

Modelling extreme ocean environments for structural design

Philip Jonathan

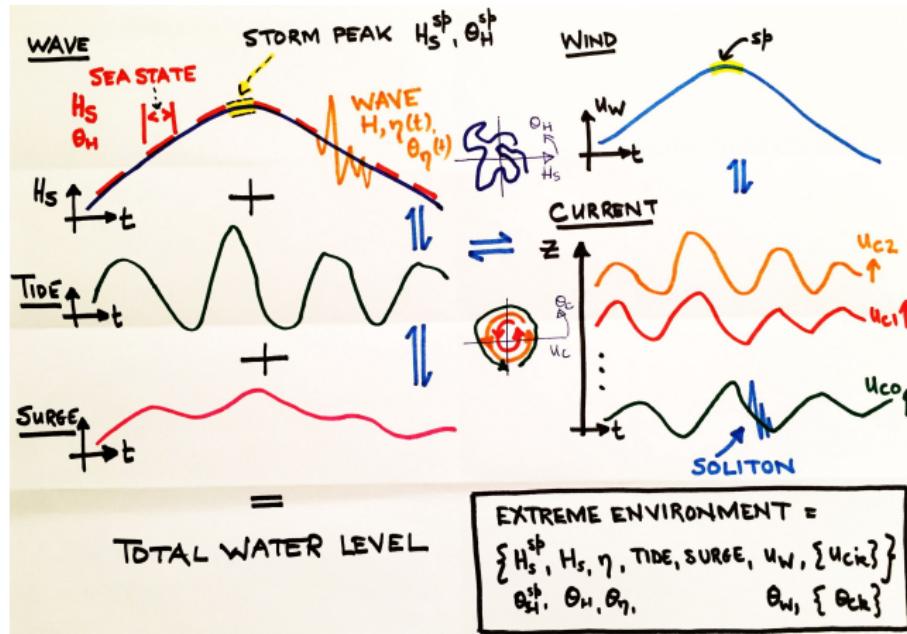
Lancaster and Shell

FORESEE, Bergen, April 2024
(Slides at www.lancs.ac.uk/~jonathan)

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

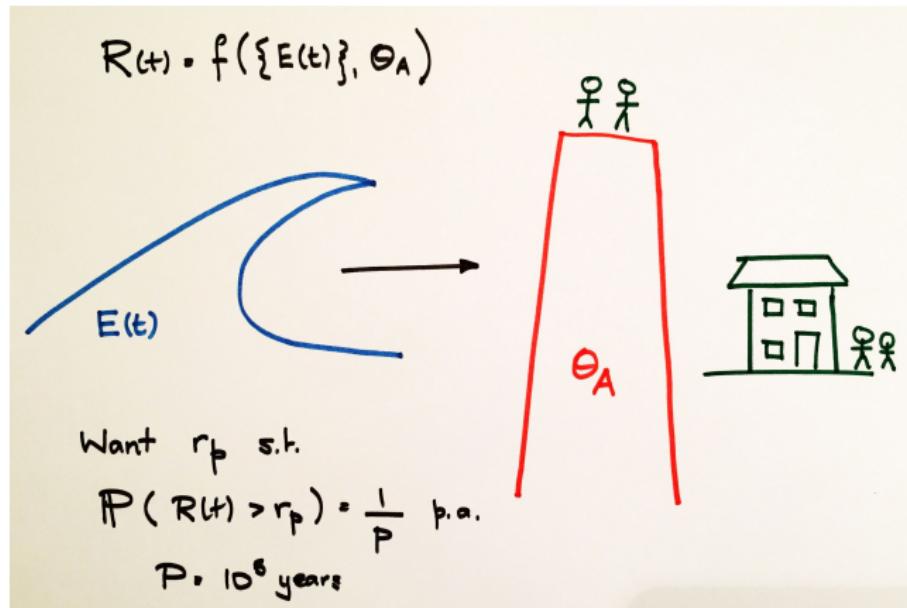


Modelling ocean storm environment



- 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

- Storm storm peak events X^{sp} dependent on covariates Θ^{sp}
- An evolving within-storm environment $\{(X_s, \Theta_s)\}_{s \in S_T}$ for storm of length T
- A structural “loading” Y
- Everything subject to sources of uncertainty Z
- $Z, \Theta^{\text{sp}}, X^{\text{sp}}, \{(X_s, \Theta_s)\}_{s \in S_T}$ and Y are **multidimensional** random variables

Unconditional distribution of loading

$$\begin{aligned}
 F_Y(y) &= \int_{\zeta} \int_{(\{(x_s, \theta_s)\}_{s \in S_T}, \tau)} \int_{X^{\text{sp}}} \int_{\Theta^{\text{sp}}} \\
 &\times F_{Y|(\{(X_s, \Theta_s)\}_{s \in S_T}, Z}(y | \{(x_s, \theta_s)\}_{s \in S_T}, \zeta) \\
 &\times f_{(\{(X_s, \Theta_s)\}_{s \in S_T}, T) | X^{\text{sp}}, \Theta^{\text{sp}}, Z} \left(\{(x_s, \theta_s)\}_{s \in S_T}, \tau | x^{\text{sp}}, \theta^{\text{sp}}, \zeta \right) \\
 &\times f_{X^{\text{sp}} | \Theta^{\text{sp}}, Z}(x^{\text{sp}} | \theta^{\text{sp}}, \zeta) \\
 &\times f_{\Theta^{\text{sp}} | Z}(\theta^{\text{sp}} | \zeta) \\
 &\times f_Z(\zeta) \\
 &\times d\theta^{\text{sp}} dx^{\text{sp}} d(\{(x_s, \theta_s)\}_{s \in S_T}, \tau) d\zeta
 \end{aligned}$$

Optimal design of marine structure

Typical

- Distribution of **annual maximum** loading
- Assumes **univariate** load

$$F_A(y) = \int_m [F_Y(y)]^m f_C(m) dm$$

- Annual rate of occurrence f_C of storms
- **Return value** for return period P years given by $F_A^{-1}(1 - 1/P)$

More generally

- Expected utility

$$\mathbb{E}(U_A|\mathcal{R}) = \int_m \int_{y_1} \dots \int_{y_m} U_A(y_1, \dots, y_m | \mathcal{R}) f_{Y_1, \dots, Y_m, M}(y_1, \dots, y_m, m) dy_1 \dots dy_m dm$$

- System annual utility $U_A(Y_1, \dots, Y_m | \mathcal{R})$ given system “strength” characteristics \mathcal{R}
- Solve for \mathcal{R} to achieve required expected annual utility

Historical approach

- Return values
- Estimators for return values from small samples
- Associated values

What is a return value?

- Random variable A represents the maximum value of some physical quantity X **per annum**
- Forget about all complicating issues like serial dependence, covariates and other sources of dependence and uncertainty
- The P -year return value x_P of X is then defined by the equation

$$F_A(x_P) = \Pr(A \leq x_P) = 1 - \frac{1}{P}$$

- Or

$$x_P = F_A^{-1}\left(1 - \frac{1}{P}\right)$$

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

An alternative definition

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

$$F_A(x_P) = 1 - \frac{1}{P}$$

$$\Rightarrow F_{A_P}(x_P) = \left(1 - \frac{1}{P}\right)^P \approx \exp(-1)$$

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

Estimating a return value

- To estimate x_P , 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_P
- 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_P 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 $F_{Y|Z}$ of random variable Y is known conditional on random variables Z , and the joint density f_Z of Z is also known, the unconditional **predictive** distribution \tilde{F}_Y can be evaluated using

$$\tilde{F}_Y(y) = \int_{\zeta} F_{Y|Z}(y|\zeta) f_Z(\zeta) d\zeta$$

- The expected value of deterministic function g of parameters Z given joint density f_Z is

$$E[g(Z)] = \int_{\zeta} g(\zeta) f_Z(\zeta) d\zeta$$

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

Different estimators of return value

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

- Various options available, including:

$$q_1 = F_{A|\mathbf{Z}}^{-1}(1 - 1/P \mid \mathbb{E}_{\mathbf{Z}}[\mathbf{Z}]) = F_{A|\mathbf{Z}}^{-1}(1 - 1/P \mid \int_{\zeta} \zeta f_{\mathbf{Z}}(\zeta) d\zeta)$$

$$q_2 = \mathbb{E}_{\mathbf{Z}}[F_{A|\mathbf{Z}}^{-1}(1 - 1/P \mid \mathbf{Z})] = \int_{\zeta} F_{A|\mathbf{Z}}^{-1}(1 - 1/P \mid \zeta) f_{\mathbf{Z}}(\zeta) d\zeta$$

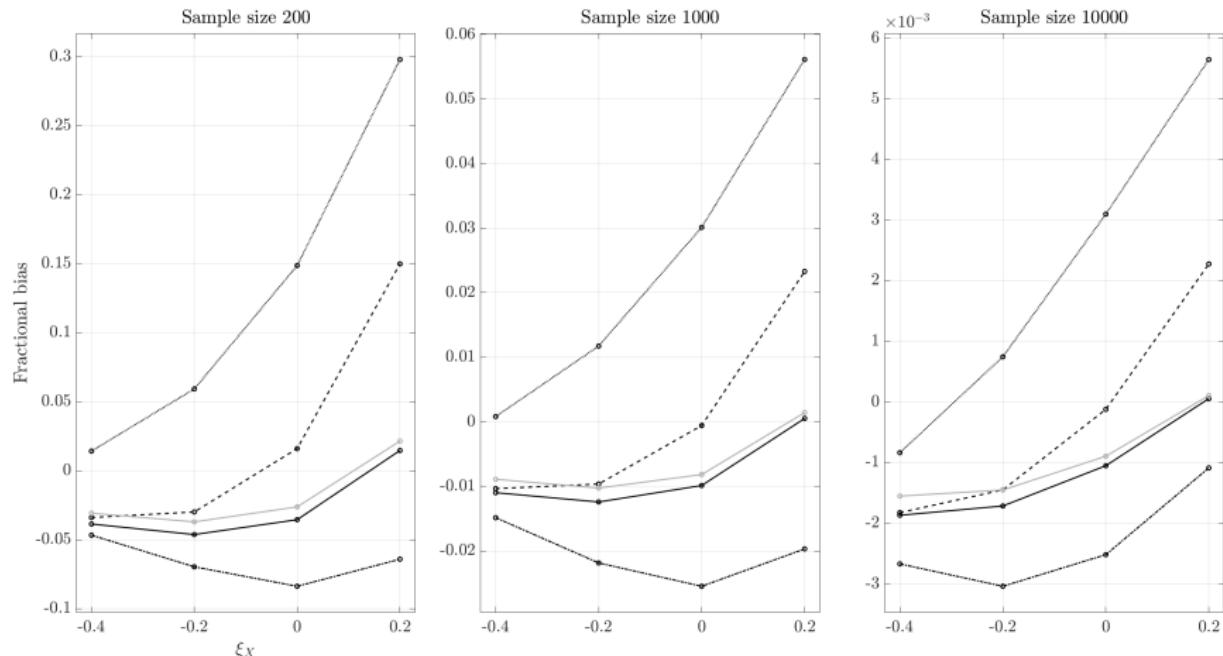
$$q_3 = \tilde{F}_A^{-1}(1 - 1/P) \text{ where } \tilde{F}_A(x) = \int_{\zeta} F_{A|\mathbf{Z}}(x \mid \zeta) f_{\mathbf{Z}}(\zeta) d\zeta$$

$$q_4 = \tilde{F}_{A_p}^{-1}(\exp(-1)) \text{ where } \tilde{F}_{A_p}(x) = \tilde{F}_A^P(x)$$

$$q_5 = \text{med}_{\mathbf{Z}}[F_{A|\mathbf{Z}}^{-1}(1 - 1/P \mid \mathbf{Z})]$$

- 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 $\mathbb{E}_Z(Y|X = q, Z)$

Estimating return values and associated values

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

Current and future approach: a full statistical framework to model the environment and fluid loading

- Model components
- Marginal modelling
- Dependence modelling

Storm peaks

Context: unconditional distribution of loading

$$\begin{aligned}
 F_Y(y) &= \int_{\zeta} \int_{(\{(x_s, \theta_s)\}_{s \in S_T}, \tau)} \int_{x^{sp}} \int_{\theta^{sp}} \\
 &\times F_{Y|(\{(x_s, \Theta_s)\}_{s \in S_T}, Z}(y | \{(x_s, \theta_s)\}_{s \in S_T}, \zeta) \\
 &\times f_{(\{(X_s, \Theta_s)\}_{s \in S_T}, T) | X^{sp}, \Theta^{sp}, Z} \left(\{(x_s, \theta_s)\}_{s \in S_T}, \tau | x^{sp}, \theta^{sp}, \zeta \right) \\
 &\times f_{X^{sp} | \Theta^{sp}, Z}(x^{sp} | \theta^{sp}, \zeta) \\
 &\times f_{\Theta^{sp} | Z}(\theta^{sp} | \zeta) \\
 &\times f_Z(\zeta) \\
 &\times d\theta^{sp} dx^{sp} d(\{(x_s, \theta_s)\}_{s \in S_T}, \tau) d\zeta
 \end{aligned}$$

Storm peaks: modelling margins and dependence

$$\begin{aligned}
 f_{X^{sp} | \Theta^{sp}, Z}(x^{sp} | \theta^{sp}, \zeta) &= \left[\prod_{j=1}^p f_{X_j^{sp} | \Theta^{sp}, Z}(x_j^{sp} | \theta^{sp}, \zeta) \right] \\
 &\times c(F_{X_1^{sp} | \Theta^{sp}, Z}(x_1^{sp} | \theta^{sp}, \zeta), \dots, F_{X_p^{sp} | \Theta^{sp}, Z}(x_p^{sp} | \theta^{sp}, \zeta) | \theta^{sp}, \zeta)
 \end{aligned}$$

More to come in a minute!

Within-storm evolution

Context: unconditional distribution of loading

$$\begin{aligned}
 F_Y(\mathbf{y}) &= \int_{\zeta} \int_{(\{\mathbf{(x}_s, \theta_s)\}_{s \in S_T}, \tau)} \int_{\mathbf{x}^{\text{sp}}} \int_{\boldsymbol{\theta}^{\text{sp}}} \\
 &\times F_{Y|(\{\mathbf{(x}_s, \theta_s)\}_{s \in S_T}, Z}(\mathbf{y} | \{\mathbf{(x}_s, \theta_s)\}_{s \in S_T}, \zeta) \\
 &\times f_{(\{\mathbf{(x}_s, \Theta_s)\}_{s \in S_T}, T) | \mathbf{x}^{\text{sp}}, \boldsymbol{\Theta}^{\text{sp}}, Z} \left(\{\mathbf{(x}_s, \theta_s)\}_{s \in S_T}, \tau | \mathbf{x}^{\text{sp}}, \boldsymbol{\theta}^{\text{sp}}, \zeta \right) \\
 &\times f_{\mathbf{x}^{\text{sp}} | \boldsymbol{\Theta}^{\text{sp}}, Z}(\mathbf{x}^{\text{sp}} | \boldsymbol{\theta}^{\text{sp}}, \zeta) \\
 &\times f_{\boldsymbol{\Theta}^{\text{sp}} | Z}(\boldsymbol{\theta}^{\text{sp}} | \zeta) \\
 &\times f_Z(\zeta) \\
 &\times d\boldsymbol{\theta}^{\text{sp}} d\mathbf{x}^{\text{sp}} d(\{\mathbf{(x}_s, \theta_s)\}_{s \in S_T}, \tau) d\zeta
 \end{aligned}$$

Models for within-storm evolution

- History matching: Feld et al. [2019], Hansen et al. [2020]
- Extreme value time-series model (and extension of conditional extremes): Tendijck et al. [2019], Tendijck et al. [2023]

Fluid loading

Context: unconditional distribution of loading

$$\begin{aligned}
 F_Y(y) &= \int_{\zeta} \int_{(\{(x_s, \theta_s)\}_{s \in S_T}, \tau)} \int_{x^{sp}} \int_{\theta^{sp}} \\
 &\times F_{Y|(\{x_s, \Theta_s\}_{s \in S_T}, Z)}(y | \{(\mathbf{x}_s, \boldsymbol{\theta}_s)\}_{s \in S_T}, \zeta) \\
 &\times f_{(\{(\mathbf{x}_s, \boldsymbol{\Theta}_s)\}_{s \in S_T}, T) | \mathbf{x}^{sp}, \boldsymbol{\Theta}^{sp}, Z}\left(\{(\mathbf{x}_s, \boldsymbol{\theta}_s)\}_{s \in S_T}, \tau | \mathbf{x}^{sp}, \boldsymbol{\theta}^{sp}, \zeta\right) \\
 &\times f_{X^{sp} | \boldsymbol{\Theta}^{sp}, Z}(x^{sp} | \boldsymbol{\theta}^{sp}, \zeta) \\
 &\times f_{\boldsymbol{\Theta}^{sp} | Z}(\boldsymbol{\theta}^{sp} | \zeta) \\
 &\times f_Z(\zeta) \\
 &\times d\boldsymbol{\theta}^{sp} dx^{sp} d(\{(\mathbf{x}_s, \boldsymbol{\theta}_s)\}_{s \in S_T}, \tau) d\zeta
 \end{aligned}$$

Models for fluid loading

- Incorporate kinematics, estimate Morison loads (e.g. LOADS, AWARE JIPs): Swan [2020], Gibson [2020]
- Interface environment and fluid loading software for full “forward model”
- Fundamentals paper: Speers et al. [2024]

Separating margins and dependence

Storm peaks: modelling margins and dependence

$$\begin{aligned} f_{\mathbf{X}^{\text{sp}} | \Theta^{\text{sp}}, \mathbf{Z}}(\mathbf{x}^{\text{sp}} | \theta^{\text{sp}}, \zeta) &= \left[\prod_{j=1}^p f_{X_j^{\text{sp}} | \Theta^{\text{sp}}, \mathbf{Z}}(x_j^{\text{sp}} | \theta^{\text{sp}}, \zeta) \right] \\ &\times c(F_{X_1^{\text{sp}} | \Theta^{\text{sp}}, \mathbf{Z}}(x_1^{\text{sp}} | \theta^{\text{sp}}, \zeta), \dots, F_{X_p^{\text{sp}} | \Theta^{\text{sp}}, \mathbf{Z}}(x_p^{\text{sp}} | \theta^{\text{sp}}, \zeta) | \theta^{\text{sp}}, \zeta) \end{aligned}$$

More to come on dependence later!

Generalised Pareto distribution

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

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

Theory

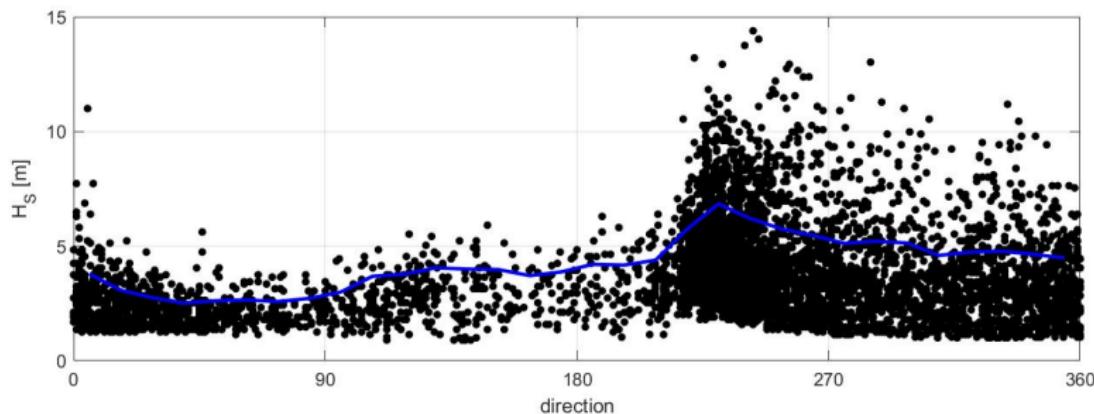
- Derived from **max-stability** of F_X
- Threshold-stability property
- "Poisson \times GP = GEV"

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

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

- 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}_θ
- Each observation belongs to exactly one θ_s
- On \mathcal{I}_θ , assume

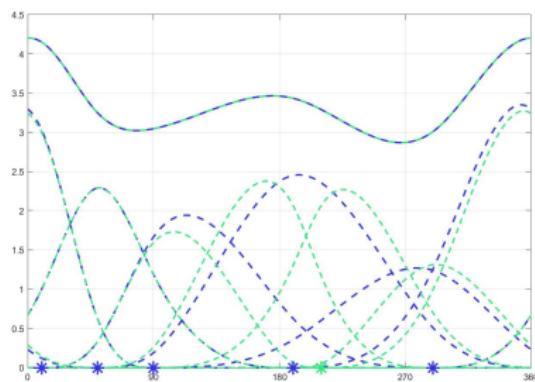
$$\begin{aligned}\eta_s &= \sum_{k=1}^n B_{sk} \beta_k, s = 1, 2, \dots, m, \text{ or} \\ \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\beta}\end{aligned}$$

- $\eta \in (\xi, \nu)$ (and similar for ρ)
- $\mathbf{B} = \{B_{sk}\}_{s=1;k=1}^{m;n}$ basis for \mathcal{D}_θ
- $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^n$ basis coefficients
- Inference reduces to estimating $n_\xi, n_\nu, \mathbf{B}_\xi, \mathbf{B}_\nu, \boldsymbol{\beta}_\xi, \boldsymbol{\beta}_\nu$ (and roughnesses λ_ξ, λ_ν)
- P-splines, BARS and Voronoi are different forms of \mathbf{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}_θ
- B consists of n B-spline bases
- Order d
- Each using $d + 1$ consecutive knot locations
- Local support
- Wrapped on \mathcal{D}_θ
- Knot locations $\{r_k\}_{k=1}^n$ vary
- Number of basis functions n varies



Periodic BARS knot birth and death

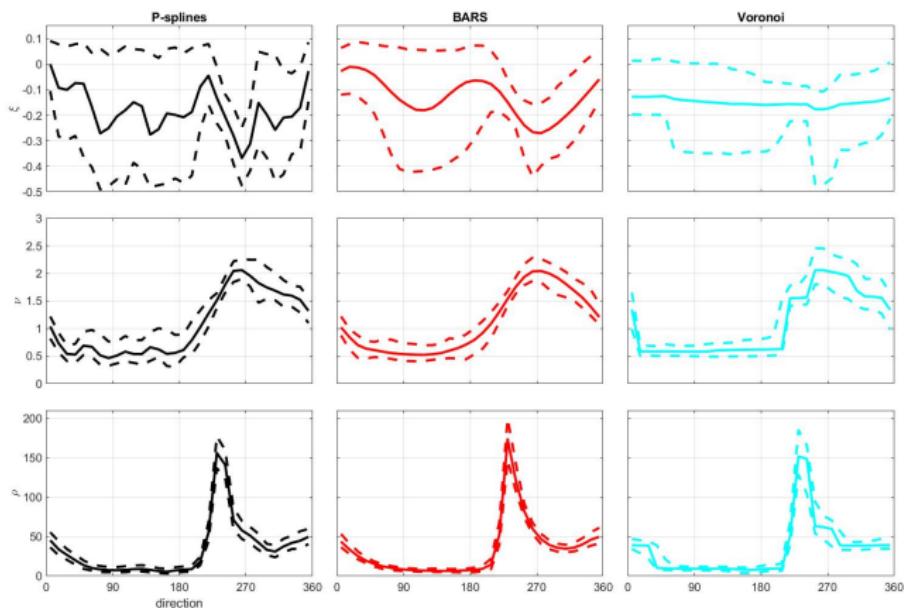
P-splines and Voronoi partition

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

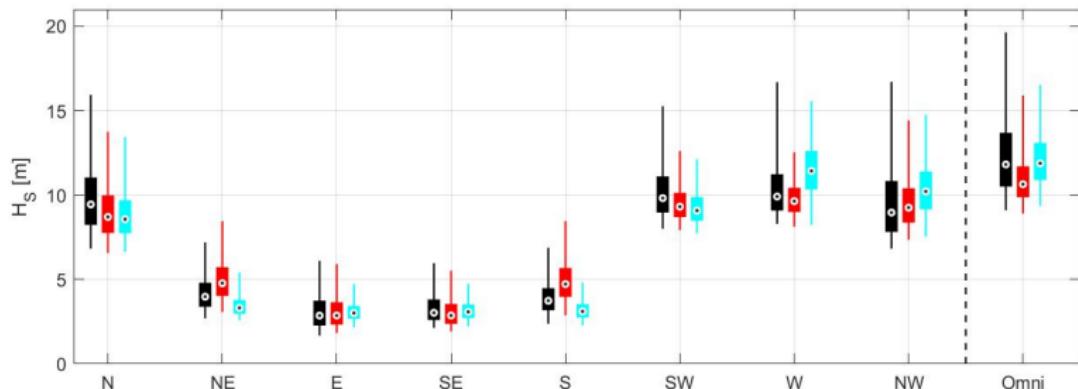
Posterior parameter estimates for ξ , ν and ρ for northern North Sea

- 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

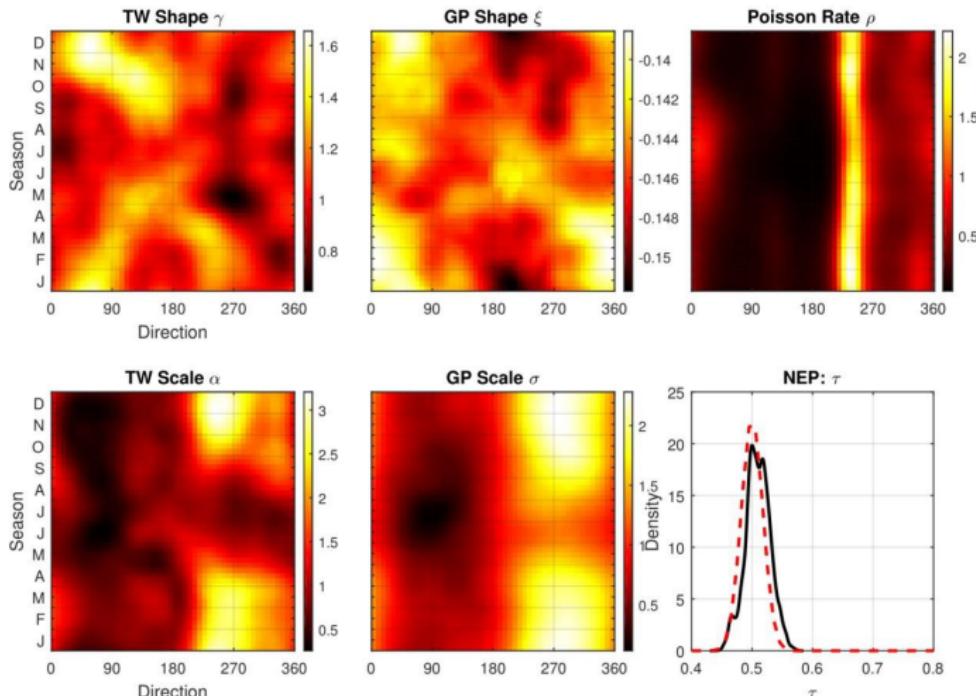


Directional posterior predictive distribution of $P = 1000$ -year maximum



- 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)

Marginal extremes

- Theory : Embrechts et al. [2003], Beirlant et al. [2004]
- Method : Coles [2001], Dey and Yan [2016]
- Motivation : Davison and Smith [1990]
- Covariate effects : Wood [2003], Chavez-Demoulin and Davison [2005], Brezger and Lang [2006], Youngman [2022]
- Practicalities : Jonathan and Ewans [2013], Feld et al. [2019], Vanem et al. [2022]
- Applied covariates : Randell et al. [2016], Zanini et al. [2020]

Conditional extremes

- Recap
- Max-stability, AD and AI
- Conditional extremes basic
- Simple application
- Multivariate spatial conditional extremes
- Time-series extremes
- Summary

Modelling margins and dependence

Context

$$F_{X^{\text{sp}}|\Theta^{\text{sp}}, Z}(x^{\text{sp}}|\theta^{\text{sp}}, \zeta) = C(F_{X_1^{\text{sp}}|\Theta^{\text{sp}}, Z}(x_1^{\text{sp}}|\theta^{\text{sp}}, \zeta), \dots, F_{X_p^{\text{sp}}|\Theta^{\text{sp}}, Z}(x_p^{\text{sp}}|\theta^{\text{sp}}, \zeta)|\theta^{\text{sp}}, \zeta)$$

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

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**: $C(\mathbf{u}) = C^k(\mathbf{u}^{1/k})$
- On Fréchet margins ($G_j(z) = \exp(-z^{-1})$), $G(z) = \exp(-V(z))$, for **exponent measure** V such that $V(rz) = r^{-1}V(z)$, 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

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

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

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

- Non-degenerate G is unknown, and estimated empirically

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

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

- $\alpha = 1, \beta = 0$: 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” $X_1, X_2, \dots, X_\tau |(X_0 = x_0), x_0 > u$
- Spatial: “spatial conditional extremes” $X_1, X_2, \dots, X_s |(X_0 = x_0), x_0 > u$

Idea

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

- Impose appropriate structure on parameters α, β and distribution of \mathbf{Z}
 - e.g. α evolves smoothly in space
 - e.g. \mathbf{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)

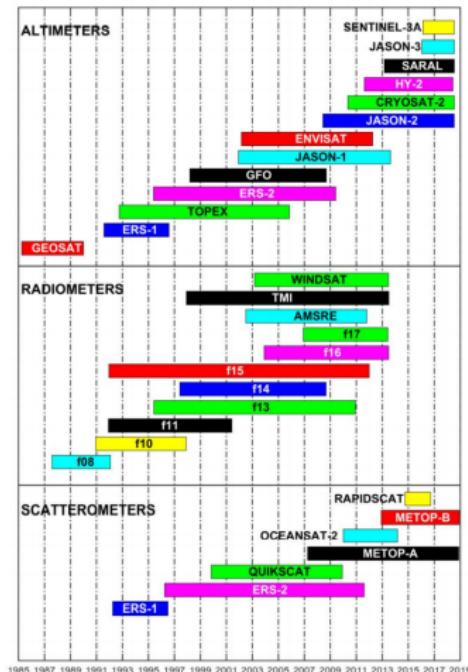
Motivation

- 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]

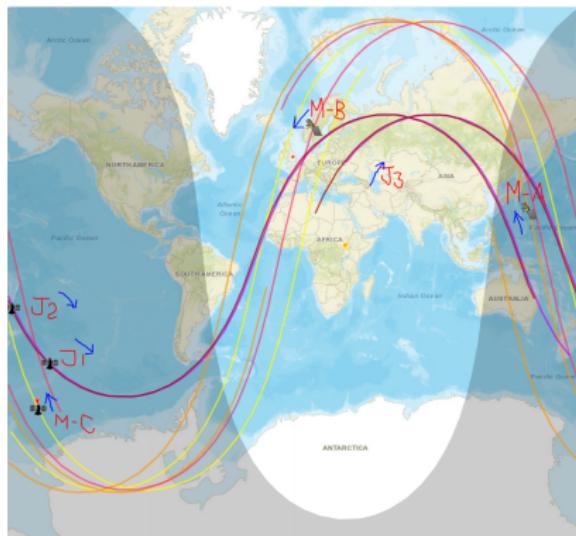
Features

- 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

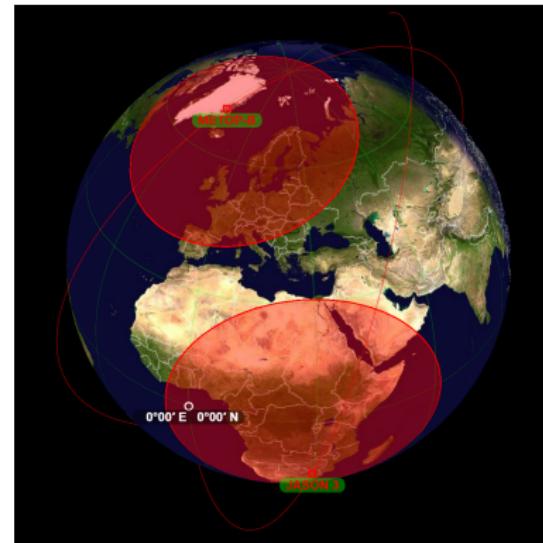
H_S : significant wave height (m)

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

JASON and METOP



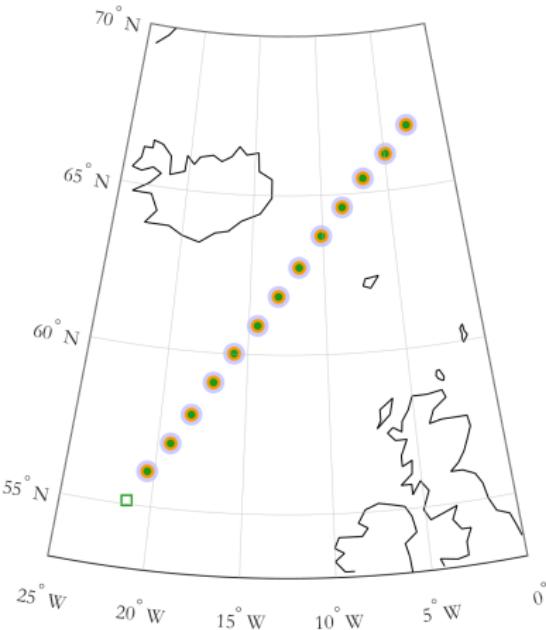
[n2yo.com, accessed 06.09.21 at around 1100UK]



[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)

$$X_{jk} \sim Lpl$$

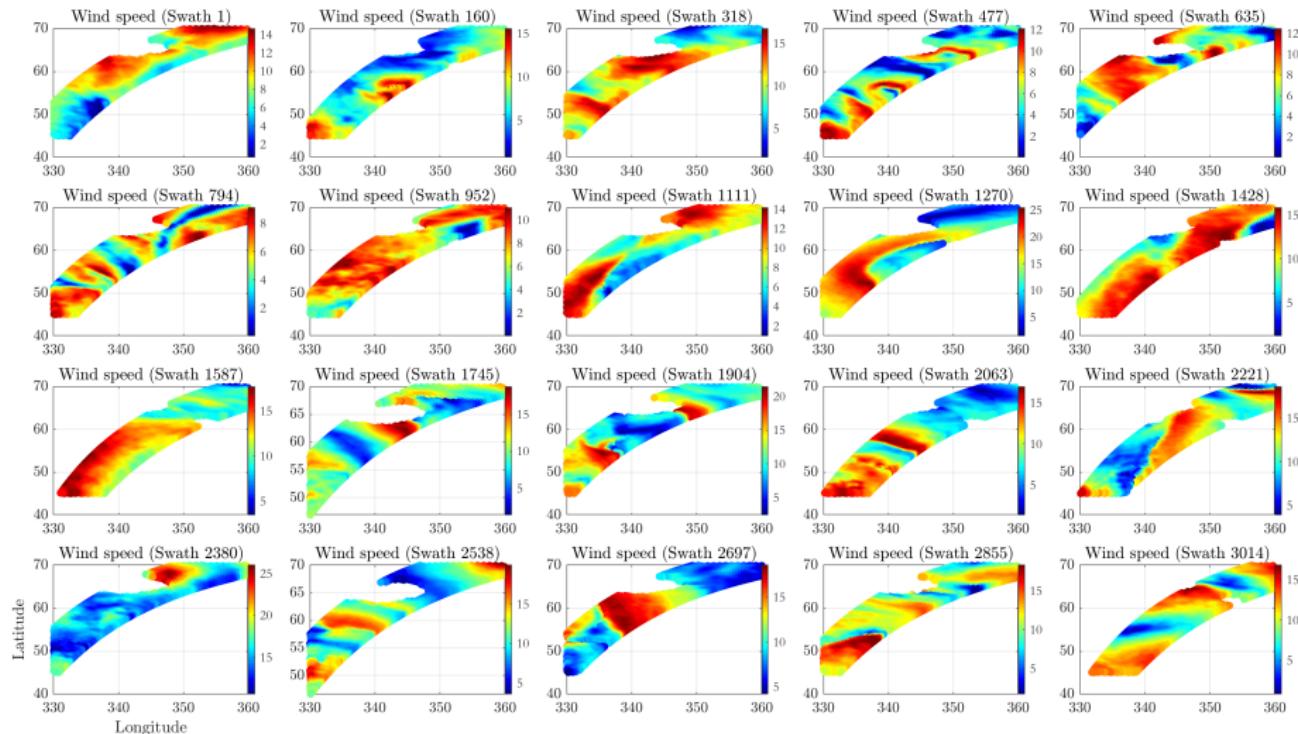
$$x > u$$

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

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

- MCMC to estimate $\alpha, \beta, \mu, \sigma, \delta$ and ρ, κ, λ
- $\alpha, \beta, \mu, \sigma, \delta$ spatially smooth for each quantity
- DL = delta-Laplace = generalised Gaussian
- Residual correlation Σ 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

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

- Delta-Laplace residual margins

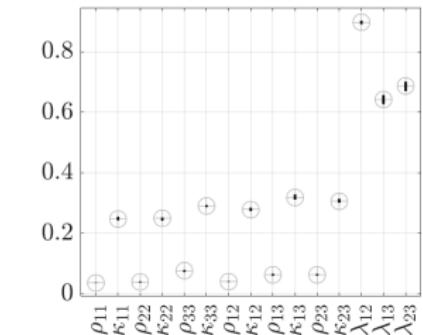
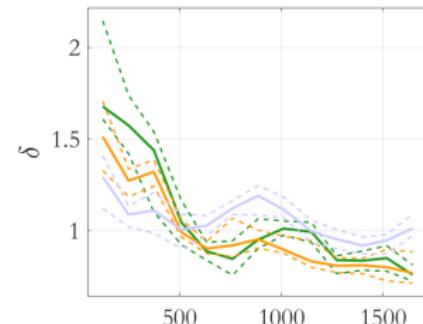
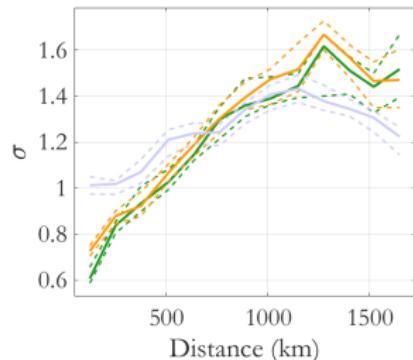
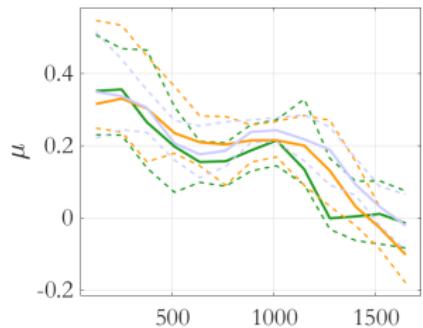
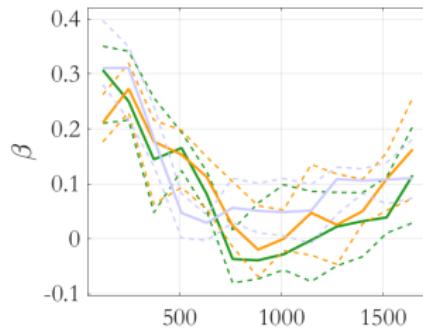
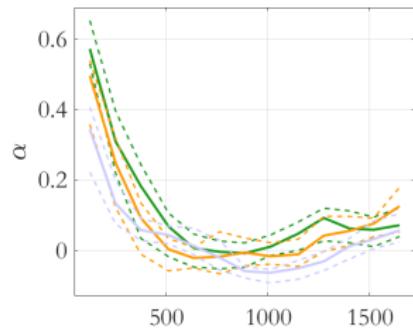
$$f_{Z_{j,k}}(z_{j,k}) = \frac{\delta_{j,k}}{2\kappa_{j,k}\sigma_{j,k}\Gamma\left(\frac{1}{\delta_{j,k}}\right)} \exp\left\{-\left|\frac{z - \mu_{j,k}}{\kappa_{j,k}\sigma_{j,k}}\right|^{\delta_{j,k}}\right\}, \quad \kappa_{j,k}^2 = \Gamma\left(1/\delta_{j,k}\right)/\Gamma\left(3/\delta_{j,k}\right)$$

- Gaussian residual dependence

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

- Piecewise linear forms for $\alpha, \beta, \mu, \sigma, \delta$ 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 $\times 100$ km), κ =exponent (need to $\times 5$), λ =cross-correlation

Conditional extremes in practice

- Non-stationary : Jonathan et al. [2014]
- Time-series : Winter and Tawn [2016], Tendijck et al. [2019], Tendijck et al. [2023]
- Mixture model : Tendijck et al. [2021]
- Spatial : Shooter et al. [2021b], Shooter et al. [2021a]
- Lots more

- **Multivariate spatial** : Shooter et al. [2022]

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 [2022]
- Limit sets and **SPAR** : Nolde and Wadsworth [2022], Simpson and Tawn [2024], Murphy-Barltrop et al. [2024]
- Metocean : Parametric conditional models (e.g. Haver 1987), design contours (e.g. Haselsteiner et al. 2021).

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
 - Limit sets (Simpson and Tawn 2024), SPAR (Murphy-Barltrop et al. 2024)
- More demanding regulatory framework

Thanks for listening / Diolch am wrando!

Pragmatic non-stationary multivariate extremes with UQ

- Fit generalised Pareto marginal models for peaks over threshold data \hat{X} and \hat{Y}
 - Physics-based identification of peaks from time-series
 - Multiple thresholds, simple piecewise constant model for covariates Θ
 - Diagnostics: threshold stability
- Transform to standard Laplace scale X and Y
 - Transform full sample
- Fit conditional extremes model $X|(Y = y)$ 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 covXtreme 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://1fenergy.org/projects/covXtreme/>, Towe et al. [2024]

References I

- J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes: theory and applications*. Wiley, Chichester, UK, 2004.
- A. Brezger and S. Lang. Generalized structured additive regression based on Bayesian P-splines. *Comput. Statist. Data Anal.*, 50:967–991, 2006.
- V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Appl. Stat.*, 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.
- R. Gibson. Extreme Environmental Loading of Fixed Offshore Structures: Summary Report, Component 2.
<https://www.hse.gov.uk/offshore/assets/docs/summary-report-component2.pdf>, 2020.
- H. F. Hansen, D. Randell, A. R. Zeeberg, and P. Jonathan. Directional-seasonal extreme value analysis of North Sea storm conditions. *Ocean Eng.*, 195:106665, 2020.
- Andreas F. Haselsteiner, Ryan G. Coe, Lance Manuel, Wei Chai, Bernt Leira, Guilherme Clarindo, C. Guedes Soares, Ásta Hannesdóttir, Nikolay Dimitrov, Aljoscha Sander, Jan Hendrik Ohlendorf, Klaus Dieter Thoben, Guillaume de Hauteclercque, Ed Mackay, Philip Jonathan, Chi Qiao, Andrew Myers, Anna Rode, Arndt Hildebrandt, Boso Schmidt, Erik Vanem, and Arne Bang Huseby. A benchmarking exercise for environmental contours. *Ocean Engineering*, 236:109504, 2021.
- S. Haver. On the joint distribution of heights and periods of sea waves. *Ocean Eng.*, 14:359–376, 1987.
- 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 modeling of spatial extremes. *WIREs Computational Statistics*, 14:e1537, 2022.
- H. Joe. *Dependence modelling with copulas*. CRC Press, 2014.
- P. Jonathan and K. 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.

References II

- A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. R. Statist. Soc. B*, 59:475–499, 1997.
- C. J. R. Murphy-Barltrop, E. Mackay, and P. Jonathan. Estimating metocean environments associated with extreme structural response. *Submitted to Extremes; pre-print at www.lancaster.ac.uk/jonathan*, 2024.
- N. Nolde and J. L. Wadsworth. Linking representations for multivariate extremes via a limit set. *Adv. Appl. Probab.*, 54:688?717, 2022.
- 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.
- F. 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.
- E. S. Simpson and J. A. Tawn. Inference for new environmental contours using extreme value analysis. *J. Agric. Biol. Environ. Stat.*, 2024. doi: 10.1007/s13253-024-00612-2.
- M. Speers, D. Randell, J. A. Tawn, and P. Jonathan. Estimating metocean environments associated with extreme structural response. *Submitted to Ocean Engineering; pre-print at www.lancaster.ac.uk/jonathan*, 2024.
- C. Swan. Extreme Environmental Loading of Fixed Offshore Structures: Summary Report, Component 1. <https://www.hse.gov.uk/offshore/assets/docs/summary-report-component1.pdf>, 2020.
- 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.
- S. Tendijck, P. Jonathan, D. Randell, and J. Tawn. Temporal evolution of the extreme excursions of multivariate kth order Markov processes with application to oceanographic data. *Accepted by Environmetrics; pre-print at https://arxiv.org/abs/2302.14501*, 2023.
- 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.

References III

- R. Towe, E. Ross, D. Randell, and P. Jonathan. covXtreme: MATLAB software for non-stationary penalised piecewise constant marginal and conditional extreme value models. *Accepted by Environmental Modelling and Software, April 2024; pre-print at www.lancaster.ac.uk/jonathan*, 2024.
- E Vanem, T Zhu, and A Babanin. Statistical modelling of the ocean environment ? a review of recent developments in theory and applications. *Marine Structures*, 86:103297, 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.
- B. Youngman. evgam: generalised additive extreme value models. <https://cran.r-project.org/package=evgam>, 2022.
- E. Zanini, E. Eastoe, M. Jones, D. Randell, and P. Jonathan. Covariate representations for non-stationary extremes. *Environmetrics*, 31:e2624, 2020.

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

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

- 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 extreme value distribution, MEVD

- $\mathbf{X}_i = (X_{i1}, \dots, X_{ij}, \dots, X_{ip})$, $i = 1, \dots, n$ iid p -vectors, distribution F
- $M_{n,j} = \max_i X_{ij}$, component-wise maximum
- The component-wise maximum is not “observed” (especially as $n \rightarrow \infty$)
- Then for $Z_{n,j} = (M_{n,j} - b_{n,j})/a_{n,j}$, normalised with scaling constants:

$$\mathbb{P}(Z \leq z) = F^n(a_n z + b_n) \rightarrow G(z) \quad \text{as } n \rightarrow \infty$$

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

$$G^k(\alpha_k z + \beta_k) = G(z)$$

- We say $F \in D(G)$
- Margins G_1, \dots, G_p are unique GEV, but $G(z)$ is **not unique**

MEVD on common margins

- On standard Fréchet margins with pseudo-polars (r, w)

$$G(\mathbf{z}) = \exp(-V(\mathbf{z}))$$

with $V(\mathbf{z}) = \int_{\Delta} \max_j \left\{ \frac{w_j}{z_j} \right\} S(d\mathbf{w}), \quad \text{on } \Delta = \{\mathbf{w} \in \mathbb{R}^p : \|\mathbf{w}\| = 1\}$

and $1 = \int_{\Delta} w_j S(d\mathbf{w}), \quad \forall j, \text{ for angular measure } S$

- Condition of **multivariate regular variation**, MRV

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

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

- Lots more

Asymptotic dependence ... admitted by MEVD

- On uniform margins

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

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

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

- $\theta = 2 - \chi$
- χ and θ describe AD
- MEVD admits AD

Asymptotic independence ... not admitted by MEVD

- On uniform margins

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

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

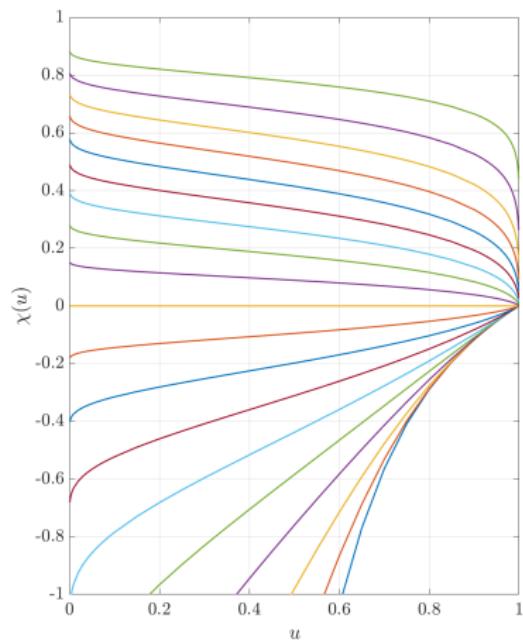
$$\frac{\mathbb{P}(Z_1 > z, Z_2 > z)}{(\mathbb{P}(Z_1 > z))^{1/\eta}} = \mathcal{L}(z)$$

where \mathcal{L} is slowly varying : $\mathcal{L}(xz)/\mathcal{L}(z) \rightarrow 1$ as $z \rightarrow \infty$

- $\bar{\chi} = 2\eta - 1$
- Idea : use non-extreme value copulas or inverted EV copulas
- Also $\mathbb{P}(Z_2 > z | Z_1 > z) \approx Cz^{1-1/\eta}$ from above
- Idea : assume a max-stable-like normalisation for conditional extremes

Extremal dependence (bivariate Gaussian)

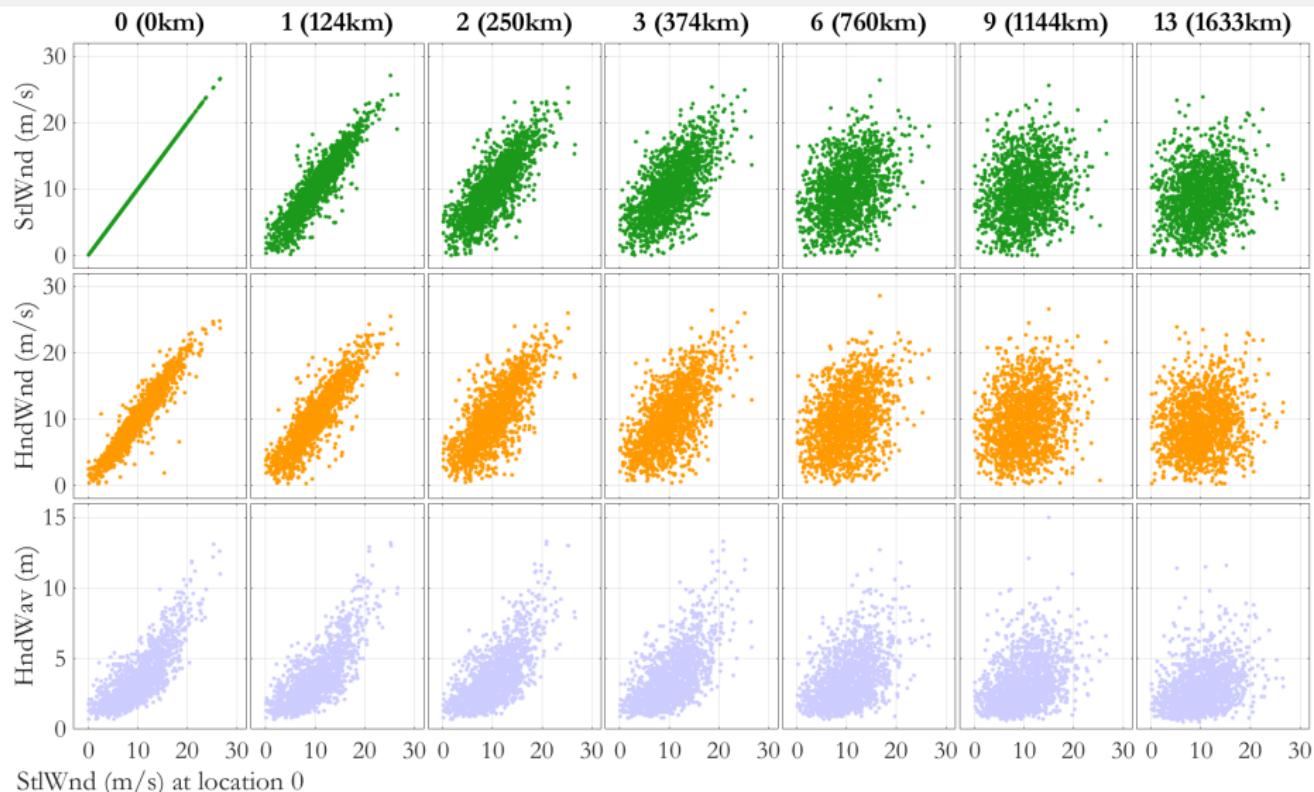
- Many (almost all?) environmental extremes problems involve asymptotic independence, at least in part ... bivariate Gaussian is one example!



$\chi(u)$ and $\bar{\chi}(u)$ for bivariate Gaussian ($\Rightarrow \chi = 0, \bar{\chi} = \rho$)

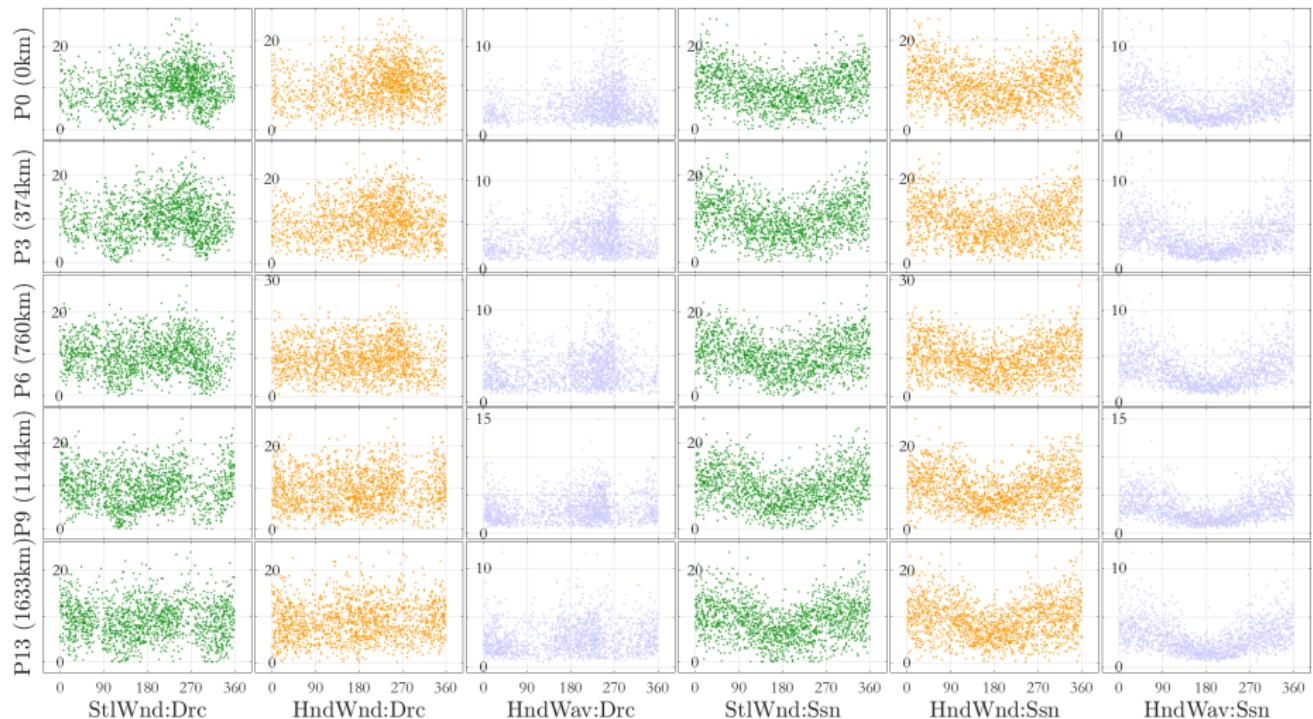
Colours are correlations ρ on -0.9, -0.8, ..., 0.9 (Recreated from Coles et al. 1999)

Scatter plots on physical scale



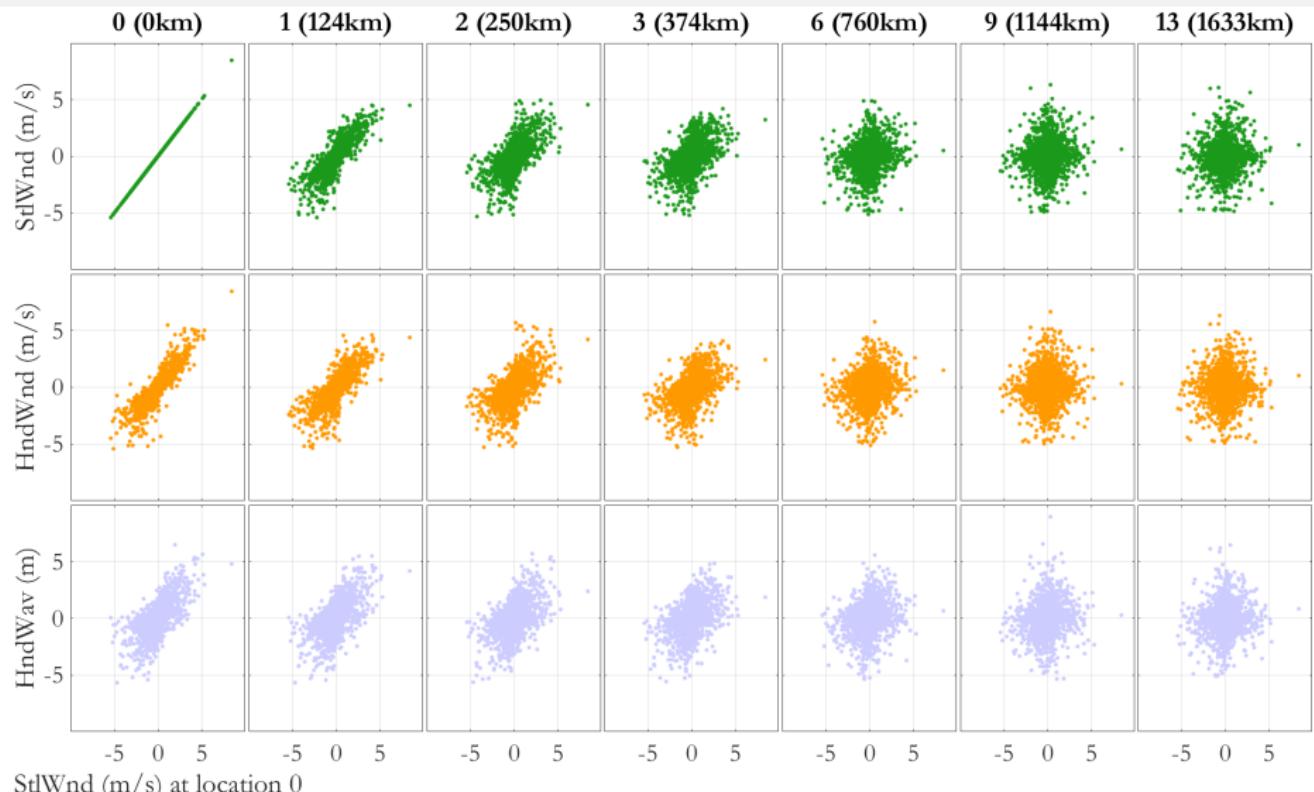
Scatter plots of registered data : StlWnd (green), HndWnd (orange), HndWav(blue)

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: StdWnd (green), HndWnd (orange), HndWav(blue)