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

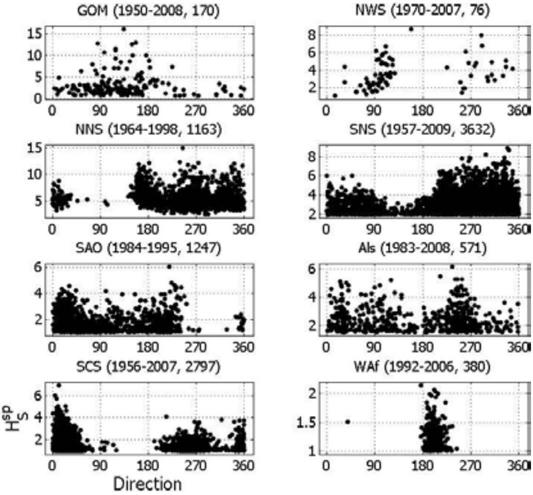
David Randell, Elena Zanini, Michael Vogel,
Kevin Ewans, Philip Jonathan
Statistics and Chemometrics

Motivation

- **Rational** design an assessment of marine structures:
  - Reducing **bias** and **uncertainty** in estimation of structural reliability
  - Improved **understanding** and communication of risk
- **Non-stationarity** with respect to **covariates** has important implications:
  - Extreme value analysis assumes stationarity
  - Typically need to incorporate covariates in extreme value models for credible design criteria
  - For storm severity, storm **direction** is typically an influential covariate

Case studies

Storm peak significant wave height H_S^{sp} on storm peak direction θ^{sp} for the 8 locations. 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)



Case studies

The physical and sample characteristics of case studies are as follows:

- GOM: Hurricanes; from Atlantic; ≈ 3 p.a.; ≈ 60 years
- NWS: Tropical cyclones; from north-east, rotation important; ≈ 2 p.a.; ≈ 40 years
- NNS: Winter storms; from Atlantic, Norwegian Sea, North Sea; ≈ 30 p.a.; ≈ 20 years
- SNS: Winter storms; from Atlantic, Norwegian Sea, North Sea; ≈ 70 p.a.; ≈ 40 years
- SAO: Extra-tropical lows; from North Atlantic, South Atlantic; ≈ 100 p.a.; ≈ 10 years
- Als: Extra-tropical lows; from Bearing Sea, Gulf of Alaska, East Siberian Sea; ≈ 20 p.a.; ≈ 20 years
- SCS: Monsoonal; from south-west and north-east; ≈ 60 p.a.; ≈ 50 years
- WAf: Swell; from south to south-west; ≈ 30 p.a.; ≈ 15 years

Questions

- Which environments are most severe?
- Which environments show greatest **variability** in extreme events?
- For which environments does incorporating non-stationarity make the biggest **difference** to estimated return values?
- Does incorporating non-stationary increase or decrease estimates return values in general?
- Does incorporating non-stationary increase or decrease spread of return values in general?

Model

- Sample $\{\dot{z}_i\}_{i=1}^{\dot{n}}$ of $\dot{n}$ **storm peak** significant wave heights observed with storm peak directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$
- Model components:
  1. **Threshold** function ψ above which observations $\dot{z}$ are assumed to be extreme estimated using quantile regression
  2. **Rate of occurrence** of threshold exceedances modelled using Poisson model with rate $\rho(\stackrel{\triangle}{=} \rho(\theta))$
  3. **Size of occurrence** of threshold exceedance using generalised Pareto (GP) model with shape and scale parameters ξ and σ

Model components

- Rate of occurrence and size of threshold exceedance functionally **independent** (Chavez-Demoulin and Davison 2005)
  - Equivalent to non-homogeneous Poisson point process model (Dixon et al. 1998)
- Smooth functions of covariates estimated using penalised B-splines (Eilers and Marx 2010)
  - Slick linear algebra (c.f. generalised linear array models, Currie et al. 2006)
- Large number of parameters to estimate
  - Computational efficiency essential

Penalised B-splines

- Physical considerations suggest model parameters ψ, ρ, ξ and σ vary smoothly with covariates θ
- Values of $(\eta =)\psi, \rho, \xi$ and σ all take the form:

$$\eta = B\beta_\eta$$

  for **B-spline** basis matrix B (defined on index set of covariate values) and some β_η to be estimated
- **Wrapped** basis for periodic directional covariate
- Roughness R_η defined as:

$$R_\eta = \beta_\eta' P \beta_\eta$$

  where effect of P is to difference neighbouring values of β_η
- Parameter **smoothness** controlled by roughness coefficient λ in roughness-penalised maximum likelihood estimation

Wrapped periodic B-spline basis

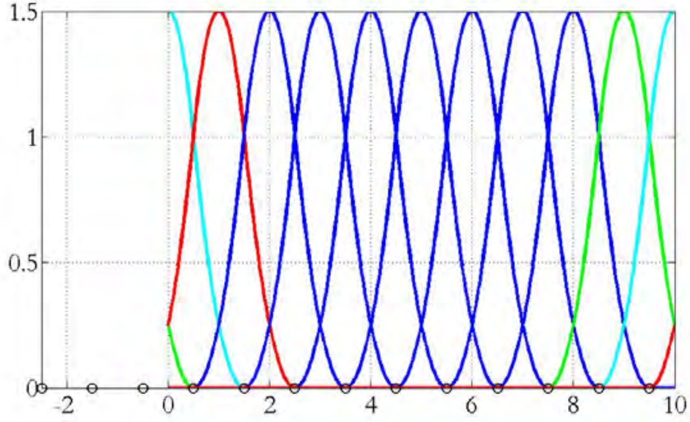


Figure: Illustrative wrapped B-spline basis on $[0, 10)$

Quantile regression model for extreme value threshold

- Estimate smooth quantile $\psi(\theta;\tau)$ for non-exceedance probability τ of z (storm peak H_S) using quantile regression by minimising **penalised** criterion ℓ^*_ψ with respect to basis parameters:

$$
\begin{aligned}
\ell^*_\psi &= \ell_\psi + \lambda_\psi R_\psi \\
\ell_\psi &= \{\tau \sum_{r_i \geq 0}^{n} |r_i| + (1-\tau) \sum_{r_i < 0}^{n} |r_i|\}
\end{aligned}
$$

  for $r_i = z_i - \psi(\theta_i, \phi_i; \tau)$ for $i = 1, 2, ..., n$, and **roughness** R_ψ controlled by roughness coefficient λ_ψ

- (Non-crossing) quantile regression formulated as linear programme (Bollaerts et al. 2006)

Poisson model for rate of threshold exceedance

- Poisson model for rate of occurrence of threshold exceedance estimated by minimising roughness penalised log likelihood:

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

- (Negative) penalised Poisson log-likelihood (and approximation):

$$\begin{aligned} \ell_\rho &= -\sum_{i=1}^{n} \log \rho(\theta_i, \phi_i) + \int \rho(\theta) d\theta dx dy \\ \hat{\ell}_\rho &= -\sum_{j=1}^{m} c_j \log \rho(j\Delta) + \Delta \sum_{j=1}^{m} \rho(j\Delta) \end{aligned}$$

- $\{c_j\}_{j=1}^m$ counts of threshold exceedances on index set of m ($>> 1$) bins partitioning covariate domain into intervals of volume Δ
- λ_ρ estimated using cross validation or similar (e.g. AIC)

GP model for size of threshold exceedance

- Generalise Pareto model for size of threshold exceedance estimated by minimising roughness penalised log-likelihood:

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

- (Negative) conditional generalised Pareto log-likelihood:

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

- Parameters: **shape** ξ, **scale** σ
- Threshold ψ set prior to estimation
- λ_ξ and λ_σ estimated using cross validation or similar. In practice set $\lambda_\xi = \kappa\lambda_\sigma$ for fixed κ

Extreme value analysis of H_S^{sp} for SAO

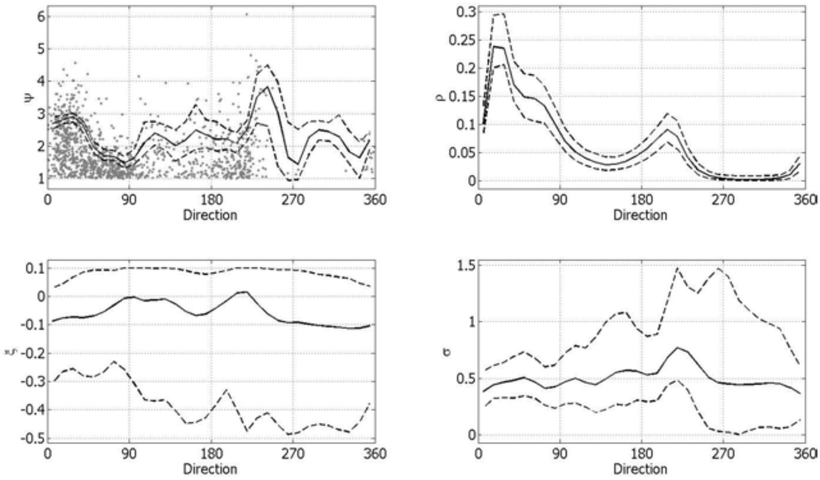


Figure: Threshold ψ, rate ρ, shape ξ and scale σ with storm peak direction θ^{sp} using $\tau = 0.8$. Bootstrap median (solid black) and 95% uncertainty band (dashed black). Sample (grey) shown with ψ

Return values

- Estimation of return values by simulation under model
- Simulate the desired return period multiple (≈ 1000) times
  - Sample number of events in period, directions of events, sizes of events
- Estimate the cumulative distribution function (CDF) for the return value of interest

- By simulating for return period equal to period of sample, can perform **model validation**

Validation of H_S^{sp} model for SAO

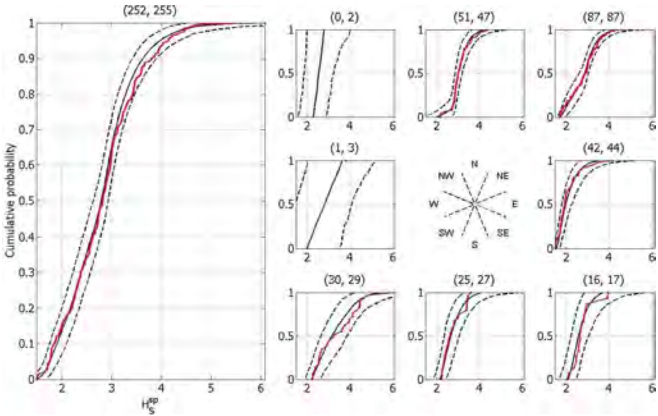


Figure: CDFs from original sample (red) and from 1000 realisations under model (black) for period of original sample, with 95% uncertainty bands. LHS: Omnidirectional. RHS: For 8 directional octants. Titles: numbers of actual and average simulated events. **Good agreement**

Comparisons of return value distributions

- Return value distributions for directional octants (centred on cardinal and semi-cardinal directions) per location
  - Identify differences in directional effects per location
- Omni-directional return value distributions per location
  - Compare environments by severity
- Centred and scaled omni-directional return value distributions per location
  - Which environments have longer tails of return value distributions?
- Assess the effect of sample size on width of return value distribution
  - How does width of return value distribution vary with sample size?
- Compare return value distributions from stationary and non-stationary models
  - How does incorporating direction change characteristics of return value distribution?

100-year H_S^{sp} for GOM, NWS, NNS and SNS

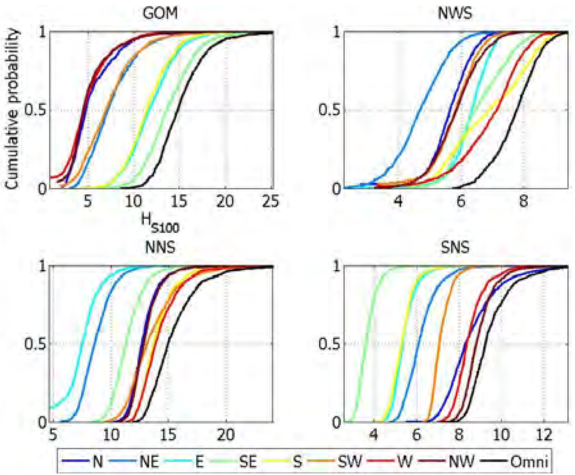


Figure: CDFs for omnidirectional (black) and directional octant (colour) return values, from simulation under directional model, incorporating uncertainty in parameter estimation using bootstrap resampling

100-year H_S^{sp} for SAO, AIs, SCS and WAf

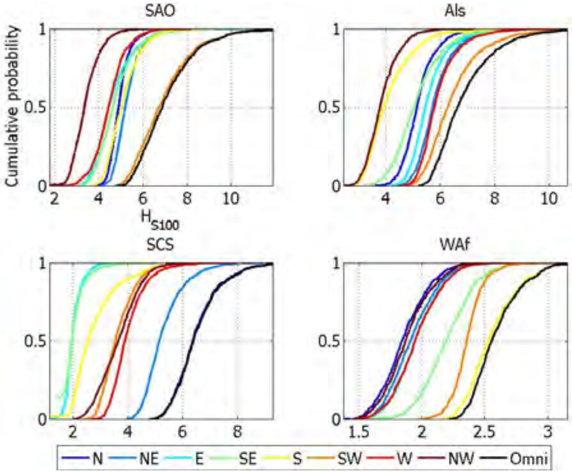


Figure: CDFs for omnidirectional (black) and directional octant (colour) return values, from simulation under directional model, incorporating uncertainty in parameter estimation using bootstrap resampling

Characteristics of 100-year H_S^{sp} distributions

- Obvious directional differences for all locations
- One directional octant dominating
  - e.g. SAO (SW), SCS (N), WAf (S)
- Sub-set of directional octants dominating
  - e.g. GOM (E, SE, S), SNS (N, NW, W)
- Obvious land shadows
  - e.g. NNS (NE,E), SNS (SE), SCS (E, SE)
- Some surprises?
  - e.g. NWS (W) due to **large rate** of occurrence

Omni-directional 100-year H_S^{sp}

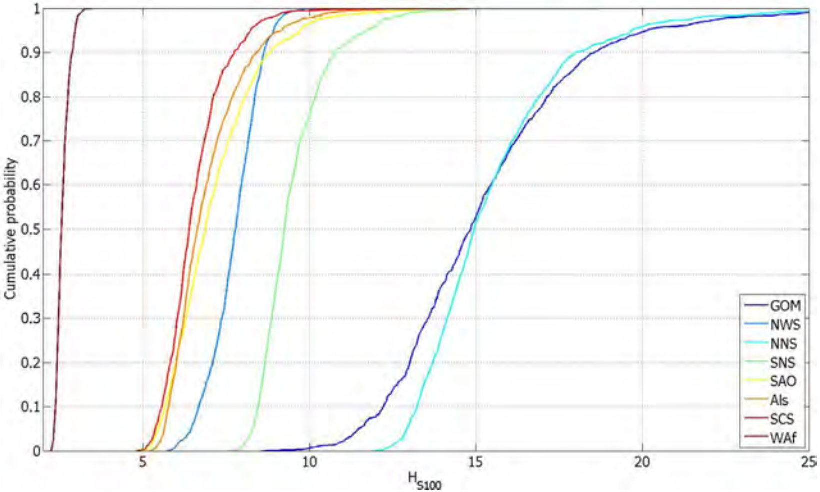


Figure: GOM and NNS are most severe, with longer tails

Centred and scaled omni-directional 100-year H_S^{sp}

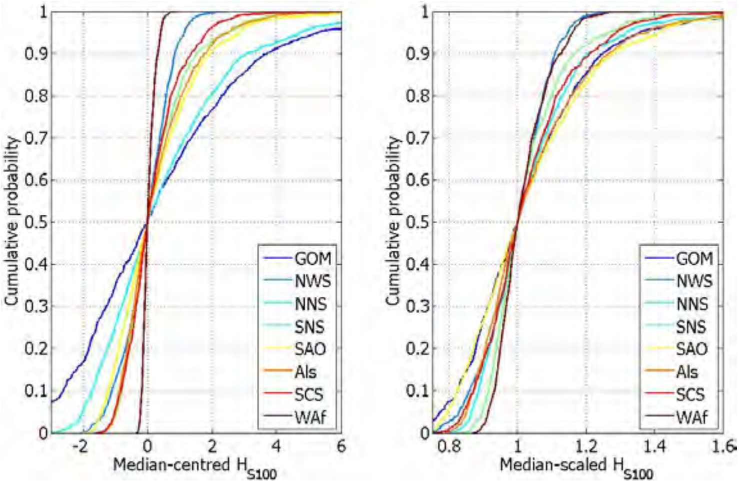


Figure: CDFs centred (LHS) and scaled (RHS) with respect to median value per sample. Once scaled, **NWS and WAf have shortest tails. Ratio of 95%ile to median ≈ 1.5 for all other locations**

IQR of 100-year H_S^{sp} distributions

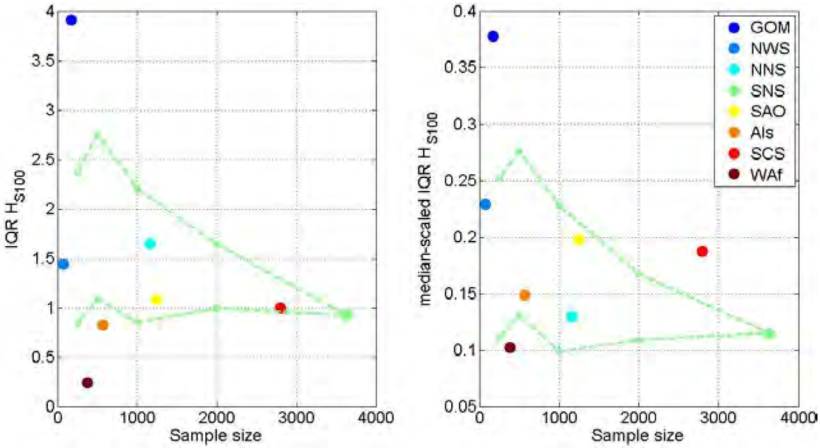


Figure: Inter-quartile range (IQR) for 100-year H_S^{sp} distributions against sample size. Dashed lines: min and max IQR from 25 randomly-chosen time-intervals (of given size) from SNS sample. LHS: IQR on original scale, RHS: median-scaled. **GOM and WAf unusual**

Comparing stationary and non-stationary CDFs

Return value distributions for 100-year H_S^{sp} estimated using three models and compared.

- Fully-directional
  - All of quantile extreme value threshold ψ, rate of threshold exceedance ρ, generalised Pareto shape ξ and scale σ are functions of direction
- "Semi-directional"
  - ψ and ρ are directional, but ξ and σ are constants with direction
- Stationary (or constant)
  - All model parameters are constant with respect to direction

Comparing stationary and non-stationary CDFs

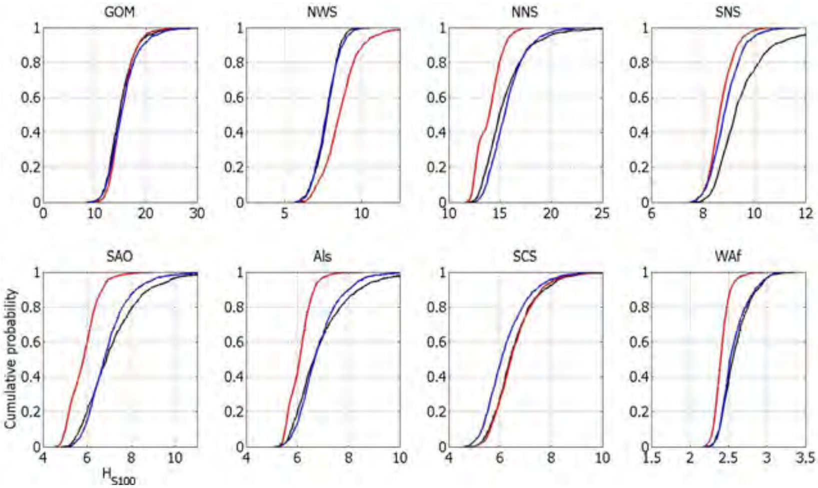


Figure: Omnidirectional 100-year H_S^{sp} CDFs from fully-directional (black), "semi-directional" (blue) and constant (or stationary, red) models

Comparing stationary and non-stationary CDFs

For the case studies examined here it appears that:

- Estimated CDFs are very similar for GOM
- CDFs from **fully-directional and "semi-directional" models agree well**
  - Accommodating covariate effects in threshold and rate is sometimes sufficient
  - Accommodating covariate effects GP shape and scale is sometimes - but not always - less important
  - SNS is an exception
- Characteristics of CDFs from constant model **unpredictable** (relative to others)
  - No systematic difference in width or median value of CDFs from constant model relative to fully-directional and "semi-directional" models
  - Not possible to predict how reliable CDFs from a constant model will be in practice

Summary

- Directional extreme value models for H_S^{sp} from 8 locations
- Return value distributions vary with direction and location in line with physical intuition
- Omni-directional return value distributions sometimes dominated by single directional octant, sometimes by sub-set of octants. Land shadow effects obvious
- Ratio 95%ile to median for distribution of 100-year $H_S^{sp} \approx 1.5$
  - NWS and WAf are exceptions, showing ratios nearer 1.2
- Distribution 100-year H_S^{sp} for GOM unusually wide
  - Width of distributions for other locations consistent with SNS
  - WAf is an exception, with unusually narrow distribution
- CDFs for 100-year H_S^{sp} from fully-directional and "semi-directional" models generally consistent
  - SNS is exception. Directional effects in GP parameters important
  - Characteristics of CDFs from constant model unpredictable relative to those from full and "semi"
- Generally important to accommodate covariate effects in threshold and rate, sometimes in GP shape and scale

References

K Bollaerts, P H C Eilers, and M Aerts. Quantile regression with monotonicity restrictions using P-splines and the L1 norm. *Statistical Modelling*, 6:189–207, 2006.

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

I. D. Currie, M. Durban, and P. H. C. Eilers. Generalized linear array models with applications to multidimensional smoothing. *J. Roy. Statist. Soc. B*, 68:259–280, 2006.

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

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