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

Callum Murphy-Barltrop

Joint work with Dr Ed Mackay (University of Exeter) and Prof Philip Jonathan (Lancaster University/Shell)

EVAN Conference 16-19 July 2024

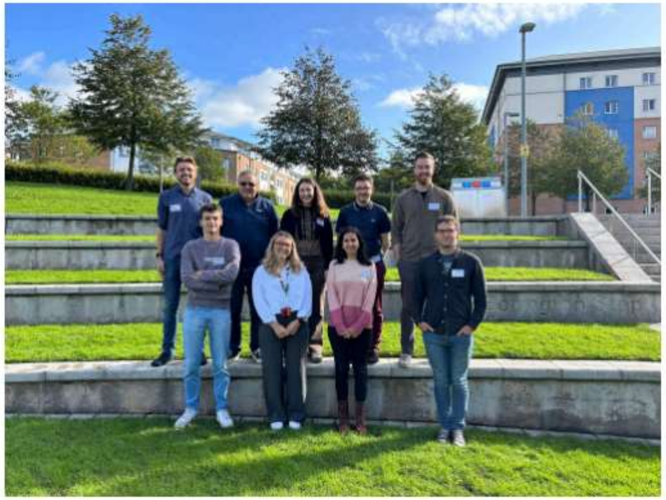
TECHNISCHE UNIVERSITÄT DRESDEN

ScaDS.AI DRESDEN LEIPZIG

Introduction

- Completed my PhD in May 2023 at Lancaster University, UK, under the supervision of Dr Jennifer Wadsworth and Dr Emma Eastoe.
- Since September 2023, I have been a postdoctoral researcher at TU Dresden, Germany.

Introduction



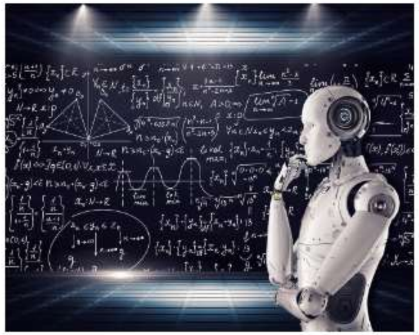
Introduction

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.
- Split between the cities of Dresden and Leipzig.

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
- Simulation study
- Case study
- Discussion

Univariate extremes

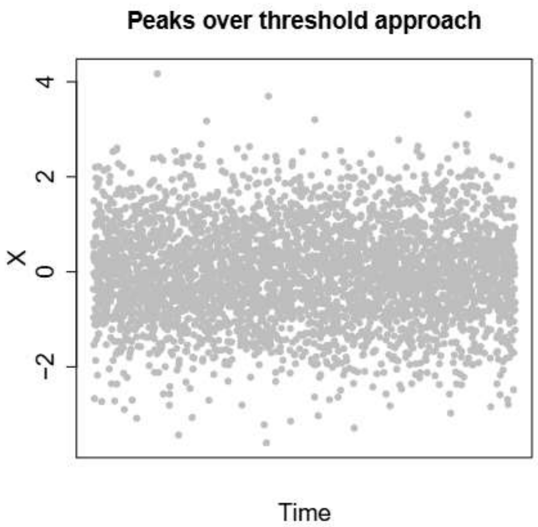
- Study of univariate extremes is very well-established (Balkema and de Haan, 1974; Davison and Smith, 1990; Coles, 2001).
- For modelling, most practitioners employ the *peaks over threshold* approach.
- A *generalised Pareto* (GP) distribution is fitted to observations exceeding some threshold.

Univariate extremes

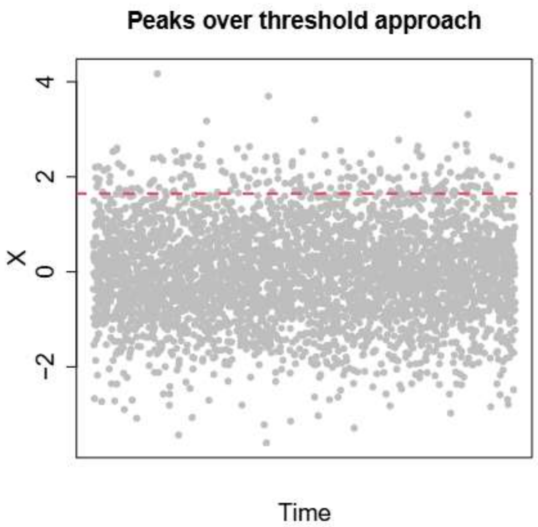
$$Y_u := (X - u \mid X > u) \sim \mathrm{GP}(\sigma, \xi)$$

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

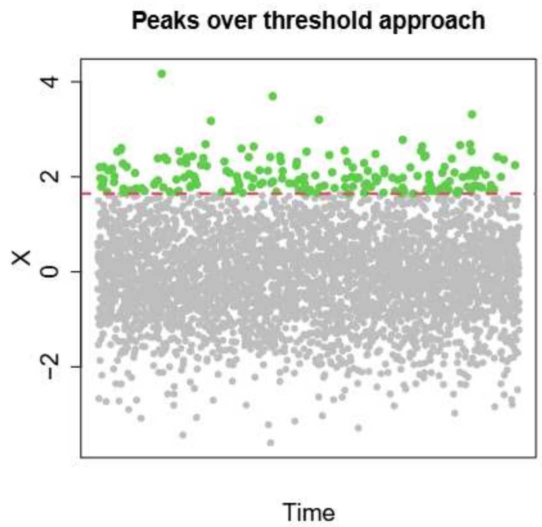
Univariate extremes



Univariate extremes



Univariate extremes



Univariate extremes

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

Dresden on a 'normal day' vs Dresden in June 2013

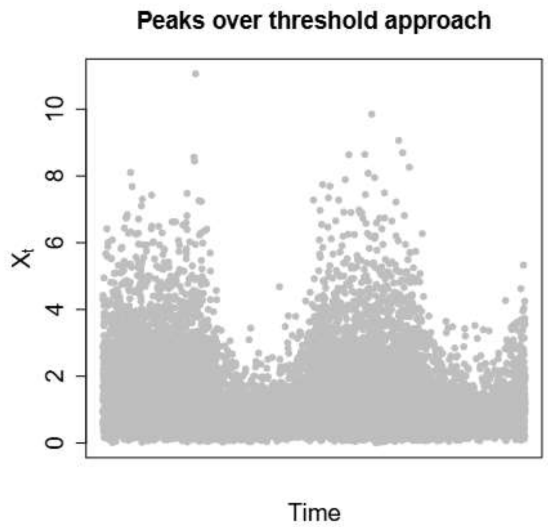
Univariate extremes

- 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

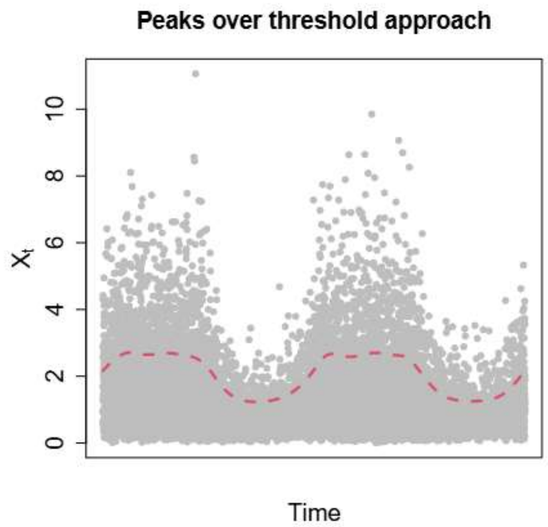
$$(X_t - u(\boldsymbol{z}_t) \mid X_t > u(\boldsymbol{z}_t)) \sim \mathrm{GP}(\sigma(\boldsymbol{z}_t), \xi(\boldsymbol{z}_t))$$

- The threshold and parameters are a **function** of the covariates.
- Note that other approaches/paradigms are available.

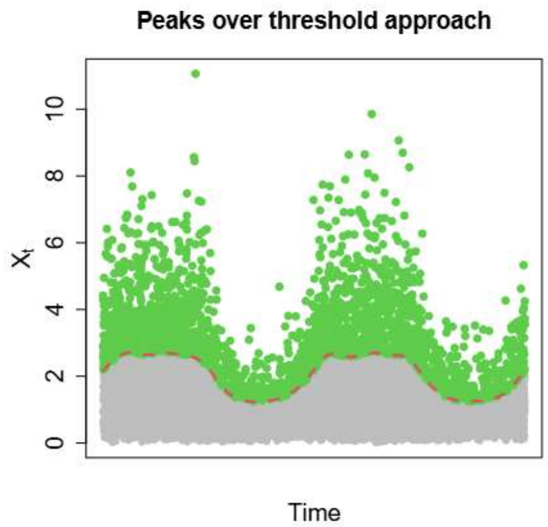
Univariate extremes



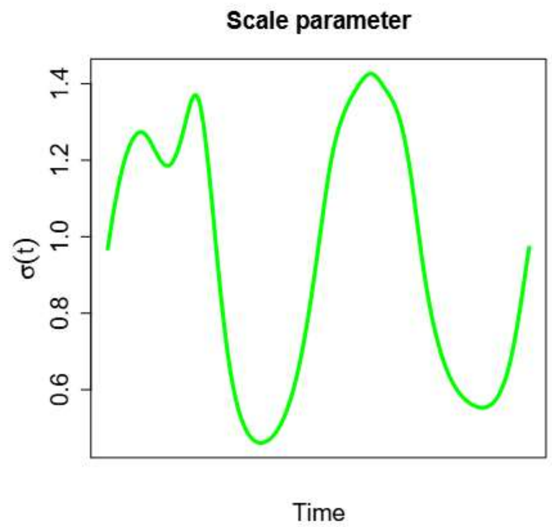
Univariate extremes



Univariate extremes



Univariate extremes

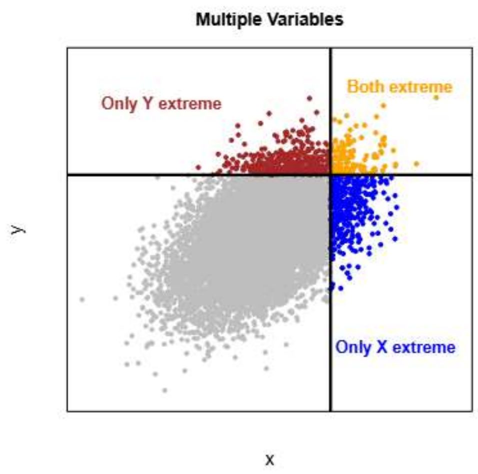


Univariate extremes

- Literature on non-stationary peaks over threshold analysis is also well established.
- Many approaches available: Davison and Smith (1990); Chavez-Demoulin and Davison (2005); Eastoe and Tawn (2009); Youngman (2019).

Multivariate extremes

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

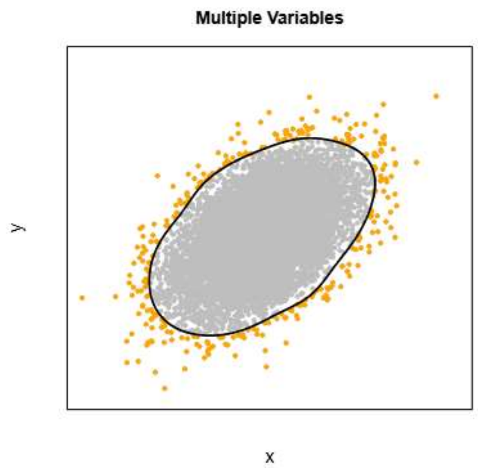


Multivariate extremes

- Various definitions of multivariate extremes have been proposed (Barnett, 1976).
- However, most classical modelling approaches have just 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 have considered *geometric representations* for modelling multivariate extremes.
- Geometric $\Rightarrow$ multivariate data is split into angular and radial components.
- See, for instance, Coles and Tawn (1991), de Haan and de Ronde (1998), Wadsworth et al. (2017) and Simpson and Tawn (2024).

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

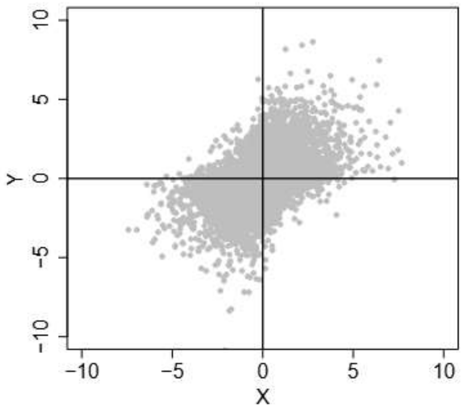
Let (X, Y) denote a continuous random vector. Then

$$R := \sqrt{X^2 + Y^2}$$
$$\Theta := \text{atan2}(Y, X).$$

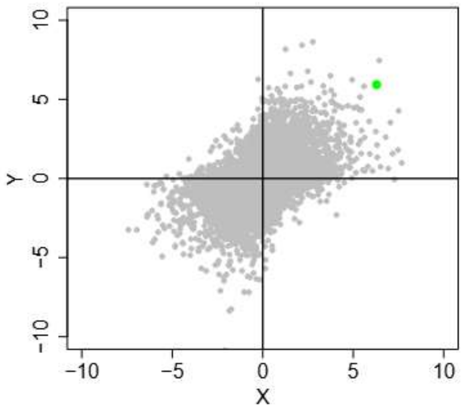
This transformation is one-to-one[1].

[1] Excluding the origin $(0, 0)$.

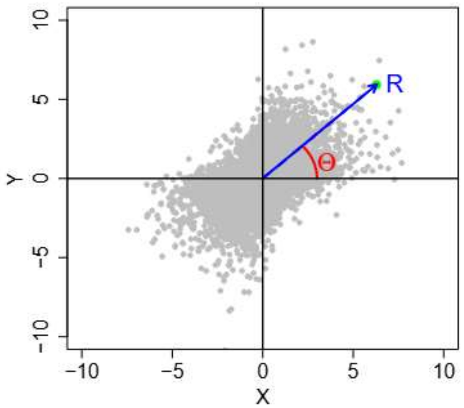
Angular-radial systems



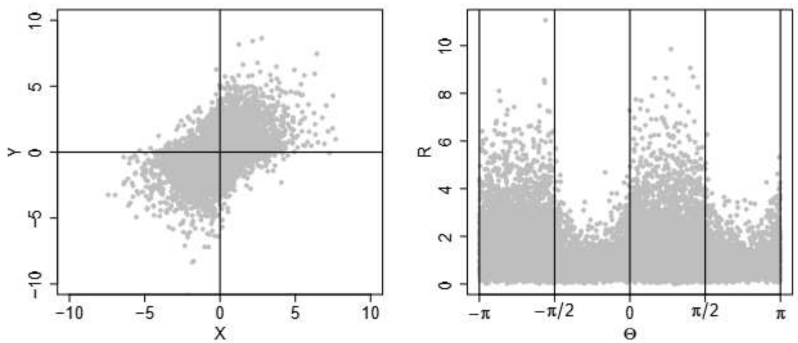
Angular-radial systems



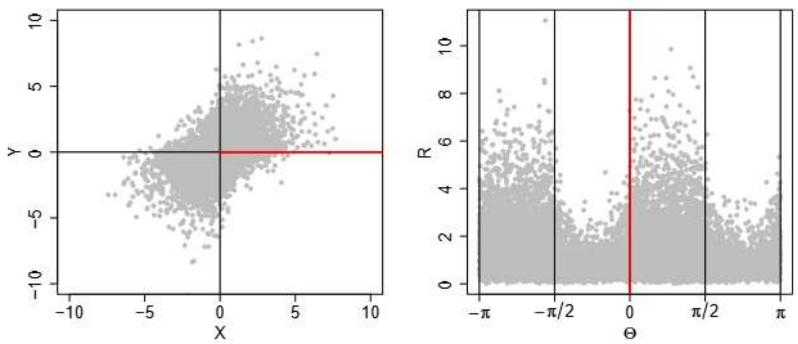
Angular-radial systems



Angular-radial systems



Angular-radial systems



Angular-radial systems

- Joint extremes can be characterised by the **stochastic behaviour** of (R, Θ).
- Θ gives the 'direction' of the event.
- R gives the 'magnitude'.

Angular-radial systems

- One can also consider alternative definitions of ‘distance’.
- For example, we could set

$$R^* := |X| + |Y|$$

  i.e., the $L1$-norm, with an angular measure defined with respect to this norm.
- However, under our modelling assumptions, the choice of angular-radial system is essentially arbitrary; see Mackay and Jonathan (2023) and Murphy-Barltrop et al. (2024).

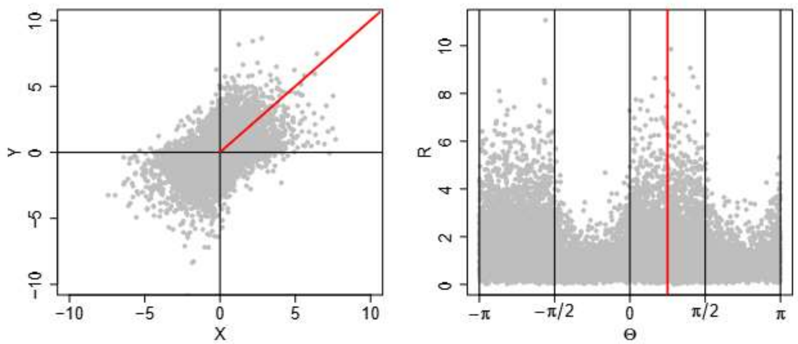
The SPAR model

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

$$(R - u(\theta) \mid \Theta = \theta, R > u(\theta)) \sim \mathrm{GP}(\sigma(\theta), \xi(\theta)),$$

  for sufficiently large $u(\theta)$.
- 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 simply assuming the variable $R \mid (\Theta = \theta)$ 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, through rigorous theoretical treatment, Mackay and Jonathan (2023) show conditional radial tails follow a GP distribution for many popular copula examples.

The SPAR model

Why should this work?

- Mackay and Jonathan (2023) also show that the SPAR model generalises the frameworks of Coles and Tawn (1991), Ledford and Tawn (1996) and Wadsworth et al. (2017).
- It is also linked to limit set theory (e.g., Nolde and Wadsworth, 2022).

The SPAR model

What's the point?

- We have the following relationship between the joint densities:

$$f_{R,\Theta}(r,\theta) = r\ f_{X,Y}(r\cos(\theta), r\sin(\theta)),$$

 where r denotes the Jacobian.
- The joint distribution of (X, Y) can be **obtained directly** from the joint distribution of (R, Θ).

The SPAR model

But what's the point?

- Applying Bayes' theorem, we can write

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

  where $\gamma = \Pr(R \leq u(\theta) \mid \Theta = \theta)$.
- Combined with a model for $f_\Theta(\theta)$, **we can model the joint tail**.

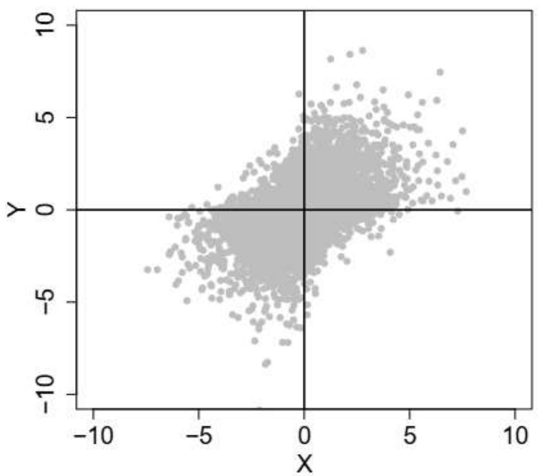
The SPAR model

- Mackay and Jonathan (2023) provide no means for inference with the SPAR model – hence this work.
- We assume semi-parametric **generalised additive model** forms for $u(\theta), \sigma(\theta)$ and $\xi(\theta)$.
- This allows us to approximate each function with minimal modelling assumptions.
- $u(\theta)$ fitted via quantile regression techniques (Koenker et al., 2017).
- $\sigma(\theta), \xi(\theta)$ fitted using the EVGAM framework (Youngman, 2019, 2020).

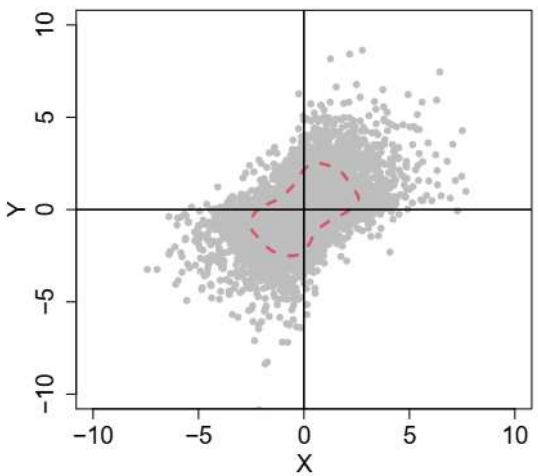
The SPAR model

- Finally, we model $f_{\Theta}(\theta)$ non-parametrically using **circular density estimation techniques**.
- See Murphy-Barltrop et al. (2024) for further details and discussion.

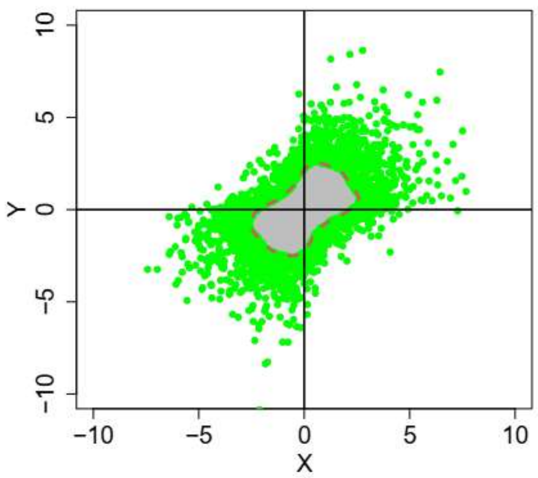
The SPAR model



The SPAR model



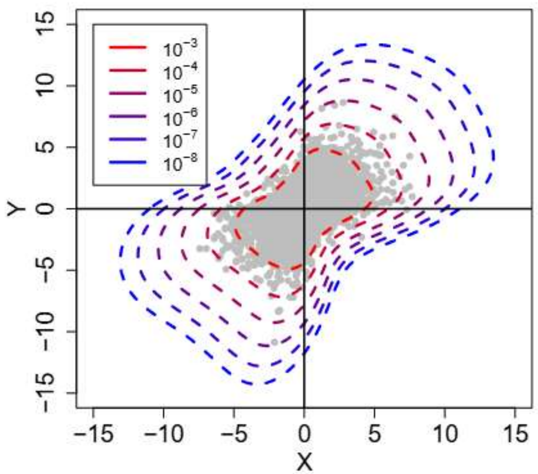
The SPAR model



The SPAR model

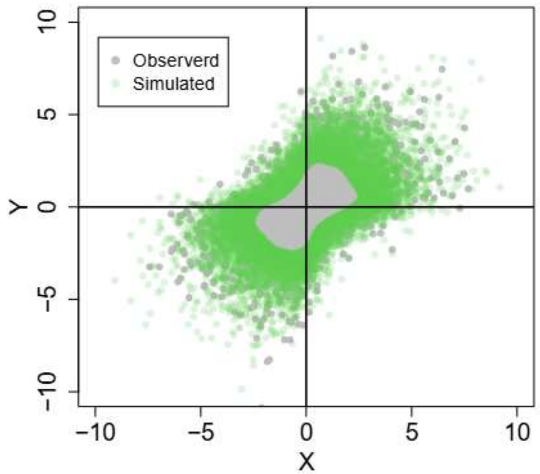
For any small value $\epsilon > 0$, isodensity contours are given by

$$\{(x, y) : f_{X,Y}(x, y) = \epsilon\}$$



The SPAR model

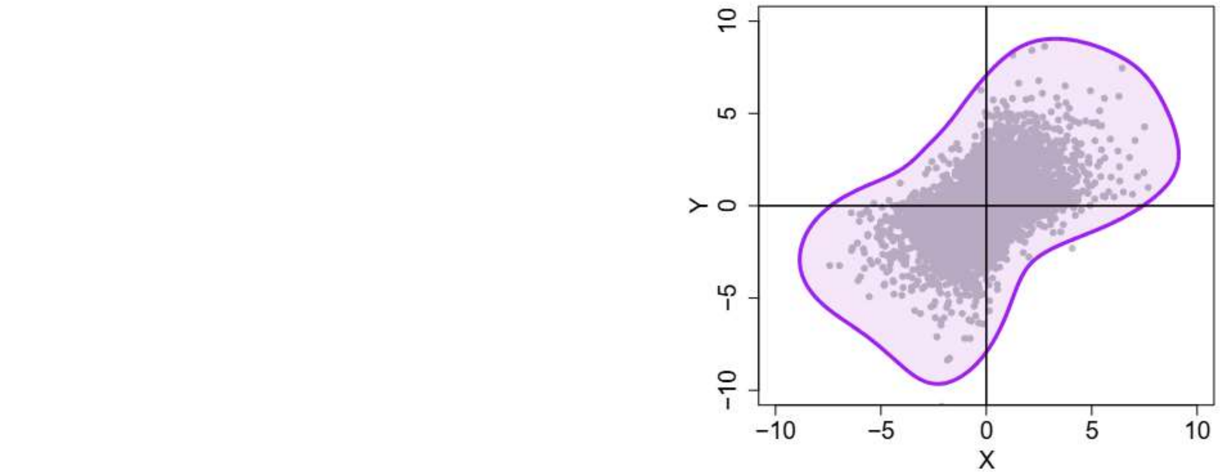
SPAR model simulations



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

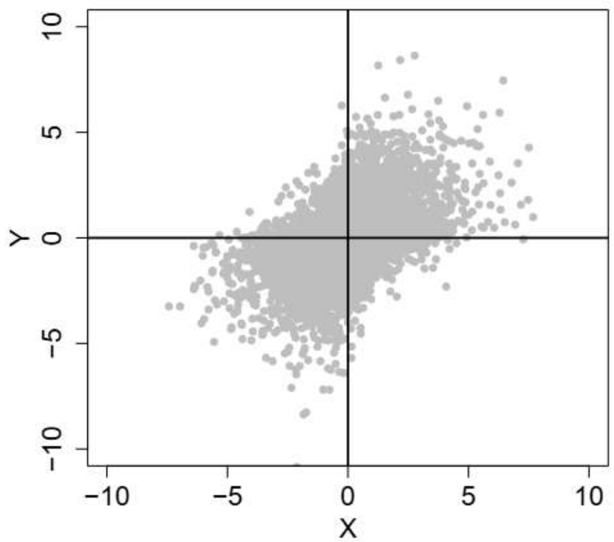
$$\Pr[(X, Y) \in \mathcal{A}_p] = p \Rightarrow \Pr[(X, Y) \in \mathcal{A}_p^c] = 1 - p.$$



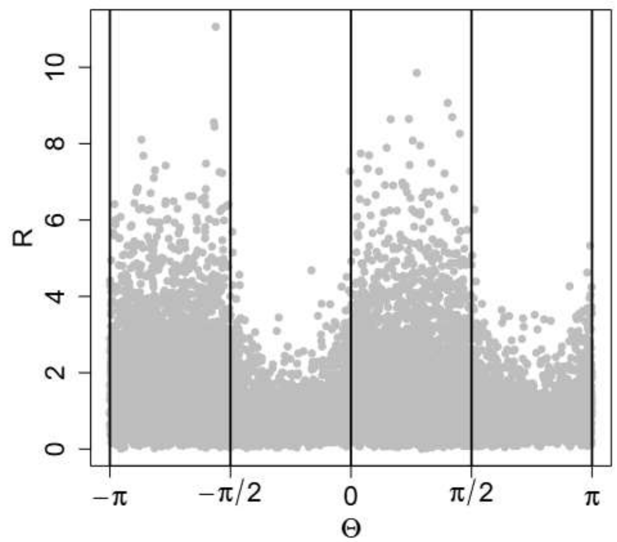
The SPAR model

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

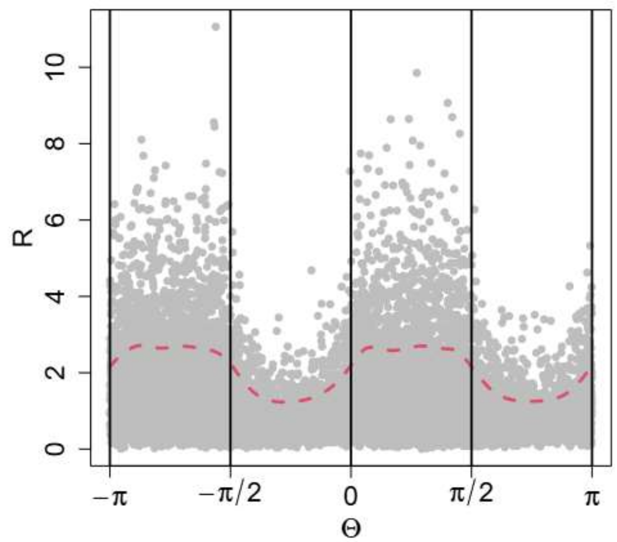
The SPAR model



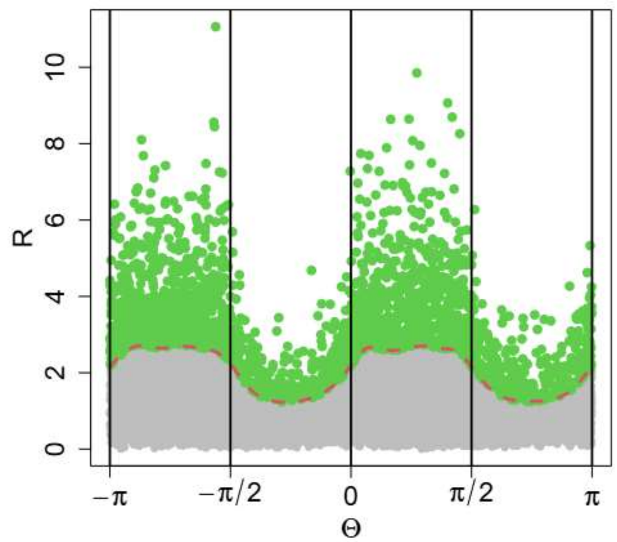
The SPAR model



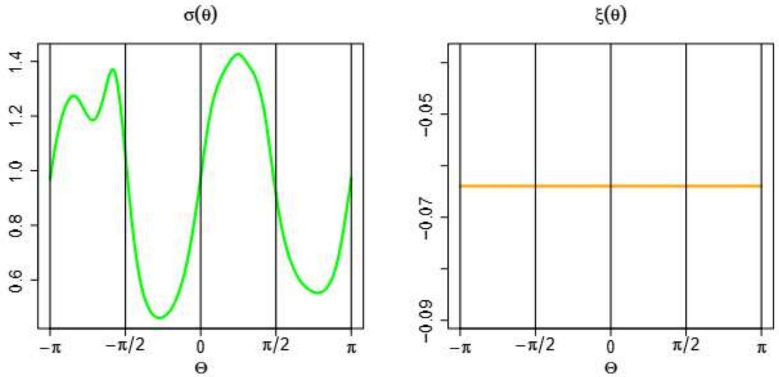
The SPAR model



The SPAR model



The SPAR model

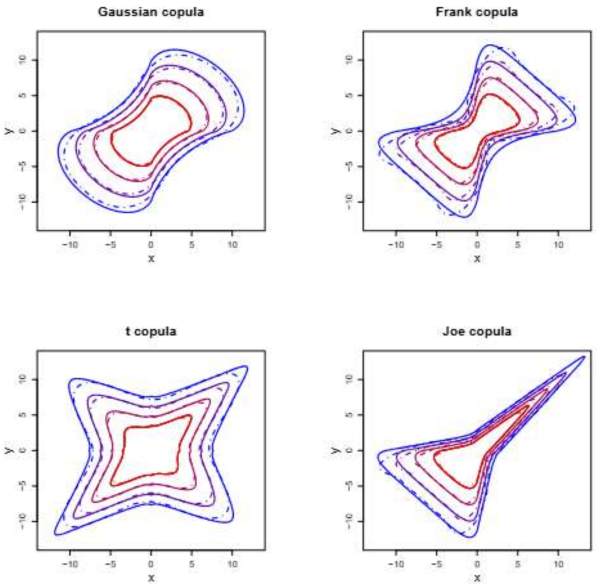


The SPAR model

Simulation study

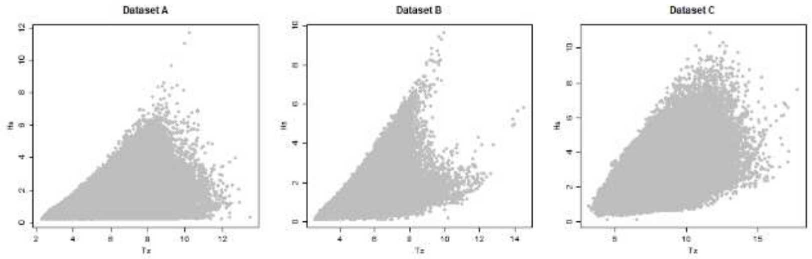
- To evaluate the performance of the SPAR model, we tested our framework over a range of popular copula examples.
- We were able to **accurately approximate** the joint tail distributions.

Simulation study



Case Study

- 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.



Case Study

- 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**.

Case Study

- 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.

Case Study

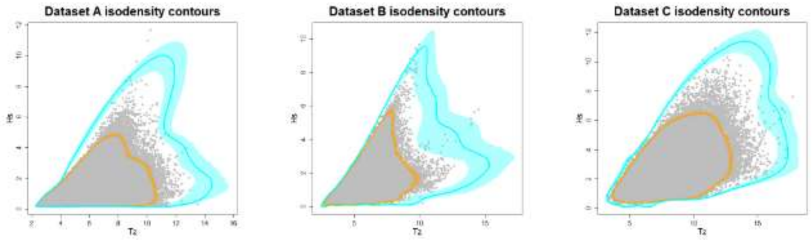


Figure: Isodensity contours at $\epsilon = 10^{-3}$ (orange lines) and $\epsilon = 10^{-6}$ (cyan lines).

Case Study

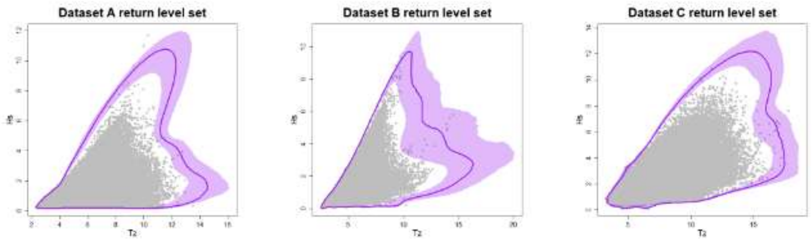


Figure: Estimated 10 year return level sets.

Discussion

- In this work we have introduced a **novel semi-parametric modelling framework** for bivariate extremes.
- Can capture the joint extremes of both asymptotically and non-asymptotically dependent data structures.
- Requires **very few assumptions** and offers a high degree of flexibility.
- We have also proposed a range of diagnostic and uncertainty quantification tools to aid with inference.

Discussion

Future work

- Further investigating sensitivity to tuning parameter choices (i.e., GAMs, circular density).
- Expanding to the general multivariate setting (ongoing work).
- Comparison to other multivariate modelling techniques.

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. arXiv preprint, arXiv:2401.07259.

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.

Thanks for listening

Any questions?

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.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 53:377–392.

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.

References II

de Haan, L. and de Ronde, J. (1998). Sea and Wind: Multivariate Extremes at Work. *Extremes*, 1:7–45.

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.

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

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

Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2024). Inference for multivariate extremes via a semi-parametric angular-radial model. *arXiv*, 2401.07259.

References III

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54:688–717.

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

Simpson, E. S. and Tawn, J. A. (2024). Inference for New Environmental Contours Using Extreme Value Analysis. *Journal of Agricultural, Biological and Environmental Statistics*.

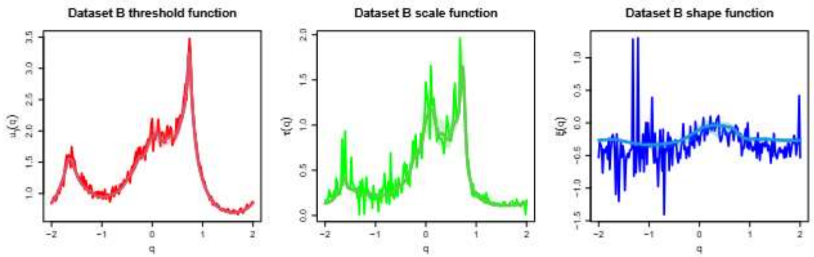
Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79:149–175.

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

References IV

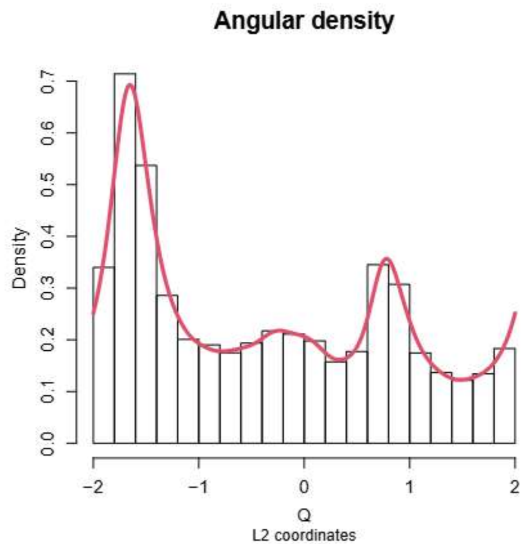
Youngman, B. D. (2019). Generalized Additive Models for Exceedances of High Thresholds With an Application to Return Level Estimation for U.S. Wind Gusts. *Journal of the American Statistical Association*, 114:1865–1879.

Appendix



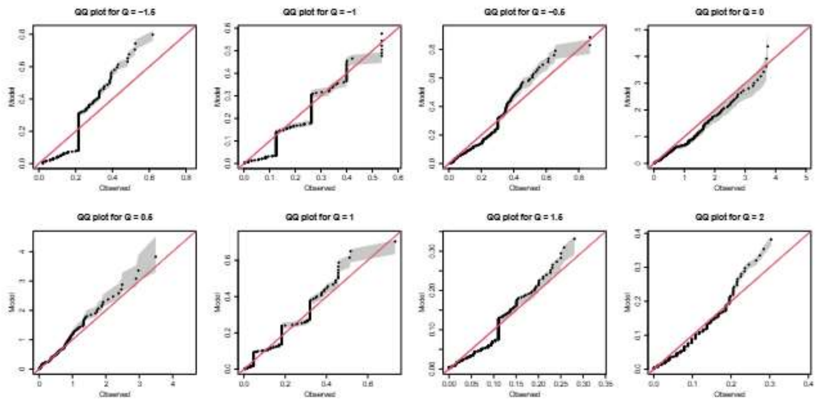
Dataset B threshold function
Dataset B scale function
Dataset B shape function

Appendix



Angular density
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−2
−1
0
1
2
Q
L2 coordinates

Appendix


QQ plot for Q = −1.5
QQ plot for Q = −1
QQ plot for Q = −0.5
QQ plot for Q = 0
QQ plot for Q = 0.5
QQ plot for Q = 1
QQ plot for Q = 1.5
QQ plot for Q = 2
Model
Observed