MULTIDIMENSIONAL COVARIATE EFFECTS IN SPATIAL AND JOINT EXTREMES

Philip Jonathan, Kevin Ewans, David Randell, Yanyun Wu
philip.jonathan@shell.com

www.lancs.ac.uk/∼jonathan

Wave Hindcasting & Forecasting Workshop,
Banff, October 27 - November 1, 2013

Outline

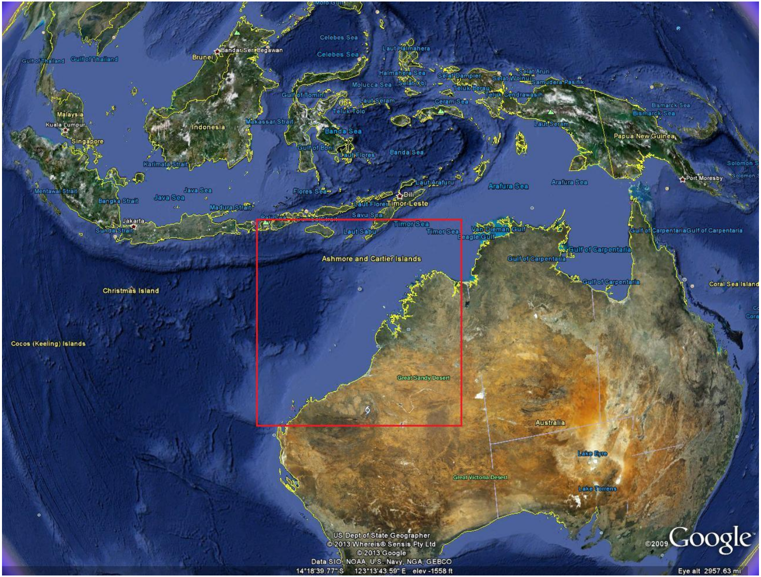
1. Background
   - Motivation
   - Australian North West Shelf
2. Extreme value analysis: challenges
3. Non-stationary extremes
   - Model components
   - Penalised B-splines
   - Quantile regression model for extreme value threshold
   - Poisson model for rate of threshold exceedance
   - Generalised Pareto model for size of threshold exceedance
   - Return values
4. Current developments

Contents

Motivation

- **Rational** design an assessment of marine structures:
  - Reducing **bias** and **uncertainty** in estimation of structural reliability
  - Improved understanding and communication of risk
  - For new (e.g. floating) and existing (e.g. steel and concrete) structures
  - Climate change
  - **Whole-basin** analysis: non-stationary analysis for 1000s of locations with covariates
- Other applied fields for extremes in industry:
  - Corrosion and fouling
  - Economics and finance

Australian North West Shelf



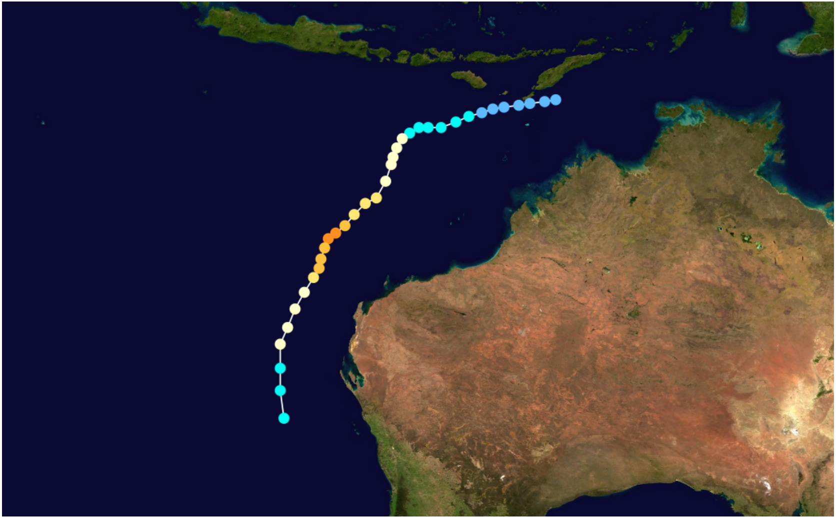
Australian North West Shelf

- Model **storm peak significant wave height** H_S
- Wave climate is dominated by westerly **monsoonal swell** and **tropical cyclones**
- Cyclones originate from Eastern Indian Ocean, Timor and Arafura Sea

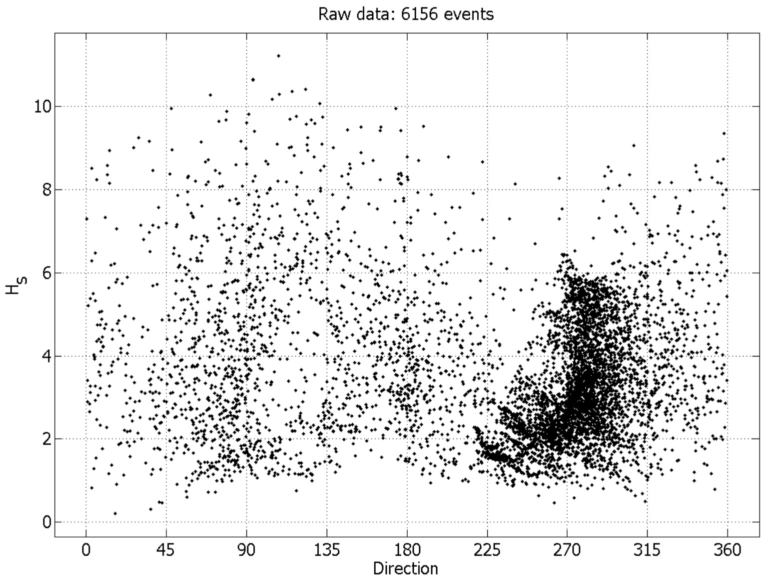
- Sample of **hindcast** storms for period 1970-2007
- 33×33 rectangular spatial grid over $4^o \times 4^o$ longitude-latitude domain
- **Spatial** and **directional** variability in extremes present
- **Marginal** spatio-directional model

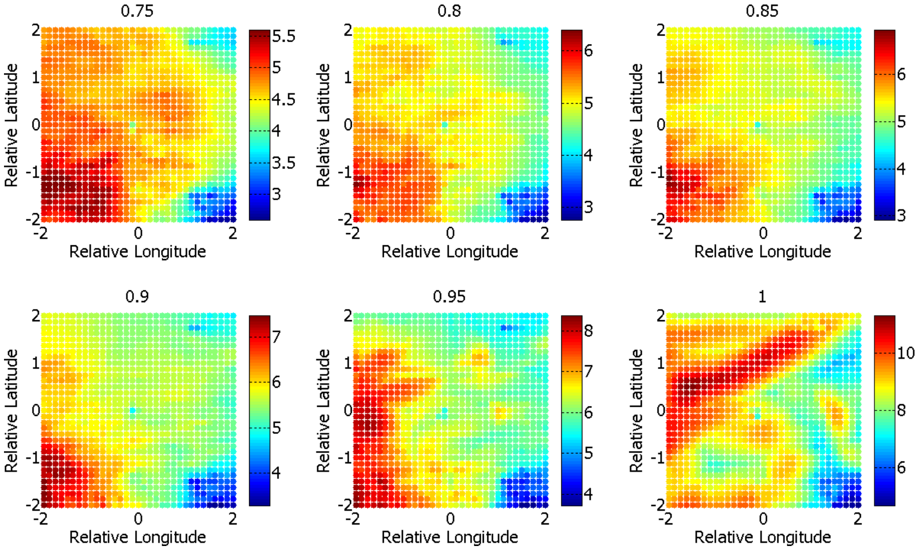
Cyclone Narelle January 2013: spatio-directional

Cyclone Narelle January 2013: cyclone track



Storm peak H_S by direction for all locations




0.75
0.8
0.85
0.9
0.95
1
Relative Latitude
Relative Longitude
5.5
5
4.5
4
3.5
3
6
5
4
3
7
8
10

Contents

Extreme value analysis: challenges

- **Covariates** and **non-stationarity**:
  - Location, direction, season, time, water depth, ...
  - Multiple / multidimensional covariates in practice
- **Cluster** dependence:
  - Same events observed at many locations (pooling)
  - Dependence in time (Chavez-Demoulin and Davison 2012)
- **Scale** effects:
  - Modelling X or $f(X)$? (Reeve et al. 2012)
- **Threshold** estimation:
  - Scarrott and MacDonald [2012]
- **Parameter** estimation
- **Measurement** issues:
  - Field measurement uncertainty greatest for extreme values
  - Hindcast data are simulations based on pragmatic physics, calibrated to historical observation

Extreme value analysis: **multivariate** challenges

- **Componentwise maxima**:
  - $\Leftrightarrow$ max-stability $\Leftrightarrow$ multivariate regular variation
  - Assumes all components extreme
  - $\Rightarrow$ Perfect independence or asymptotic dependence **only**
  - Composite likelihood for spatial extremes (Davison et al. 2012)
- **Extremal dependence**: (Ledford and Tawn 1997)
  - Assumes regular variation of joint survivor function
  - Gives more general forms of extremal dependence
  - $\Rightarrow$ Asymptotic dependence, asymptotic independence (with +ve, -ve association)
  - Hybrid spatial dependence model (Wadsworth and Tawn 2012)
- **Conditional extremes**: (Heffernan and Tawn 2004)
  - Assumes, given one variable being extreme, convergence of distribution of remaining variables
  - Allows some variables not to be extreme
  - Not equivalent to extremal dependence
- Application:
  - ... *a huge gap in the theory and practice of multivariate extremes* ... (Beirlant et al. 2004)

Contents

Model components

- Sample $\{\dot{z}_i\}_{i=1}^{\dot{n}}$ of $\dot{n}$ storm peak significant wave heights observed at locations $\{\dot{x}_i, \dot{y}_i\}_{i=1}^{\dot{n}}$ with storm peak directions $\{\dot{\theta}_i\}_{i=1}^{\dot{n}}$
- Model components:
  1. **Threshold** function ϕ 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 $\rho(\stackrel{\triangle}{=} \rho(\theta, x, y))$
  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)
- $\sim 4 \times 33 \times 33 \times 32 \sim 10^5$ parameters to estimate
  - Computational efficiency essential

Penalised B-splines

- Physical considerations suggest model parameters ϕ, ρ, ξ and σ vary smoothly with covariates θ, x, y
- Values of $(\eta =)\phi, \rho, \xi$ 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
- Multidimensional basis matrix B formulated using Kronecker products of marginal basis matrices:

$$B = B_\theta \otimes B_x \otimes B_y$$

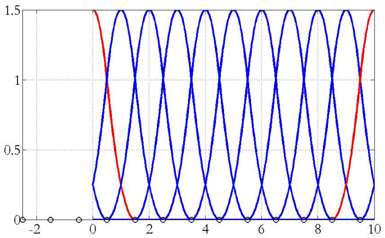
- Roughness R_η defined as:

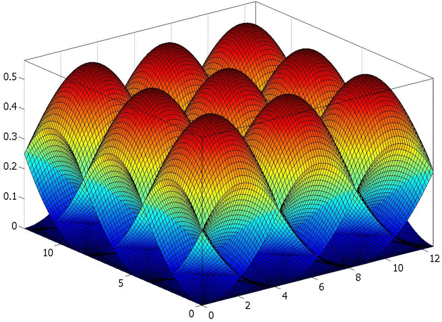
$$R_\eta = \beta_\eta' P \beta_\eta$$

  where effect of P is to difference neighbouring values of β_η

Penalised B-splines

- **Wrapped** bases for periodic covariates (seasonal, direction)
- **Multidimensional** bases easily constructed. **Problem size** sometimes prohibitive
- Parameter **smoothness** controlled by roughness coefficient λ: **cross validation** or similar chooses λ optimally





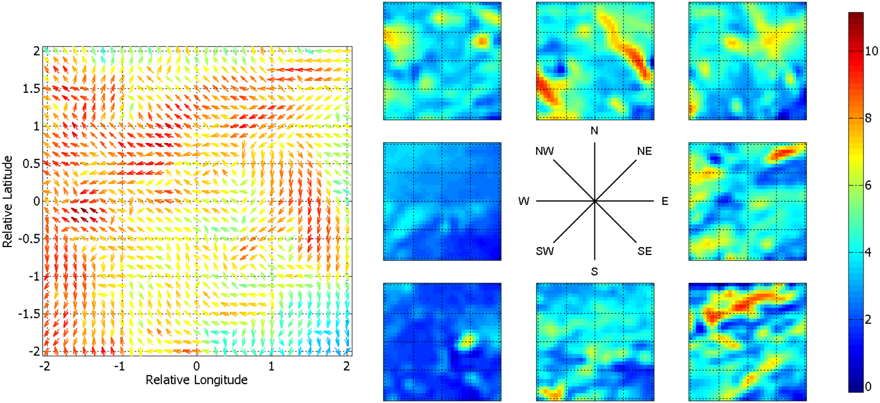
Quantile regression model for extreme value threshold

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

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

  for $r_i = z_i - \phi(\theta_i, x_i, y_i; \tau)$ for $i = 1, 2, ..., n$, and **roughness** R_ϕ controlled by roughness coefficient λ_ϕ
- (Non-crossing) quantile regression formulated as linear programme (Bollaerts et al. 2006)

Spatio-directional 50% quantile threshold



lhs: direction of highest threshold per location
rhs: spatial threshold for 8 (semi-) cardinal directions

Poisson model for rate of threshold exceedance

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

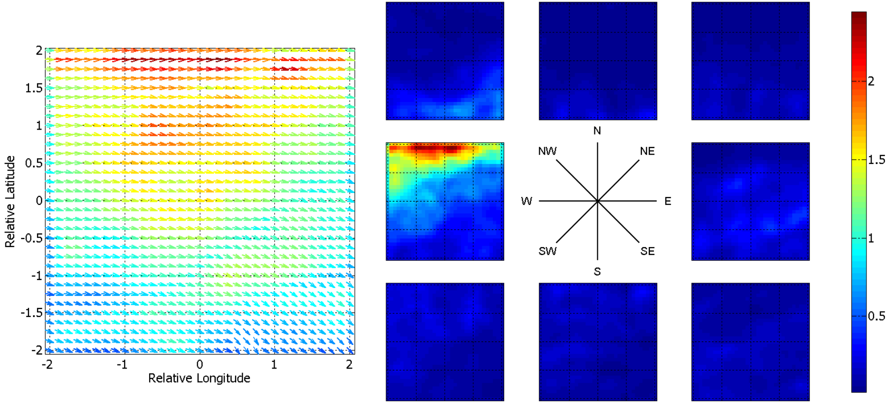
$$\ell_\rho^* = \ell_\rho + \lambda_\rho R_\rho$$

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

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

- $\{c_j\}_{j=1}^m$ 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)

Spatio-directional rate of threshold exceedances



lhs: direction of highest rate per location
rhs: spatial rate for 8 (semi-) cardinal directions

Generalised Pareto 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:

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

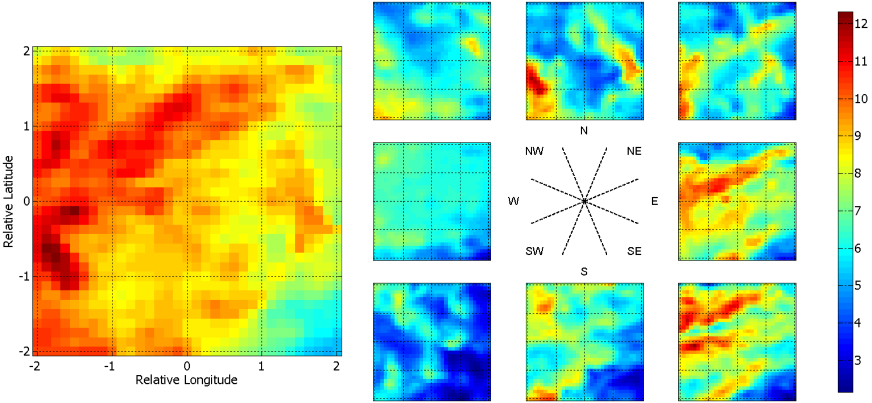
- Parameters: **shape** ξ, **scale** σ
- Threshold ϕ set prior to estimation
- λ_ξ and λ_σ estimated using cross validation or similar. In practice set $\lambda_\xi = \kappa\lambda_\sigma$ for fixed κ

Return values

- Estimation of return values by simulation under model
  - Spatio-directional octants
  - Sample number of events in period, directions of events, sizes of events
- Alternative: closed form function of parameters
  - Return value z_T of storm peak significant wave height corresponding to return period T (years) evaluated from estimates for ϕ, ρ, ξ and σ:

$$z_T = \phi - \frac{\sigma}{\xi}(1 + \frac{1}{\rho}(\log(1 - \frac{1}{T}))^{-\xi})$$

  - Interpretation **problematic**
- z_{100} corresponds to 100–year return value, denoted H_{S100}



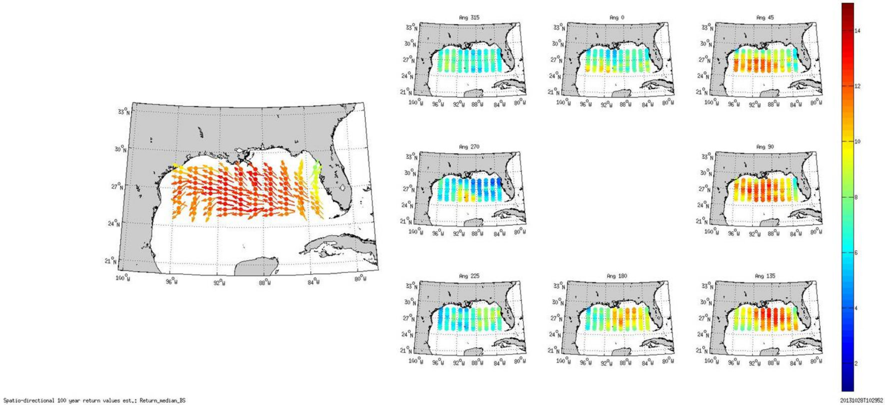
lhs: omni-directional spatial; rhs: directional octant spatial

Contents

Current developments

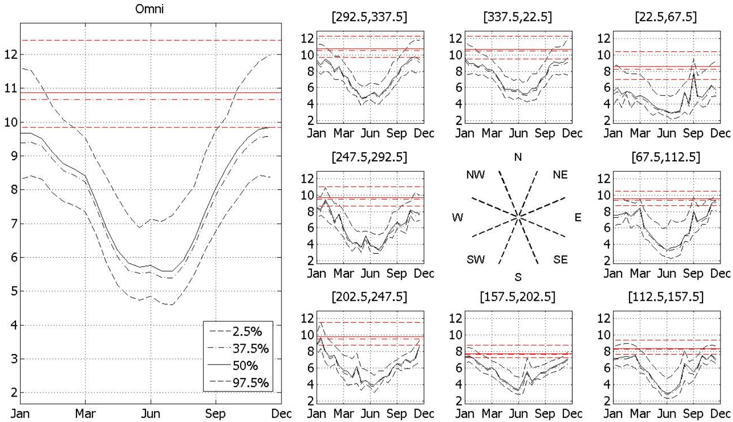
- Non-stationary **marginal** extremes
  - Spatio-directional, seasonal-directional and spatio-seasonal-directional
- Computational efficiency
  - **Sparse** and **slick** matrix manipulations (e.g. linear array methods)
  - **Parallel** implementation
- Incorporating uncertainty
  - **Bootstrapping** including threshold uncertainty
  - **Bayesian** penalised B-splines(Nasri et al. 2013, Oumow et al. 2012)
- Spatial dependence
  - Composite likelihood: model componentwise maxima
  - Censored likelihood: block maxima $\rightarrow$ threshold exceedances
  - Hybrid model: **full range** of extremal dependence
- Interpretation within **structural design framework**
- Non-stationary **conditional** extremes

Spatio-directional 100-year H_S for GoM



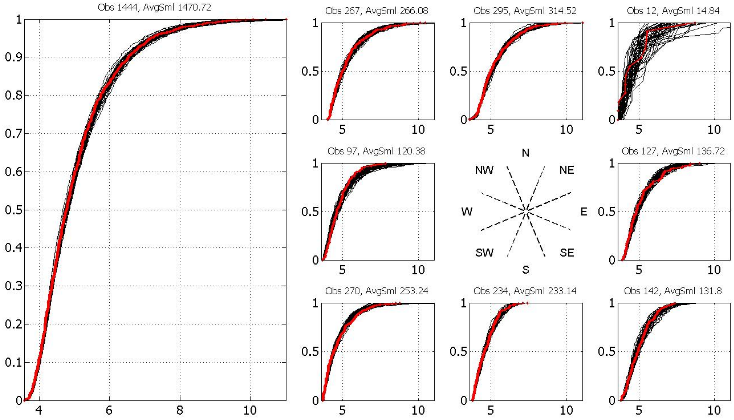
lhs: omni-directional spatial; rhs: spatial for 8 (semi-) cardinal directions

Seasonal-directional 100-year H_S for North Sea



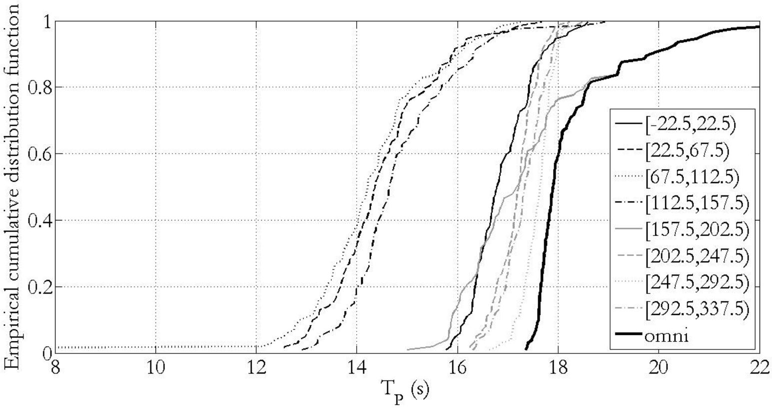
lhs: omni-directional seasonal; rhs: seasonal for 8 directional sectors
bootstrap uncertainty intervals encompass all analysis steps

Seasonal-directional H_S diagnostics



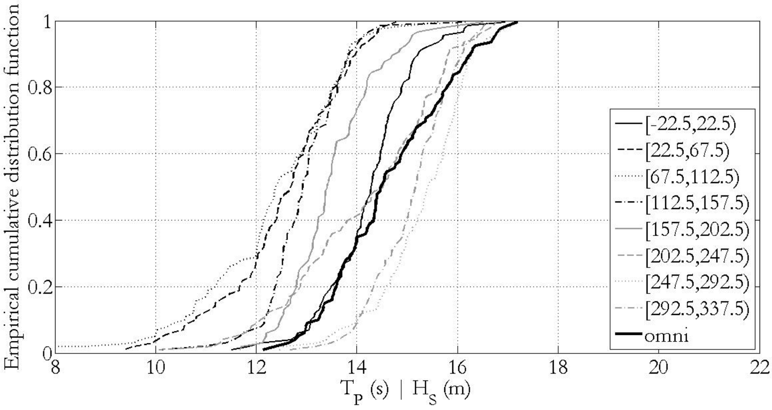
For period of data, comparison of actual (red) and multiple simulated (omni-) directional cumulative distribution functions

Directional conditional extremes of T_P given 100-year H_S for North Sea



Omni-directional and sector **marginal** distributions of 100-year T_P independent of 100-year H_S

Directional conditional extremes of T_P given 100-year H_S for North Sea



Omni-directional and sector **conditional** distributions of T_P **given** 100-year H_S using extension of model of Heffernan & Tawn

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.

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

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.

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

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.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497, 2004.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. R. Statist. Soc. B*, 59:475–499, 1997.

B. Nasri, S. El Adlouni, and T. B. M. J. Ouarda. Bayesian estimation for GEV B-Spline model. *Open Journal of Statistics*, 3: 118–128, 2013.

B. Oumow, M. de Carvalho, and A. C. Davison. Bayesian P-spline mixture modeling of extreme forest temperatures. *Available at www.mat.puc.cl/∼ mdecarvalho*, 2012.

D.T. Reeve, D. Randell, K.C.Ewans, and P. Jonathan. Accommodating measurement scale uncertainty in extreme value analysis of North Sea storm severity. *Ocean Eng.*, 53:164–176, 2012.

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

J.L. Wadsworth and J.A. Tawn. Dependence modelling for spatial extremes. *Biometrika*, 99:253–272, 2012.