Inference for multivariate extremes via a semi-parametric angular-radial model

Callum Murphy-Barltrop

Joint work with Dr Ed Mackay (University of Exeter), Dr Jordan Richards (University of Edinburgh) and Prof Philip Jonathan (Lancaster University)

CASE Workshop 30.06.2025

TECHNISCHE UNIVERSITÄT DRESDEN

ScaDS.AI DRESDEN LEIPZIG

Introduction

- September 2019 - May 2023: PhD, Lancaster University.
- September 2023 - Present: Postdoc, TU Dresden/ScaDS.AI.

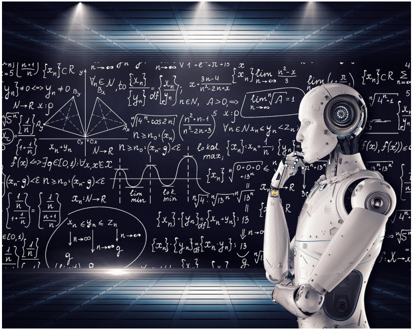
Introduction

Introduction

Introduction

- ScaDS.AI = Center for **Sca**lable **D**ata Analytic**s** and **A**rtificial **I**ntelligence
- One of the five new centres in Germany funded under the government's AI strategy.

Introduction



- My current research focus is on employing state-of-the-art **artificial intelligence** and **machine learning** techniques for modelling **multivariate extremes**.

Overview

- Univariate extremes
- Multivariate extremes
- Angular-radial systems
- The SPAR model
- GAM approach
- Deep learning approach
- Discussion

Univariate extremes

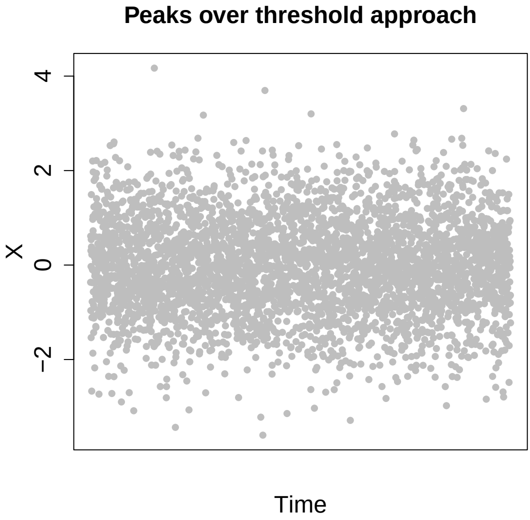
- Study of univariate extremes is very well-established.
- For modelling, most practitioners employ the **peaks over threshold** approach.
- A **generalised Pareto** (GP) distribution is fitted to observations exceeding some threshold.

Univariate extremes

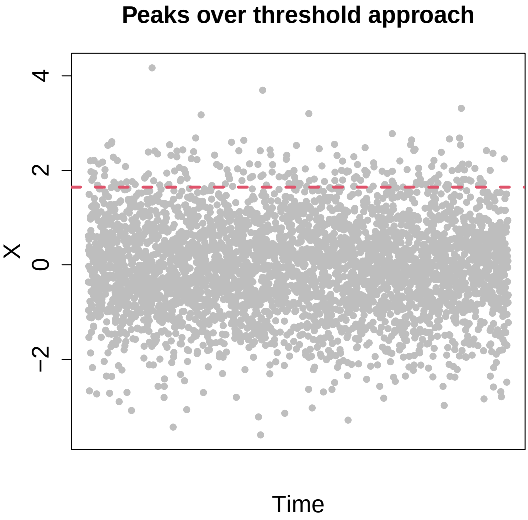
$$Y_u := (X - u \mid X > u) \sim \mathrm{GP}(\sigma, \xi)$$

$$F_{Y_u}(y) = 1 - \left\{1 + \frac{\xi y}{\sigma}\right\}_+^{-1/\xi}$$

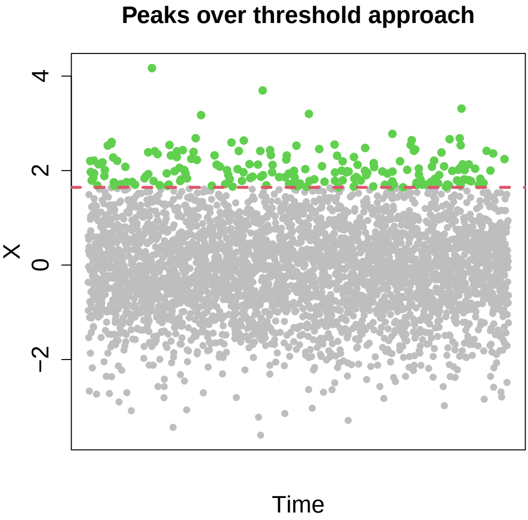
Univariate extremes



Univariate extremes



Univariate extremes



Univariate extremes

- Most approaches assume that the variable X is **stationary** (i.e., not changing in time).
- However, this is unrealistic in most applications.

NATURE AND ENVIRONMENT | EUROPE

Amid EU's hottest year cities face more floods, extreme heat

Tim Schauenberg
04/15/2025

Europe saw its hottest year on record in 2024 — and the fallout was alarming. But there was also a flicker of hope, according to the EU's new Copernicus climate report.

Major flooding in Spain's Valencia region in late 2024 killed more than 220 people and destroyed homes, cars and infrastructure
Image: Emilio Morenatti/AP/dpa/picture alliance

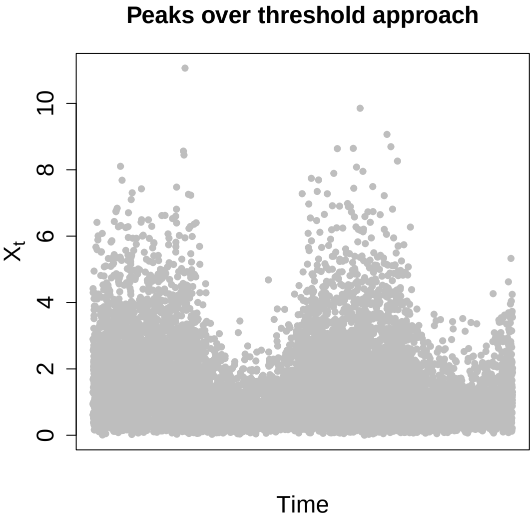
Univariate extremes

- In such cases, we must allow for **non-stationarity** in the modelling framework.
- Given a variable X_t and some covariates $\boldsymbol{Z}_t$, with t denoting time, one could consider the model

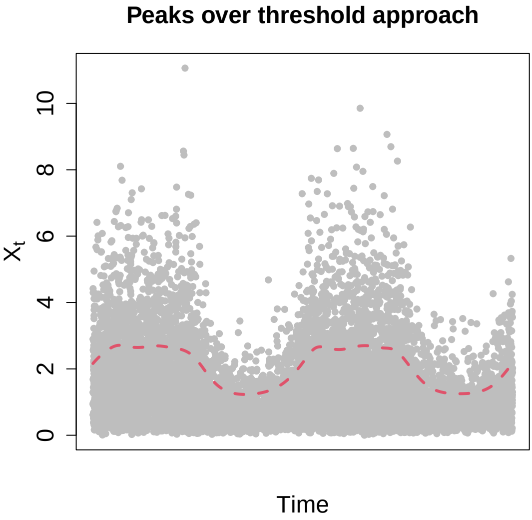
$$(X_t - u(\boldsymbol{z}_t) \mid X_t > u(\boldsymbol{z}_t)) \sim \mathrm{GP}(\sigma(\boldsymbol{z}_t), \xi(\boldsymbol{z}_t))$$

- The threshold and parameters are a **function** of the covariates.

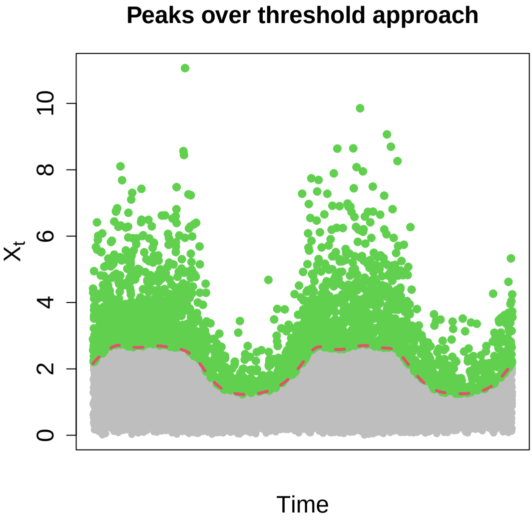
Univariate extremes



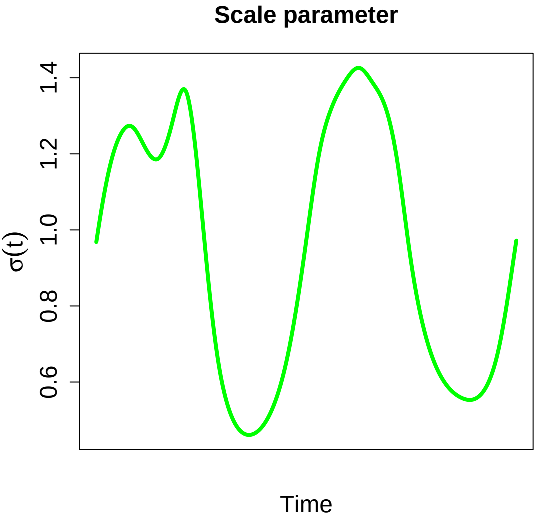
Univariate extremes



Univariate extremes



Univariate extremes

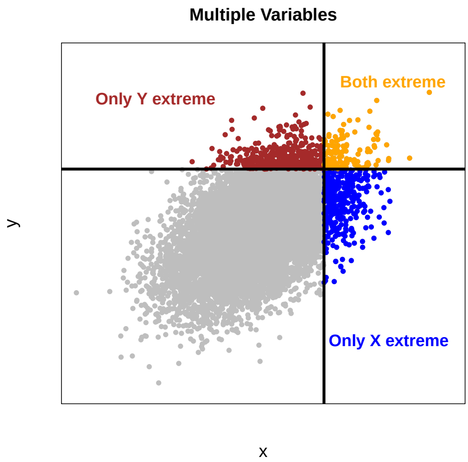


Univariate extremes

- Literature on non-stationary peaks over threshold analysis is **also well established**.
- Many approaches available:
  1. Parametric models: Davison and Smith (1990); Eastoe and Tawn (2009).
  2. Generalised additive models (GAMs): Chavez-Demoulin and Davison (2005); Youngman (2019).
  3. Deep learning: Pasche and Engelke (2024); Richards and Huser (2022)

Multivariate extremes

Consider the question: what is 'extreme' for multiple variables?

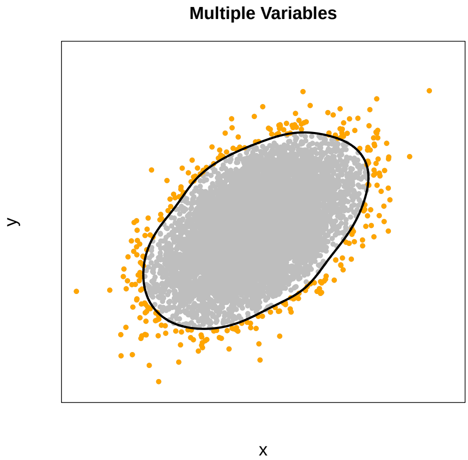


Multivariate extremes

- Various definitions of multivariate extremes have been proposed (Barnett, 1976).
- However, many classical approaches have focused on the orange region.
- This analysis is very limited – and misses important information in the joint tail.

Multivariate extremes

In this work, we define multivariate extremes in a **more general manner**.



Multivariate extremes

- Many approaches use **geometric representations** for modelling multivariate extremes.
- Geometric $\Rightarrow$ multivariate data is split into angular and radial components.

Angular-radial systems

- How do we define ‘angular’ and ‘radial’ components?
- This all depends on how we define ‘distance’.
- The standard approach is to use the **Euclidean norm**.

Angular-radial systems

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ denote a continuous random vector. Then

$$R := \|\mathbf{X}\| = \sqrt{\sum_i X_i^2}$$

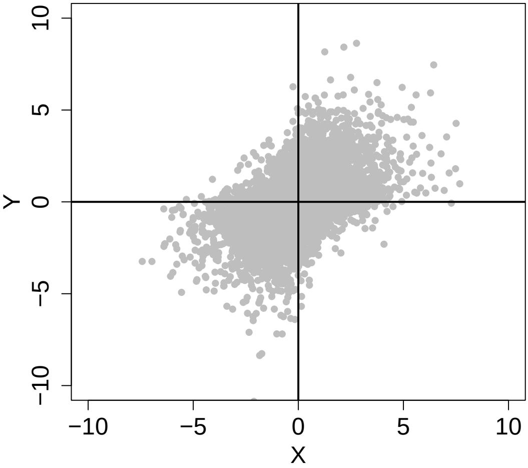
$$\boldsymbol{W} := \frac{\mathbf{X}}{R}$$

This transformation is one-to-one[1].

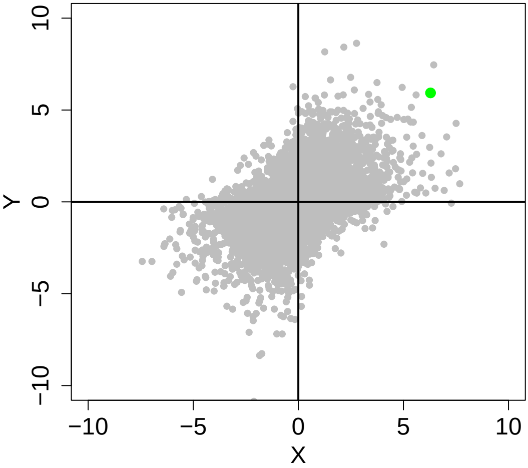
$$\boldsymbol{W} \in \mathcal{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}.$$

[1]Excluding the origin **0**.

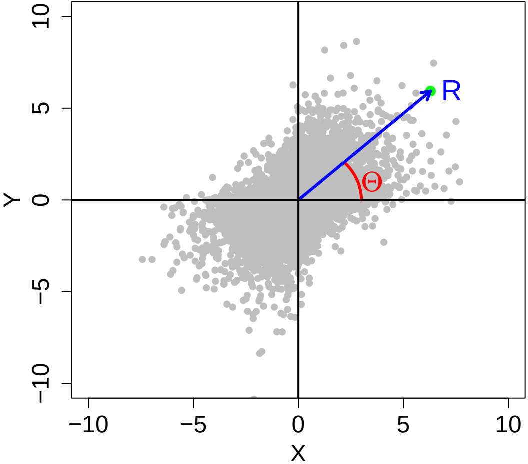
Angular-radial systems



Angular-radial systems

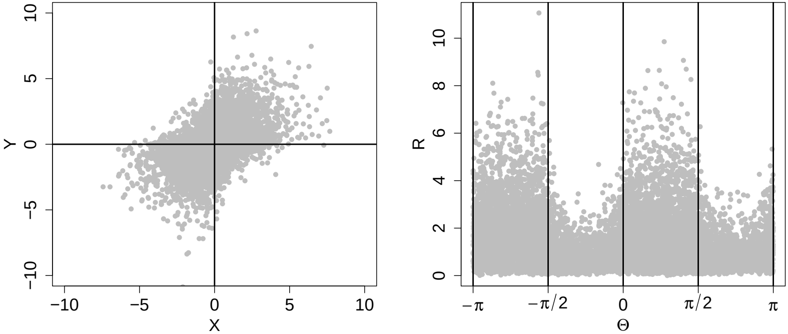


Angular-radial systems

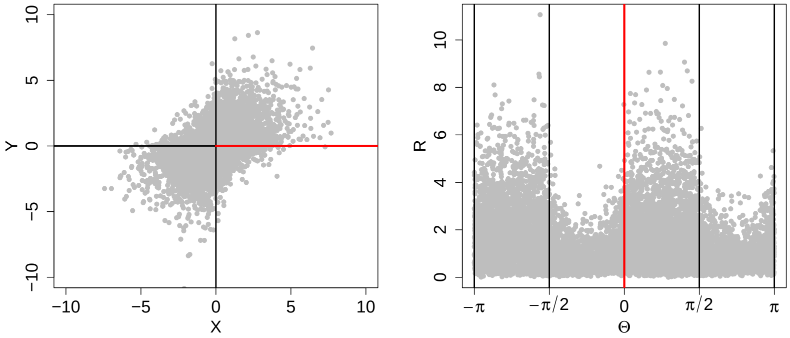


$$\Theta = \text{atan2}(Y, X) = \text{atan2}(W_2, W_1),\ \boldsymbol{W} = (\cos(\Theta), \sin(\Theta)).$$

Angular-radial systems



Angular-radial systems



Angular-radial systems

- Joint extremes can be characterised by the **stochastic behaviour** of $(R, \boldsymbol{W})$.
- $\boldsymbol{W}$ gives the ‘direction’ of the event.
- R gives the ‘magnitude’.

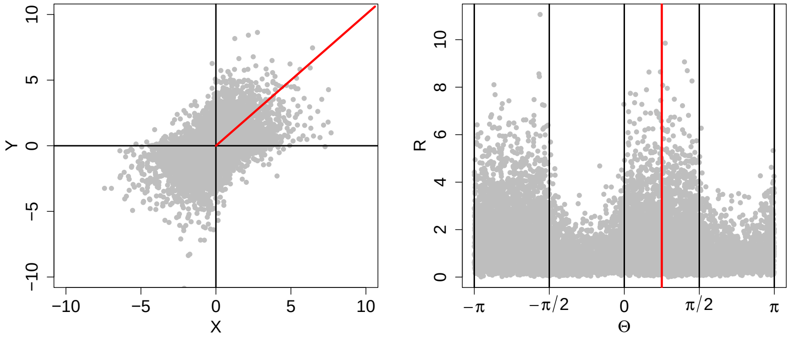
The SPAR model

- SPAR model = **s**emi-**p**arametric **a**ngular-**r**adial model.
- This model, proposed by Mackay and Jonathan (2023), assumes the following:

$$(R - u(\boldsymbol{w}) \mid \boldsymbol{W} = \boldsymbol{w}, R > u(\boldsymbol{w})) \sim \text{GP}(\sigma(\boldsymbol{w}), \xi(\boldsymbol{w})),$$

  for sufficiently large $u(\boldsymbol{w})$.
- Conditional on a **fixed angle**, the corresponding radial tails follow a GP distribution.

The SPAR model



The SPAR model

Why should this work?

- We are assuming the variable $R \mid (\boldsymbol{W} = \boldsymbol{w})$ is in the **domain of attraction** of an extreme value distribution.
- These conditions are very **general**, and hold for many univariate variables (Balkema and de Haan, 1974; Pickands, 1975).
- Furthermore, Mackay and Jonathan (2023) show these assumptions are valid for many popular copula examples (and different marginal scales).

The SPAR model

What's the point?

- We have the following relationship between the joint densities:

$$f_{R,\boldsymbol{W}}(r, \boldsymbol{w}) = r^{d-1} \, f_{\boldsymbol{X}}(r\boldsymbol{w}),$$

  where r denotes the Jacobian.
- The joint distribution of $\boldsymbol{X}$ can be **obtained directly** from the joint distribution of $(R, \boldsymbol{W})$.

The SPAR model

What's the point?

- Applying Bayes' theorem, we can write

$$\begin{aligned} f_{R,\boldsymbol{W}}(r,\boldsymbol{w}) &= f_{\boldsymbol{W}}(\boldsymbol{w}) f_{R|\boldsymbol{W}=\boldsymbol{w}}(r \mid \boldsymbol{w}) \\ &\approx (1-\gamma) f_{\boldsymbol{W}}(\boldsymbol{w}) f_{GP}(r - u(\boldsymbol{w}) \mid \sigma(\boldsymbol{w}), \xi(\boldsymbol{w})), \end{aligned}$$

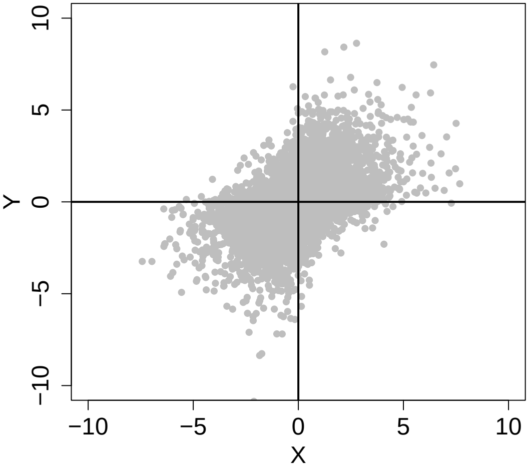
  where $\gamma = \Pr(R \leq u(\boldsymbol{w}) \mid \boldsymbol{W} = \boldsymbol{w})$.
- Combined with a model for $f_{\boldsymbol{W}}(\boldsymbol{w})$, **we can model the joint tail** of $\boldsymbol{X}$.
- Note: this does not require a pre-specified set of margins.

The SPAR model

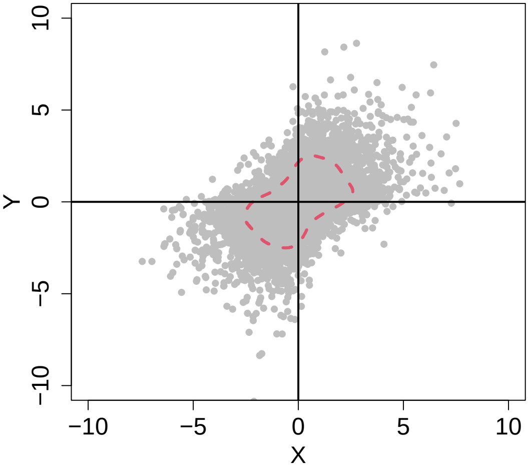
Additional assumptions:

- $f_{\boldsymbol{w}}$, u, σ and ξ are finite and continuous.
- The data cloud is star-shaped at $\mathbf{0}$.

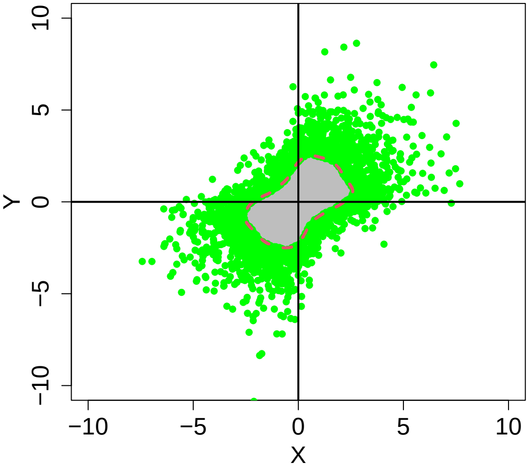
The SPAR model



The SPAR model



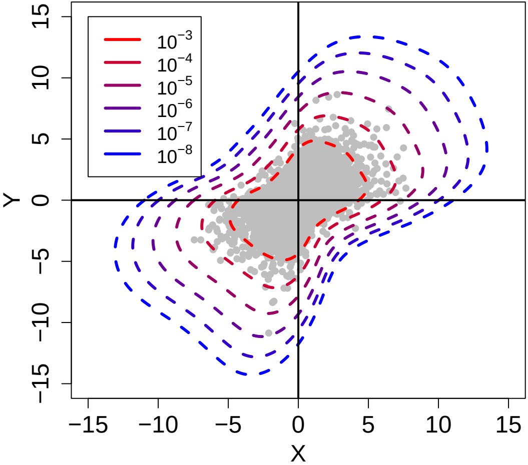
The SPAR model



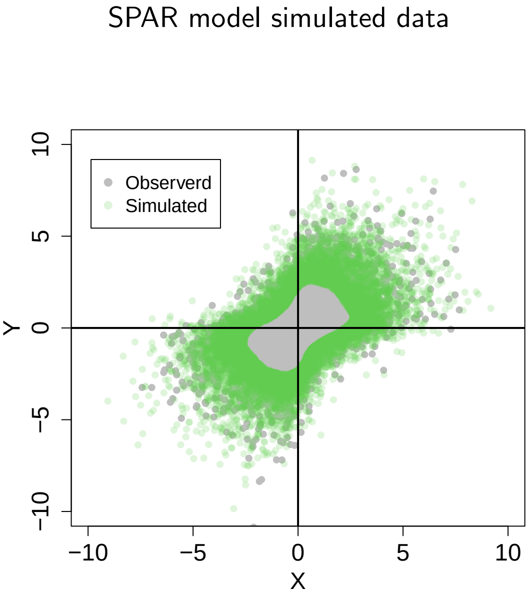
The SPAR model

For any small value $\epsilon > 0$

$$\{\boldsymbol{x} : f_{\boldsymbol{X}}(\boldsymbol{x}) = \epsilon\}$$



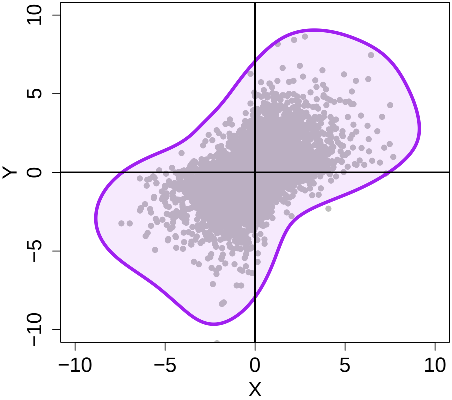
The SPAR model



The SPAR model

Given a probability $p \in [0,1]$ close to 1, a **return level set** (environmental contour) is defined as a set $\mathcal{A}_p$ such that

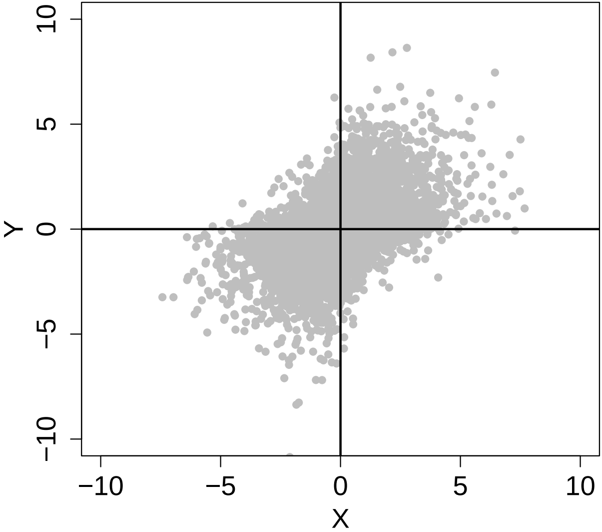
$$\Pr[\boldsymbol{X} \in \mathcal{A}_p] = p$$



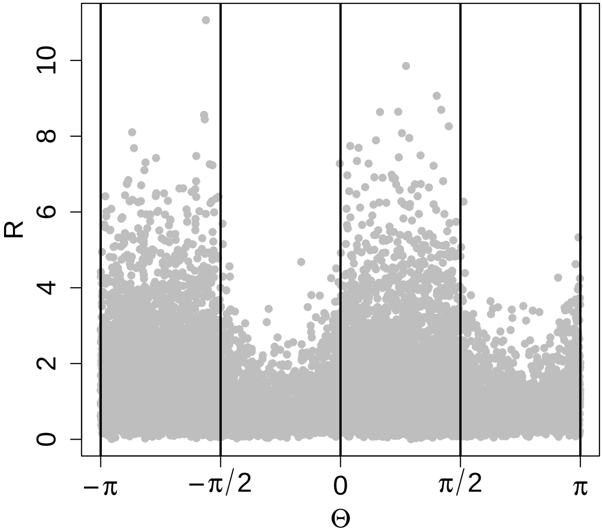
The SPAR model

But did you spot the **sleight of hand**?

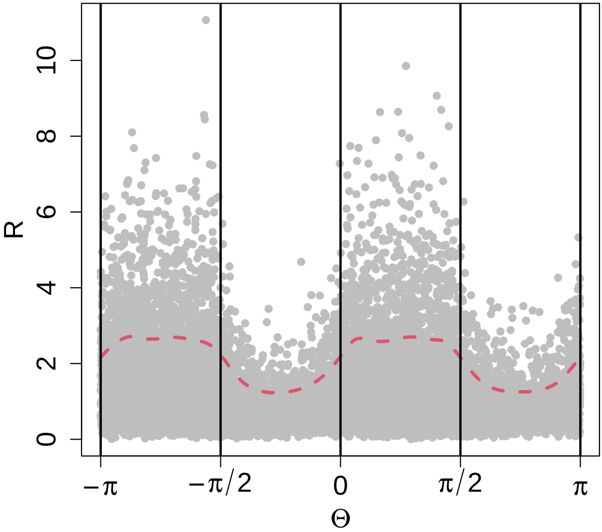
The SPAR model



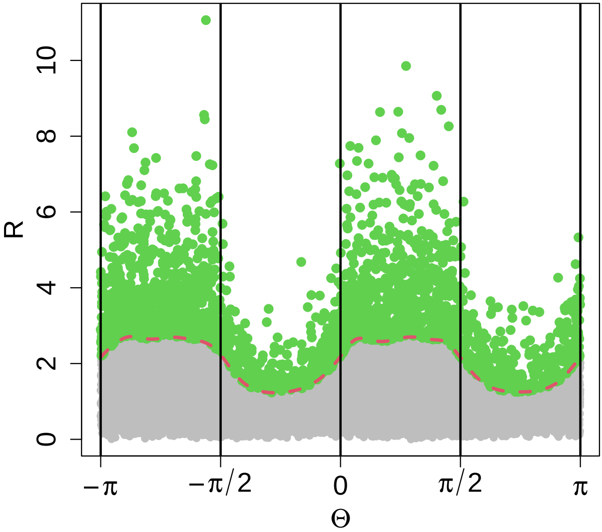
The SPAR model



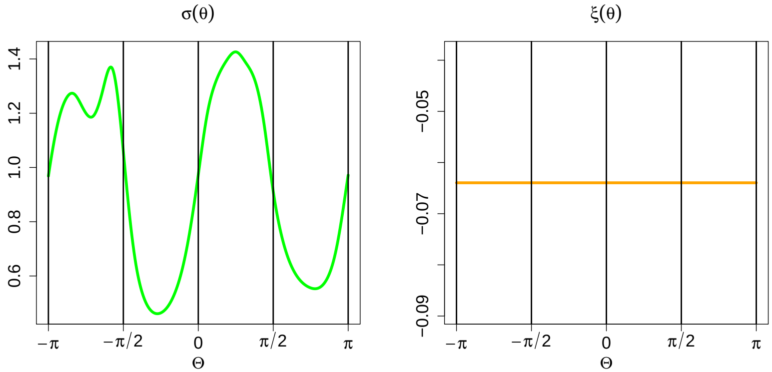
The SPAR model



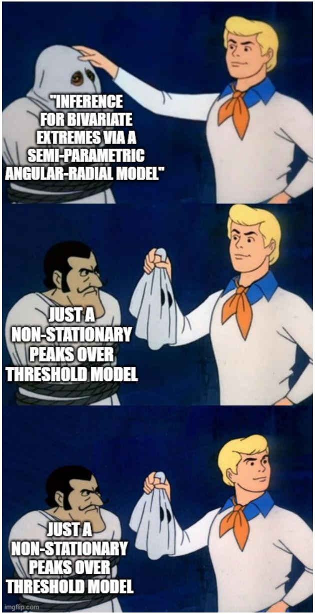
The SPAR model



The SPAR model



The SPAR model



The SPAR model

Regular variation:

$$\lim_{r\to\infty} \Pr(\boldsymbol{W} \in B, R > sr \mid R > r) = H(B)s^{-1}, s \geq 1.$$

$\boldsymbol{W}$ and R independent in limit.
$\boldsymbol{W} \mid R > r \sim H$.

The SPAR model

Regular variation:

$$\lim_{r\to\infty} \Pr(\boldsymbol{W} \in B, R > sr \mid R > r) = H(B)s^{-1}, s \geq 1.$$

$\boldsymbol{W}$ and R independent in limit.
$\boldsymbol{W} \mid R > r \sim H$.

SPAR method:

$$(R, \boldsymbol{W}) \mid R > u(\boldsymbol{w})$$

Not independent - we assume angular dependence.
$\boldsymbol{W} \mid R > u(\boldsymbol{w}) \stackrel{d}{=} \boldsymbol{W}$.

The SPAR model

- Mackay and Jonathan (2023) provide no means for inference with the SPAR model.
- **Our contribution:** turn the model into a working inference framework.
- Note we aren't really doing anything novel:

$$X_t \mid (\boldsymbol{Z}_t = \boldsymbol{z}_t)$$

The SPAR model

- Mackay and Jonathan (2023) provide no means for inference with the SPAR model.
- **Our contribution:** turn the model into a working inference framework.
- Note we aren't really doing anything novel:

$$X_t \mid (\boldsymbol{Z}_t = \boldsymbol{z}_t)$$

$$R \mid (\boldsymbol{W} = \boldsymbol{w})$$

GAM approach

- For our first collaboration, we restricted attention to the **bivariate setting** and used **GAMs** to model $u(\boldsymbol{w}), \sigma(\boldsymbol{w})$ and $\xi(\boldsymbol{w})$.
- GAMs provide smooth approximations for functions with high flexibility.

$$g(\theta) = \beta_0 + \sum_{j=1}^{k} B_j(\theta)\beta_j, \tag{1}$$

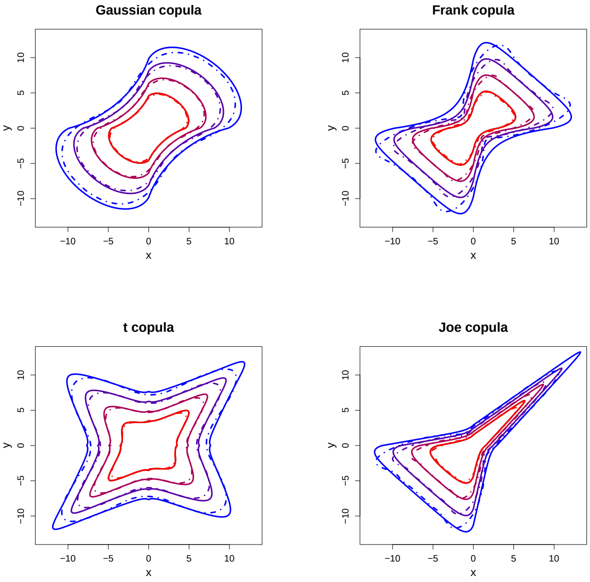
GAM approach

- All functions approximated using **cyclic cubic splines** with significant numbers of knots.
- $u(\boldsymbol{w})$ fitted via quantile regression techniques (Koenker et al., 2017).
- $\sigma(\boldsymbol{w}), \xi(\boldsymbol{w})$ fitted using restricted maximum likelihood (Wood, 2011).
- EVGAM framework used for estimation (Youngman, 2020).

GAM approach

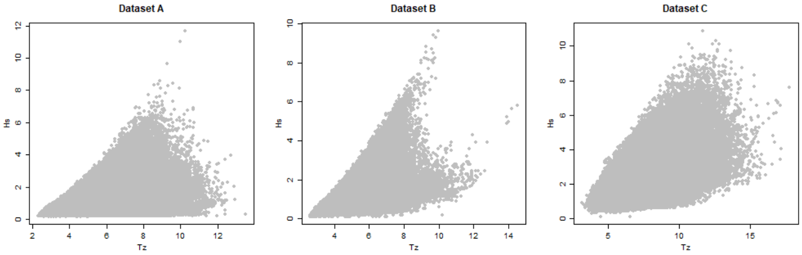
- Finally, we model $f_{\boldsymbol{W}}(\boldsymbol{w})$ non-parametrically using **kernel density estimation** techniques (von-Mises kernel).
- See Murphy-Barltrop et al. (2024) for further details.

GAM approach



GAM approach

- To demonstrate the **utility** of the SPAR framework, we consider three bivariate **metocean time series** made up of zero-up-crossing period, T_z, and significant wave height, H_s, observations.



GAM approach

- Understanding the joint extremes of metocean variables is important for **assessing the reliability** of offshore structures, e.g., wind turbines.
- Extreme responses can occur with either short- or long-period waves – so it is **necessary** to characterise the joint distribution in **both of these ranges**.

GAM approach

- We fit the SPAR model **without** marginal transformation, i.e., on the observed scale.
- Uncertainty quantified using block bootstrapping techniques.
- Diagnostics indicate **reasonable model fits**.

GAM approach

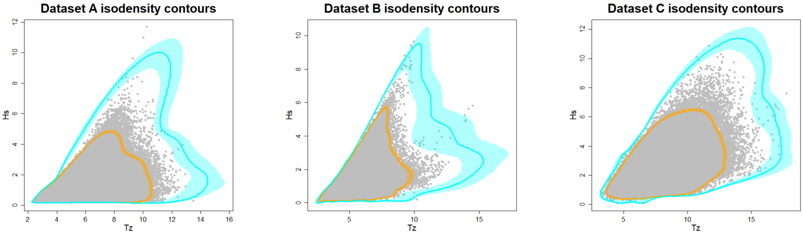


Figure: Isodensity contours at $\epsilon = 10^{-3}$ (orange lines) and $\epsilon = 10^{-6}$ (cyan lines).

GAM approach

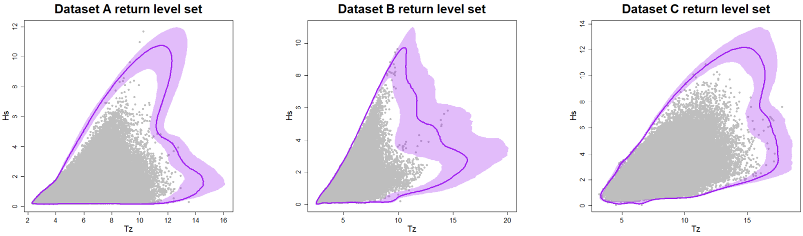


Figure: Estimated 10 year return level sets.

Shameless self-promotion

Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2024). Inference for bivariate extremes via a semi-parametric angular-radial model. Extremes: 1-30.

Mackay, E., **Murphy-Barltrop, C. J. R.**, and Jonathan, P. (2024). The SPAR model: a new paradigm for multivariate extremes. Application to joint distributions of metocean variables. OMAE2024/130932.

R package: `https://github.com/callumbarltrop/SPAR`

Deep learning approach

- The GAM framework **does not easily extend** to higher dimensions (splines on hyperspheres?).
- How can we fit SPAR in $d > 2$ in a highly flexible manner?

Deep learning approach

- The GAM framework **does not easily extend** to higher dimensions (splines on hyperspheres?).
- How can we fit SPAR in $d > 2$ in a highly flexible manner?
- Answer: **deep learning**

Deep learning approach

- We **multilayer perceptrons** to model $u(\boldsymbol{w}), \sigma(\boldsymbol{w})$ and $\xi(\boldsymbol{w})$.
- Fully-connected, feed forward neural network.
- Step 1: pinball loss used to estimate $u(\boldsymbol{w})$.
- Step 2: negative log-likelihood used to estimate $\sigma(\boldsymbol{w}), \xi(\boldsymbol{w})$.

Deep learning approach

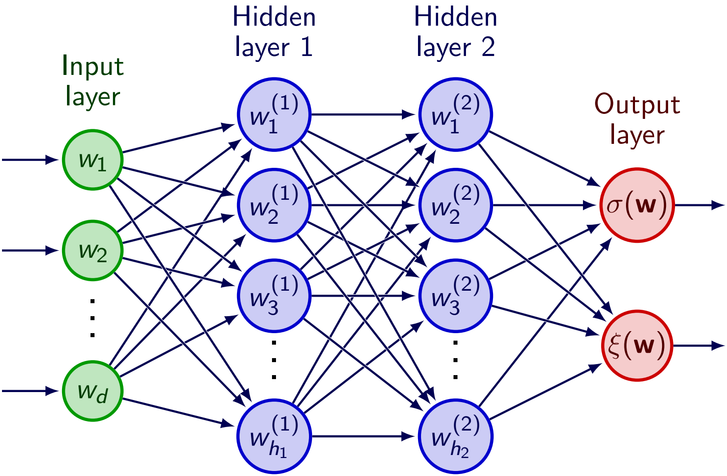


Figure: Just a fancy, highly parameterised function.

Deep learning approach

- Again, this is nothing new: we are **adapting** the approaches of Pasche and Engelke (2024), Richards and Huser (2022), Richards and Huser (2024) (among others).
- Loss surface for GPD **highly irregular**. Training procedure adapted to account for convergence issues – see Mackay et al. (2024) for further details.
- Keras/Tensorflow used for model fitting.

Deep learning approach

- Angular distribution $f_{\boldsymbol{w}}$ modelled again using KDE techniques.
- Power spherical distribution kernel used, with bandwidth selected using cross validation.
- Easy to simulate from.

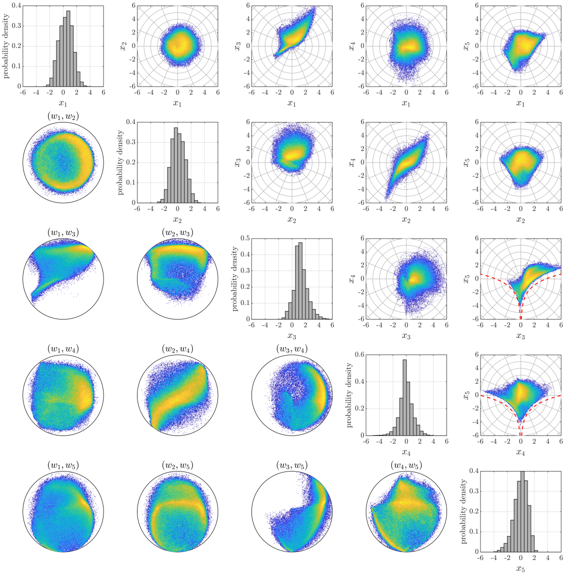
Deep learning approach

- We apply the framework to 31-years of hindcast dataset from the South West of the UK.
- This area has **high potential** for floating wind farms.

Deep learning approach

- We consider the x- and y-components of **wave height** and **wind speed**, alongside mean **wave period** (i.e., 5D case).
- These variables all influence the **motion** and **loading** of floating wind turbines.
- Robust risk analysis of compound extremes **crucial** for design engineering.

Deep learning approach



Deep learning approach

- We compare a range of **architectures** and **tuning parameters**.
- **Optimise** using diagnostics.
- **Simulate new data** from the model – both in the joint body and tail.

Deep learning approach

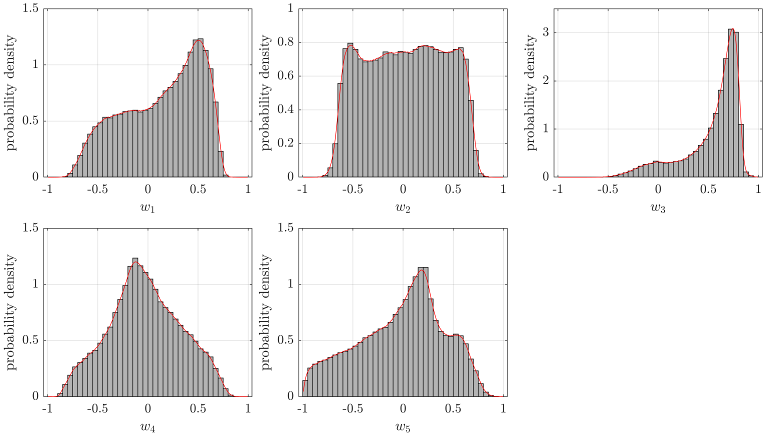


Figure: Angular histograms

Deep learning approach

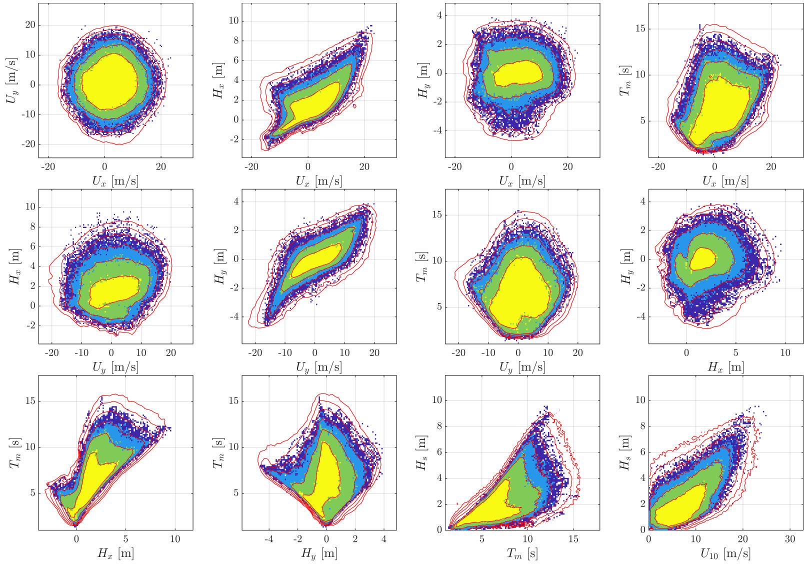


Figure: Pairwise contour plots

Deep learning approach

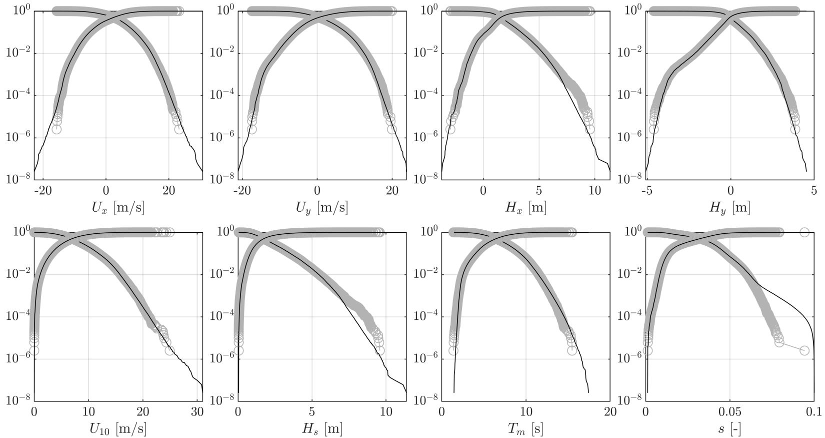


Figure: Marginal tails assessment.

Deep learning approach

- We have proposed a novel, deep learning approach for modelling complex metocean variables.
- The fitted model **respects physical constraints**, without us having to specify these constraints.
- Appears **flexible** and **accurate**.

Shameless self-promotion

Mackay, E., **Murphy-Barltrop, C. J. R.**, Richards, J., and Jonathan, P. (2025). Deep learning joint extremes of metocean variables using the SPAR model. OMAE-25-1104.

Journal paper + R packages to appear soon!

Discussion

- In this work, we have introduced a semi-parametric modelling framework for multivariate extremes.
- This framework adapts **existing univariate approaches**.
- Can capture the joint extremes of **many dependent structures**, requires **few assumptions** and offers a **high degree** of flexibility.

Discussion

Future work

- Further investigating **sensitivity** to tuning parameter choices (i.e., GAMs, circular density, architectures).
- **Comparison** to other multivariate modelling techniques (e.g., regular variation, conditional extremes, geometric extremes).
- Optimal **placement of origin**.
- More applications (German river data, Dutch flood modelling)

Thanks for listening!

Any questions? :)

callum.murphy-barltrop@tu-dresden.de

References I

Balkema, A. A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2:792–804.

Barnett, V. (1976). The Ordering of Multivariate Data. *Journal of the Royal Statistical Society. Series A (General)*, 139:318–344.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54:207–222.

Davison, A. C. and Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 52:393–425.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 58:25–45.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. Chapman and Hall/CRC.

Mackay, E. and Jonathan, P. (2023). Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv*, 2310.12711.

References II

Mackay, E., Murphy-Barltrop, C., Richards, J., and Jonathan, P. (2024). Deep learning joint extremes of metocean variables using the SPAR model. *arXiv*, 2412.15808.

Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2024). Inference for bivariate extremes via a semi-parametric angular-radial model. *Extremes*.

Pasche, O. C. and Engelke, S. (2024). Neural networks for extreme quantile regression with an application to forecasting of flood risk. *The Annals of Applied Statistics*, 18:1–27.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3:119–131.

Richards, J. and Huser, R. (2022). A unifying partially-interpretable framework for neural network-based extreme quantile regression. *arXiv*, pages 1–50.

Richards, J. and Huser, R. (2024). Extreme quantile regression with deep learning. *arXiv*, 2404.09154.

Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73:3–36.

Youngman, B. (2020). evgam: Generalised Additive Extreme Value Models. *R Package*.

References III

Youngman, B. D. (2019). Generalized Additive Models for Exceedances of High Thresholds With an Application to Return Level Estimation for U.S. Wind Gusts. *Journal of the American Statistical Association*, 114:1865–1879.