Accommodating measurement scale uncertainty in extreme value analysis of northern North Sea storm severity

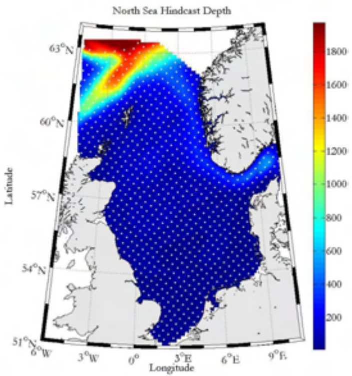
Yanyun Wu, David Randell, Daniel Reeve
Philip Jonathan, Kevin Ewan
yanyun.wu@shell.com

July 2011

Motivation

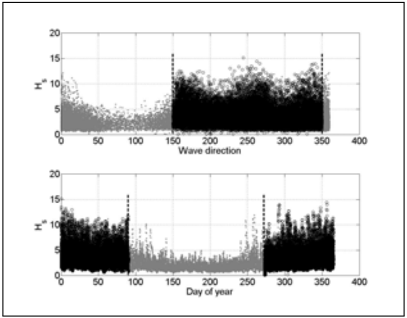
- Modelling **storm severity** is critical to the design and reliable operation of marine structures
- The extreme value analysis of **significant wave height**, denoted H_S is one of many cases where data relating to the same physical process may be measured on more than one scale; e.g. H_S^2 is proportional to the drag force induced by the waves on a structure
- Return value estimates obtained from the square root of an extreme value model fitted to H_S^2 data **differ** from estimates fitted to H_S data, since different tail behaviour is indicated by the two sets of parameter estimates.

Data



- **Significant wave height** (H_S) can be estimated as four times the standard deviation of displacement from mean sea level and is a measure of ocean energy
- Hindcast data:
  - **Location**: northern North Sea, 50 locations, and extended to 800 locations
  - **Period**: 1st October 1964 to 31st March 1995 inclusive
  - **Frequency**: continuously at 3-hour intervals for each location
- Having selected a more homogeneous sample in an attempt to reduce covariate effects, we have 3000 observations per location

Data Omitted



- Significant wave height is a function of covariates such as **wave direction** & **day of the year**
- Certain months and directions were omitted to try to create a more homogenous data set
  - 31st March to 1st October from each year
  - 150 – 350^o (modified for the extended analysis)

Generalised Extreme Value Distribution

- Extreme value analysis uses sample data from rare events and via a process of extrapolation attempts to make rational predictions about the probable outcome of future events (Coles 2001)
- The three parameter **Generalised Extreme Value Distribution** (GEV) has the cumulative distribution function shown, where μ,σ and ξ are the location, shape and scale parameters respectively.

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\};$$

$$\sigma > 0; \mu, \xi \in \mathbb{R}; z_{+} = \max(z, 0)$$

Generalised Extreme Value Distribution

- For a data set divided into block maxima, the GEV is used to model the distribution of the maximum values of each block (e.g. the weekly maxima of 3 hourly H_S values)

$$\mathbb{P}\left(\frac{M_{X,n} - b_{X,n}}{a_{X,n}} \leqslant x\right) \xrightarrow{Dist^n} G_X(x) \;\; as \;\; n \to \infty$$

$$where \;\; M_{X,n} = \max\{X_1, ..., X_n\}\,; \;\; G(x) \sim \text{GEV}(0, 1, \xi_X)$$

- Smith (1987) has shown that

$$b_{X,n} = F_X(1 - 1/n); \;\; a_{X,n} = h_X(b_{X,n}); \;\; \xi_X = \lim_{x \to x^F} h'_X(x)$$

$$where \;\; h(x) = \frac{1 - F_X(x)}{f_X(x)}$$

- With penultimate approximation

$$\xi_{X,n} \approx h'_X(b_{X,n})$$

- Hence approximately

$$M_{X,n} \sim \text{GEV}(a_{X,n}, b_{X,n}, \xi_{X,n})$$

Box-Cox

- Following Wadsworth et al. (2010) we incorporate scale selection into the model through a fourth parameter, λ, using a Box-Cox transformation (Box and Cox, 1964).

$$M_{Y,n} = \frac{M_{X,n}^{\lambda} - 1}{\lambda} \sim \text{GEV}(a_{Y,n}, b_{Y,n}, \xi_{Y,n})$$

$$\textit{where} \quad b_{Y,n} = \frac{b_{X,n}^{\lambda} - 1}{\lambda}$$

$$a_{Y,n} = a_{X,n} b_{X,n}^{\lambda-1}$$

$$\xi_{Y,n} = \xi_{X,n} + \frac{a_{X,n}}{b_{X,n}}(\lambda - 1)$$

NHPP

- We model threshold exceedences using a Non Homogeneous Poisson Process (NHPP) with intensity

$$\frac{1}{\sigma_y}\left[1+\xi_y\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^{-\frac{1}{\xi_y}-1}$$

- The likelihood can be calculated on both X scale and Y scale. On Y (transformed) scale:

$$L_{ppy}=K_y exp\left\{-\left[1+\frac{\xi_y(u_y-\mu_y)}{\sigma_y}\right]^{-\frac{1}{\xi_y}}\right\}\prod\frac{x_i^{\lambda-1}}{\xi_y}\left[1+\frac{\xi_y(y_i-\mu_y)}{\sigma_y}\right]$$

where $K_y = Constant$; $u_y=\frac{u_x^{\lambda}-1}{\lambda}$; $y_i=\frac{x_i^{\lambda}-1}{\lambda}$
With the relationship:

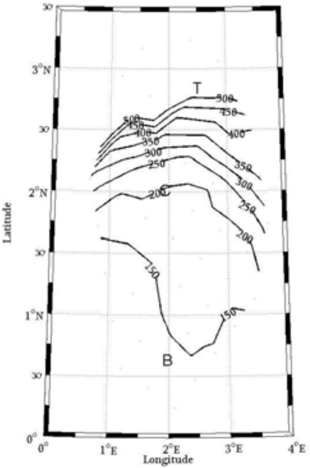
$$\mu_Y=\frac{\mu_x^{\lambda}-1}{\lambda};\;\; \sigma_y=\sigma_x\mu_x^{\lambda-1};\;\; \xi_y=\xi_x+c(\lambda-1)$$

(where c is estimated empirically)

MCMC

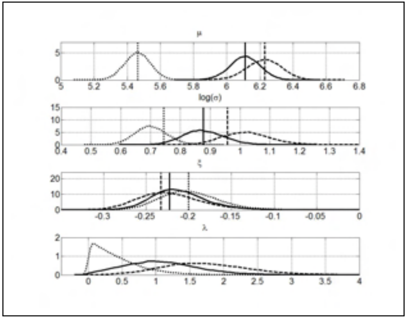
- **Prior**:
  - $\lambda \sim$ Uniform
  - $\mu_x \sim$ Normal
  - $log \sigma_x \sim$ Normal
  - $\xi_x \sim$ Normal
- **Likelihood**: on the Y scale
- Analysis is carried out using MCMC

Northern North Sea Data



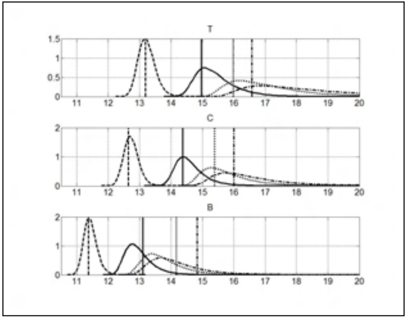
- The data are fitted using both the 3-Parameter (without scale parameter) and 4-Parameter (with scale parameter) point process models
- Return level are denoted $Q_{0.1}$, Q_1, Q_{10} and Q_{100} respectively, where the Q_i event corresponds to an event with return period equal to i times that of the data sample.

Posterior Distribution of 4 Parameter Model



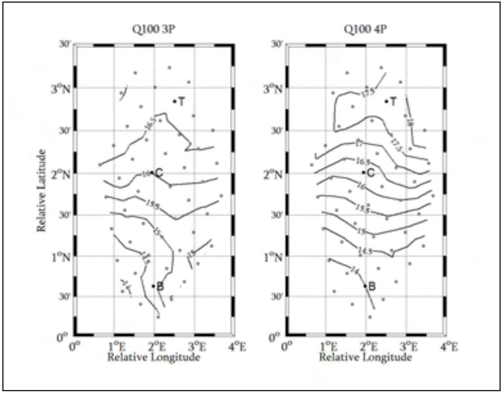
- Posterior distributions for 4-Parameter model parameters at location T (top, dashed line), location C (centre, solid line) and location B (bottom, dotted line).
- 3-Parameter maximum likelihood estimates for, μ, $log(\sigma)$ and ξ are also indicated as vertical lines for comparison. Results are for 75% threshold and 48 hour blocked data

Posterior Distribution of Quantiles



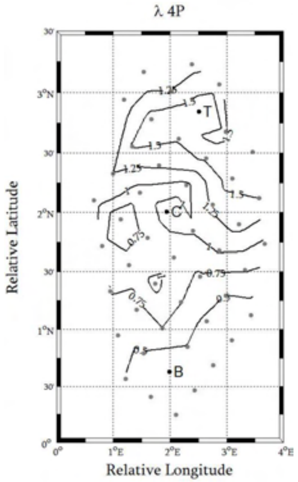
- Posterior distributions for $Q_{0.1}$ (dashed line), Q_1 (solid line), Q_{10} (dotted line) and Q_{100} (dot-dashed line) return values for 4-Parameter model at locations T (top), C (centre) and B (bottom).
- 3-Parameter model maximum likelihood estimates are also shown as vertical lines for comparison. Results are for 75% threshold and 48 hour blocked data.

Q_{100} Median Return Value



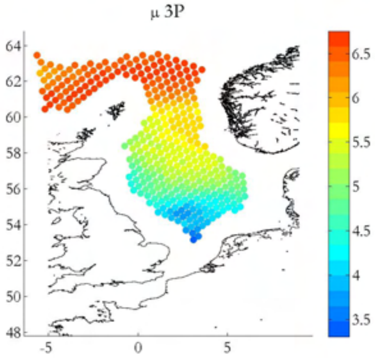
- Q100 median return values for the 3- and 4-Parameter models
- The contours show that the Q_{100} median return level increases towards deeper, more exposed northern sites

Scale λ

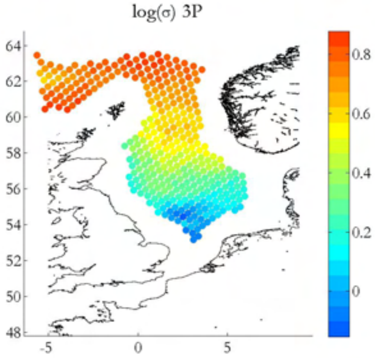


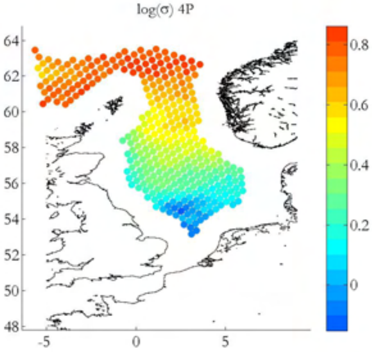
- Median values of scale parameter λ from 4-Parameter model as contours over locations.
- The value of λ varies from approximately 0.4 to 1.8 and increases moving northwards into deeper water.
- The data scale transformations from the square root to the square of the data may be appropriate to enhance the extreme value model.

Extended Data μ

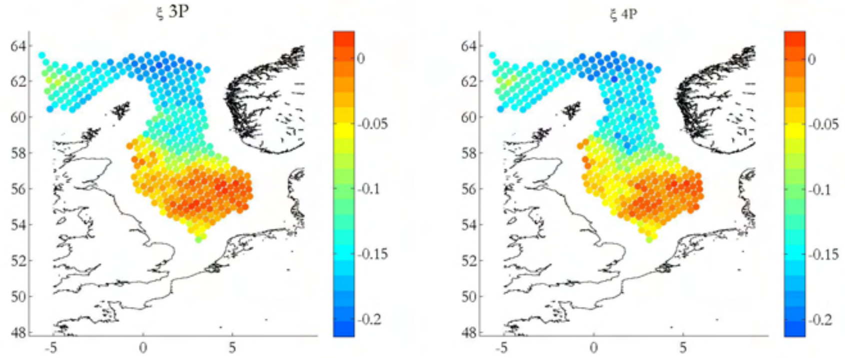


Extended Data $log(\sigma)$

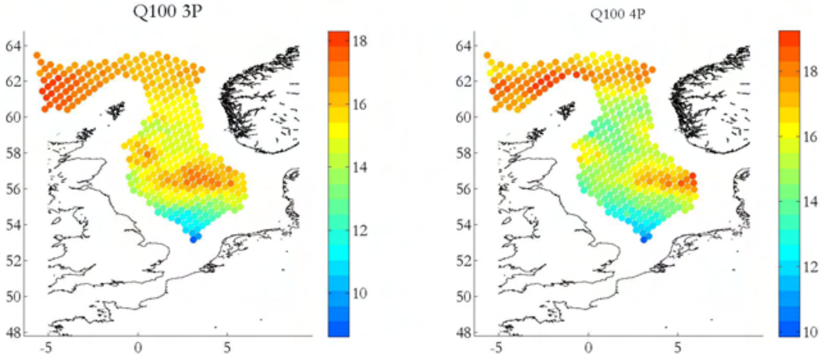




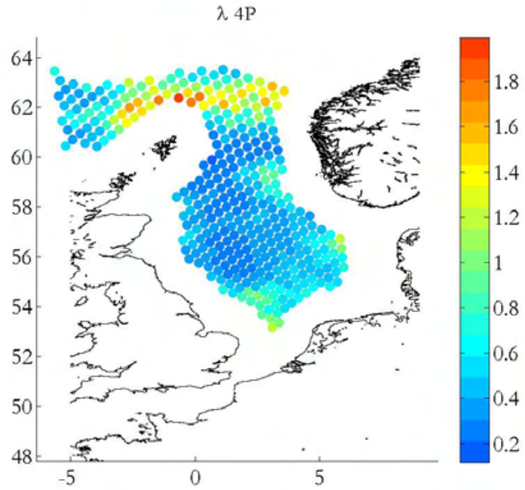
Extended Data ξ



Extended Data $Q100$



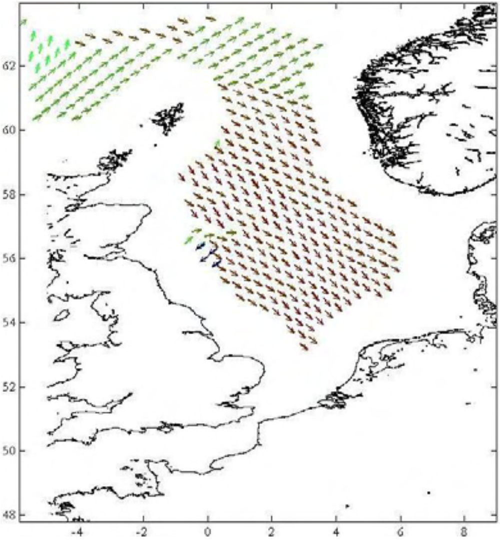
Extended Data λ



Summary of results

- Analysis suggests that measurement scale is related to location, particularly to latitude
  - Northerly locations, exposed to longer fetches, favour linear to square scales
  - More southerly locations, with limited fetches, favour smaller values of scale parameter
- The 4-parameter model allows us to explore scale effects and quantify their influence on estimation of return levels
- For the present North Sea application, it appears that scale effects do not influence extreme quantiles materially

Next Step



- Spatial model (spline, BHM)
- Directional model

Thanks for listening!