Uncertainty due to choice of measurement scale in Extreme Value modelling of North Sea Storm Severity

D. T. Reeve

University of Lancaster, U.K.

D. Randell

Shell Projects and Technology, U.K.

K. C. Ewans

Shell Sarawak, Malaysia.

P. Jonathan*

Shell Projects and Technology, U.K.

Published. Ocean Engineering 53 (2012) 164–176.

Abstract

Modelling extreme storm severity is critical to design and reliable operation of marine structures. Extreme hindcast storm peak significant wave heights (H_S) for 816 locations throughout the North Sea are modelled, using the four parameter Poisson point process model of Wadsworth et al. (2010), incorporating measurement scale variability via a Box-Cox transformation. The model allows estimation of the posterior distribution for measurement scale parameter and point process parameters within a Bayesian framework. The effect of measurement scale on return values of significant wave height (H_S) is quantified by comparison with a three parameter Poisson point process model ignoring measurement scale uncertainty. It is found that the median value (over all locations) of the median posterior Box-Cox parameter (per location) is approximately 0.7, suggesting that the appropriate measurement scale for extreme value analysis is $H_S^{0.7}$. The value of the median Box-Cox parameter (per location) varies considerably between locations, with a 90% uncertainty band of approximately $(0.2, 2.2)$ and quartiles of 0.4 and 1.2; the value of Box-Cox parameter is also influenced by threshold choice for extreme value analysis in particular. The ratio

*philip.jonathan@shell.com, Tel:+44 151 272 5421, Fax:+44 151 +44 5384

(over all locations) of the (posterior median) return value from the four parameter model, to the return value from the three parameter model (and a return period of 100 times the period of the hindcast) has a median value of 0.92, suggesting that median return values may be reduced for this data set by better modelling of measurement scale effects. The ratio of return values has a 90% uncertainty band of approximately $(0.72, 1.37)$, illustrating the extra variability in return values that incorporation of measurement scale uncertainty introduces.

---

1. Introduction

Typical results of an extreme value analysis on a sample of data are generally not invariant to measurement scale. In an ocean engineering context, we might perform extreme value modelling of data for significant wave height H_S to estimate a return value corresponding to some return period. We might also choose to perform extreme value modelling on the squared values of the data, motivated by the fact that wave forcing on a drag-dominated structure is proportional to the square of wave height (Tromans and Vanderschuren, 1995), then estimate the return value and square-root it. Estimates for return values from inferences on the linear and square scales will be different in general.

A simple simulation study illustrates this feature. Samples of size 500, 1000, 10000 and 100000 are generated at random from a generalised Pareto distribution with parameters typical of North Sea storm conditions with negative shape parameter (resulting in a finite upper end point to the distribution). Return values (on the linear scale) corresponding to different return periods are then estimated from the samples by extreme value analysis on linear and square scales. Figure 1 compares estimates of mean return values from 1000 realisations of random samples of different sizes (broken lines) with the known values (solid). Estimates on the linear scale are close to the true curve. However, estimates from analysis on the square scale show considerable bias, which reduces as sample size increases. Inspection of diagnostic plots, such as the mean excess or shape parameter with threshold, does not typically permit the selection of a preferred modelling scale. In this example, data are simulated on the linear scale, the scale that is usually tacitly assumed (in this case correctly) by practitioners; but the example demonstrates the need for development of extreme value models which are less sensitive or invariant to measurement scale.

[Figure 1 about here.]

The reason for the apparent inconsistency in estimation of return values on different measurement scales is poor convergence of the distribution of sample maxima (or threshold exceedances) to the corresponding asymptotic distributional forms used to model the sample, on one or both of the measurement scales. Modelling extreme values using an asymptotic distributional form requires an assumption that the sample data are sufficiently large that the asymptotic distributional form is valid. In practice, we can never be sure of this; bias introduced by assuming an asymptotic distributional form, when this form is not appropriate, can be large, as is evident in Figure 1, especially for small samples. However, the asymptotic assumption will be better on some measurement scales than others.

The rate of convergence of maxima or threshold exceedances from the distribution from which the data are drawn (henceforth the *parent* distribution to avoid ambiguity) to the asymptotic form can be improved by judicious choice of measurement scale, thereby reducing bias in estimation of return values. Improving the rate of convergence is therefore a strong motivator to consider scale transformation prior to extreme value inference. For example, Cook (1982) proposes extreme value modelling of dynamic pressure (linearly related to the square of wind speed) instead of wind speed itself. Assuming that wind speed (for the United Kingdom) is (approximately) Rayleigh distributed, the square of wind speed is exponential distributed. Cook (1982) demonstrates that the rate of convergence of maxima from the distribution of wind speed to the asymptotic Gumbel form is slower than the rate of convergence of maxima from the distribution of squared wind speed to the same asymptotic form. Moreover, convergence of maxima from the distribution of wind speed is from below (in the sense that Gumbel fitting would over-estimate return values), whereas convergence of maxima from the distribution of squared wind speed is from above (so that Gumbel fitting under-estimates return values). These ideas are extended by Cook and Harris (2004) and Harris (2004). Following Cook (1982), Harris (1998) proposes that maxima of squared wind speeds should be modelled using the Gumbel distribution. Taylor and Goh (2000) investigate measurement scale effects in extreme value analysis of the most probable maximum wave height in a storm event, H_{MP}. They suggest performing extreme value analysis on the square scale; squared observations of H_{MP} are found to be approximately exponentially distributed. Two simple illustrations of the benefit of scale transformation are given in Appendix 1, for Weibull and generalised Pareto parent distributions.

Importantly, once convergence is effectively achieved on a particular measurement scale, we can

estimate return values reliably on that scale, and thereby estimate return values consistently on the original measurement scale using the inverse measurement scale transformation. Were we to achieve convergence on a number of measurement scales, there would be consistency between the estimates of return values on all of those scales.

In the statistics literature, Smith (1987) introduces so-called penultimate approximations of the nth power of any parent distribution, and the corresponding conditional distribution for excesses of a high threshold from the parent distribution, provided that the parent distribution is in the domain of attraction of a max-stable distribution. These approximations motivate the adoption of generalised extreme value (GEV) and generalised Pareto asymptotic forms respectively for statistical inference. Moreover, Smith (1987) evaluates the rate of convergence of the penultimate approximations to the corresponding asymptotic forms. Teugels and Vanroelen (2004) and Wadsworth et al. (2010) demonstrate that the rate of convergence of penultimate approximations can be improved by measurement scale transformation using the Box-Cox transformation. A brief statistical motivation is provided in Appendix 2.

Physical considerations also motivate measurement scale transformation in some instances. As already mentioned, since the square of wave height is linearly related to drag force on an offshore structure, some practitioners prefer this scale for extreme value analysis.

In practical application, it is not possible to specify a particular best measurement scale for extreme value analysis, since the parent distribution is generally unknown. To improve convergence and reduce bias in these circumstances, Wadsworth et al. (2010) suggest a four parameter Poisson point process model for extreme values in which measurement scale is treated as a model parameter, removing the need to select a particular measurement scale. Measurement scale variability is incorporated using the Box-Cox transformation ($y = \frac{x^\lambda - 1}{\lambda}$, see Box and Cox (1964), with $\lambda > 0$ here), the Box-Cox parameter λ being estimated in the modelling. Specifically, within a Bayesian framework, the joint posterior distributions of point process location, scale and shape, and (Box-Cox) measurement scale parameter is estimated using Markov Chain Monte Carlo (MCMC) simulation.

The objective of the current work is to explore the application of the four parameter Poisson point process model to hindcast storm peak significant wave height data for the North Sea, in particular to compare estimates of return values with those from a three parameter Poisson point process model fitted on the linear scale, thereby quantifying the effect of measurement scale uncertainty on return values. The extreme value analysis of significant wave heights is probably the most

important application of extremal analyses in the offshore industry; it is almost certainly the most frequently applied extreme value analysis in this field. For most offshore structures, environmental loading is dominated by wave loading; accurate estimates of low probability extreme wave events are necessary to ensure the structure's design is adequate and that the annual probability of failure of the facility due to environmental loading is at the appropriate level. Sea states from the North Sea are a good choice for examining measurement scale effects, due to widely varying wave climate from north to south. The northern North Sea is open to windows with very long fetches to the west and is exposed to severe winter storms moving from west to east at northern latitudes. The southern North Sea is relatively sheltered by adjacent coast lines, is more distant from severe winter storms, and wave heights are limited by shallow-water dissipation effects. Thus, the northern North Sea has significantly more extreme wave conditions than the southern North Sea, and this might be expected to be reflected in measurement scale variation.

The paper is arranged as follows. In Section 2 we introduce the North Sea hindcast data, and describe data pre-processing performed prior to extreme value analysis. This is followed by an overview of the four parameter point process model in Section 3. In Section 4, approximately 30 years of hindcast data for 816 locations in the North Sea are modelled. Return values for periods ranging from 0.1 to 100 times the period of the hindcast estimated and measurement scale uncertainty quantified by comparing estimates with a three parameter Poisson point process model. Spatial variability of parameter estimates and return values is also explored. Section 5 provides a discussion of results, conclusions, and suggestions for further work. Appendix 1 illustrates the usefulness of measurement scale transformation in two elementary examples. Appendix 2 outlines the theory of penultimate approximation underpinning the four parameter model. We emphasise that the objective and focus of the current work is to quantify the effect of incorporating measurement scale variability in the estimation of extreme value models, rather than to estimate particular extreme quantiles for offshore design purposes.

2. Data

The data examined are significant wave heights (H_S) from the NEXTRA North Sea hindcast for the period 1st October 1964 to 31st March 1995 inclusive, sampled continuously at 3h intervals for a total of 816 locations. It is essential to accommodate covariate effects in extreme value analysis (Jonathan et al., 2008). To create a homogeneous sample for subsequent extreme value analysis,

the effects of seasonal and directional covariates need to be eliminated as far as practicable. To reduce seasonal effects, since the most severe storms occur during the winter period, observations per location corresponding to the six month period from 1st April to 30th September were omitted from each year. Reduction of directional effects was achieved by selecting data corresponding to a $\pm 30^o$ angular sector around the direction associated with the largest values of H_S per location, henceforth referred to as the peak direction, as illustrated in Figure 2.

[Figure 2 about here.]

The peak direction associated with the largest value of H_S per location is shown spatially in Figure 3. In most regions, there is local consistency between values of peak direction. However, it is interesting that peak direction changes markedly in some regions, particularly around Orkney and Shetland. Further, there appear to be approximately linear boundaries between regions of consistent peak direction, for example, running north-east from Aberdeen, running south-east from Aberdeen, and running south-east from Shetland. We assume that these boundaries are the result of most severe storm events with different directional characteristics, at least in part due to the influence of nearby land masses. In the northern North Sea, peak direction corresponds to waves from the Atlantic, whereas in the central North Sea, peak direction corresponds to waves from the North.

[Figure 3 about here.]

To eliminate serial correlation, block maxima for contiguous 48 hour intervals of the reduced data are then isolated. Block maxima exceeding the median block maximum are retained and used for all subsequent analysis. The selection of observations by direction and season reduces the number of (block maxima) available for analysis. The proportion of block maxima (for all directions and seasons) over median threshold (of the reduced data only) retained for extreme value analysis per location is shown in Figure 4, and is seen to vary from approximately 20% to 60%. An interval of 48 hours for blocking is considered reasonable since this corresponds to the period of longer North Sea storms.

[Figure 4 about here.]

We explored the characteristics of the four parameter model as a function of blocking period (examining periods of 24, 48 and 72 hours), and found that estimates were reasonably stable. We judged that adoption of a median threshold (per location) for rejection of smaller values of block maxima was appropriate by examining the stability of the four parameter model for other thresholds (including 60%ile, 70%ile, 80%ile and 90%ile). Although there was general consistency (for example in the estimate for extreme quantiles) using different thresholds, sample size limitations became particularly problematic at locations for which directional-seasonal sub-sampling results in elimination of a large part of the original data.

3. Method

We follow the modelling procedure described by Wadsworth et al. (2010). We assume that independent threshold exceedances follow a Poisson process model (see, e.g. Coles (2001)) on a Box-Cox transformed (Y) measurement scale with Box-Cox parameter λ, rather than the original (X) scale. For given measurement scale parameter λ, the sample likelihood is written in terms of transformed exceedances $\{y_i\}_{i=1}^m$ of threshold u_Y as:

$$\exp\left\{-m(1-\frac{\xi_Y}{\sigma_Y}(u_Y-\mu_Y))^{-1/\xi_Y}\right\}\prod_{i=1}^{m}\frac{1}{\sigma_Y}\left(1-\frac{\xi_Y}{\sigma_Y}(y_i-\mu_Y)\right)^{-1/\xi_Y-1},$$

where μ_Y, σ_Y and ξ_Y are Poisson process location, scale and shape parameters respectively on the transformed (Y) scale. Note that the factor m in the first (exponential) term of the likelihood is arbitrary. As explained in Wadsworth et al. (2010), a factor equal to m is used to preserve the close correspondence between the form of the likelihood and that of the corresponding generalised extreme value model for a sample of m independent observations. (A factor equal to m is also found to improve the mixing in the Markov chain Monte Carlo (MCMC) inference (see, e.g. Gamerman and Lopes 2006) used for parameter estimation.)

We estimate the parameters of the Poisson process model on the original rather than the transformed scale, in terms of location, scale and shape parameters μ_X, σ_X, ξ_X respectively and λ, since model parameters μ_Y, σ_Y, ξ_Y and λ on the transformed scale exhibit strong dependence. To achieve this, we need to relate quantities on the Y scale with their counterparts on the X scale. The transformed sample is related to the original by the Box-Cox transformation with parameter λ:

$$y_i = \frac{x_i^\lambda - 1}{\lambda}$$

for $i = 1, 2, ..., m$.

Threshold u_Y on the transformed scale is therefore related to threshold u_X on the original scale by $u_Y = \frac{u_X^\lambda - 1}{\lambda}$. Using the theory of penultimate approximation (Smith 1987, outlined in Appendix 2), Wadsworth et al. (2010) show that the scale parameter on the transformed scale, σ_Y, is related to corresponding parameter on the original scale, σ_X, by:

$$\sigma_Y = \sigma_X \mu_X^{\lambda - 1}.$$

Similarly, the location parameter on the transformed scale, μ_Y, is related to location on the original scale, μ_X, by:

$$\mu_Y = \frac{\mu_X^\lambda - 1}{\lambda}.$$

The relationship between the shape parameter on the transformed scale, ξ_Y, and its equivalent on the original scale, ξ_X, is somewhat more complicated. Its form follows:

$$\xi_Y = \xi_X + c(\lambda - 1)$$

for some unknown constant c. Following Wadsworth et al. (2010) we choose to estimate c up-front using the profile likelihood of ξ_Y as a function of λ, as illustrated in Appendix 2.

Thus, on the original X scale, the transformed likelihood of exceedances $\{x_i\}_{i=1}^m$ of threshold u_X is:

$$L(\mu_X, \sigma_X, \xi_X, \lambda) = \exp\left\{-m(1 - \frac{\xi_Y}{\sigma_Y}(u_Y - \mu_Y))^{-1/\xi_Y}\right\} \prod_{i=1}^{m} \frac{x_i^{\lambda-1}}{\sigma_Y} \left(1 - \frac{\xi_Y}{\sigma_Y}(\frac{x_i^\lambda - 1}{\lambda} - \mu_Y)\right)^{-1/\xi_Y - 1},$$

the factor $x_i^{\lambda-1}$ inside the product arising from the Jacobian of the Box-Cox transformation. The posterior density for μ_X, σ_X, ξ_X and λ is given by:

$$f(\mu_X, \sigma_X, \xi_X, \lambda | \{x_i\}_{i=1}^n) = KL(\mu_X, \sigma_X, \xi_X, \lambda) f(\mu_X) f(\sigma_X) f(\xi_X) f(\lambda)$$

with constant of integration, K. We generate a random sample from the posterior using MCMC inference, assuming vague Gaussian priors for μ_X, $\log_e(\sigma_X)$ and ξ_X, and a uniform prior for λ. We use Gaussian random walk Metropolis-Hastings acceptance sampling to update the four parameters μ_X, σ_X, ξ_X and λ in turn. As is conventional in MCMC inference, we incorporate a simulation burn-in period and subsequent sequence thinning. We also found it useful to vary the standard deviation of the Gaussian proposal distribution automatically during MCMC simulation to maintain the acceptance rate of the Metropolis-Hastings step at values between 0.25 and 0.50. To preserve positivity of σ_X, we also choose to model in terms of $\log_e(\sigma_X)$ rather than σ_X.

Estimates for posterior medians and 95% credible intervals for μ_X, σ_X, ξ_X and λ for a reference central North Sea location are given in Figure 5, as a function of threshold for extreme value modelling. We observe that parameter estimates are relatively constant with threshold, implying reasonable model fit, until we exceed the 80% threshold. We conclude that, for this location at least, any threshold between the median and 80% quantile would be reasonable. We select the median threshold to retain as much data as possible consistent with the Poisson process model.

[Figure 5 about here.]

Figure 6 shows the posterior median and 95% credible interval for the Q_{100} extreme quantile (corresponding to a return period of 100 times the period of the hindcast). We calculate the return value Q_T corresponding to a return period of T times the period of the original hindcast, using the standard expression for return value on the transformed scale, applying the inverse Box-Cox transformation to the result:

$$1 - \frac{1}{mT} = \exp\left\{ - \left(1 + \xi_Y \frac{\frac{Q_T^\lambda - 1}{\lambda} - \mu_Y}{\sigma_Y} \right)^{-1/\xi_Y} \right\}.$$

For the reference central North Sea location, the Q_{100} quantile estimate appears stable until considerably beyond the 80% threshold.

[Figure 6 about here.]

4. Application

We apply the four parameter Poisson process model to block maxima from directionally and seasonally sub-sampled data which exceed the corresponding median value per location, independently for each of 816 grid locations in the North Sea. To illustrate the general characteristics of parameter estimates, Figure 7 gives posterior distributions for the four parameter model for reference locations in the northern, central and southern North Sea, together with three parameter model maximum likelihood point estimates for comparison. We observe that the posterior distributions for location parameter μ_X are approximately symmetric at each location, with modes corresponding to the three parameter model point estimates; the location parameter in the southern North Sea is seen to be smaller than in the central North Sea, which in turn is smaller than in the northern North Sea. The same trend is seen in the estimates of scale parameter σ_X. Interestingly, the posterior mode for scale is lower than the three parameter point estimate at each location. The posterior modes for shape parameter ξ_X are similar to the three parameter point estimate for the central and northern North Sea, but lower for the southern North Sea. Finally, the posterior distribution of measurement scale parameter λ is similar for all locations, with mode considerably below unity (corresponding to the original scale), with relatively little mass above unity.

[Figure 7 about here.]

The corresponding posterior distributions for return values $Q_{0.1}, Q_1, Q_{10}$ and Q_{100} (corresponding to return periods of $0.1, 1, 10$ and 100 times the period of the hindcast) are shown in Figure 8 as dashed, solid, dotted and dot-dashed lines respectively. The modes of the posteriors in each case are lower than, but still generally consistent with, the corresponding three parameter point estimates.

[Figure 8 about here.]

In applying the four parameter model at all locations in turn, we find that for some locations, estimation is particularly problematic. We assess lack of fit by inspecting the median likelihood value for the model fit per location from the MCMC simulation. Large likelihoods suggest generally good model fit. As seen in Figure 9, fit is particularly poor, for example, at two locations south-east of Shetland, and four locations south-east of Aberdeen. On further investigation, it was found that at these and a few other locations with low median likelihoods, storm direction fluctuates

considerably (see, e.g. Figure 4). As a consequence, it is possible that directional sub-sampling would not perform reasonably; it was decided to eliminate a total of 11 locations from subsequent analysis.

[Figure 9 about here.]

Posterior median parameter estimates for the four parameter Poisson model at the remaining 805 locations are summarised on the left hand side in Figures 10-12, and compared with point estimates (on the right hand side) from the three parameter maximum likelihood estimation. There is excellent agreement between estimates for location μ_X; there is a trend of decreasing μ_X with decreasing latitude. There is also good agreement between estimates for scale σ_X; larger values are observed to the North and particularly to the West of Shetland. There is a band of larger σ_X values extending down the North Sea, somewhat nearer to Norway (and then Denmark) than to the U.K.

[Figure 10 about here.]

[Figure 11 about here.]

Agreement between posterior median values from the four parameter Poisson process model and the maximum likelihood estimates from the three parameter model are also good for shape ξ_X. Large values of shape parameter are observed between Shetland and Norway as might be expected for both models, since these locations are exposed to long fetches. It is curious however that large values of shape are observed off the English coast in the region of Newcastle-upon-Tyne for both models (but note that the corresponding value of scale, σ_X, is small at these locations).

[Figure 12 about here.]

Figure 13 illustrates the values of posterior median measurement scale parameter λ per location, for the four parameter model. Larger, darker circles correspond to larger values of λ. There are clear trends in λ with location. In the northern North Sea, north-east of Shetland, values between 0.3 and 0.9 are observed away from coastal regions. Larger values are observed north-west of Shetland, and south of Shetland. Particularly small values are observed in the region of 55^oN, 1^oE. Larger values

are observed off the coast of Denmark. There are peculiarly large values in the English Channel near Dover; these correspond to locations with different peak storm direction characteristics to their neighbours (c.f. Figure 3).

Maps showing the spatial variation of extreme quantile Q_{100} are given in Figure 14, from maximum likelihood estimation using the three parameter model on the left hand side, and as the median value from the MCMC trace for the four parameter on the right. The two figures are generally consistent; larger values of Q_{100} are observed in the northern North Sea, and reduce with reducing latitude. In the left hand figure, larger values of Q_{100} extend somewhat further down the North Sea. Posterior median return values differ somewhat from the corresponding three parameter maximum likelihood estimates at each location. To further aid comparison, Figure 15 maps the ratio of the median four parameter Q_{100} return levels to the corresponding return levels from the three parameter point process model parameter estimates per location. The ratio of return values has a median value of 0.92 across all locations and a 90% uncertainty band of approximately $(0.72, 1.37)$. This illustrates the extra variability in return values that measurement scale uncertainty introduces.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

5. Discussion

The four parameter Poisson point process model introduced by Wadsworth et al. (2010) provides a means to reduce bias in extreme value modelling without knowledge of the parent distribution (from which the sample is drawn) by exploiting transformation of measurement scale, without the requirement to specify a particular scale transformation. For a given application, the joint posterior distribution of the measurement scale parameter λ, together with the Poisson point process location μ_X, scale σ_X and shape ξ_X parameters is estimated. Note that, although parameters μ_X, σ_X and ξ_X are referred to as being on the original scale, they might more strictly be thought of as part of a re-parameterisation. In particular, the estimate of ξ_X will be affected by the estimate of c, the slope in the profile likelihood (see Appendix 2).

Figure 13 shows that the measurement scale parameter λ varies considerably spatially, yet Figures 14-15 suggest that the impact of this variation on return values is perhaps smaller than might be expected. As illustrated by Cook (1982), measurement scale influences rate of convergence to asymptotic form. Lack of convergence may result in bias, in the form of under- or over- estimation of return values, depending on the characteristics of the parent distribution. As illustrated in Figure 5, estimates for measurement scale λ and other model parameters for a reference location in the central North Sea are relatively insensitive to threshold choice, for thresholds below the 80%ile. Above this level, estimates for λ in particular increase and become more uncertain due to sample size reduction. (Interestingly, once convergence is achieved approximately for some set of values of λ, the actual value of λ estimated (from within that set) is arbitrary, since any of the values are then appropriate; hence increasing variability of λ at convergence is expected!) The corresponding return value estimate Q_{100} is stable up to a 90%ile threshold. Above this level, estimates of Q_{100} increase and become more uncertain. These trends are typical of the North Sea data examined. Threshold selection is of considerable concern to realistic modelling. Recent work by MacDonald et al. (2011) and Wadsworth and Tawn (2012) provide approaches to overcome difficulties in threshold selection, but these methods are applicable to homogeneous samples only. Stability of shape parameter or of mean excess (e.g. Coles 2001) as a function of threshold is critical to reliable modelling. Indications of lack of threshold stability suggest poor model fit. For example, upward trends in these diagnostics with threshold can be indicative of underlying Normal (Fisher and Tippett 1928) or Weibull (Cook and Harris (2004)) distributions respectively, even though the shape parameter estimate remains negative.

Figures 13 and 15 are visually very similar, suggesting that locations where scale differs from unity also yield Q_{100} estimates that differ most from the 3-parameter model analyses. Inspection of Figure 9 shows that median likelihoods, indicating quality of fit, are lower for locations off Denmark and north of Aberdeen, suggesting relatively poor model fit. Referring to Figure 13, these regions also correspond to larger posterior median estimates for Box-Cox parameter λ, perhaps suggesting that estimates for λ are less reliable for these cases. This observation is further supported by the trend in estimated λ with quantile threshold level shown in Figure 5; as quantile threshold increases, median λ also increases. We propose that, at locations of good model fit (see Figure 9), the four parameter model offers realistic estimates of measurement scale uncertainty and improves estimates of return values. For example, off the Northumberland coast, model fit is good, estimated λ is smaller than unity and 100-year return levels from the four parameter model are considerably smaller than those

provided by the three parameter model (see Figures 14 and 15).

For the current hindcast data, the median value (over all locations) of the median posterior λ (per location) is approximately 0.7, with a 90% uncertainty band of approximately $(0.2, 2.2)$, suggesting that accommodating spatial variation is important. Estimation of the corresponding statistics for hindcasts of other ocean basins, and for measured data, would provide an interesting comparison and additional experience with the approach for practical applications.

In this work, since our aim is to investigate the effect of measurement scale, we avoid the need to consider the effect of covariates (such as storm direction and season) on the extreme value analysis, by judiciously selecting data corresponding to a narrow directional sector in winter months. This ensures the sub-sample used is approximately homogeneous. Estimates of λ show some sensitivity to the size of the directional sector used. A better approach, certainly for serious application of the method to structural design, would be to parameterise the extreme value model in terms of covariates, thereby avoiding the need to sub-sample, following for example Coles and Walshaw (1994), Cook et al. (2003) or Jonathan and Ewans (2007).

We use the median posterior values of parameter and quantiles from the four parameter model to compare with maximum likelihood estimates from the three parameter model. In Wadsworth et al. (2010), the posterior predictive return value is used in place of the median posterior quantile; for the current application, comparisons (not shown) confirm that these two quantities yield very similar estimates.

We choose to model individual locations independently for simplicity. In reality, the marginal characteristics of neighbouring locations will be similar. This suggests that spatially smooth representations for the parameters of both three and four parameter models should be considered. Often, joint return values for two or more environmental variables (such as H_S and the peak wave period T_P) are sought. Extension of the current approach to joint modelling is an area for future investigation.

In this work, we quantify the uncertainty in estimates of extreme quantiles for ocean design due to arbitrary choice of measurement scale. We find that 100-year return value estimates can vary by as much as $\pm 30\%$, specifically that the ratio of median return values per location from the four parameter (incorporating measurement scale uncertainty) and three parameter models has a 90% uncertainty band of approximately (0.72, 1.37). We also find that the median value of this

ratio is approximately 0.92, suggesting somewhat lower return values using the four parameter model in this application. The four parameter model is more sensitive to small sample size, due to its extra flexibility, and should therefore be used with care when data are limited. The current implementation of the four parameter model also relies on directional data pre-processing, itself sensitive to sample size; an improved approach would incorporate directional and other covariates explicitly within the extreme value model.

Acknowledgements

We acknowledge numerous useful discussions with Jonathan Tawn and Jenny Wadsworth at Lancaster and Yanyun Wu at Shell. We are grateful for financial support from the U.K. Engineering and Physical Sciences Research Council and Shell.

Appendix 1: Motivating scale transformation

The purpose of this appendix is to demonstrate that in particular circumstances, judicious choice of measurement scale improves the quality of extreme value inference.

Weibull maxima

Suppose that we know that the parent distribution of a set of n independent random values $\{X_i\}_{i=1}^n$ is a Weibull distribution with exponent κ:

$$\Pr(X_i \leq x) = 1 - e^{-x^\kappa} \ \forall\, i, \ \ x > 0 \ .$$

Suppose further that we are prepared to transform measurement scale using the power transformation $z = x^s$ $(s > 0)$ and perform extreme value modelling on the set $\{Z_i\}_{i=1}^n$, the distribution function of which is:

$$F(z) = \Pr(Z_i \leq z) = \Pr(X_i^s \leq z) = \Pr(X_i \leq z^{1/s}) = 1 - e^{-z^{(\kappa/s)}} \ \forall\, i \ .$$

We define the n-event return value z_n by:

$$F(z_n) = 1 - \frac{1}{n} \text{ or } n(1 - F(z_n)) = 1.$$

The distribution of the maximum $M = \max_{i=1,\ldots,n}\{Z_i\}$ is therefore:

$$\begin{aligned}
\Pr(M \leq z) &= F^n(z) \\
&= (1 - (1 - F(z)))^n \\
&= \left(1 - \frac{1 - F(z)}{n(1 - F(z_n))}\right)^n \\
&\sim \exp\left\{-\frac{1 - F(z)}{(1 - F(z_n))}\right\}
\end{aligned}$$

for large n, according to the Cauchy limit $\lim_{n\to\infty}(1 + x/n)^n = e^x$. Substituting the form of F into the above yields:

$$\Pr(M \leq z) = e^{-e^{-y}}$$

where:

$$y = z^{\kappa/s} - z_n^{\kappa/s}.$$

Following Cramer (1946) and Cook and Harris (2004), Taylor expansion of $z^{\kappa/s}$ about $z_n^{\kappa/s}$ gives:

$$y = \frac{\kappa}{s} z_n^{(\frac{\kappa}{s}-1)}(z - z_n)(1 + \frac{1}{2}(\frac{\kappa}{s} - 1)\frac{(z - z_n)}{z_n} + \frac{1}{6}(\frac{\kappa}{s} - 1)(\frac{\kappa}{s} - 2)\frac{(z - z_n)^2}{z_n} + ...) \ .$$

The leading order term of y is linear in z, indicating that the appropriate asymptotic limiting form for the distribution of M is the Gumbel distribution (i.e. Fisher-Tippett Type 1). Moreover the rate of convergence of the distribution of the maximum to the Gumbel is determined by the leading error terms.

When the error terms are large, the difference between the distribution of the maximum M and the limiting Gumbel form will be large, and inference based on Gumbel fitting will be suspect. Best inferences from Gumbel fitting will be possible when the error terms are small. In the case $s = \kappa$, all error terms are zero identically (even for non-integer κ), suggesting that convergence will be rapid

(with rate determined by the rate of convergence of the Cauchy approximation above only). This observation strongly suggests transforming the scale of data (from X to $Z = X^{\kappa}$) to improve rate of convergence, thereby reducing bias in parameter estimation using the (correct) limiting Gumbel distribution.

For example, when the parent distribution is Rayleigh ($\kappa = 2$), bias can be reduced by Gumbel fitting squared data ($s = \kappa = 2$). Rayleigh (or near-Rayleigh) behaviour is typical of many environmental phenomena, such as wind speed and individual ocean wave heights.

In practice, however, we are rarely confident of the true value of κ, nor indeed of the form of the parent distribution (nor its asymptotic limiting form). For example, Simiu et al. (2001) cast doubt on the appropriateness of the Gumbel distribution for modelling squared wind speed, since the parent distribution of wind speed is not generally Rayleigh.

Note that, were it possible to achieve approximate convergence on both scales, consistent estimates for return value would be obtained. From the definition of z_n above and the known Weibull parent distribution:

$$z_n = F^{-1}(1 - \frac{1}{n}) = (\log n)^{s/\kappa}$$

from which:

$$x_n = z_n^{1/s} = (\log n)^{1/\kappa}$$

independent of measurement scale s. In reality, rate of convergence would vary as a function of s; we would therefore base our return value estimation on the scale s with best convergence characteristics.

Generalised Pareto exceedances

More generally, suppose we have a sample of n independent threshold exceedances $\{X_i\}_{i=1}^{n}$ which (exactly) follow the generalised Pareto distribution:

$$\Pr(X_i > x | X_i > u) = \left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-1/\xi} \quad \forall\, i$$

from which the unconditional distribution function is:

$$\begin{aligned}\Pr(X_i < x) &= 1 - \Pr(X_i > x | X_i > u) \Pr(X_i > u) \\ &= 1 - p(1 + \frac{\xi}{\sigma}(x-u))^{-1/\xi} \; \forall \, i, \;\; x > u,\end{aligned}$$

for some threshold u, where $\Pr(X_i > u) = p$. The n-event return value x_n, defined as before using:

$$\Pr(X_i < x) = 1 - 1/n$$

is given by:

$$x_n = \frac{\sigma}{\xi}\left((np)^{\xi} - 1\right) + u.$$

Suppose that we choose (misguidedly) to model the sample on a power-transformed scale $z = x^s$. For the transformed sample:

$$\begin{aligned}\Pr(Z_i \leq z) = \Pr(X_i^s \leq z) &= \Pr(X_i \leq z^{1/s}) \\ &= 1 - p\left(1 + \frac{\xi}{\sigma}(z^{1/s} - u)\right)^{-1/\xi} \; \forall \, i \; .\end{aligned}$$

This distribution is clearly not generalised Pareto. Therefore, fitting threshold exceedances of the transformed sample using the generalised Pareto distribution would yield biased estimates. We can perform a Taylor expansion of $z^{1/s}$ about u^s to find a generalised Pareto approximation to the distribution, however. Since:

$$z^{1/s} = u + \frac{1}{s}u^{1-s}(z - u^s) + \frac{1}{2s^2}(1-s)(z-u^s)^2 + O((z-u^s)^3)$$

we obtain:

$$\Pr(Z_i \leq z) = 1 - p\left[1 + \frac{\xi}{\sigma s u^{s-1}}(z-u^s)\left\{1 + \frac{1}{2}(1 - \frac{1}{s})u^{s-1}(z-u^s) + O((z-u^s)^2)\right\}\right]^{-1/\xi} \; \forall \, i.$$

For sufficiently large u, we can neglect all but the leading term, showing asymptotically that the transformed data are still generalised Pareto distributed, with the same shape ξ but different scale

$\sigma s u^{s-1}$. The rate of convergence of the threshold exceedances on the transformed scale to the generalised Pareto limit is determined by the second term in general. We note that when $s = 1$ this term is zero (since in this case we would return to the original scale), and rate of convergence is optimal. Otherwise, as s increases beyond 1, the rate of convergence decreases, and is determined by the size of u^{s-1}. As s decreases below 1 to 0, the rate of convergence decreases due to the $(1 - 1/s)$ factor.

Using the leading term above only, we can derive the analogous expression for the n-event return value z_n:

$$z_n = \frac{\sigma s u^{s-1}}{\xi} \left\{ (np)^{\xi} - 1 \right\} + u^s.$$

To see the correspondence between return values estimated on the original and transformed scales, we note that:

$$x_n^s = u^s (1 + \frac{\sigma}{\xi u} \left\{ (np)^{\xi} - 1 \right\})^s.$$

Binomial expansion yields:

$$x_n^s = u^s \left[1 + \frac{\sigma s}{\xi u} \left\{ (np)^{\xi} - 1 \right\} + \frac{s(s-1)}{2} \left\{ \frac{\sigma}{\xi u} \left((np)^{\xi} - 1 \right) \right\}^2 + O(u^{-3}) \right].$$

The leading two terms correspond to the expression for z_n above, demonstrating that for sufficiently large u, $x_n \sim z_n^{(1/s)}$. Rate of convergence is determined by the third term above. In the case $s = 1$ we note that the third and subsequent terms are zero.

The current work, motivated by that of Wadsworth et al. (2010), seeks to exploit scale transformation to reduce bias in extreme value modelling without knowledge of the parent distribution or the requirement to specify a particular scale transformation. For a given application, the statistical characteristics of the appropriate scale transformation parameter are estimated; once convergence is effectively achieved on transformed measurement scale, we can estimate return values on that scale, and thereby estimate return values consistently on the original measurement scale using the inverse Box-Cox transformation.

Appendix 2: Penultimate approximations and scale transformation

We motivate the modelling strategy employed in Section 3 using block maxima, a common alternative to modelling exceedances, following Wadsworth et al. (2010).

Suppose we have a set of n independent, identically distributed random variables $\{X_i\}_{i=1}^n$ (in the domain of attraction of a *max-stable* random variable) with distribution function F_X, density f_X and (possibly infinite) upper end point x^F, whose maximum is $M_{X,n}$. Since the distribution is max-stable, we know there exists a sequence of normalising constants $a_{X,n}(> 0)$ and $b_{X,n}$ such that:

$$\Pr(\frac{M_{X,n} - b_{X,n}}{a_{X,n}} \leq x) \to G_X(x) \text{ as } n \to \infty$$

where $G_X(x) = \exp\{-(1 + \xi_X x)^{-1/\xi_X}\}$, denoted $GEV(0, 1, \xi_X)$, where convergence is in distribution.

Smith (1987) showed that suitable normalising constants a_X, b_X (henceforth we omit subscript n in notation) could be derived from the reciprocal hazard function:

$$h_X(x) = \frac{1 - F_X(x)}{f_X(x)}$$

of the form:

$$b_X = F_X(1 - \frac{1}{n}), \quad a_X = h_X(b_X).$$

Moreover, the value of ξ_X is given in terms of the first derivative of the reciprocal hazard function, h'_X:

$$\xi_X = \lim_{x \to x^F} h'_X \ .$$

Smith (1987) further showed that:

$$F_X(a_X x + b_X)^n = \exp\{-(1 + h'_X(z)x)^{-1/h'_X(z)}\}$$

for some *unknown* value $z = b_X + \epsilon_X, \epsilon_X \in [\min\{a_X x, 0\}, \max\{a_X x, 0\}]$. This suggests the approximation:

$$\xi_X^* = h'_X(b_X)$$

as a so-called *penultimate approximation* for ξ_X. Notably, when estimating the shape parameter assuming normalising constants a_X, b_X, the estimated value of the shape parameter will correspond to the value of the derivative of the reciprocal hazard function *near, but not at* b_X.

On the transformed scale:

$$Y = \frac{X^\lambda - 1}{\lambda} \text{ for } \lambda > 0$$

we can easily relate the value of the sample maximum $M_{Y,n}$ to $M_{X,n}$ using:

$$M_Y = \frac{M_X^\lambda - 1}{\lambda} \ .$$

Some further straight-forward algebra yields similar relationships for the normalising constants:

$$b_Y = \frac{b_X^\lambda - 1}{\lambda}, \text{ and } a_Y = a_X b_X^{\lambda - 1}$$

and for the derivative of the reciprocal hazard function:

$$h'_Y(y(x)) = h'_X(x) + \frac{h_X(x)}{x}(\lambda - 1) \ .$$

The last expression leads (with $x = b_X$) to:

$$\xi_Y^* = \xi_X^* + \frac{a_X}{b_X}(\lambda - 1)$$

relating the penultimate approximations for ξ_X, and (in the limit as $x \to x^F$):

$$\xi_Y = \xi_X + \lim_{x \to x^F} \frac{h_X(x)}{x}(\lambda - 1)$$

relating the limiting shape parameters.

When we estimate the GEV model on original and transformed scales, we substitute the GEV location and scale parameters (μ_X, σ_X on the original scale, and μ_Y, σ_Y on the transformed scale) with b_X, a_X, b_Y and a_Y respectively. With these substitutions, we can relate the GEV parameter on the original and transformed scales to facilitate MCMC inference. However, the corresponding GEV shape parameters γ_X, γ_Y will not correspond to either the penultimate approximations (ξ_X^*, ξ_Y^*) or the limiting forms (ξ_X, ξ_Y). Unfortunately, therefore, the relationship between ξ_X and ξ_Y cannot be obtained analytically. Instead, following Wadsworth et al. (2010), we assume that the relationship between ξ_Y and $\lambda - 1$ will be linear, and estimate the gradient c of the relationship by evaluating the profile likelihood of ξ_Y with respect to λ as a function of λ. For given data, the value of c needs be estimated once-only prior to MCMC simulation. Moreover, at $\lambda = 1$ we have $\xi_Y = \xi_X$. Hence, for MCMC inference, we use:

$$\xi_Y = \xi_X + c(\lambda - 1).$$

Figure 16 illustrates the estimation of c based on the profile likelihood of ξ_Y with respect to ξ.

[Figure 16 about here.]

References

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistics Society. Series B (Methodological)*, 26:211–252.

Coles, S. (2001). *An introduction to statistical modelling of extreme values*. Springer, London.

Coles, S. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Applied Statistics*, 43:139–157.

Cook, N. J. (1982). Towards better estimation of extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 9:295–323.

Cook, N. J. and Harris, R. I. (2004). Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents. *Structural Safety*, 26:391–420.

Cook, N. J., Harris, R. I., and Whiting, R. (2003). Extreme wind speeds in mixed climated revisited. *Journal of Wind Engineering and Industrial Aerodynamics*, 91:403–422.

Cramer, H. (1946). *Mathematical methods in statistics*. Princeton.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, 24:180–190.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Taylor & Francis.

Harris, R. I. (1998). Estimation of long return period design values for wind speeds. *Journal of Engineering Mechanics*, 124:252–259.

Harris, R. I. (2004). Extreme value analysis of epoch maxima–convergence, and choice of asymptote. *Journal of Wind Engineering and Industrial Aerodynamics*, 92:897918.

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

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

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis*, 55:2137–2157.

Simiu, E., Heckert, N. A., Filliben, J. J., and Johnson, S. K. (2001). Extreme wind load estimates based on the gumbel distribution of dynamic pressures: an assessment. *Structural Safety*, 23:221–229.

Smith, R. L. (1987). Approximations in extreme value theory. *University of North Carolina, Department of Statistics, Technical Report No. 205.*

Taylor, P. H. and Goh, G. Y. K. (2000). Extrapolation of long-term storm statistics, the P-O-T method and data transformation. *Proc. 19th International Conference on Offshore Mechanics and Arctic Engineering, New Orleans.*

Teugels, J. L. and Vanroelen, G. (2004). Box-Cox transformations and heavy-tailed distributions. *J. Appl. Probab.*, 41:213–227.

Tromans, P. S. and Vanderschuren, L. (1995). Risk based design conditions in the North Sea: Application of a new method. *Offshore Technology Confernence, Houston (OTC–7683)*.

Wadsworth, J. L. and Tawn, J. A. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B. In press.*

Wadsworth, J. L., Tawn, J. A., and Jonathan, P. (2010). Accounting for choice of measurement scale in extreme value modelling. *Annals of Applied Statistics*, 4:1558–1578.

List of Figures

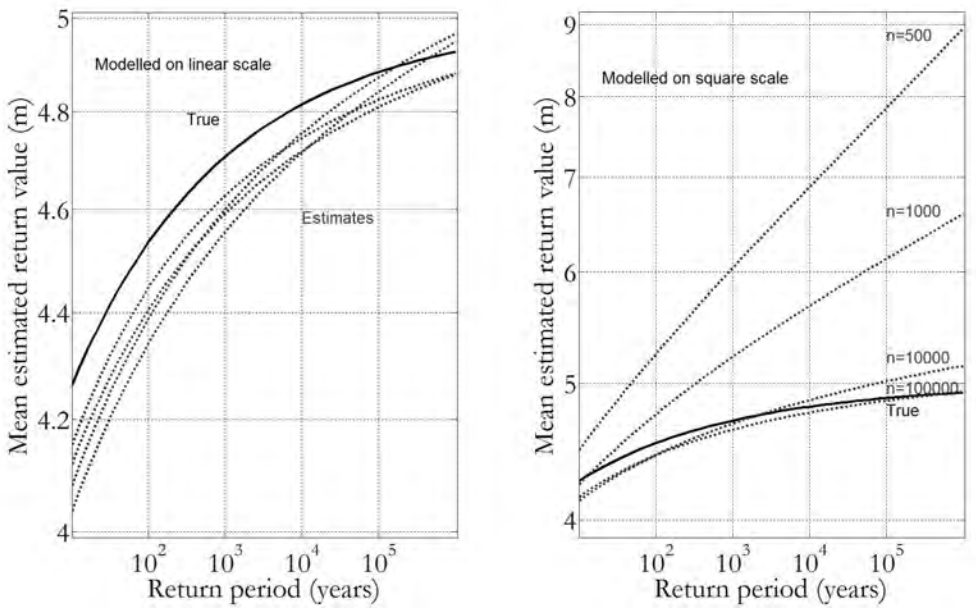

Figure 1: Mean return values estimated using generalised Pareto data (GP, with shape -0.2, scale 1 and threshold 0) simulated on a linear scale, and modelled on linear (left) and square (right) scales, as a function of return period. Solid lines give true return values. Broken lines are return values estimated using GP fitting on the largest 100 individuals in the sample, for different sample sizes, n, of 500, 1000, 10000 and 100000, assuming 1000 occurrences per annum. Return values estimated on the (correct) linear scale are all relatively close to the true curve. Return values estimated on the square scale show systematic bias (due to poor convergence to the asymptotic distributional form) which reduces with increasing sample size.

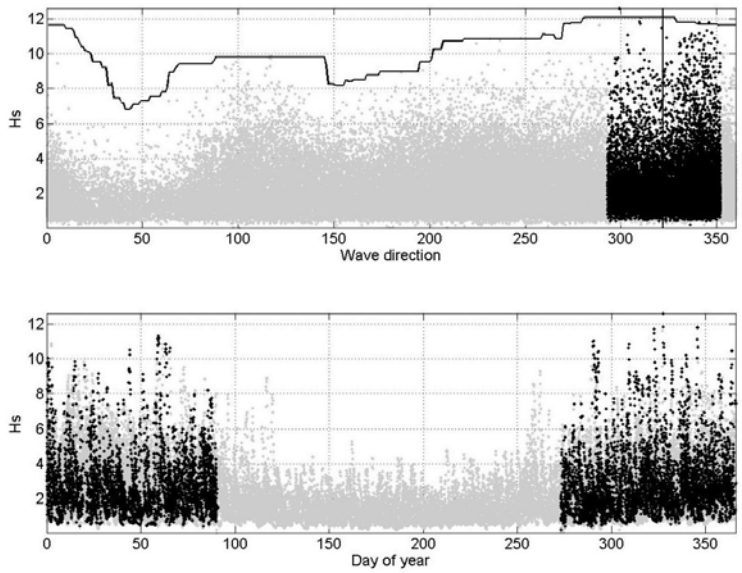

Figure 2: Significant wave height as a function of wave direction (upper) and day of the year (lower) for a reference central North Sea location. The local maximum value of H_S with direction is estimated using a moving median (solid line in upper plot). Observations within $\pm 30^o$ of the direction corresponding to the maximum of the moving median with direction (vertical line, upper plot) which occur during the winter period (1st October to 31st March) are retained for extreme value analysis (and shown in black). Observations not retained are shown in grey.

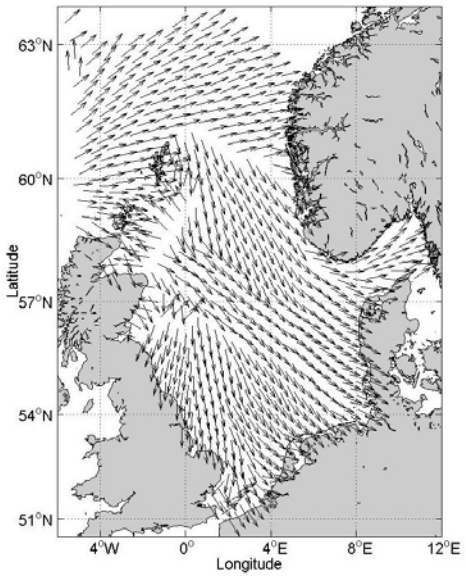

Figure 3: Peak direction per location. Most regions show local consistency of peak direction. Some regions show considerable local variability from location to location. There are also approximately linear boundaries between regions of consistent peak direction.

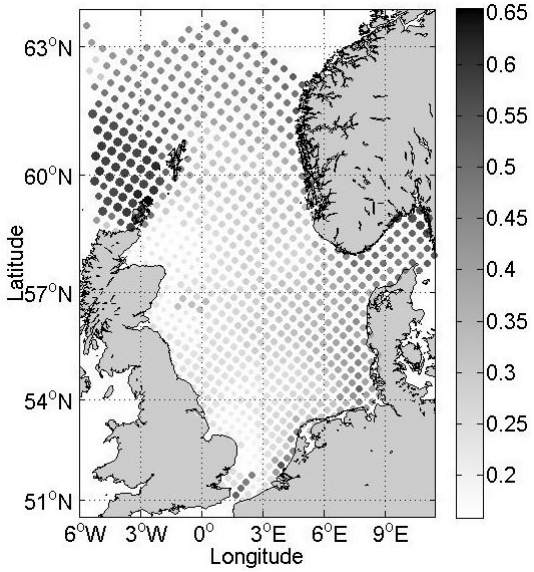

Figure 4: Proportion of block maxima events retained as a result of selection of data from winter months corresponding to directional sector of $\pm 30^o$ with respect to the peak direction per location.

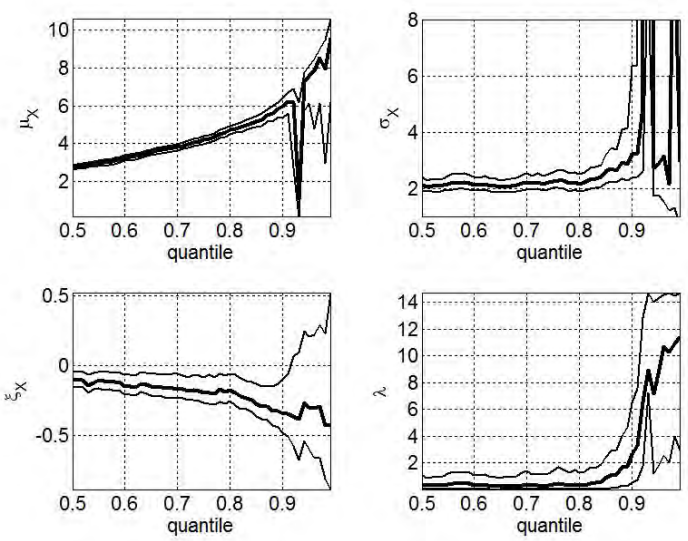

Figure 5: Median parameter estimates with associated 95% credible intervals as a function of threshold for a reference location in the central North Sea. Parameter values remain relatively stable up to an 80% quantile after which parameter instability occurs due to a lack of data, at which point estimates for λ become unstable.

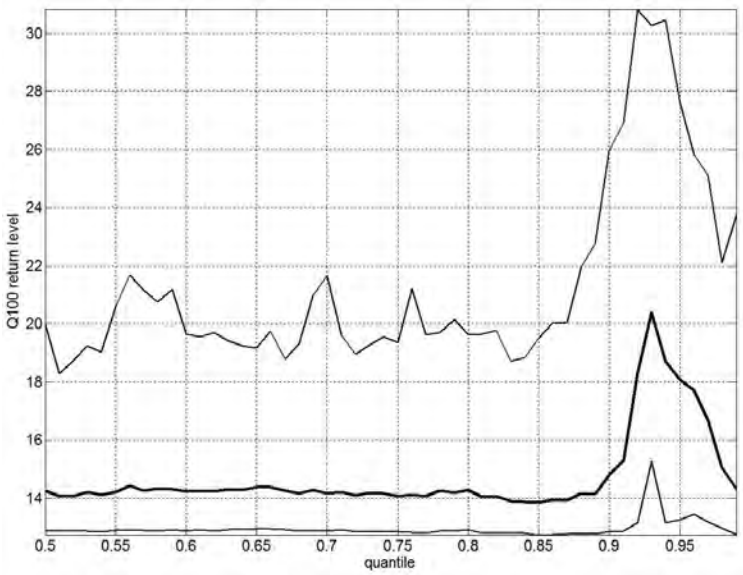

Figure 6: Q_{100} posterior median return values with associated 95% uncertainty bands as a function of threshold for the reference location in the central North Sea. Return value remains stable up to 80% threshold.

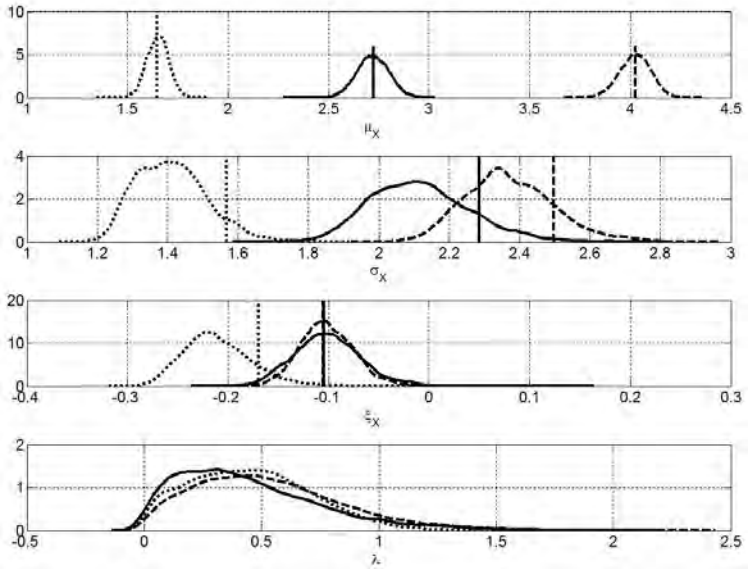

Figure 7: Posterior distributions for four parameter model parameters at reference locations northern North Sea (dashed line), central North Sea (solid line) and southern North Sea (dotted line), with three parameter maximum likelihood point estimates for μ_X, σ_X and ξ_X (vertical lines).

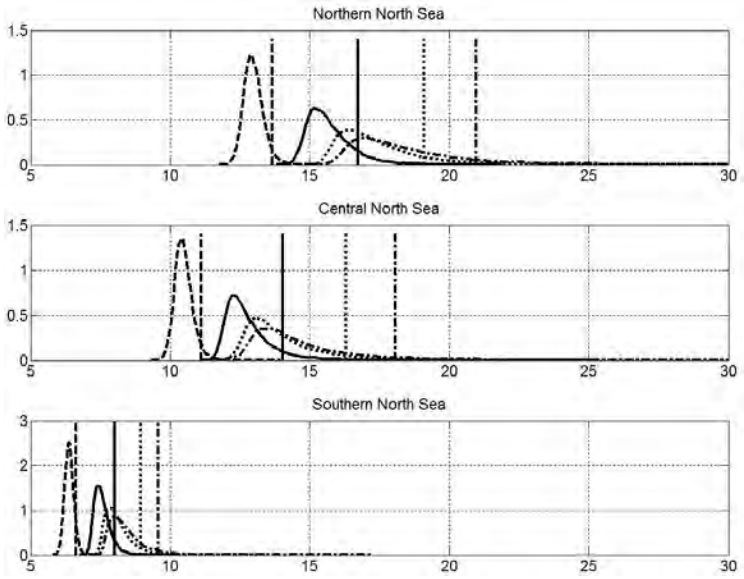

Figure 8: Posterior distributions for $Q_{0.1}$ (dashed line), Q_1 (solid line), Q_{10} (dotted line) and Q_{100} (dot-dashed line) return values for four parameter model at three reference locations. Maximum likelihood point estimates are also shown (vertical lines).

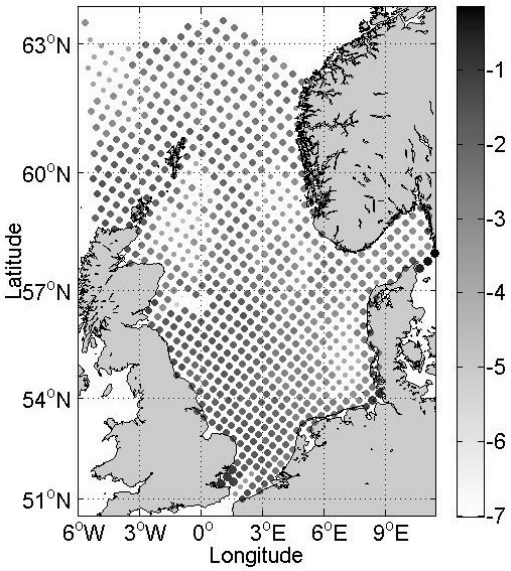

Figure 9: Map of median likelihood per location from the Monte Carlo Markov Chain. Larger and darker points indicate a better fit. Very small likelihoods are taken as an indication of poor fit with the data. Fit is particularly poor, for example, at two locations south-east of Shetland, and four locations south-east of Aberdeen.

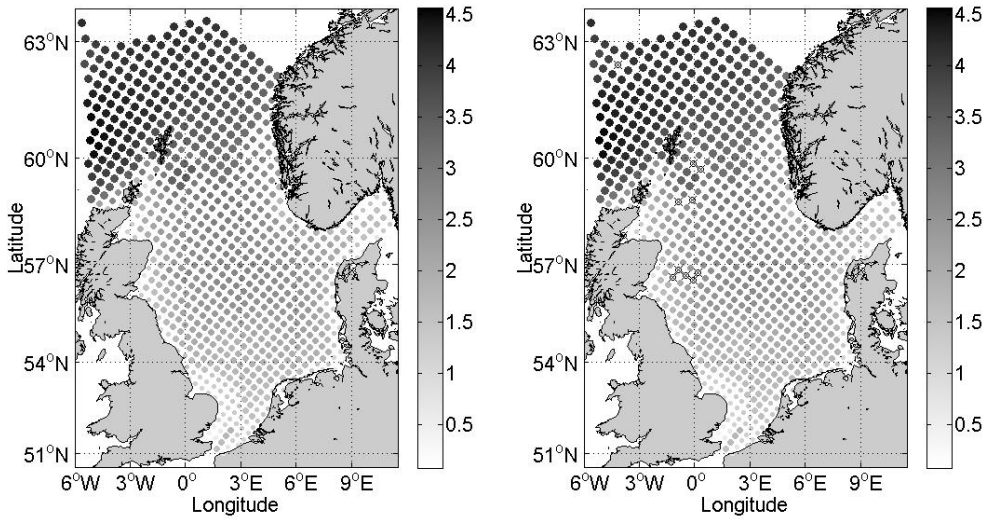

Figure 10: Maps showing parameter estimates for μ_X with three parameter maximum likelihood estimates (left) and four parameter median posteriors from the MCMC (right). Larger and darker points correspond to larger parameter estimates. Locations which have very small likelihoods for the four parameter model are removed and plotted as a circle with a cross.

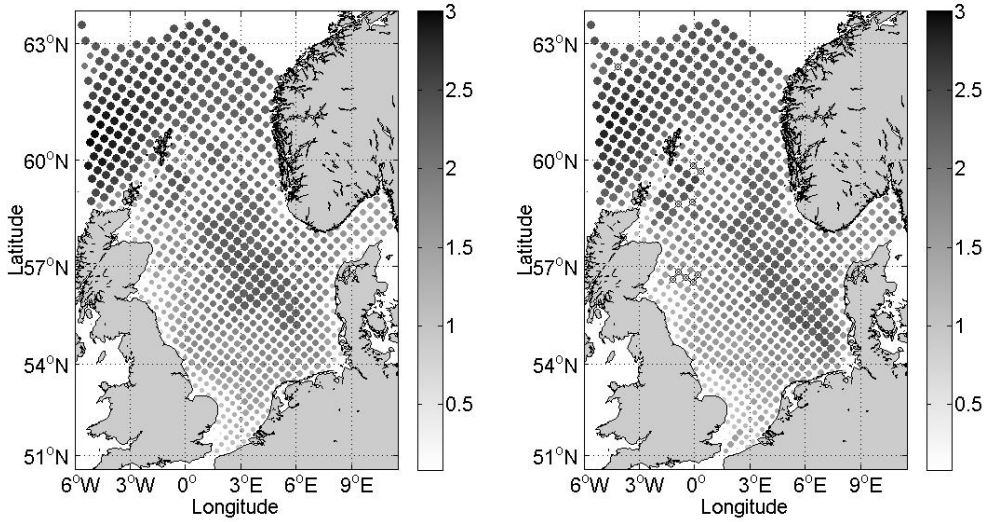

Figure 11: Maps showing parameter estimates for σ_X with three parameter maximum likelihood estimates (left) and four parameter median posteriors from the MCMC (right). Larger and darker points correspond to larger parameter estimates.

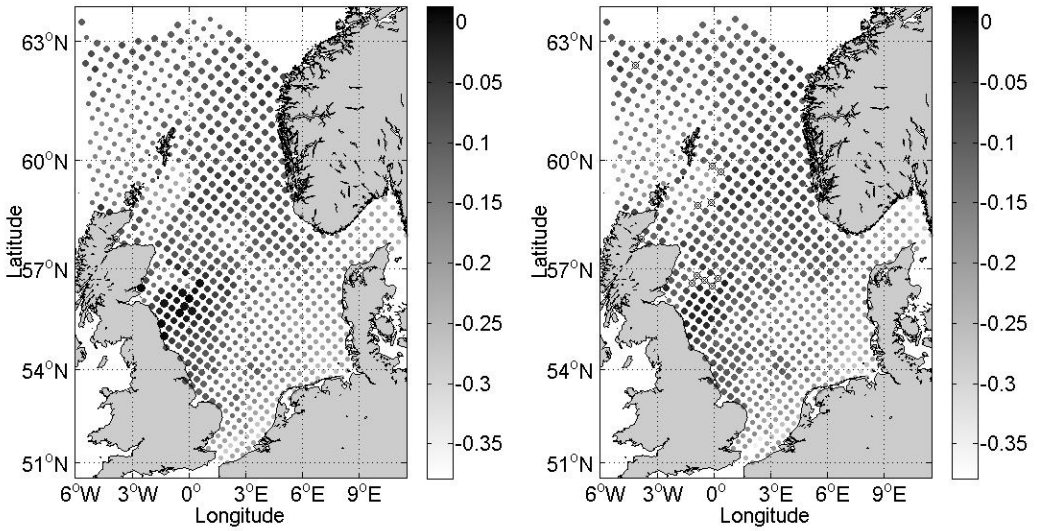

Figure 12: Maps showing parameter estimates for ξ_X with three parameter maximum likelihood estimates (left) and four parameter median posteriors from the MCMC (right). Larger and darker points correspond to larger parameter estimates.

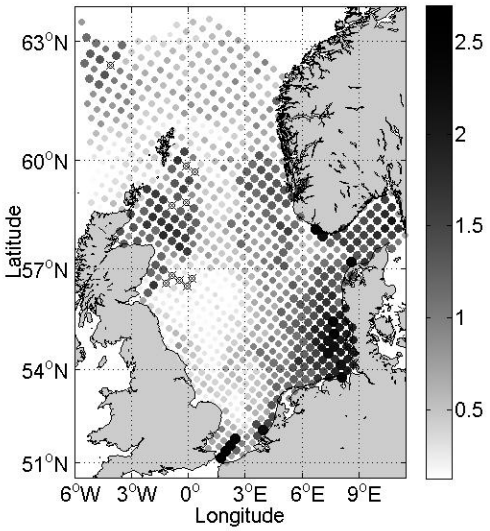

Figure 13: Map illustrating posterior medians for λ from the four parameter model. Larger and darker points correspond to larger parameter estimates.

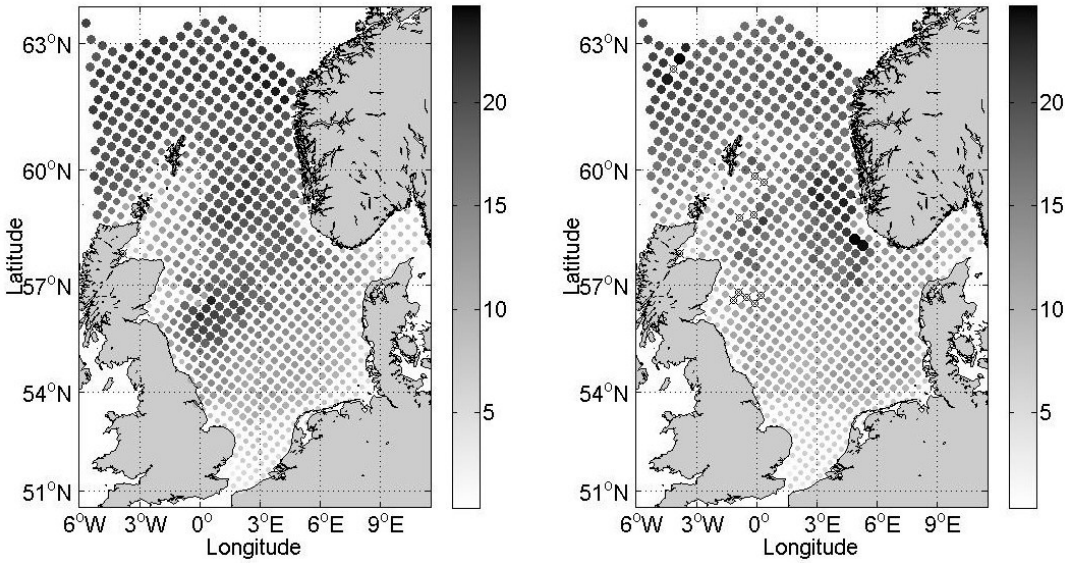

Figure 14: Maps showing Q_{100} return value estimates based on three parameter maximum likelihood estimates (left) and four parameter posterior medians (right). Larger and darker points correspond to larger return values.

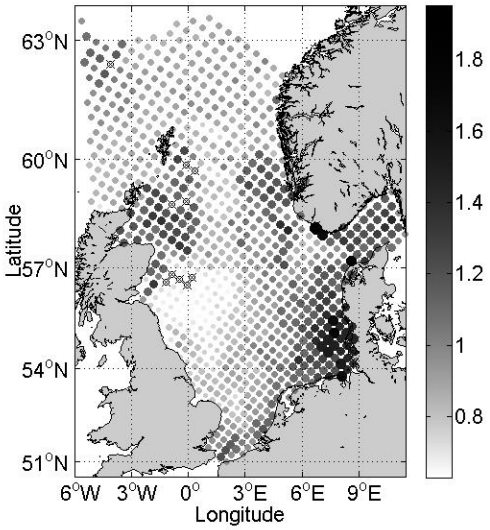

Figure 15: Map showing the ratio of median four parameter Q_{100} to the corresponding three parameter maximum likelihood estimates per location. Larger and darker points correspond to more severe estimates from the four parameter model.

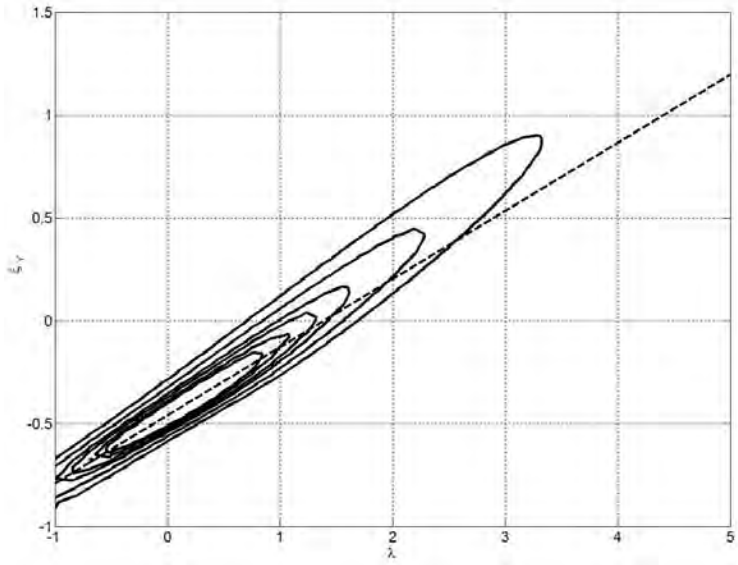

Figure 16: Profile likelihood for ξ_Y against λ based on three parameter model for reference location in the central North Sea and linear fit (dashed line). The gradient of the dashed line is used as an estimate for c, relating the shape parameter ξ_Y on the transformed scale to the measurement scale λ. The value of the dashed line at $\lambda = 1$ is an estimate for ξ_X.