ is again evaluated using cross-validation, and Poisson (negative log) likelihood is given by

$$\ell_\rho = -\sum_{i=1}^{n} \log \rho(\theta_i) + \int \rho(\theta) d\theta$$

For computational ease, ρ is estimated on a partition of the covariate domain into m equally-sized contiguous intervals as explained in Chavez-Demoulin and Davison [2005]. The generalised Pareto model of size of threshold exceedance is estimated in a similar manner by minimising the roughness penalised (negative log) GP likelihood

$$\ell_{\xi,\sigma}^* = \ell_{\xi,\sigma} + \lambda_\xi R_\xi + \lambda_\sigma R_\sigma$$

where R_ξ and R_σ are parameter roughnesses with respect to direction, λ_ξ and λ_σ are evaluated using cross-validation, and GP (negative log) likelihood is given by

$$\ell_{\xi,\sigma} = \sum_{i=1}^{n} \log \sigma_i + (\frac{1}{\xi_i} + 1) \log(1 + \frac{\xi_i}{\sigma_i}(z_i - \psi_i))$$

where $\psi_i = \psi(\theta_i)$, $\xi_i = \xi(\theta_i)$ and $\sigma_i = \sigma(\theta_i)$, and a similar expression is used when $\xi_i = 0$ (see Jonathan and Ewans 2013). In practice, we set $\lambda_\xi = \kappa \lambda_\sigma$ for prespecified constant κ, so that only one cross-validation loop is necessary. The value of κ is estimated by inspection of the relative smoothness of ξ and σ with respect to covariates.

3.3 Parameter smoothness

Physical considerations suggest that we should expect the model parameters ψ, ρ, ξ and σ to vary smoothly with respect to direction θ. For estimation, this can be achieved by expressing each parameter in terms of an appropriate basis for the directional domain $D = [0, 360)$. We calculate a periodic marginal B-spline basis matrix B_θ for an index set of 32 directional knots, so that the linear combination $B\beta$ provides a flexible description of any function varying with θ on the index set, for appropriate choice of coefficient vector β. We also allocate observations to 32 directional bins centred at the spline knots, based on (circular) directional distance. Model estimation therefore reduces to estimating appropriate basis coefficients for each of ψ, ρ, ξ and σ. The roughness R of any function can also be easily evaluated on the index set. Following the approach of Eilers and Marx (see, for example, Eilers and Marx 2010), we define roughness using

$$R = \beta' P \beta$$

where the penalty matrix P is easily evaluated for the directional domains. The form of P is motivated by taking differences of neighbouring values of β, penalising lack of local smoothness.

3.4 Uncertainty quantification

Bootstrap resampling is used for uncertainty quantification. 95% bootstrap uncertainty bands are estimated by repeating the full extreme value analysis for 1000 resamples of the original storm peak sample. In particular, estimation of optimal roughness penalties is performed independently for each bootstrap resample, so that uncertainty bands also reflect variability in these choices. It was also confirmed that 1000 resamples was sufficient to ensure stability of bootstrap confidence intervals. Experience of similar applications suggests that the naive bootstrap used here provides reasonable estimates of confidence intervals, but we note that other more sophisticated bootstrap schemes are available, including the bias-corrected accelerated bootstrap (BCa, Efron 1987, DiCiccio and Efron 1996). Note that bootstrap uncertainty is also fully incorporated within the estimation of return value distributions in Section 4 in the spirit of the weighted likelihood bootstrap of Newton and Raftery [1994].

3.5 Estimated parameters

For all models discussed here, extreme value threshold corresponding to a non-exceedance probability of 0.8 were set. As an illustration, Figure 2 shows plots for ψ, ρ, ξ and σ with direction, for non-exceedance probability 0.8 of H_S^{sp}, in the case of the South Atlantic Ocean (SAO) sample. The directional extreme value threshold ψ is largest for storms peaks directions from the southwest, then for storms peaks directions from the northwest. In the directional interval $[240, 350)$ corresponding to the land-shadow of Brazil, since no observations are present to be allocated to directional knots, the directional threshold is set to the global sample median; this has no effect on the analysis, since there are no data present for these directions. The rate ρ of threshold exceedance is highest for storms from the northeast, then for storms from the southwest. From inspection of the 95% uncertainty bands in the figure, it is clear that there is significant directional variability in both ψ and ρ. The
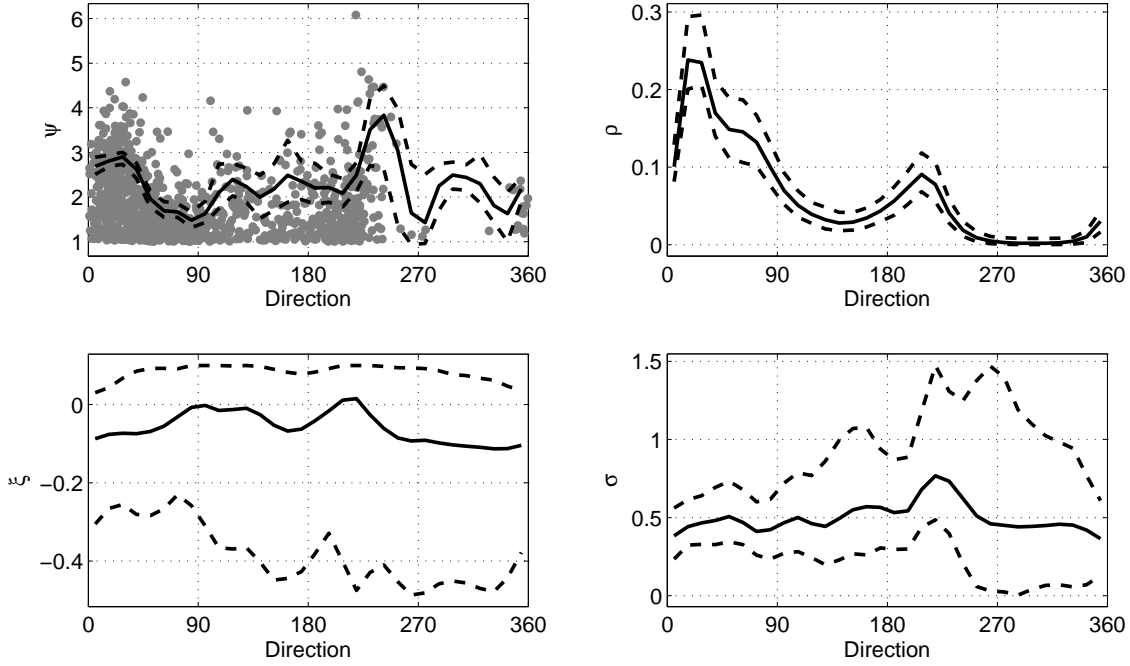
Fig. 2. Extreme value analysis of H_S^{sp} for the South Atlantic Ocean (SAO) sample: Extreme value threshold ψ, rate of threshold exceedance ρ, generalised Pareto shape ξ and scale σ with storm peak direction θ^{sp} using a threshold non-exceedance probability of 0.8. Each panel shows bootstrap median threshold (solid black) on storm peak direction θ^{sp} with 95% bootstrap uncertainty band (dashed black). Also shown, for ψ, is the actual sample (grey).

same cannot be said for the generalised Pareto shape ξ and scale σ estimates, which are highly uncertain with direction as illustrated by the wide bootstrap uncertainty bands. There is evidence for a peak of both ξ and σ at around $200°$.

Model diagnostics are essential to demonstrate adequate model fit. For the current applications, we are particularly concerned that estimated storm peak extreme value models generate directional distributions of H_S^{sp} consistent with observed storm peak data. To quantify this, we use a simulation procedure (described below in Section 4) to generate 1000 realisations of storms, each realisation for the same period as the original data. We then construct 95% uncertainty bands for cumulative distribution functions (cdfs) of H_S^{sp}, omnidirectionally and partitioned by direction. Then we confirm that empirical cdfs for the actual data are consistent with the simulated cdfs. Figure 3 illustrates this the South Atlantic Ocean (SAO) sample. The left-hand panel shows the omnidirectional cdf for the original sample (red), the corresponding median from simulation (solid black), together with 2.5%ile and 97.5%ile from simulation (both dashed). The right hand panel compares cdfs for directional octants in the same way. There is good agreement for all samples.

4 Estimation of return values

Return values corresponding to some return period P of interest are estimated by simulation under the model developed in Section 3. The procedure is as follows, for each of a large number of realisations of storms:

1. Select a bootstrap resample and the corresponding estimated directional extreme value model for storm peak significant wave height.
2. For each directional covariate bin, estimate the number of storm peak realisations to be drawn at random from a Poisson distribution, using the estimated directional rate of threshold exceedance, ρ, for that bin, scaled to return period, P. If T is the period of the original sample, the scaled rate for period P is $\rho \times P/T$). Then, for each storm peak realisation:

   (a) Draw a pair of values for storm peak direction θ^{sp*} at random from the directional bin.
   (b) Draw a value of storm peak significant wave height H_S^{sp*} corresponding to θ^{sp*} at random from corresponding the generalised Pareto model.

3. Accumulate maximum values for H_S^{sp*} per directional bin.

Empirical cumulative distribution functions for storm peak events for return period P are then trivially estimated by sorting the values for simulated storm peaks for arbitrary combinations of directional bins. Since realisations based on models from different bootstrap resamples of the original sample are used, the resulting cumulative distribution functions
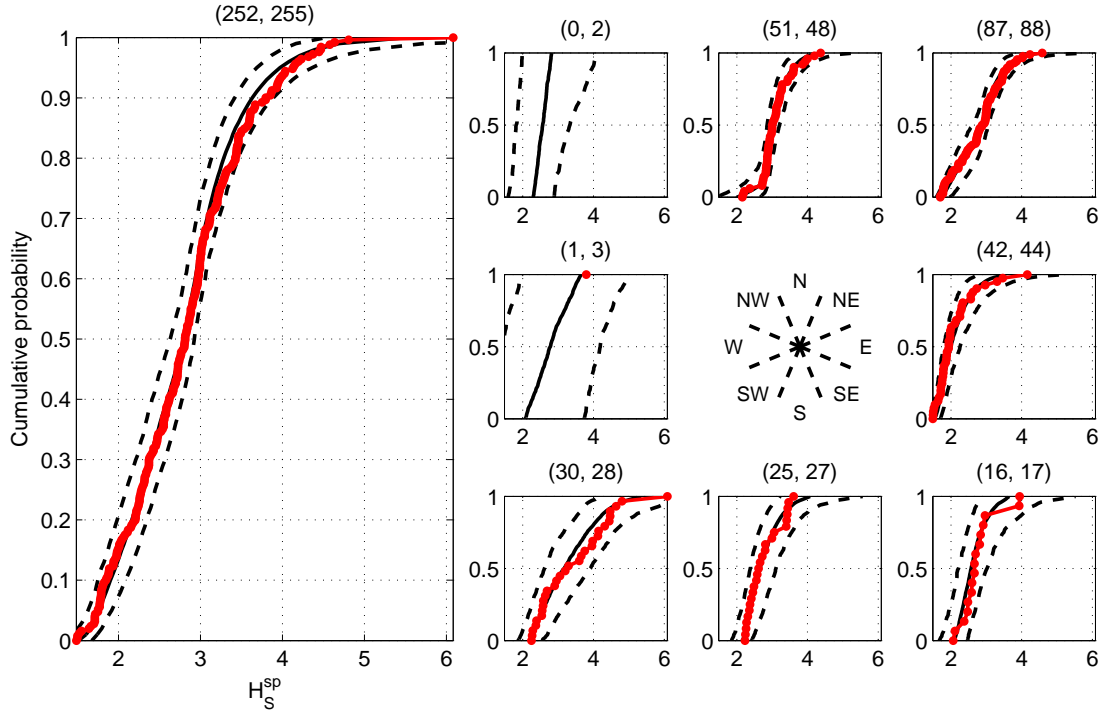
Fig. 3. Validation of directional model for H_S^{sp} for the South Atlantic Ocean (SAO) sample, by comparison of cumulative distribution functions (cdfs) for original sample with those from 1000 sample realisations under the model corresponding to the same time period as the original sample. The left-hand panel shows the omnidirectional cdf for the original sample (red), the corresponding median from simulation (solid black), together with 2.5%ile and 97.5%ile from simulation (both dashed). The right hand panel compares cdfs for 8 directional octants (centred on cardinal and semi-cardinal directions) in the same way. Titles for plots, in brackets, are the number of actual events and the average number of simulated events in each directional sector.

incorporate both the (aleatory) inherent randomness of return values and the extra (epistemic) uncertainty introduced by model parameter estimation from the sample data.

A return value (such as "a 100-year event") is itself a random quantity with a probability distribution. Engineering design convention is to summarise the distribution of a return value using one number: its most probable value, corresponding to the mode of the probability density of the return value. In a stationary case, when the generalised Pareto shape parameter of the corresponding extreme value model is known to be identically zero, the mode of the return value distribution corresponds to the return value with non-exceedance probability $\exp(-1) \approx 0.37$. That is, there is a probability of ≈ 0.63 that the largest event in the return period will exceed the (most probable = engineering) return value in reality! In general, the non-exceedance probability of the mode of the return value distribution depends on the values of all of ψ, ρ, ξ and σ, which are themselves functions of θ in this article. More importantly, the width of the return value distribution (e.g. its inter-quartile range) also depends on all model parameters in general. In order to understand the relative severities of extreme events at different locations, it is wise therefore to consider the complete distribution of return value, rather than just its most probable value. This is the purpose of the simulation performed here.

Figures 4 and 5 show directional 100-year return value distributions for all samples in terms of cumulative distribution functions for individual directional octants and the omnidirectional case. Systematic variation of return values with direction is clear.

Figure 6 compares omnidirectional 100-year return value distributions for all samples. The severest environments are seen to be the northern North Sea (NNS) and Gulf of Mexico (GOM). It is apparent that the width of the distribution of return value is also considerably larger for these two locations. To investigate this effect further, Figure 7 compares omnidirectional 100-year return value distributions for all samples first centred with respect to the median 100-year return value for the sample, and then scaled with respect to it. On the original scale, NNS has the longest right-hand tail, and West Africa (WAf) the shortest. When scaled with respect to the median of the distribution, the distributions of return values are also more consistent. WAf remains the shortest tailed, but now the South Atlantic Ocean (SAO) is the longest tailed.

5 Discussion

Uncertainty in a return value such as H_{S100} at some location is determined by (a) the physical environment (producing storms whose severities are inherently random - this natural variability cannot be reduced) and (b) the sample (the uncertainty
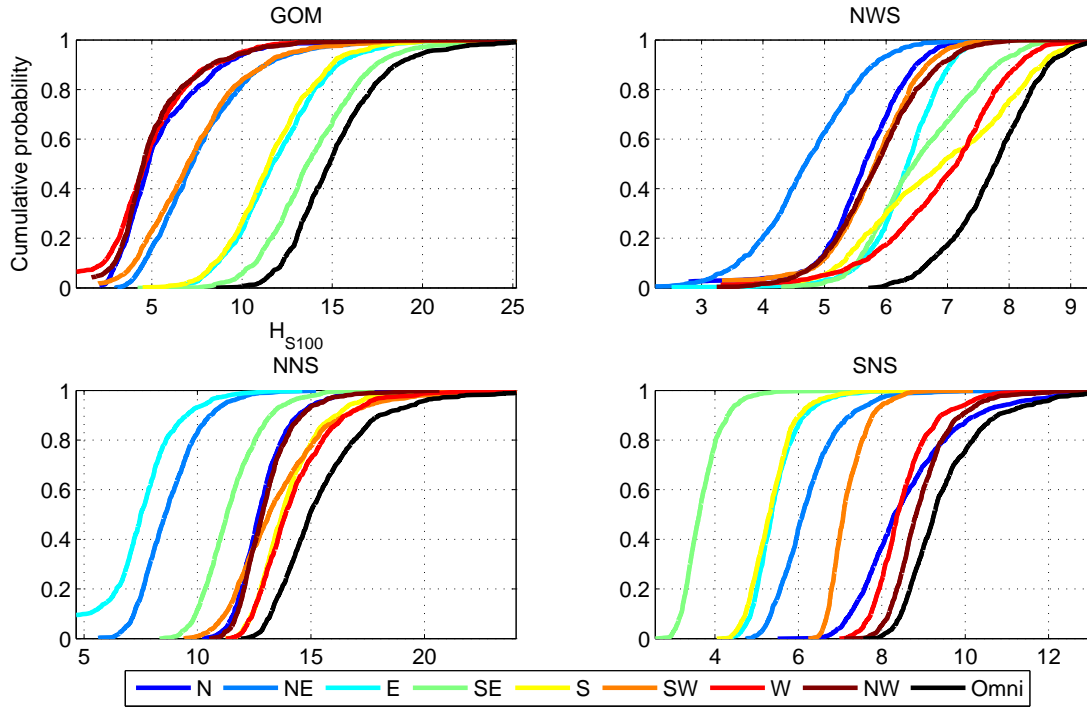
Fig. 4. Directional 100-year return value distributions for GOM, NWS, NNS and SNS samples. Each panel shows cumulative distribution functions for individual directional octants (as labelled beneath the figure) and the omnidirectional case (black). Return value distributions are estimated by simulation under the directional model, incorporating uncertainty in parameter estimation using bootstrap resampling. Directional octants are centred on cardinal and semi-cardinal directions, and labelled clockwise from North with colours from dark to light blue through yellow to light then dark red.The omnidirectional estimate is the rightmost curve.

due to which can be reduced in principle by increasing sample size). We might expect to see some evidence for the relative impact of natural and sample uncertainty in the estimated cumulative distribution functions of H_{S100} in Figures 4-7. To aid discussion, Figure 8 gives inter-quartile ranges (IQR, that is, the difference between the 75%ile and 25%ile) for H_{S100} distributions from different samples as a function of sample size. The left hand plot shows IQR on the original scale, and the right hand plot shows median-scaled IQR (that is, IQR divided by the median of the corresponding cumulative distribution function). The lines in Figure 8 represent minimum and maximum inter-quartile ranges estimated from models of 25 randomly-chosen time-intervals (of given sizes) of the Southern North Sea (SNS) sample. These show that inter-quartile range reduces as a function of sample size as expected. Relative to the SNS, therefore, it would appear that the IQR for the Gulf of Mexico (GOM) location is large on the original scale (probably due to the small sample size, see Figure 1), and that for West Africa (WAf) is particularly small (probably due to the near unidirectionality of the sample). The median-scaled IQR for the South China Sea (SCS) sample is also large compared to SNS (probably due to the physical environment). IQRs for the North-West Shelf of Australia (NWS), the Northern North Sea (NNS), South Atlantic Ocean (SAO) and Alaska (Als) show reasonable consistency with SNS in both original and median-scaled plots.

Jonathan et al. [2008] demonstrates that fitting a non-stationary sample using a stationary model leads to biased estimates of return values, whereas adoption of models which incorporate covariate effects provide relatively unbiased estimates. For data exhibiting directional effects, a directional extreme value model generally explains observed variation signicantly better than a model which ignores directionality, and that omnidirectional criteria developed from a directional model are different from those generated when directionality is not accounted for. They further show, for simulated directional data with known extreme value characteristics, that omnidirectional criteria derived from a directional model are more accurate and should be preferred in general over those based on models which ignore directional effects. It is interesting therefore to compare cumulative distribution functions for H_{S100} using different model which incorporate directional modelling to different degrees. Figure 9 compares omnidirectional 100-year return value distributions from stationary and non-stationary models for all samples.

For each location, cumulative distribution functions (cdfs) in black correspond to a fully directional model in which all of quantile extreme value threshold ψ, rate of threshold exceedance ρ, generalised Pareto shape ξ and scale σ (see Section 3) are functions of direction. Cdfs corresponding to a "semi-directional" model are in blue, in which ψ and ρ are directional, but ξ and σ are constants with direction). Cdfs in red correspond to a constant model, in which all model parameters are fixed
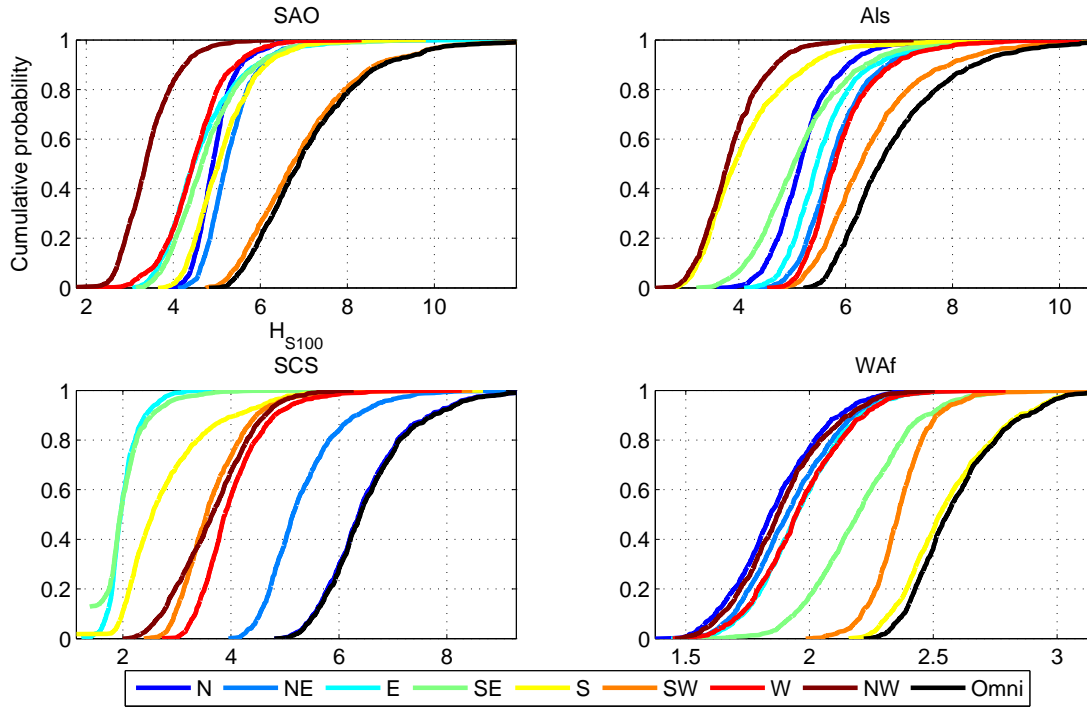
Fig. 5. Directional 100-year return value distributions for SAO, Als, SCS and WAf samples. See Figure 4 for further description.

with respect to direction. Estimated cdfs are very similar for GOM, but there is considerable variation between estimates for most locations. In general, for these samples, we see that (a) fully-directional estimates agree well with "semi-directional" estimates, for all but the SNS sample; (b) the characteristics of cdfs from the constant model are unpredictable with respect to those from the fully-directional and "semi-directional" model; (c) there is no systematic difference in width or median value of cdfs from the fully-directional and "semi-directional" models compared to the constant model. From (a) we infer, as is intuitively obvious, that accommodating covariate effects in extreme value threshold and threshold exceedance rate estimation is sometimes sufficient for a good extreme value model, and that accommodating covariate effects in models for sizes of threshold exceedances is sometimes - but not always - less important. From (b) and (c) we infer that a constant model gives unreliable estimates with unpredictable behaviour with respect to full-directional and "semi-directional" models when covariate effects are present. For all models and all locations discussed in this work, an extreme value threshold corresponding to a non-exceedance probability of 0.8 is adopted. That is, only the largest 20% of storm peak significant wave height events are considered for extreme value modelling. Clearly it might be the case that different threshold levels should be considered for different model forms at different locations. In ongoing work, the effect of threshold selection (in particular) on the relative characteristics of return value distributions is being examined more thoroughly. We are also exploring, both theoretically and by means of further extreme value modelling of hindcast samples, the impact of location characteristics on the widths (for example, inter-quartile ranges) of return value distributions in particular.

In this work, we estimate omnidirectional return value distributions for eight geographic locations in different ocean basins using a non-stationary extreme value model. Physical environment has a large effect on estimated distributions of 100-year return values; the most severe environments of those considered are the Gulf of Mexico and northern North Sea. However, when return value distributions are normalised relative to their median values, the (normalised) return value distributions for all locations considered are remarkably similar. Moreover, once the effect of sample size is accounted for, the widths of return value distributions (quantified in terms of the inter-quartile range) are also remarkably consistent. The effect on estimated return value distributions of neglecting the influence of non-stationarity at different stages of the extreme value modelling procedure is unpredictable; a fully non-stationary model is recommended. In general, accommodating non-stationarity in extreme value threshold and rate of occurrence of threshold exceedance appears most critical.
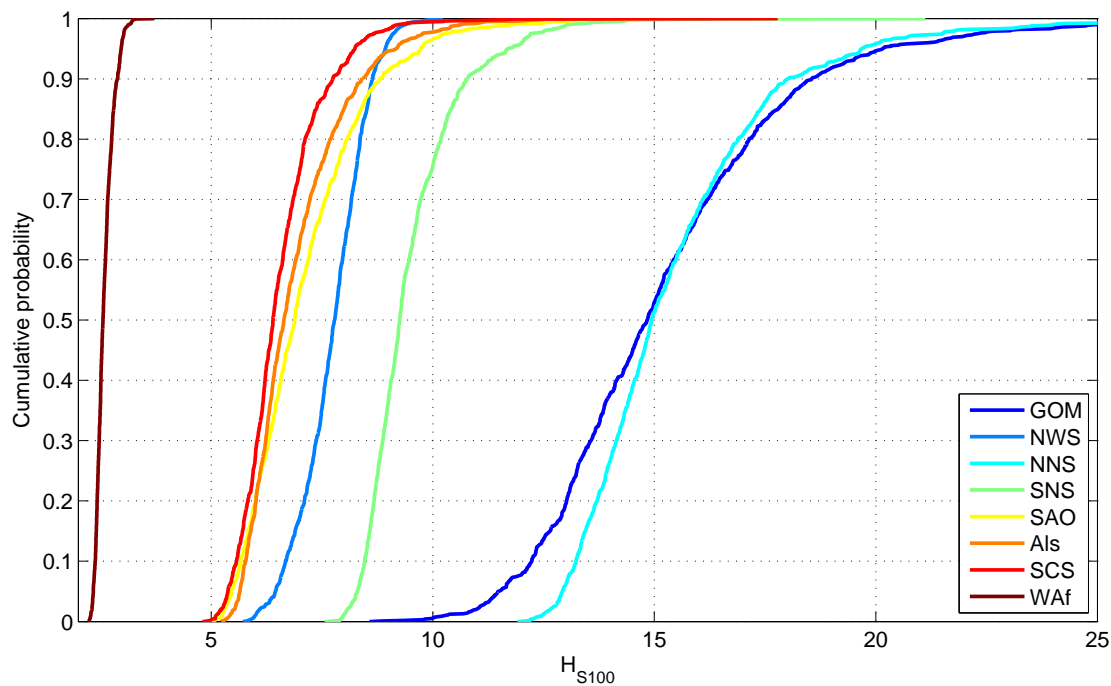
Fig. 6.   Omni-directional 100-year return value distributions for all samples (as labelled), estimated by simulation under the directional model, incorporating uncertainty in parameter estimation using bootstrap resampling.

References

BOMOS. *Brazil Offshore Metocean Study Joint Industry Project*. Oceanweather Inc.

Y. Cai and D. E. Reeve. Extreme value prediction via a quantile function model. *Coastal Eng.*, 77:91–98, 2013.

F. Calderon-Vega, A. O. Vazquez-Hernandez, and A. D. Garcia-Soto. Analysis of extreme waves with seasonal variation in the Gulf of Mexico using a time-dependent GEV model. *Ocean Eng.*, 73:68–82, 2013.

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207, 2005.

V. Chavez-Demoulin and A.C. Davison. Modelling time series extremes. *REVSTAT - Statistical Journal*, 10:109–133, 2012.

A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statistical Science*, 27:161–186, 2012.

A.C. Davison and R. L. Smith. Models for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393, 1990.

T. J. DiCiccio and B. Efron. Bootstrap confidence intervals. *Statist. Sci.*, 11:189–228, 1996.

J. M. Dixon, J. A. Tawn, and J. M. Vassie. Spatial modelling of extreme sea-levels. *Environmetrics*, 9:283–301, 1998.

E.F. Eastoe and J.A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Biometrika, doi: 10.1093/biomet/asr078*, 2012.

B. Efron. Better bootstrap confidence intervals. *J. Am. Statist. Soc.*, 82:171–185, 1987.

P H C Eilers and B D Marx. Splines, knots and penalties. *Wiley Interscience Reviews: Computational Statistics*, 2:637–653, 2010.

K. C. Ewans and P. Jonathan. The effect of directionality on northern North Sea extreme wave design criteria. *J. Offshore Mechanics Arctic Engineering*, 130:10, 2008.

L. Fawcett and D. Walshaw. Improved estimation for temporally clustered extremes. *Environmetrics*, 18:173–188, 2007.

C. A. T. Ferro and J. Segers. Inference for clusters of extreme values. *J. Roy. Statist. Soc. B*, 65:545–556, 2003.

G. Z. Forristall. On the use of directional wave criteria. *J. Wtrwy., Port, Coast., Oc. Eng.*, 130:272–275, 2004.

O. P. Francis and D. E. Atkinson. Synoptic forcing of wave states in the southeast Chukchi Sea, Alaska, at an offshore location. *Nat. Hazards*, 62:1169–1189, 2012.

GOMOS. *Gulf of Mexico hindcast study*. Oceanweather Inc., 2008.

C. G. Graham. Problems with the design statistical approach. In *Extreme Value Analysis of Directional Wind and Wave Data, Int. Conf. on Wave and Wind Directionality, Paris.*, pages 379–401, 1981.

GROW-FINE. *GROW Fine Arctic Hindcast Study*. Oceanweather Inc.

P. Jonathan and K. C. Ewans. Statistical modelling of extreme ocean environments with implications for marine design : a
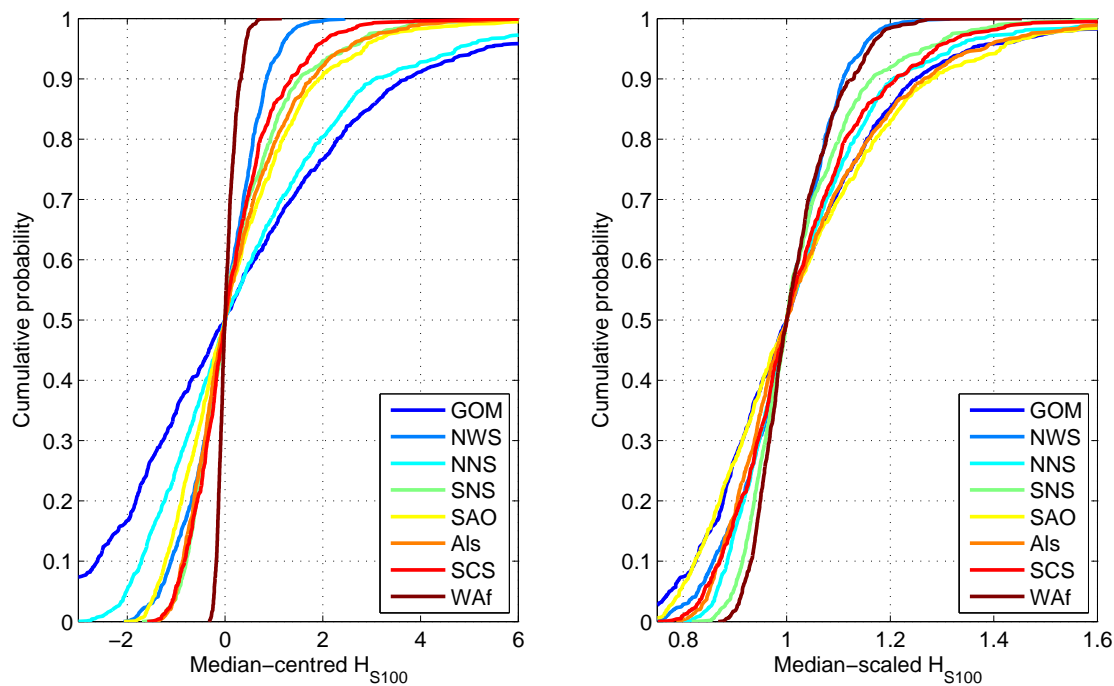
Fig. 7. Omni-directional 100-year return value distributions for all samples (as labelled) centred on the median 100-year return value for each sample (left) and scaled with respect to it (right), estimated by simulation (see Figure 6).

review. *Ocean Engineering*, 62:91–109, 2013.

P. Jonathan, K. C. Ewans, and G. Z. Forristall. Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects. *Ocean Eng.*, 35:1211–1225, 2008.

P. Jonathan, D. Randell, Y. Wu, and K. Ewans. Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *(Accepted by Ocean Engineering July 2014, draft at www.lancs.ac.uk/∼jonathan)*, 2014.

E. B. L. Mackay, P. G. Challenor, and A. S. Bahaj. On the use of discrete seasonal and directional models for the estimation of extreme wave conditions. *Ocean Eng.*, 37:425–442, 2010.

A. K. Magnusson and M. A. Donelan. Omae2012-84099: The Andrea wave characteristics of a measured North Sea rogue wave. *Proc. 31st Int. Conf. of the OMAE, Rio de Janeiro, Brazil*, 2012.

F J Mendez, M Menendez, A Luceno, R Medina, and N E Graham. Seasonality and duration in extreme value distributions of significant wave height. *Ocean Eng.*, 35:131–138, 2008.

F.J. Mendez, M. Menendez, A. Luceno, and I.J. Losada. Estimation of the long-term variability of extreme significant wave height using a time–dependent pot model. *Journal of Geophysical Research*, 11:C07024, 2006.

G. Muraleedharan, Claudia Lucas, C. Guedes Soares, N. Unnikrishnan Nair, and P.G. Kurup. Modelling significant wave height distributions with quantile functions for estimation of extreme wave heights. *Ocean Eng.*, 54:119–131, 2012.

NAMOS. *North Australia Metocean Study Joint Industry Project*. Oceanweather Inc.

M A Newton and A E Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. B*, 56:3–48, 1994.

D. Randell, Y. Wu, P. Jonathan, and K. C. Ewans. Omae2013-10187: Modelling covariate effects in extremes of storm severity on the Australian North West Shelf. *Proc. 32nd Conf. Offshore Mech. Arct. Eng.*, 2013.

M Reistad, O Breivik, H Haakenstad, O J Aarnes, and B R. Furevik. A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea and the Barents Sea. *Norwegian Meteorological Institute*, 2009.

P Ruggiero, P D Komar, and J C Allan. Increasing wave heights and extreme value projections: The wave climate of the US pacific northwest. *Coastal Eng.*, 57:539–522, 2010.

C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10:33–60, 2012.

M.G. Scotto and C. Guedes-Soares. Modelling the long-term time series of significant wave height with non-linear threshold models. *Coastal Eng.*, 40:313, 2000.

M.G. Scotto and C. Guedes-Soares. Bayesian inference for long-term prediction of significant wave height. *Coastal Eng.*, 54:393, 2007.
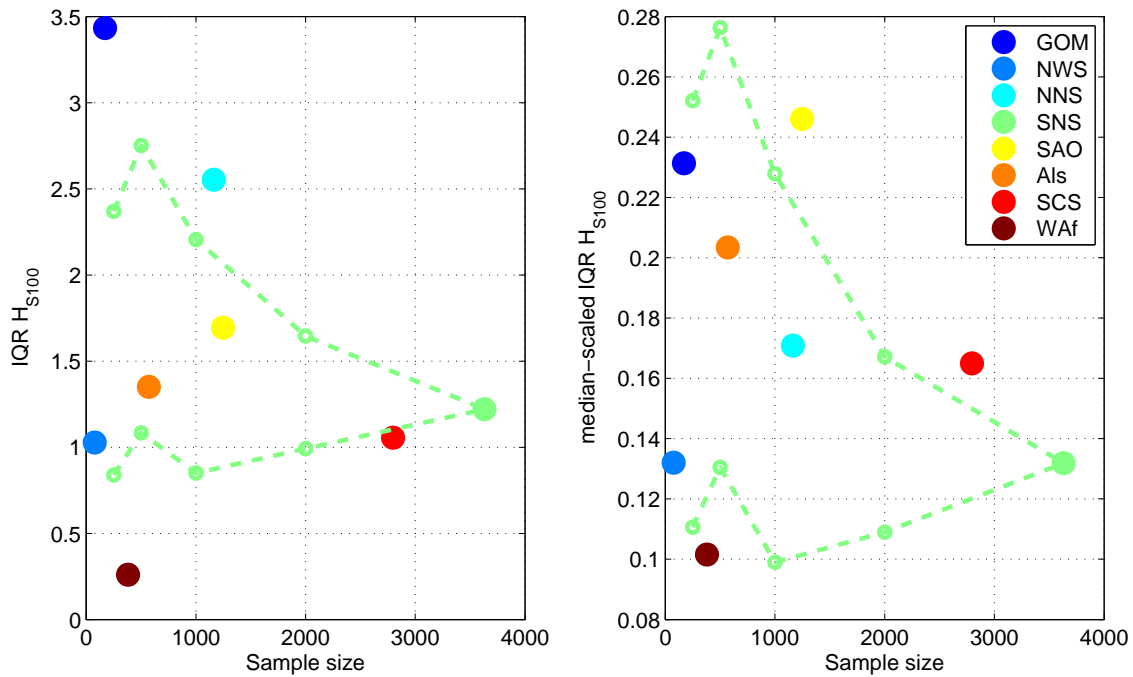
Fig. 8. Inter-quartile range (IQR) of the H_{S100} distributions (from Figure 6 or 7, coloured dots) as a function of sample size. The dashed lines are minimum and maximum inter-quartile ranges estimated using 25 randomly-chosen time-intervals (of given size) of the Southern North Sea (SNS) sample. The left hand plot shows IQR on the original scale, whereas IQRs are median-scaled on the right hand side.

SEAFINE. *South East Asia Meteorological and Oceanographic Hindcast (SEAMOS) Fine Grid Hindcast Study Joint Industry Project*. Oceanweather Inc.

A. Tancredi, C.W. Anderson, and A. O'Hagan. Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9:87–106, 2006.

P Thompson, Y Cai, D Reeve, and J Stander. Automated threshold selection methods for extreme wave analysis. *Coastal Eng.*, 56:1013–1021, 2009.

P. Thompson, Y. Cai, R. Moyeed, D. Reeve, and J. Stander. Bayesian nonparametric quantile regression using splines. *Computational Statistics and Data Analysis*, 54:1138–1150, 2010.

J. L. Wadsworth and J. A. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. Roy. Statist. Soc. B*, 2012.

WANE. *West Africa Extremes Joint Industry Project*. Oceanweather Inc.

R. Weinzapfel, G. Harvey, J. Andrews, L. Clamp, and J. Dykas. Winds, waves and sea ice during the 1999-2007 open water seasons of the Beaufort and Chukchi Seas. *Offshore Technology Conference, OTC 21782*, 2011.
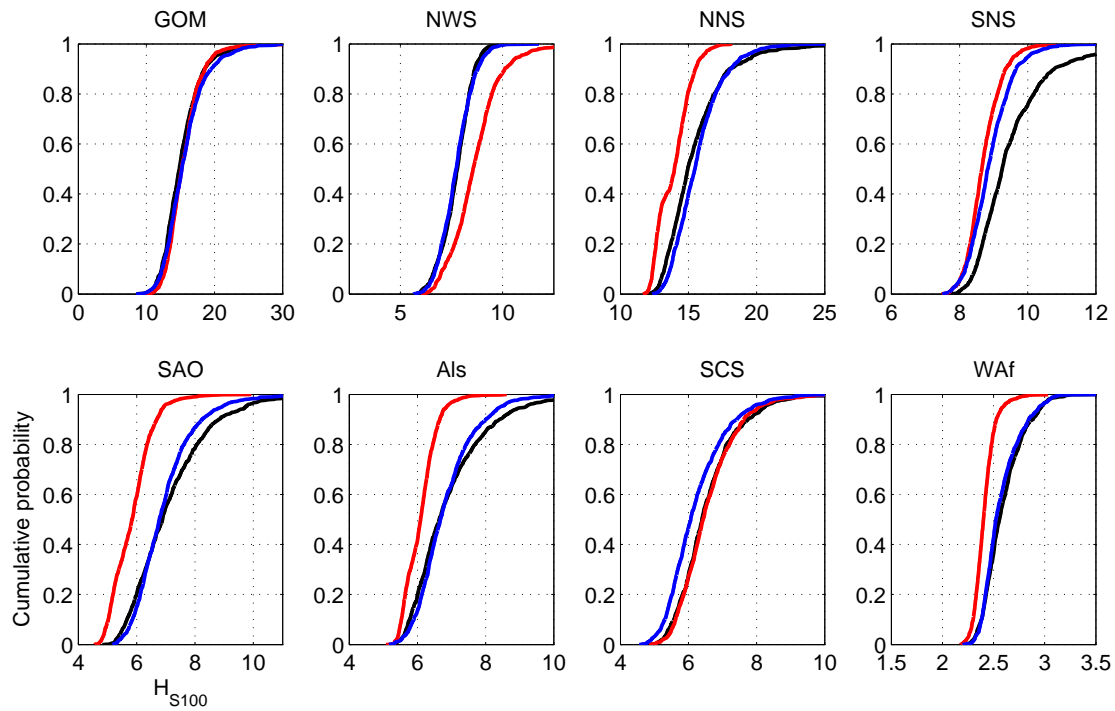
Fig. 9. Comparison of omnidirectional 100-year return value distributions from stationary and non-stationary models, for all samples. Each panel shows omnidirectional cumulative distribution functions for the corresponding sample. These are estimated by simulation under a fully-directional model (in black, in which all of quantile extreme value threshold ψ, rate of threshold exceedance ρ, generalised Pareto shape ξ and scale σ are functions of direction), a "semi-directional" model (in blue, in which ψ and ρ are directional, but ξ and σ are constants with direction) and a constant or stationary model (in red, in which no model parameters vary with direction).