Ocean extremes: environmental risk, marginal and multivariate conditional extremes

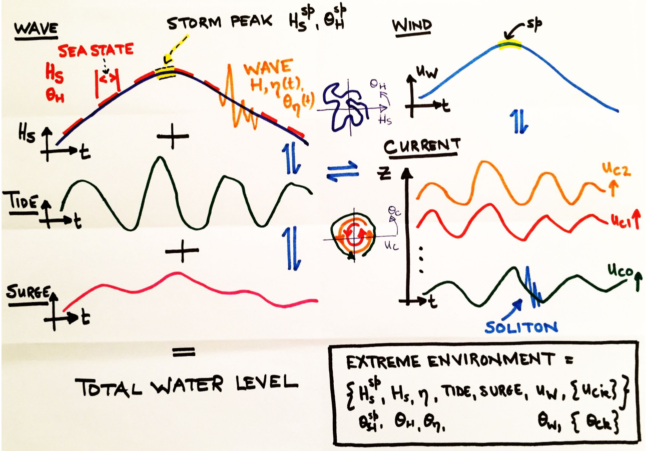
Philip Jonathan

Lancaster University, Department of Mathematics and Statistics

MetOffice / Plymouth
(Slides at *www.lancs.ac.uk/∼jonathan*)

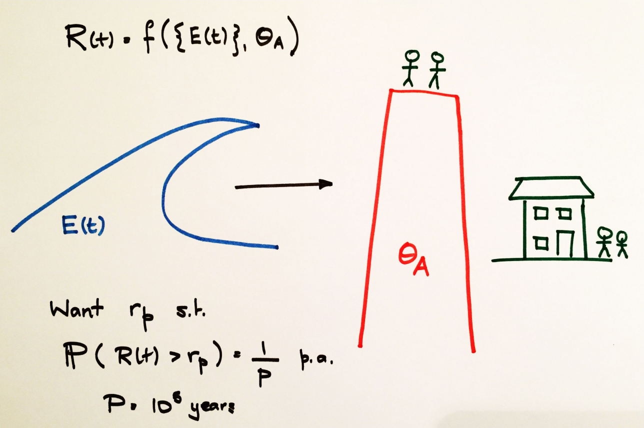
... with thanks to colleagues at Lancaster, Shell and elsewhere

Modelling ocean storm environment



- Multiple coupled physical processes
- Rare, extreme events

Modelling structural risk



- Ocean environment is harsh
- Marine structures at risk of failure
- Reliability standards must be met

Optimal design of marine structure

Set-up

- A marine system with "strength" specifications $\boldsymbol{\mathcal{S}}$
- An ocean environment $\boldsymbol{X}$ dependent on covariates $\boldsymbol{\Theta}$
- A structural "loading" $\boldsymbol{Y}$ as a result of environment X and covariates $\boldsymbol{\Theta}$
- System utility (or risk) $U(Y|\boldsymbol{\mathcal{S}})$ for loading $\boldsymbol{Y}$ and specification $\boldsymbol{\mathcal{S}}$
- Desired U typically specified in terms of annual probability of failure
- $\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Theta}$ and $X|\boldsymbol{\Theta}$ (and U?) subject to uncertainty $\boldsymbol{Z}$
- $\boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{X}, \boldsymbol{Y}$ are multidimensional random variables

Optimal design

- A model $f_{X|\Theta,Z}$ for the environment
- A model $f_{Y|X,\Theta,Z}$ for environment-structure interaction
- A model $f_{\Theta|Z}$ for the covariates

$$\mathbb{E}[U|\boldsymbol{\mathcal{S}}] = \int_\zeta \int_y \int_x \int_\theta U(\boldsymbol{y}|\boldsymbol{\mathcal{S}},\zeta) f_{Y|X,\Theta,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta},\zeta) f_{X|\Theta,Z}(\boldsymbol{x}|\boldsymbol{\theta},\zeta) f_{\Theta|Z}(\boldsymbol{\theta}|\zeta) f_Z(\zeta)\, d\boldsymbol{\theta}\, d\boldsymbol{x}\, d\boldsymbol{y}\, d\zeta$$

$\Rightarrow$ solve for $\boldsymbol{\mathcal{S}}$ to achieve required (safety) utility

Conventional approach: environmental return values

- Estimating $\mathbb{E}[U|\boldsymbol{\mathcal{S}}]$ is difficult
- Ignore the structural response Y

- Design to extreme quantile of marginal annual distribution of single X instead

$$F_A(x) = \int_{\boldsymbol{Z}} \int_{\theta} \sum_k F^k_{X|\boldsymbol{\Theta},\boldsymbol{Z}}(x|\boldsymbol{\theta},\boldsymbol{\zeta}) f_{C|\boldsymbol{\Theta},\boldsymbol{Z}}(k|\boldsymbol{\theta},\boldsymbol{\zeta}) f_{\boldsymbol{\Theta}|\boldsymbol{Z}}(\boldsymbol{\theta}|\boldsymbol{\zeta}) f_{\boldsymbol{Z}}(\boldsymbol{\zeta})\, d\boldsymbol{\theta}\, d\boldsymbol{\zeta}$$

  where $f_{C|\boldsymbol{\Theta},\boldsymbol{Z}}$ is the density of annual rate of events given covariate Θ.
- Set the return value x_T (for $T = 1000$ years say) such that

$$F_A(x_T) = 1 - \frac{1}{T}$$

- Specify conditional return values for other Xs given $X = x_T$
- Potentially as a function of covariates
- Ambiguous ordering of expectation operators ...

What is a return value?

- Random variable A represents the maximum value of some physical quantity X per annum
- The N-year return value x_N of X is then defined by the equation

$$F_A(x_N) = \Pr(A \leq x_N) = 1 - \frac{1}{N}$$

- Or

$$x_N = F_A^{-1}(1 - \frac{1}{N})$$

- Typically $N \in [10^2, 10^8]$ years

An alternative definition

- Random variable A_N represents the N-year maximum value of X
- The N-year return value x'_N of X can be found from F_{A_N} for large N, assuming independent annual maxima since

$$\begin{aligned} F_A(x_N) &= 1 - \frac{1}{N} \\ \Rightarrow F_{A_N}(x_N) &= \left(1 - \frac{1}{N}\right)^N \approx \exp(-1) \end{aligned}$$

- Use $F_{A_N}(x'_N) = \exp(-1)$ to define an alternative return value x'_N

Estimating a return value

- To estimate x_N, we need knowledge of the distribution function F_A of the annual maximum
- We might estimate F_A using extreme value analysis on a sample of independent observations of A
- Typically more efficient to estimate the distribution $F_{X|X>\psi}$ of threshold exceedances of X above some high threshold ψ using a sample of independent observations of X, and use this in turn to estimate F_A and x_N
- How is this done?

Estimating a return value

- Asymptotic theory suggests for high threshold $\psi \in (-\infty, \infty)$ that

$$F_{X|X>\psi}(x|\psi, \sigma, \xi) = 1 - \left(1 + \frac{\xi}{\sigma}(x - \psi)\right)_+^{-1/\xi}$$

  for $x > \psi$, shape $\xi \in (-\infty, \infty)$ and scale $\sigma \in (0, \infty)$
- The full distribution of X is $F_X(x) = \tau + (1-\tau)F_{X|X>\psi}(x)$ where $\tau = \Pr(X \leq \psi)$
- Thus

$$F_A(x) = \Pr(A \leq x) = \sum_{k=0}^{\infty} f_C(k) F_X^k(x)$$

  where C is the number of occurrences of X per annum, with probability mass function f_C to be estimated (say with a Poisson model with parameter λ)

- So what's the problem?

Parameter uncertainty

- x_N can be estimated easily in the absence of uncertainty

- In reality, we estimate parameters λ, ψ, σ and ξ from a sample of data, and we cannot know their values exactly
- How does this epistemic uncertainty affect return value estimates?

- A number of different plausible estimators for return values under uncertainty
- Different estimators perform differently (bias and variance)
- Which estimators are likely to perform reasonably in fairly general circumstances?

- Is it even **sensible or desirable** to estimate return values?

Incorporating uncertainty

- If a distribution $F_{Y|\boldsymbol{Z}}$ of random variable Y is known conditional on random variables $\boldsymbol{Z}$, and the joint density $f_{\boldsymbol{Z}}$ of $\boldsymbol{Z}$ is also known, the unconditional predictive distribution $\tilde{F}_Y$ can be evaluated using

$$\tilde{F}_Y(y) = \int_{\zeta} F_{Y|\boldsymbol{Z}}(x|\zeta) f_{\boldsymbol{Z}}(\zeta)\, d\zeta$$

- Th expected value of deterministic function g of parameters $\boldsymbol{Z}$ given joint density $f_{\boldsymbol{Z}}$ is

$$E[g(\boldsymbol{Z})] = \int_{\zeta} g(\zeta) f_{\boldsymbol{Z}}(\zeta)\, d\zeta$$

- $\zeta = (\lambda, \psi, \sigma, \xi)$, $Y = A$ (or $Y = A_N$)

Different estimators of return value

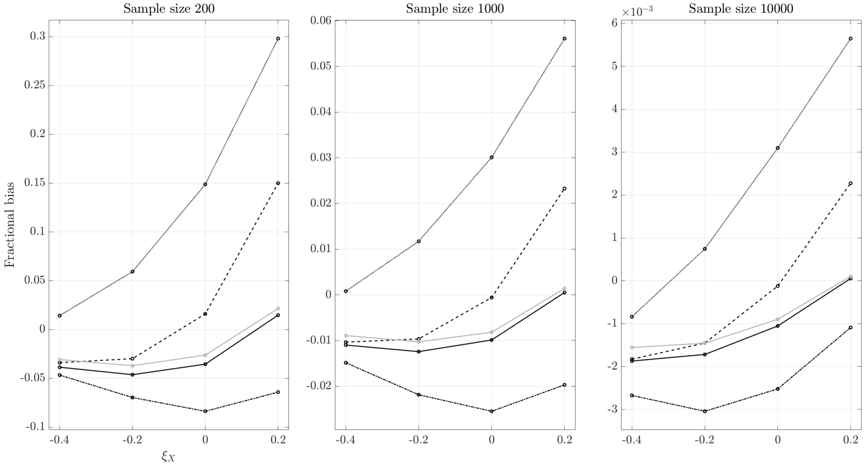
- Uncertain estimates of GP model parameters from fit to sample represented by random variables $\boldsymbol{Z}$
- Estimate distribution $F_{A|\boldsymbol{Z}}$ of annual maximum event using $\boldsymbol{Z}$
- Estimate N-year return value by finding the $1 - 1/N$ quantile of $F_{A|\boldsymbol{Z}}$

- Various options available, including:

$$\begin{aligned}
q_1 &= F_{A|\boldsymbol{Z}}^{-1}(1 - 1/N \mid \mathbb{E}_{\boldsymbol{Z}}[\boldsymbol{Z}]) = F_{A|\boldsymbol{Z}}^{-1}(1 - 1/N \mid \int_{\zeta} \zeta f_{\boldsymbol{Z}}(\zeta) d\zeta) \\
q_2 &= \mathbb{E}_{\boldsymbol{Z}}[F_{A|\boldsymbol{Z}}^{-1}(1 - 1/N \mid \boldsymbol{Z})] = \int_{\zeta} F_{A|\boldsymbol{Z}}^{-1}(1 - 1/N \mid \zeta) f_{\boldsymbol{Z}}(\zeta) d\zeta \\
q_3 &= \tilde{F}_A^{-1}(1 - 1/N) \text{ where } \tilde{F}_A(x) = \int_{\zeta} F_{A|\boldsymbol{Z}}(x \mid \zeta) f_{\boldsymbol{Z}}(\zeta) d\zeta \\
q_4 &= \tilde{F}_{A_N}^{-1}(\exp(-1)) \text{ where } \tilde{F}_{A_N}(x) = \tilde{F}_A^N(x) \\
q_5 &= \text{med}_{\boldsymbol{Z}}[F_{A|\boldsymbol{Z}}^{-1}(1 - 1/N \mid \boldsymbol{Z})]
\end{aligned}$$

- For small samples, these have very different properties

Fractional bias of return value estimators



Fractional bias of return value estimates from different estimators using maximum likelihood, as a function of sample size and true GP shape ξ.
LHS top to bottom: q_3, q_2, q_5, q_1, q_4.

- Knock-on effects for associated values of the form $\mathbb{E}_{\boldsymbol{Z}}(Y|X = q, \boldsymbol{Z})$

Modelling the non-stationary multivariate extreme environment

- Expected utility and return values are dominated by extreme environments
- Have to estimate tails of distributions well

- Think of a simple $\boldsymbol{Z}$-free 2-D environment with stationary dependence. Then

$$F_{\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{Z}}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\zeta}) = C\Big(F_{X_1|\Theta}(x_1|\theta), F_{X_2|\Theta}(x_2|\theta)\Big) \text{ and so}$$

$$\begin{aligned} f_{\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{Z}}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\zeta}) &= f_{X_1,X_2|\Theta}(x|\theta) \\ &= f_{X_1|\Theta}(x_1|\theta) f_{X_2|\Theta}(x_2|\theta) \times c\Big(F_{X_1|\Theta}(x_1|\theta), F_{X_2|\Theta}(x_2|\theta)\Big) \text{ typically} \end{aligned}$$

- Marginal models (non-stationary, extreme) $f_{X_1|\Theta}(x_1|\theta)$, $f_{X_2|\Theta}(x_2|\theta)$
- Multivariate model on standard marginal scale (stationary, "extreme") $c(u_1, u_2)$

Generalised Pareto distribution

- Suppose we have an exceedance X of high threshold $\psi \in \mathbb{R}$
- The Pickands-Balkema-De Haan theorem states

$$\begin{aligned} \lim_{\psi \to \infty} \mathbb{P}[X \leq x | X > \psi] &= \lim_{\psi \to \infty} \frac{F_X(x)}{1 - F_X(\psi)} \\ &= \mathrm{GP}(x|\xi, \sigma, \psi) \\ &= 1 - \left(1 + \frac{\xi}{\sigma}(x - \psi)\right)_+^{-1/\xi}, \quad \sigma > 0, \quad \xi \in \mathbb{R} \end{aligned}$$

Theory

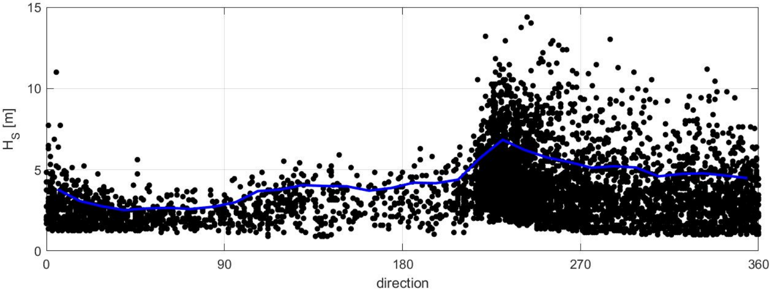
- Derived from max-stability of F_X
- Threshold-stability property
- "Poisson $\times$ GP $=$ GEV"

Practicalities

- How to isolate independent threshold exceedances from observed time-series?
- How to specify extreme threshold ψ?
- ξ, σ, ψ functions of covariates

Motivation

- Environmental extremes vary smoothly with multidimensional covariates
- Generic modelling framework for different covariate representations
- Statistical and computational efficiency for n-D covariates
- Thorough Bayesian uncertainty quantification



Typical data for northern North Sea. Storm peak H_S on direction, with $\tau = 0.8$ extreme value threshold.
Rate and size of occurrence varies with direction.

Model for size of occurrence

- Sample of storm peaks Y over threshold $\psi_\theta \in \mathbb{R}$, with 1-D covariate $\theta \in \mathcal{D}_\theta$
- Extreme value threshold ψ_θ assumed known
- Y assumed to follow generalised Pareto distribution with shape ξ_θ, (modified) scale ν_θ

$$f_{GP}(y|\xi_\theta, \nu_\theta) = \frac{1}{\sigma_\theta}\left(1 + \frac{\xi_\theta}{\sigma_\theta}(y - \psi_\theta)\right)_+^{-1/\xi_\theta - 1} \text{ with } \nu_\theta = \sigma_\theta(1 + \xi_\theta)$$

- Shape parameter $\xi_\theta \in \mathbb{R}$ and scale parameter $\nu_\theta > 0$
- (Non-stationary Poisson model for rate of occurrence, with rate $\rho_\theta \geq 0$)

Covariate representations in 1-D

- Index set $\mathcal{I}_\theta = \{\theta_s\}_{s=1}^m$ on periodic covariate domain $\mathcal{D}_\theta$
- Each observation belongs to exactly one θ_s
- On $\mathcal{I}_\theta$, assume

$$\eta_s = \sum_{k=1}^{n} B_{sk}\beta_k, s = 1, 2, ..., m, \text{ or}$$
$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\beta}$$

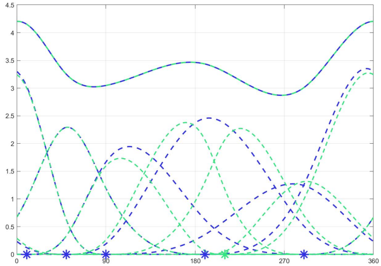
- $\eta \in (\xi, \nu)$ (and similar for ρ)
- $\boldsymbol{B} = \{B_{sk}\}_{s=1;k=1}^{m;n}$ basis for $\mathcal{D}_\theta$
- $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^{n}$ basis coefficients
- Inference reduces to estimating $n_\xi, n_\nu, \boldsymbol{B}_\xi, \boldsymbol{B}_\nu, \boldsymbol{\beta}_\xi, \boldsymbol{\beta}_\nu$ (and roughnesses λ_ξ, λ_ν)

- P-splines, BARS and Voronoi are different forms of $\boldsymbol{B}$
- Tensor products and slick GLAM algorithms for n-D covariate representations

Basis representations ... BARS and others

Bayesian adaptive regression splines (BARS)

- n irregularly-spaced knots on $\mathcal{D}_\theta$
- $\boldsymbol{B}$ consists of n B-spline bases
- Order d
- Each using $d+1$ consecutive knot locations
- Local support
- Wrapped on $\mathcal{D}_\theta$
- Knot locations $\{r_k\}_{k=1}^{n}$ vary
- Number of basis functions n varies



Periodic BARS knot birth and death

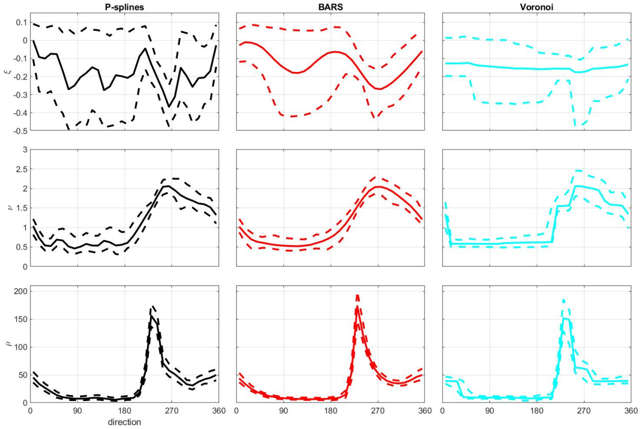
P-splines and Voronoi partition

- P-splines use fixed number of regularly-spaced knots
- Voronoi partition uses piecewise-constant representation, trivially extended to n-D

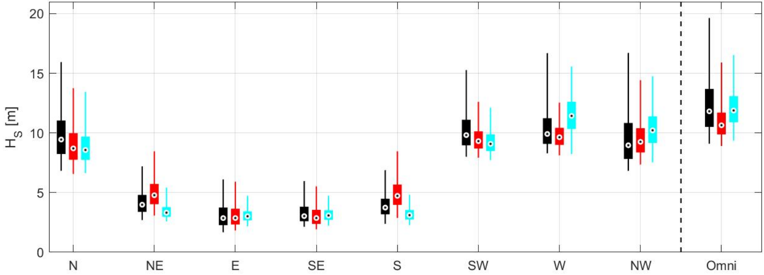
Posterior parameter estimates for ξ, ν and ρ for northern North Sea

- MCMC inference (Gibbs sampling, reversible jump, etc.)

- Note colour scheme
- Rate ρ and ν very similar
- Voronoi gives almost constant ξ
- Voronoi piecewise constant
- Land shadow effects
- General agreement
- ... for other parameters also



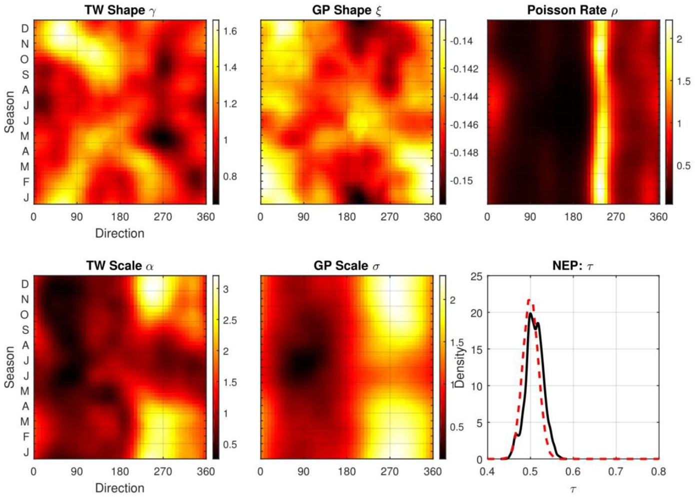
Directional posterior predictive distribution of $T = 1000$-year maximum



- Box-whiskers with 2.5%, 25%, 50%, 75% and 97.5% percentiles
- General agreement

- This is more-or-less what the engineer needs to design a "compliant" structure

Extension to 2D : directional-seasonal



- 2-D tensor product P-spline bases for same northern North Sea location
- Marginal posterior median estimates (plus posterior density for τ)

Recap: model the non-stationary multivariate extreme environment

- Expected utility dominated by extreme environments

$$\mathbb{E}[U|\boldsymbol{\mathcal{S}}] = \int_\zeta \int_y \int_x \int_\theta U(\boldsymbol{y}|\boldsymbol{\mathcal{S}},\zeta) f_{Y|X,\Theta,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta},\zeta) f_{X|\Theta,Z}(\boldsymbol{x}|\boldsymbol{\theta},\zeta) f_{\Theta|Z}(\boldsymbol{\theta}|\zeta) f_Z(\zeta)\, d\boldsymbol{\theta}\, d\boldsymbol{x}\, d\boldsymbol{y}\, d\zeta$$

- Copulas (suppressing $\boldsymbol{Z}$ for clarity)

$$F_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{\theta}) = C\Big(F_{X_1|\boldsymbol{\Theta}}(x_1|\boldsymbol{\theta}), F_{X_2|\boldsymbol{\Theta}}(x_2|\boldsymbol{\theta}), ..., F_{X_p|\boldsymbol{\Theta}}(x_p|\boldsymbol{\theta})|\boldsymbol{\theta}\Big)$$

- We already have marginal models $F_{X_j|\boldsymbol{\Theta}}(x_j|\boldsymbol{\theta})$, $j = 1, 2, ..., p$
- Now we need a dependence model or copula $C = C(u_1, u_2, ..., u_p)$

Which dependence function?

Max-stability == multivariate extreme value distribution, MEVD

- The copula is not unique
- Max-stability is one popular assumption, which itself involves a common but often unrealistic assumption of component-wise maxima
- On uniform margins, extreme value copula: $C(\boldsymbol{u}) = C^k(\boldsymbol{u}^{1/k})$
- On Fréchet margins ($G_j(z) = \exp\left(-z^{-1}\right)$), $G(\boldsymbol{z}) = \exp\left(-V(\boldsymbol{z})\right)$, for exponent measure V such that $V(r\boldsymbol{z}) = r^{-1}V(\boldsymbol{z})$, homogeneity order -1
- Rich spatial extensions to max-stable processes, MSPs
- Multivariate generalised Pareto distribution, MGPD

AD and AI

- All MEVD distributions exhibit asymptotic dependence (AD)
- Many common distributions (e.g. the multivariate Gaussian) exhibit asymptotic independence (AI)
- So extreme value copulas are not general enough to describe extremal dependence in nature
- Other copula forms do exhibit AI
- The conditional extremes model admits AD (on the boundary) and AI

Conditional extremes ... moving beyond component-wise maxima

- $\boldsymbol{X} = (X_1, ..., X_j, ..., X_p)$
- Each X and Y have standard Laplace margins ($f(x) = \exp(-|x|)/2, x \in \mathbb{R}$)
- Seek a model for $\boldsymbol{X}|(Y = y)$ for $y > u$

- Assume we can find p-dimensional scaling $\boldsymbol{a}, \boldsymbol{b} > \boldsymbol{0}$ such that

$$\mathbb{P}(\boldsymbol{Z} \leq \boldsymbol{z}|Y = y) \quad \rightarrow \quad G(\boldsymbol{z}) \quad \text{as} \quad u \rightarrow \infty$$
$$\text{for} \quad \boldsymbol{Z} \quad = \quad \frac{\boldsymbol{X} - \boldsymbol{a}(y)}{\boldsymbol{b}(y)}$$

- Non-degenerate G is unknown, and estimated empirically

- Typical scaling is $\boldsymbol{a} = \boldsymbol{\alpha} y$ and $\boldsymbol{b} = y^{\boldsymbol{\beta}}$, $\boldsymbol{\alpha} \in [-1, 1]^p$, $\boldsymbol{\beta} \in (-\infty, 1]^p$
- So simply fit regression model

$$\boldsymbol{X}|(Y = y) = \boldsymbol{\alpha} y + y^{\boldsymbol{\beta}} \boldsymbol{Z}, \text{ for } y > u$$

- $\alpha = 1, \beta = 0$: perfect dependence and AD, and $\alpha \in (0, 1)$: AI

- Heffernan and Tawn [2004] find choices for α and β for popular bivariate cases
- Bivariate Gaussian : $\alpha = \rho^2, \beta = 1/2$

Developments of the conditional extremes model

Canonical extensions

- Basic: $X|(Y = y), y > u$
- Temporal: "heatwave model" $X_1, X_2, ..., X_\tau|(X_0 = x_0), x_0 > u$
- Spatial: "spatial conditional extremes" $X_1, X_2, ..., X_s|(X_0 = x_0), x_0 > u$

Idea

$$X_1, X_2, ..., X_p|(Y = y) = \boldsymbol{\alpha} y + y^{\boldsymbol{\beta}} \boldsymbol{Z}$$

- Impose appropriate structure on parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and distribution of $\boldsymbol{Z}$
  - e.g. $\boldsymbol{\alpha}$ evolves smoothly in space
  - e.g. $\boldsymbol{Z}$ follows a multivariate Gaussian or extension thereof with appropriate mean and covariance forms
- Make a simplifying assumption
  - e.g. apply a low-order model repeatedly $X_{t+1}, X_{t+2}|(X_t = x) = [\alpha_1 \alpha_2]x + x^{[\beta_1 \beta_2]}[Z_1 Z_2]$

Further extensions

- Non-stationary and multivariate temporal and spatial models

Multivariate spatial conditional extremes (MSCE)

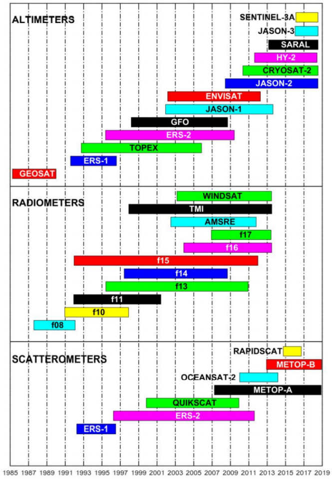
Motivation

- How useful are satellite observations of ocean waves and winds?
- Could they become the primary data source for decisions soon?
- What are the spatial characteristics of extremes from satellite observations?

Overview

- A look at the data : satellite wind, hindcast wind, hindcast wave
- Brief overview of methodology
- Results for joint spatial structure of extreme scatterometer wind speed, hindcast wind speed and hindcast significant wave height in the North Atlantic
- Implications for future practical applications

Satellite observation



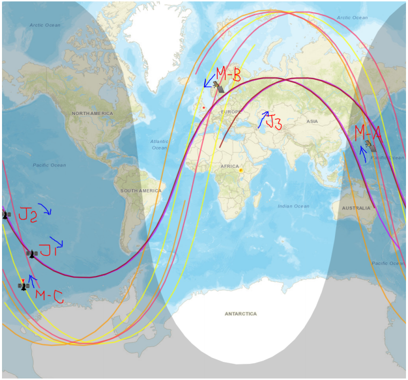
[Ribal and Young 2019]

Features

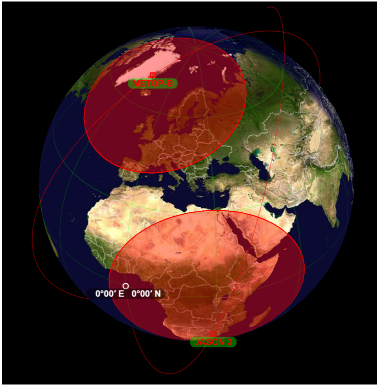
- Altimetry: H_S and U_{10}
- Scatterometry: best for U_{10} and direction
- $>$ 30 years of observations
- Spatial coverage is by no means complete: one observation daily if all well
- Calibration necessary (to buoys and reanalysis datasets, Ribal and Young 2020)
- METOP(-A,-B,-C) since 2007

H_S: significant wave height (m)
U_{10}: wind speed (ms^{-1}) at 10m (calibrated to 10-minute average wind speed)

JASON and METOP



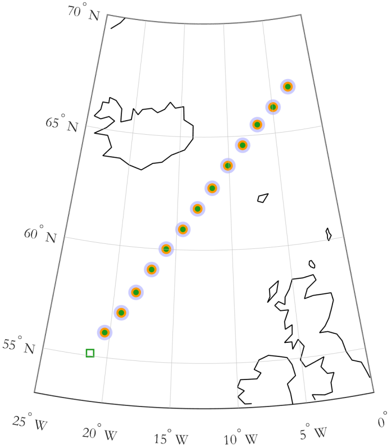
[n2yo.com, accessed 06.09.21 at around 1100UK]



[stltracker.github.io, accessed 27.08.2021 at around 1235UK]

- JASON and METOP similar polar orbits
- JASON all ascending, METOP all descending over North Atlantic
- Joint occurrence of JASON and METOP over North Atlantic rare

In a nut-shell



- Transform to standard margins using independent non-stationary GP models
- Condition on large value x of first quantity X_{01} at one location $j = 0$ (green square)
- Estimate "conditional spatial profiles" for $m > 1$ quantities $\{X_{jk}\}_{j=1,k=1}^{p,m}$ at $p > 0$ other locations (green, orange and blue circles)

$$X_{jk} \sim \text{Lpl}$$

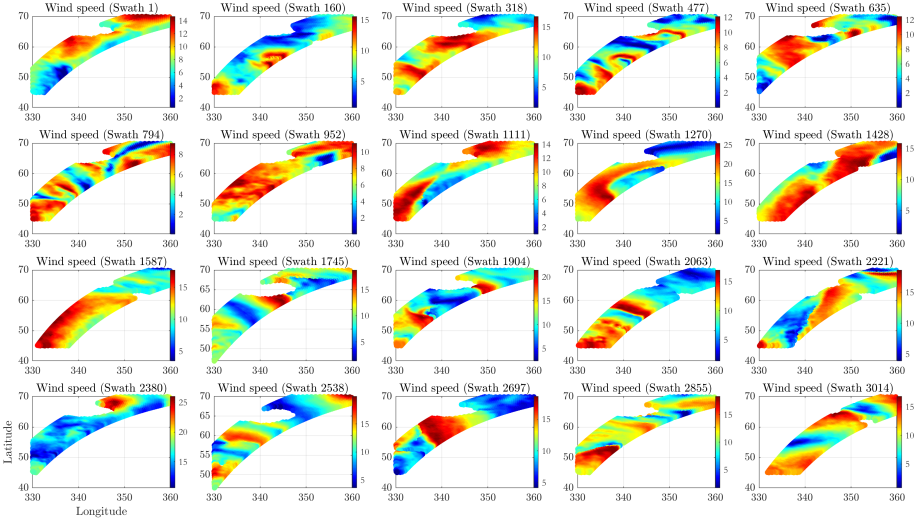
$$x > u$$

$$\boldsymbol{X}|\{X_{01} = x\} = \boldsymbol{\alpha} x + x^{\boldsymbol{\beta}} \boldsymbol{Z}$$

$$\boldsymbol{Z} \sim \text{DL}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\delta}; \boldsymbol{\Sigma}(\boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\kappa}))$$

- MCMC to estimate $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\delta}$ and $\boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\lambda}$
- $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\delta}$ spatially smooth for each quantity
- DL = delta-Laplace = generalised Gaussian
- Residual correlation $\boldsymbol{\Sigma}$ for conditional Gaussian field, powered-exponential decay with distance

Swath wind speeds



Daily descending METOP swaths. Satellite swath location changes over time. Spatial structure evident

Inference

$$X_{jk} \sim \text{Lpl}, \quad x > u, \quad \boldsymbol{X} | \{X_{01} = x\} = \boldsymbol{\alpha} x + x^{\boldsymbol{\beta}} \boldsymbol{Z}, \quad \boldsymbol{Z} \sim \text{DL}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\delta}; \boldsymbol{\Sigma}(\boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\kappa}))$$

- Delta-Laplace residual margins

$$f_{Z_{j,k}}(z_{j,k}) = \frac{\delta_{j,k}}{2\kappa_{j,k}\sigma_{j,k}\Gamma\left(\frac{1}{\delta_{j,k}}\right)} \exp\left\{ - \left| \frac{z - \mu_{j,k}}{\kappa_{j,k}\sigma_{j,k}} \right|^{\delta_{j,k}} \right\}, \quad \kappa_{j,k}^2 = \Gamma\left(1/\delta_{j,k}\right) / \Gamma\left(3/\delta_{j,k}\right)$$

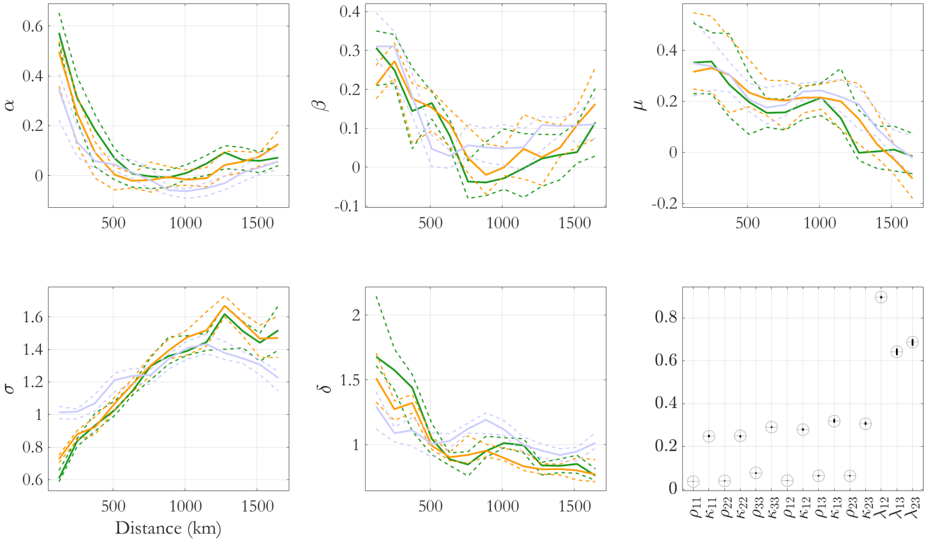
- Gaussian residual dependence

$$\boldsymbol{\Sigma}^*_{\mathcal{A}^*(j,k)\mathcal{A}^*(j',k')} = \lambda_{k,k'}^{|k-k'|} \exp\left(- \left(\frac{\text{dist}(r_j, r_{j'})}{\rho_{k,k'}} \right)^{\kappa_{k,k'}} \right)$$

- Piecewise linear forms for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\delta}$ with distance using n_{Nod} spatial nodes

- Adaptive MCMC, Roberts and Rosenthal [2009]
- Total of $m(5n_{\text{Nod}} + (3m+1)/2)$ parameters
- Rapid convergence, 10k iterations sufficient

Parameter estimates



Estimates for α, β, μ, σ and δ with distance, and residual process estimates ρ, κ and λ. Model fitted with $\tau = 0.75$
StlWnd (green), HndWnd (orange), HndWav(blue)
Residual Gaussian field : ρ=scale (need to $\times$100km), κ=exponent (need to $\times$5), λ=cross-correlation

Pragmatic non-stationary multivariate extremes with UQ

- Fit generalised Pareto marginal models for peaks over threshold data $\dot{\boldsymbol{X}}$ and $\dot{Y}$
  - Physics-based identification of peaks from time-series
  - Multiple thresholds, simple piecewise constant model for covariates $\boldsymbol{\Theta}$
  - Diagnostics: threshold stability
- Transform to standard Laplace scale $\boldsymbol{X}$ and Y
  - Transform full sample
- Fit conditional extremes model $\boldsymbol{X}|(Y = y)$ for $y > u$
  - Multiple thresholds, simple piecewise constant covariate model for α
  - Diagnostics: threshold stability, residual structure
- Calculate probabilities of extreme sets
  - MC simulation, importance sampling
  - Estimate environmental contours

- Free PPC software for MATLAB does all of above
  - UQ: incorporates epistemic uncertainty using bootstrapping cradle to grave
  - Model averaging: incorporates multiple models for different threshold combinations
  - Multidimensional X and covariates
  - Cross-validation for optimal parameter roughness in marginal and dependence models
  - Careful return value and associated value definitions
  - `https://github.com/ECSADES/ecsades-matlab`

Summary

Why?

- Careful quantification of “rare-event” risk
- Characterise tails of (multivariate) distributions
- Limited observations
- Combine solid theory and pragmatic application, UQ
- Immediate real-world consequences

The next 10 years?

- Univariate : fuller covariate descriptions, exploit measurement scale / sub-asymptotics, UQ, provide real-world decision-support
- Multivariate : theoretical development, computational tractability, expansion in scope (time-series, spatial), serious real-world applications
- More demanding regulatory framework

Thanks for listening / Diolch am wrando!

References

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes: theory and applications*. Wiley, Chichester, UK, 2004.

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207–222, 2005.

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

S Coles, J Heffernan, and J Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:339–365, 1999.

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

D. Dey and J. Yan, editors. *Extreme value modeling and risk analysis: methods and applications*. CRC Press, Boca Raton, USA, 2016.

P. Embrechts, C. Klueppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer-Verlag, 2003.

G. Feld, D. Randell, E. Ross, and P. Jonathan. Design conditions for waves and water levels using extreme value analysis with covariates. *Ocean Eng.*, 173: 851–866, 2019.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497–546, 2004.

R. Huser and J. L. Wadsworth. Advances in statistical modelling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2020. doi: 10.1002/wics.1537.

H. Joe. *Dependence modelling with copulas*. CRC Press, 2014.

P. Jonathan and K. C. Ewans. Statistical modelling of extreme ocean environments with implications for marine design : a review. *Ocean Eng.*, 62:91–109, 2013.

P. Jonathan, K. C. Ewans, and D. Randell. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188, 2014.

P. Jonathan, D. Randell, J. Wadsworth, and J.A. Tawn. Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution. *Ocean Eng.*, 220:107725, 2021.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

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

D. Randell, K. Turnbull, K. Ewans, and P. Jonathan. Bayesian inference for non-stationary marginal extremes. *Environmetrics*, 27:439–450, 2016.

A. Ribal and I. R. Young. 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Sci. Data*, 6:77, 2019.

A. Ribal and I. R. Young. Global calibration and error estimation of altimeter, scatterometer, and radiometer wind speed using triple collocation. *Remote Sens.*, 12:1997, 2020.

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *J. Comp. Graph. Stat.*, 18:349–367, 2009.

Francesco Serinaldi. Dismissing return periods! *Stoch. Env. Res. Risk A.*, 29:1179–1189, 2015.

R. Shooter, E Ross, A. Ribal, I. R. Young, and P. Jonathan. Spatial conditional extremes for significant wave height from satellite altimetry. *Environmetrics*, 32: e2674, 2021a.

R Shooter, J A Tawn, E Ross, and P Jonathan. Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24:241–265, 2021b.

R. Shooter, E Ross, A. Ribal, I. R. Young, and P. Jonathan. Multivariate spatial conditional extremes for extreme ocean environments. *Ocean Eng.*, 247:110647, 2022.

S. Tendijck, E. Ross, D. Randell, and P. Jonathan. A non-stationary statistical model for the evolution of extreme storm events. *Environmetrics*, 30:e2541, 2019.

S Tendijck, E Eastoe, J Tawn, D Randell, and P Jonathan. Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *J. Am. Statist. Soc.*, 2021. doi: 10.1080/01621459.2021.1996379.

R Towe, D Randell, J Kensler, G Feld, and P Jonathan. Estimation of associated values from conditional extreme value models. *Ocean Eng., under review*, 2022.

J L Wadsworth, J A Tawn, A C Davison, and D M Elton. Modelling across extremal dependence classes. *J. Roy. Statist. Soc. C*, 79:149–175, 2017.

H. C. Winter and J. A. Tawn. Modelling heatwaves in central France: a case-study in extremal dependence. *J. Roy. Statist. Soc. C*, 65:345–365, 2016.

S. N. Wood. Thin plate regression splines. *J. Roy. Statist. Soc. B*, 65:95–114, 2003.

Backup

Marginal extremes

- Theory : Embrechts et al. [2003], Beirlant et al. [2004]
- Method : Coles [2001], Dey and Yan [2016]
- Motivation : Davison and Smith [1990], Chavez-Demoulin and Davison [2005]

- Practicalities : Jonathan and Ewans [2013], Feld et al. [2019]
- Semi-parametric : Randell et al. [2016], Zanini et al. [2020]
- Lots more : Wood [2003]

Generalised extreme value distribution

- F_X^n is the distribution of the maximum of n independent draws of X
- If F_X^n "looks like" $F_X^{n'}$, we say F_X is max-stable

- More formally, F_X is max-stable if there exist sequences of constants $a_n > 0$, b_n, and non-degenerate G_ξ such that

$$\lim_{n \to \infty} F_X^n(a_n x + b_n) = G_\xi(x)$$

- We say $F_X \in D(G_\xi)$ or that F_X lies in the max-domain of attraction of G_ξ
- The Fisher–Tippett–Gnedenko theorem states that G_ξ is the generalised extreme value distribution with parameter ξ

$$G_\xi(y) = \exp\left(-(1+\xi y)^{-1/\xi}\right), \quad \xi \in \mathbb{R}$$

- For large n, makes sense to model block maxima of n iid draws using G_ξ (with $(x-\mu)/\sigma$ in place of y above)

Multivariate extremes

- Theory : Beirlant et al. [2004]
- Copulas : Joe [2014]
- Method : Dey and Yan [2016]
- Key ideas in AI : Ledford and Tawn [1996], Ledford and Tawn [1997], Coles et al. [1999], Heffernan and Tawn [2004]
- Modelling across dependence classes : Wadsworth et al. [2017], Huser and Wadsworth [2020]

Multivariate extreme value distribution, MEVD

- $\boldsymbol{X}_i = (X_{i1}, ..., X_{ij}, ..., X_{ip}), i = 1, ..., n$ iid p-vectors, distribution F
- $M_{n,j} = \max_i X_{ij}$, component-wise maximum
- The component-wise maximum is not "observed" (especially as $n \to \infty$)

- Then for $Z_{n,j} = (M_{n,j} - b_{n,j})/a_{n,j}$, normalised with scaling constants:

$$\mathbb{P}(\boldsymbol{Z} \leq \boldsymbol{z}) = F^n(\boldsymbol{a}_n \boldsymbol{z} + \boldsymbol{b}_n) \to G(\boldsymbol{z}) \quad \text{as} \quad n \to \infty$$

- Non-degenerate $G(\boldsymbol{z})$ must be max-stable, so $\forall k \in \mathbb{N}, \exists\, \boldsymbol{\alpha_k} > \boldsymbol{0}, \boldsymbol{\beta_k}$ s.t.

$$G^k(\boldsymbol{\alpha}_k \boldsymbol{z} + \boldsymbol{\beta}_k) = G(\boldsymbol{z})$$

- We say $F \in D(G)$
- Margins $G_1, ..., G_p$ are unique GEV, but $G(\boldsymbol{z})$ is not unique

MEVD on common margins

- On standard Fréchet margins with pseudo-polars $(r, \boldsymbol{w})$

$$
\begin{aligned}
G(\boldsymbol{z}) &= \exp\left(-V(\boldsymbol{z})\right) \\
\text{with } V(\boldsymbol{z}) &= \int_\Delta \max_j \{\frac{w_j}{z_j}\}\, S(d\boldsymbol{w}), \quad \text{on } \Delta = \{\boldsymbol{w} \in \mathbb{R}^p : ||\boldsymbol{w}|| = 1\} \\
\text{and } 1 &= \int_\Delta w_j\, S(d\boldsymbol{w}), \quad \forall j, \text{ for angular measure } S
\end{aligned}
$$

- Condition of multivariate regular variation, MRV

$$\frac{1 - F(t\boldsymbol{x})}{1 - F(t\boldsymbol{1})} \to \lambda(\boldsymbol{x}) \text{ as } t \to \infty, \boldsymbol{x} \in \mathbb{R}^p$$

useful to prove that $F \in D(G)$ for some MEVD G

- Lots more

Asymptotic dependence ... admitted by MEVD

- On uniform margins

$$\chi(u) = \frac{\mathbb{P}(U > u, V > u)}{\mathbb{P}(U > u)} = \frac{\bar{C}(u,u)}{1-u} \quad \to \chi \text{ as } u \to 1$$

- $\chi = 1$ perfect dependence
- $\chi \in (0,1)$ asymptotic dependence, AD
- $\chi = 0$ perfect independence

$$\theta(u) = \frac{\log \mathbb{P}(U \leq u, V \leq u)}{\log \mathbb{P}(U \leq u)} = \frac{\log C(u,u)}{\log u} \quad \to \theta \text{ as } u \to 1$$

- $\theta = 2 - \chi$

- χ and θ describe AD
- MEVD admits AD

Asymptotic independence ... not admitted by MEVD

- On uniform margins

$$\bar{\chi}(u) = 2\frac{\log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > u)} - 1 = 2\frac{\log(1-u)}{\log \bar{C}(u,u)} - 1 \quad \to \bar{\chi} \text{ as } u \to 1$$

- $\bar{\chi} = 1$ perfect dependence and AD
- $\bar{\chi} \in (0,1)$ asymptotic independence, AI
- $\bar{\chi} = 0$ perfect independence
- On Fréchet margins ($F(z) = \exp\left(-z^{-1}\right)$), assume

$$\frac{\mathbb{P}(Z_1 > z, Z_2 > z)}{\left(\mathbb{P}(Z_1 > z)\right)^{1/\eta}} = \mathcal{L}(z)$$

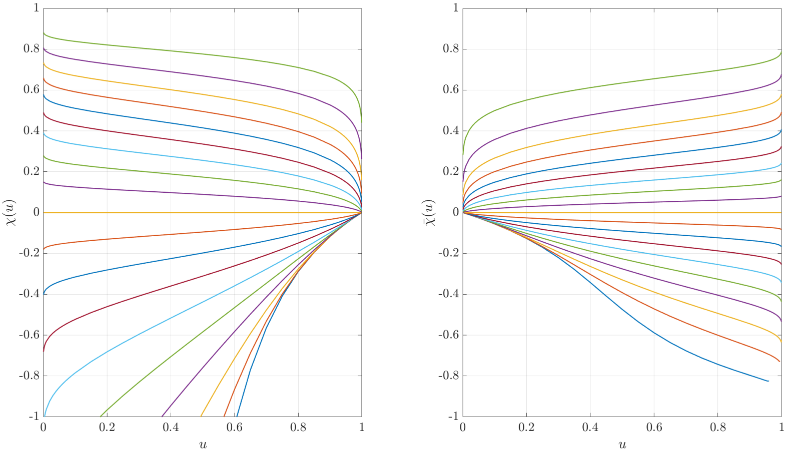
where $\mathcal{L}$ is slowly varying : $\mathcal{L}(xz)/\mathcal{L}(z) \to 1$ as $z \to \infty$
- $\bar{\chi} = 2\eta - 1$

- Idea : use non-extreme value copulas or inverted EV copulas

- Also $\mathbb{P}(Z_2 > z | Z_1 > z) \approx C z^{1-1/\eta}$ from above
- Idea : assume a max-stable-like normalisation for conditional extremes

Extremal dependence (bivariate Gaussian)

- Many (almost all?) environmental extremes problems involve asymptotic independence, at least in part ... bivariate Gaussian is one example!



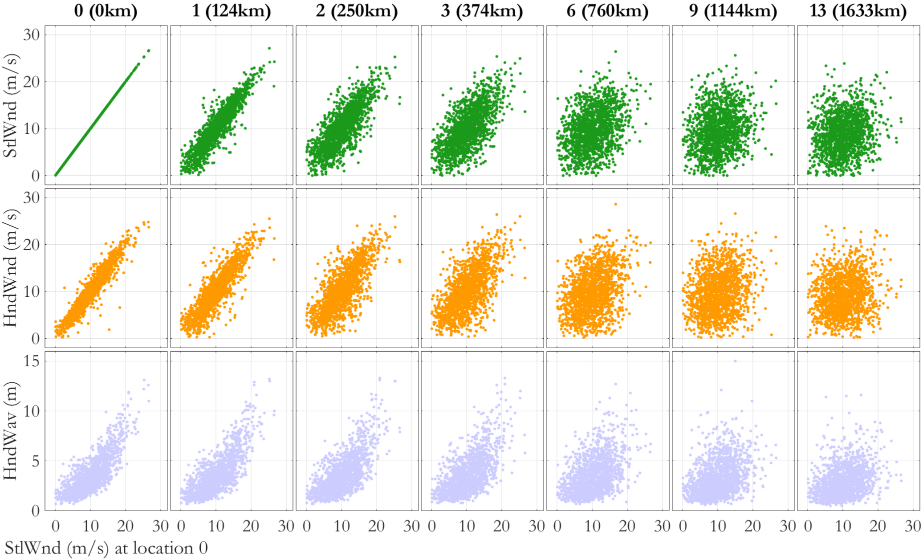
$\chi(u)$ and $\bar{\chi}(u)$ for bivariate Gaussian ($\Rightarrow \chi = 0, \bar{\chi} = \rho$)
Colours are correlations ρ on -0.9, -0.8, ..., 0.9 (Recreated from Coles et al. 1999)

Conditional extremes in practice

- Non-stationary : Jonathan et al. [2014]
- Time-series : Winter and Tawn [2016], Tendijck et al. [2019]
- Mixture model : Tendijck et al. [2021]
- Spatial : Shooter et al. [2021b], Shooter et al. [2021a]
- Lots more

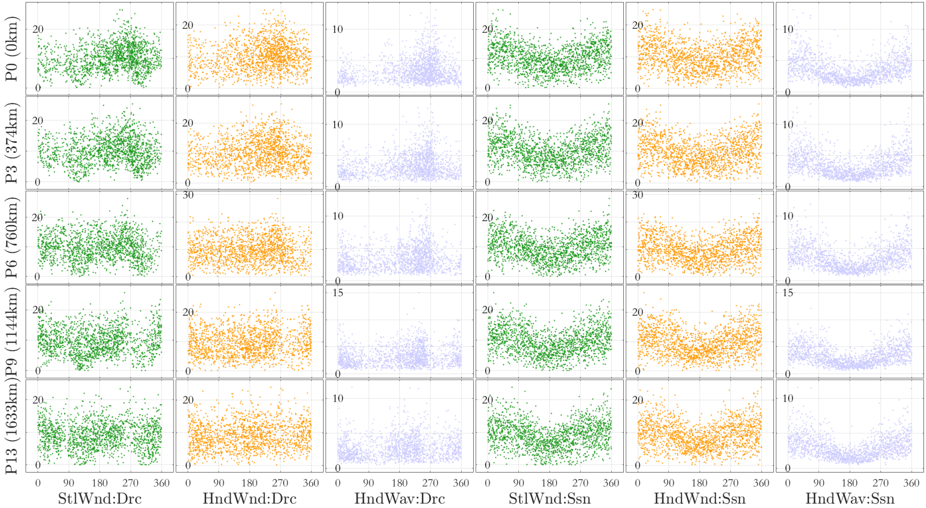
- Multivariate spatial : Shooter et al. [2022]

Scatter plots on physical scale



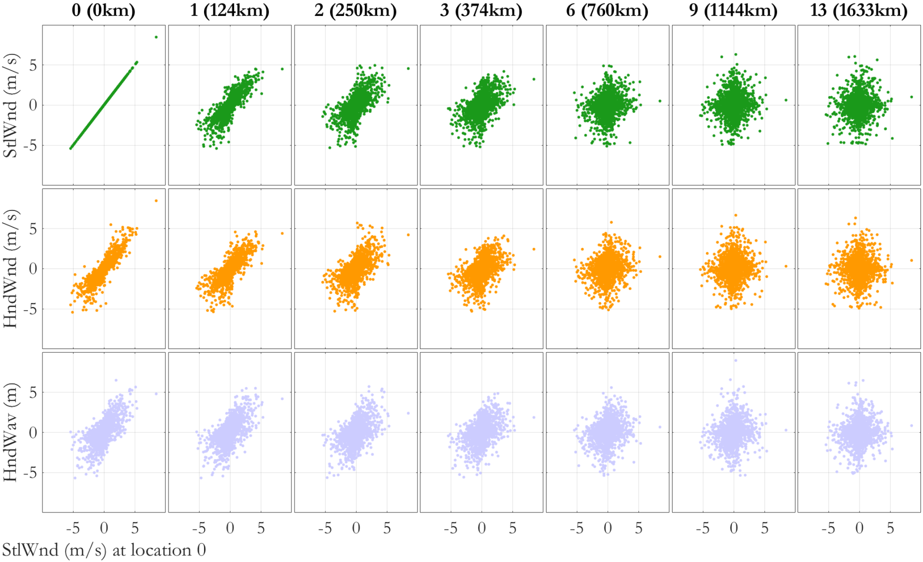
Scatter plots of registered data : StlWnd (green), HndWnd (orange), HndWav(blue)

Covariate dependence on physical scale



Directional and seasonal dependence. "Direction" is that from which fluid flows measured clockwise from North
StlWnd (green), HndWnd (orange), HndWav(blue)

Scatter plots on Laplace scale



Registered data on Laplace scale: StlWnd (green), HndWnd (orange), HndWav(blue)

Estimating return values and associated values

- Return values: Serinaldi [2015], Jonathan et al. [2021]
- Associated values : Towe et al. [2022]