Environmental decision support:
rare events, monitoring and inversion, and uncertainty quantification

Slides at *www.lancs.ac.uk/∼jonathan*

Philip Jonathan & Matthew Jones

Acknowledgement and overview

Acknowledgement

- Shell: Graham Feld, Rutger Ijzermans, Bas v.d. Kerkhof, David Randell, Emma Ross, Ross Towe
- Lancaster
- Cambridge
- Durham, Melbourne, UK MetOffice

Overview

- This is a talk about environmental applications
- Extremes (non-stationary marginal, multivariate spatial conditional extremes)
- Remote sensing (airborne, line-of-sight, satellite)

Extremes

Non-stationary marginal extreme value analysis

- Environmental extremes of Y vary continuously with multidimensional covariates Ω
- Asymptotic theory gives form of distribution of exceedances of high threshold ψ

$$Y|(\Omega, Y > \psi) \sim GP(\xi, \sigma, \psi), \text{ generalised Pareto with parameters } \xi, \sigma, \psi$$

- Inferences should reflect sources of uncertainty fairly
- Need statistical and computational efficiency

- **Predict extreme quantiles** of Y
- **Assess risk** (or expected loss $\mathbb{E}(L)$) for system $S = s_0$ due to Y and structural response R

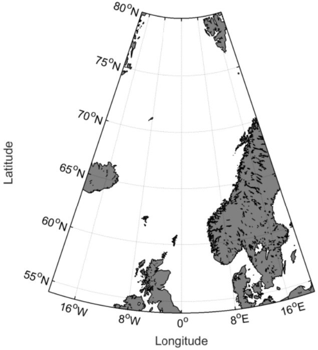
$$\mathbb{E}(L|S = s_0) = \int_r \int_y \int_\omega L(r|S = s_0) f_{R|Y}(r|y) f_{Y|\Omega}(y|\omega) f_\Omega(\omega) d\omega dy dr$$

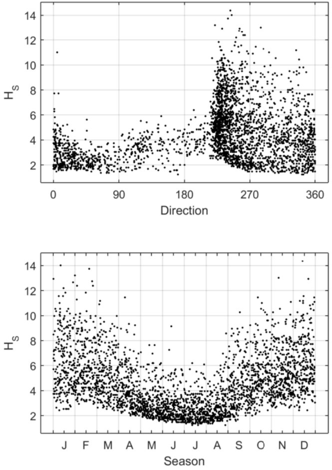
- Use cases: Offshore and coastal design, weather windows and alerts

- Jones et al. [2018], Hansen et al. [2020] ,Towe et al. [2021]

Directional-seasonal: Application

Storm peak significant wave height at northern North Sea location; clear directional and seasonal variability in storm severity; directional variability more dramatic at around 225°; seasonal variability more gradual.





Directional-seasonal: The model

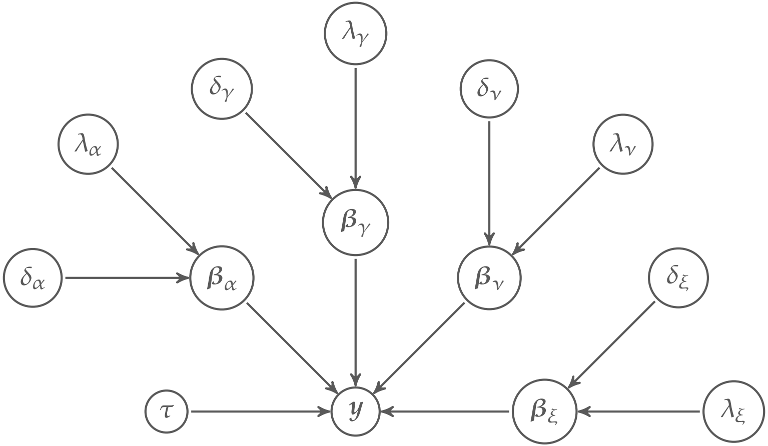
Density

- $f(y|\xi,\sigma,\alpha,\gamma,\psi,\tau) = \begin{cases} \tau \times f_{TW}(y|\alpha,\gamma) & \text{for } y \le \psi, \text{ a truncated Weibull (or similar)} \\ (1-\tau) \times f_{GP}(y|\xi,\sigma) & \text{for } y > \psi \end{cases}$
- Threshold non-exceedance probability τ to be inferred
- Physics suggests parameters $\alpha, \beta, \rho, \xi, \sigma, \psi$ and τ vary smoothly with covariates θ, ϕ
- Randell et al. [2016]

Covariate representation: B-splines

- Values of $\eta \in \{\alpha, \beta, \rho, \xi, \sigma, \psi, \tau\}$ on some index set of covariates take the form $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\beta}_\eta$
- $\boldsymbol{B}$ takes the form $\boldsymbol{B}_\phi \otimes \boldsymbol{B}_\theta$, GLAMs provide efficient manipulation
- Spline roughness penalty is **quadratic form** $\boldsymbol{\beta}_\eta' \boldsymbol{P}_\eta \boldsymbol{\beta}_\eta$ $\Rightarrow$ movitates prior for $\boldsymbol{\beta}_\eta$
- $\boldsymbol{P}_\eta = \lambda_{\eta\theta}\boldsymbol{P}_{\eta\theta} + \lambda_{\eta\phi}\boldsymbol{P}_{\eta\phi}$, includes **stochastic roughness penalties** $\{\delta_{\eta\theta}, \delta_{\eta\phi}\}$
- Brezger and Lang [2006], Currie et al. [2006], Eilers and Marx [2010], Zanini et al. [2020]

Directional-seasonal: DAG for size of threshold exceedances



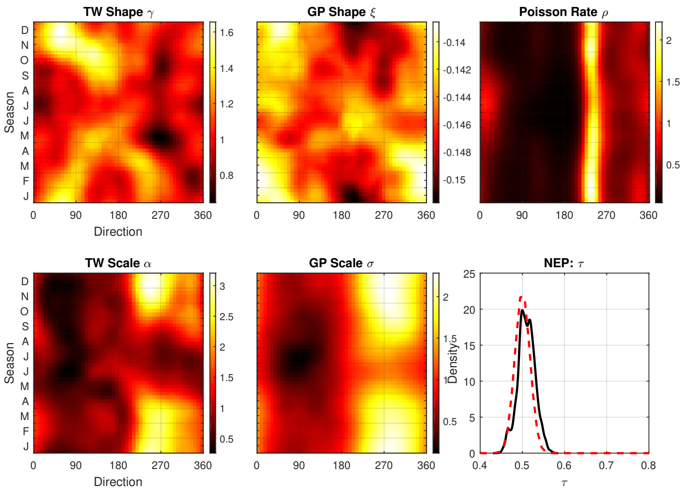
Directional-seasonal: Inference

- Sampling from full conditionals
- Gibbs sampling when full conditionals available in closed form
- Metropolis-Hastings (MH) within Gibbs otherwise, using suitable proposal mechanisms (mMALA)

- Roberts and Stramer [2002], Girolami and Calderhead [2011], Xifara et al. [2014]

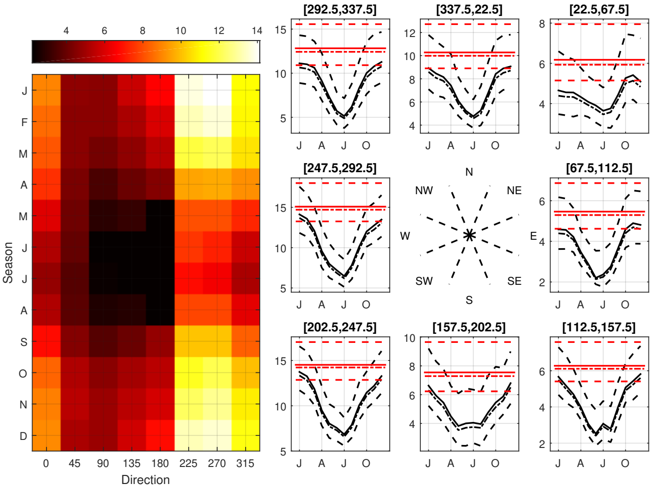
Directional-seasonal: Parameter estimates

Prior for τ in red, posterior in black; all parameters except ξ and τ suggest strong directional variation; seasonal variation less pronounced but clear for α, σ (and ρ); ξ effectively constant; sample not informative about τ.



Directional-seasonal: Return values

Predictive distribution of the 100-year maximum (in metres); directional and seasonal variability of the median estimate on lhs; seasonal variation of predictive distribution for directional octants (2.5%, 37%, median and 97.5% values) in black; corresponding omni-seasonal estimates in red; large difference between S and SW; smooth seasonal variation.



Multivariate extremes

Multivariate extremes

Use cases

- Rare events from **multivariate** distributions
- Spatially-dependent rare events
- Structure of a time-series near an extreme occurrence

Models

- Max-stable processes (MSPs), copulas (e.g. `www.lancs.ac.uk/~jonathan/EVAN17.pdf`)
- Conditional extremes: $Y|(X = x, x > \psi)$
  - Spatial: $Y(s)|(Y(0) = y, y > \psi), s \in \mathcal{N}_0$

- Ross et al. [2017], Tendijck et al. [2019], Tendijck et al. [2021]

A note on extremal dependence

- Dependence **in body** and dependence **in tail** are **different**
- $[X, Y] \sim N(0, [1\ \rho; \rho\ 1]), \rho < 1, \lim_{x \to \infty} \Pr(Y > x | X > x) = 0$

Multivariate spatial conditional extremes (MSCE)

Context for study

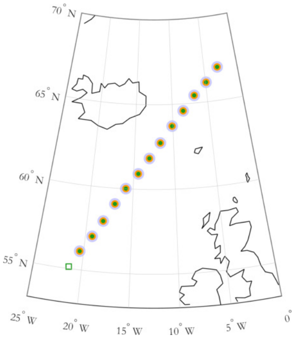
- Motivation: Understand **spatial characteristics of extremes** from satellite observations and hindcast computer model output
- Application: Coastal defences, unmanning, wind farm design and maintenance
- Spatial conditional extremes: Shooter et al. [2019, 2021d,b]
- Competitors: MSPs, hierarchical MSPs and multivariate MSPs

Key underpinning result

$$Y|\{X = x\} = \alpha x + x^{\beta} Z$$

- Asymptotically-motivated, Heffernan and Tawn [2004]
- $X \sim \text{Lpl}$, $Y \sim \text{Lpl}$, and $x > \psi$; $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1]$
- Z is unknown residual process, $\sim N(\mu, \sigma^2)$ for estimation

MSCE : Methodology



$\{X_{jk}\}$
locations j
quantities k

- Condition on **large value** x of **first quantity** X_{01} at **location** $j = 0$
- Estimate "conditional spatial profiles" for $m > 1$ **quantities** $\{X_{jk}\}_{j=1,k=1}^{p,m}$ at $p > 0$ **other locations**

$$X_{jk} \sim \text{Lpl}, \quad x > \psi$$

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

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

- DL is delta-Laplace, or generalised Gaussian
- 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
- Residual correlation $\boldsymbol{\Sigma}$ for conditional Gaussian field, powered-exponential decay with distance
- Shooter et al. [2021c,a]

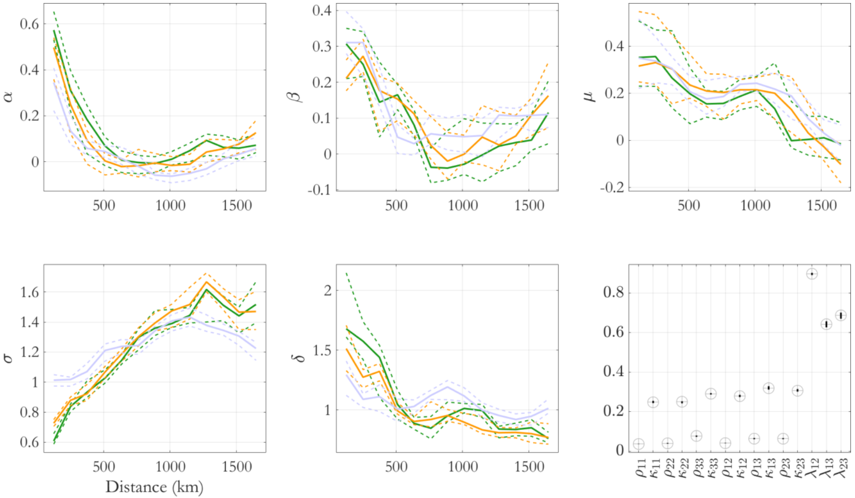
Data sources

- METOP scatterometer directional U_{10} (wind speed)
- NORA10 hindcast directional H_S and directional U_{10}

Inference

- Adaptive MCMC, Roberts and Rosenthal [2009]
- Piecewise linear forms for all parameters with distance

MSCE: Parameter estimates for North Atlantic application



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)

Remote sensing

Remote sensing of gaseous and particulate emissions

Airborne
Hirst et al. [2013]

Sander Geophysics

Line of sight
Hirst et al. [2017], Hirst et al. [2020]

Boreal Laser

Remote sensing of gaseous and particulate emissions

Drone

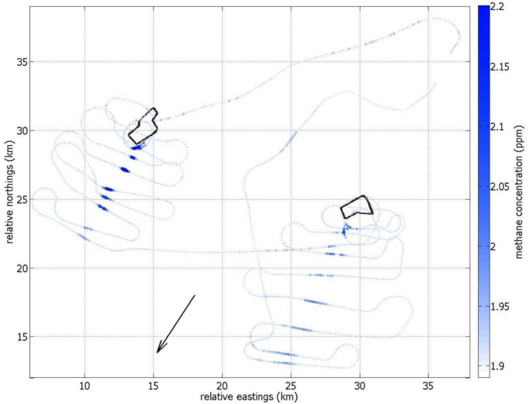
Scientific Aviation

Satellite

Sentinel5 TROPOMI

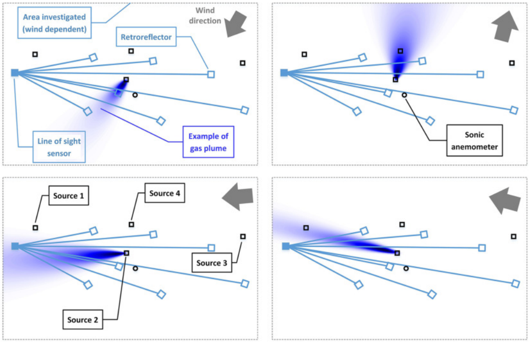
Remote sensing: Examples

Airborne
Hirst et al. [2013]



Canadian land-fill (actual, CH_4)

Line of sight
Hirst et al. [2020]



Chilbolton Observatory (schematic)

Remote sensing: Model specification

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{b} + \mathbf{d} + \boldsymbol{\epsilon}$$

- $\mathbf{y}$ observations
- $\mathbf{b}$ background concentrations
- $\mathbf{s}$ source emission rates
- $\mathbf{d}$ calibration offsets
- $\boldsymbol{\epsilon}$ measurement errors
- λ measurement error precision (e.g. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \lambda^{-2}\boldsymbol{I})$)

Coupling matrix $\mathbf{A}$ from suitable **dispersion model**

- Gaussian plume dispersion model used in most applications (steady-state)
- More complex dispersion model more appropriate in some applications

Remote sensing: Inversion

$$f(\mathbf{s}, \mathbf{b}, \mathbf{d}, \lambda | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{s}, \mathbf{b}, \mathbf{d}, \lambda) f(\mathbf{s}) f(\mathbf{b}) f(\mathbf{d}) f(\lambda)$$

- $f(\mathbf{b})$: imposes **smooth** spatio-temporal evolution of $\mathbf{b}$
- $f(\mathbf{s})$: imposes **sparsity** of source map

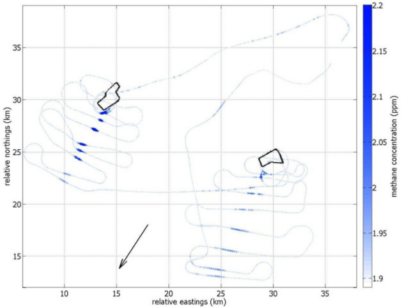
Source representation

- Hirst et al. [2013]: "free" sources ($\Rightarrow$ RJ-MCMC)
- Hirst et al. [2020]: fixed grid of candidate sources

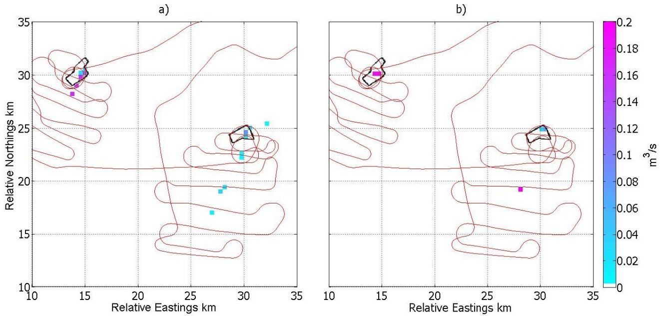
Inference

- Full conditionals where possible
- MH with gradient-based (mMALA) proposals otherwise

Remote sensing: Source estimates for airborne



observations



a) starting solution, (b) posterior median

Remote sensing: Satellite sensing of gaseous and particulate emissions

Data sources

- ESA Sentinel 5P **TROPOMI** instrument (data publicly available)
- Private by commission (e.g. GHGSat)

Pros and cons:

- Daily measurements, globally
- Direct quantification of column-integrated concentrations
- Sensor limitations: oceans, cloud, albedo / reflection, **striping**
- Smallest source detectable: TROPOMI $\approx$ 5 T/hr, GHGSat $\approx$ 100 kg/hr
- Spatial resolution: TROPOMI $\approx$ 5km, GHGSat $\approx$ 50m

Remote sensing: NO_2 as surrogate for CH_4

NO_2 is a good surrogate

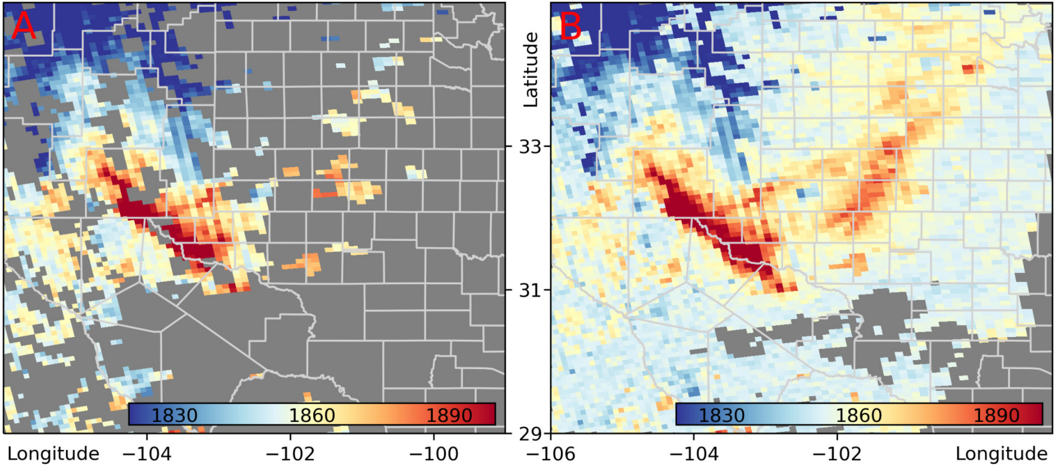
- More easily detected, better spatial coverage than CH_4
- Half life of days (better temporal source identification)

Model

$$
\begin{aligned}
&\textit{Observation} \\
N^o_{id} &= N_{id} + \epsilon_{Nid}\sigma_{Nid}, \quad d = 1, 2, ..., n_{\text{Day}}, i = 1, 2, ..., n_{\text{Obs}(d)} \\
C^o_{id} &= C_{id} + \epsilon_{Cid}\sigma_{Cid} \\
&\textit{System} \\
C_{id} &= \alpha_d + \beta_d N_{id} + \epsilon_{Tid}\sigma_{Td}
\end{aligned}
$$

- Bayesian inference as before

Remote sensing: NO_2 as surrogate for CH_4



Left: original CH_4 map. Right: inferred CH_4 map (both ppb by volume)

Thanks to **Clay Roberts**, Oli Shorttle and Kaisey Mandel at IoA, Cambridge

Summary

- Coupling of **physical** and **statistical** knowledge within an **appropriate framework** for inference
- Exploit growing sources of data from **direct observation** and **physical models**
- Careful **uncertainty quantification** for better decisions

Diolch & Thank-you!

www.lancs.ac.uk/∼jonathan

References

A. Brezger and S. Lang. Generalized structured additive regression based on Bayesian P-splines. *Comput. Statist. Data Anal.*, 50:967–991, 2006.

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.

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

M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. Roy. Statist. Soc. B*, 73:123–214, 2011.

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.

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

B. Hirst, P. Jonathan, F. González del Cueto, D. Randell, and O. Kosut. Locating and quantifying gas emission sources using remotely obtained concentration data. *Atmospheric Environ.*, 74:141–158, 2013.

B. Hirst, D. Randell, M. Jones, P. Jonathan, B. King, and M. Dean. A new technique for monitoring the atmosphere above onshore carbon storage projects that can estimate the locations and mass emission rates of detected sources. *Energy Procedia*, 114:3716–3728, 2017.

B. Hirst, D. Randell, M. Jones, J. Chu, A. Kannath, N. Macleod, M. Dean, and D. Weidmann. Methane emissions: Remote mapping and source quantification using an open-path laser dispersion spectrometer. *Geophys. Res. Let.*, 47:e2019GL086725, 2020.

M. J. Jones, H. F. Hansen, A. R. Zeeberg, D. Randell, and P. Jonathan. Uncertainty quantification in estimation of ocean environmental return values. *Coastal Eng.*, 141:36–51, 2018.

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

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

G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.*, 4:337–358, 2002.

E Ross, M Kereszturi, M van Nee, D Randell, and P Jonathan. On the spatial dependence of extreme ocean storm seas. *Ocean Eng.*, 145:359–372, 2017.

R. Shooter, E. Ross, J. A. Tawn, and P. Jonathan. On spatial conditional extremes for ocean storm severity. *Environmetrics*, 30:e2562, 2019.

R. Shooter, E Ross, and P. Jonathan. Multivariate spatial conditional extremes. `https://github.com/ygraigarw/MultivariateSpatialConditionalExtremesPublic`, 2021a.

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, 2021b.

R. Shooter, E Ross, A. Ribal, I. R. Young, and P. Jonathan. Multivariate spatial conditional extremes for extreme ocean environments. *Submitted to Ocean Eng., draft at www.lancs.ac.uk/∼jonathan*, 2021c.

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

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. Modelling the extremes of bivariate mixture distributions with application to oceanographic data. *Accepted for publication in JASA, draft at www.lancs.ac.uk/∼jonathan*, 2021.

R. Towe, E. Zanini, D. Randell, G. Feld, and P. Jonathan. Efficient estimation of distributional properties of extreme seas from a hierarchical description applied to calculation of un-manning and other weather-related operational windows. *Ocean Eng.*, 238:109642, 2021.

T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, and M Girolami. Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Stat. Probabil. Lett.*, 91(2002):14–19, 2014.

E. Zanini, E. Eastoe, M. Jones, D. Randell, and P. Jonathan. Covariate representations for non-stationary extremes. *Environmetrics*, 31:e2624, 2020.