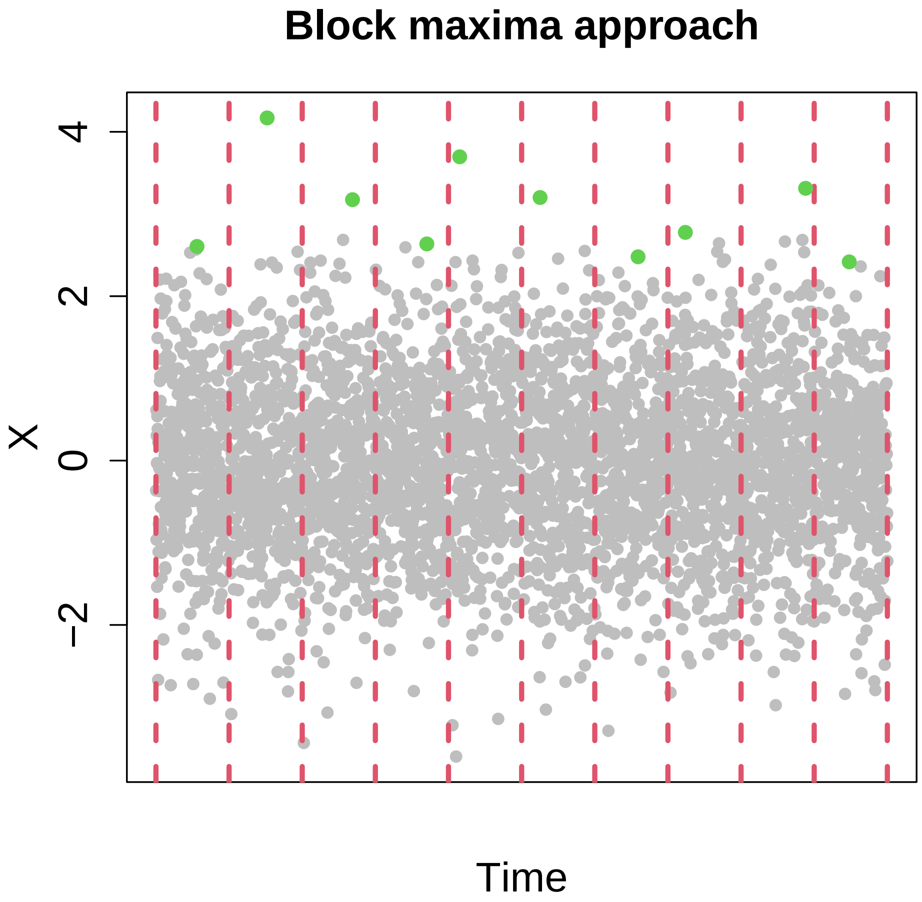
Practical methodologies for multivariate extremes

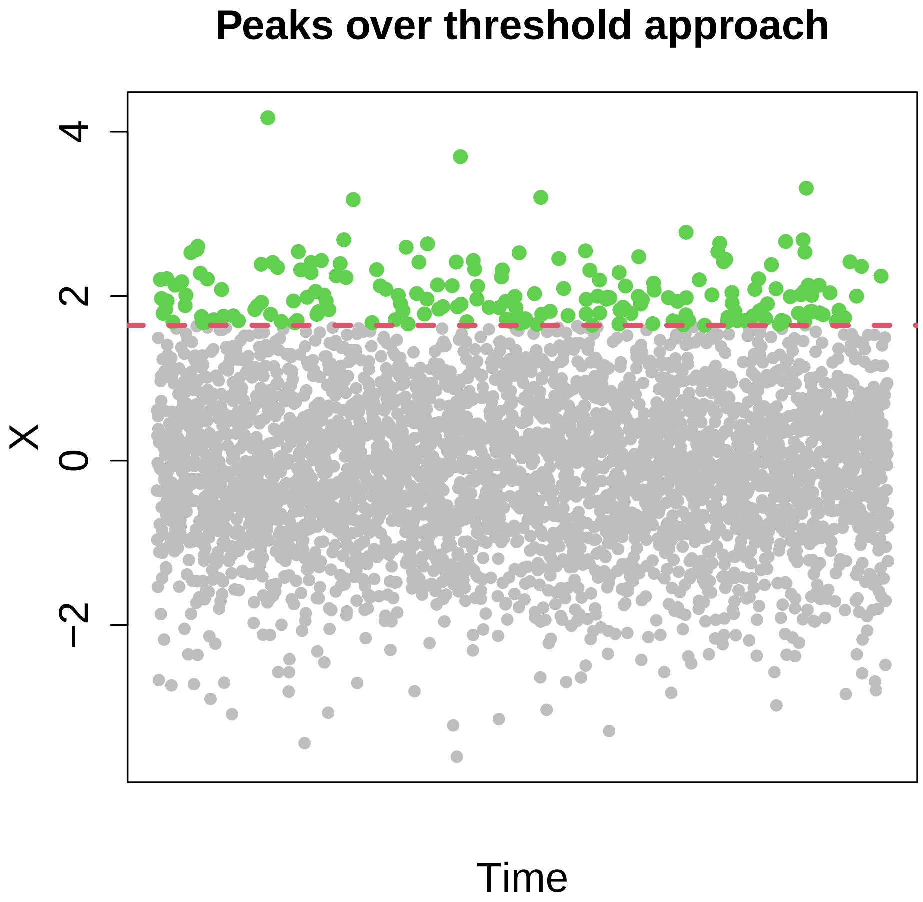
Jordan Richards

With talks from Jenny Wadsworth, Emma Simpson, and Ed Mackay

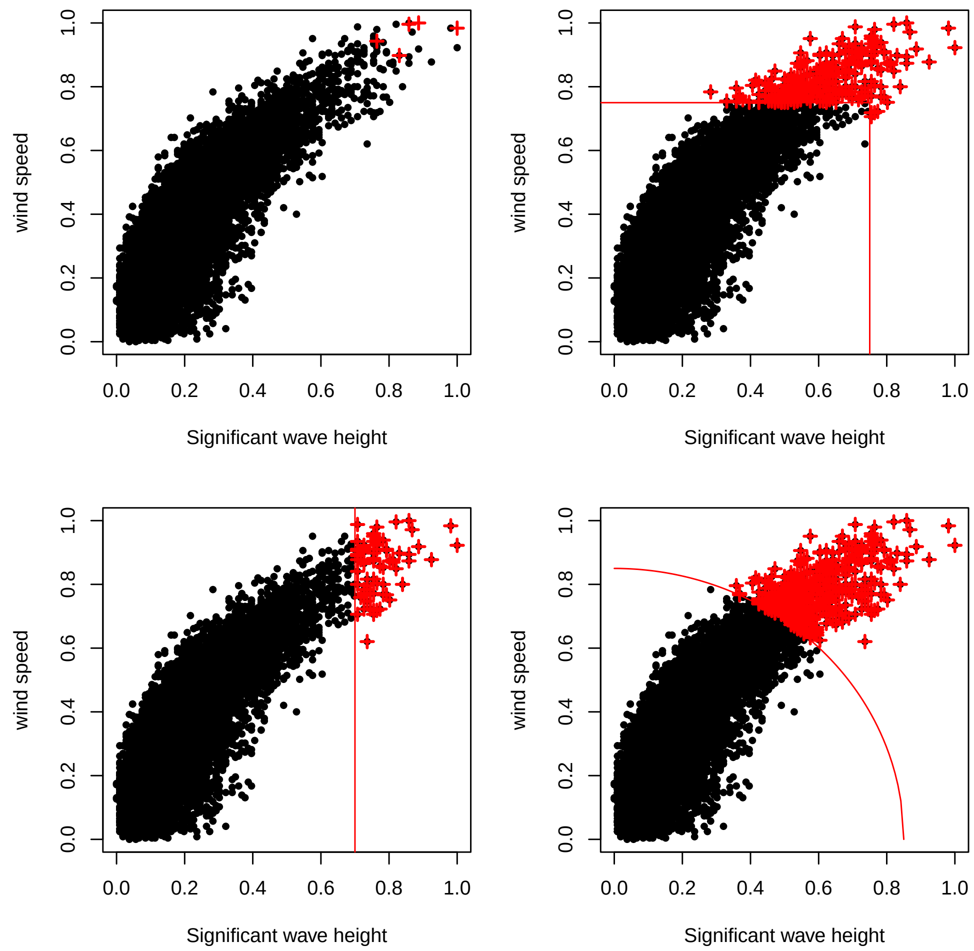
RSS 2025

Defining multivariate extremes



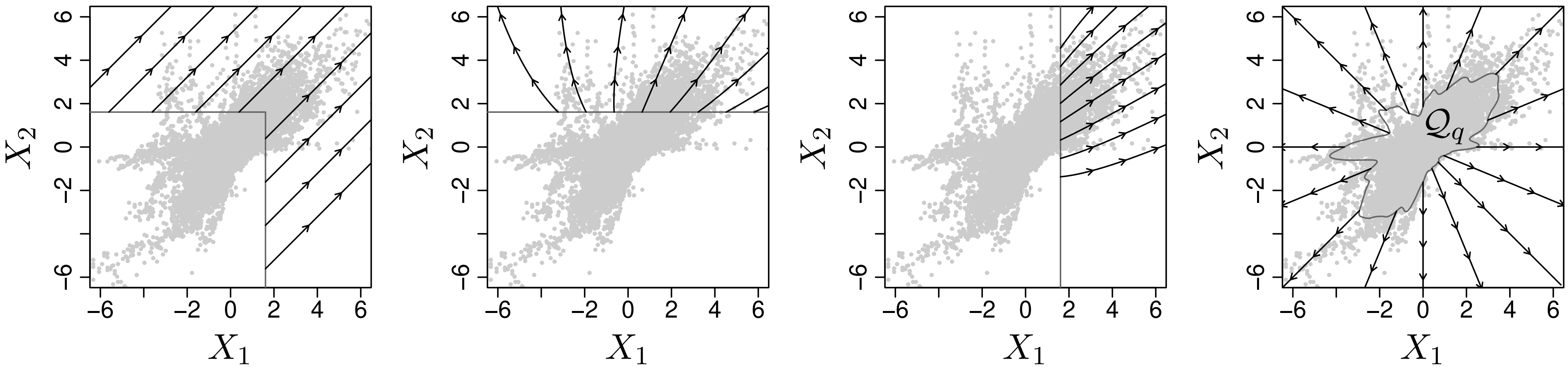


Defining multivariate extremes



Gimeno-Sotelo, L., Richards, J., Hazra, A., Mhalla, L., and de Zea Bermudez, P. (2025). A Review of Applications of Extreme Value Theory to Environmental Risk Assessment. In *Environmental Statistics: Innovative Methods and Applications*.

Extrapolation

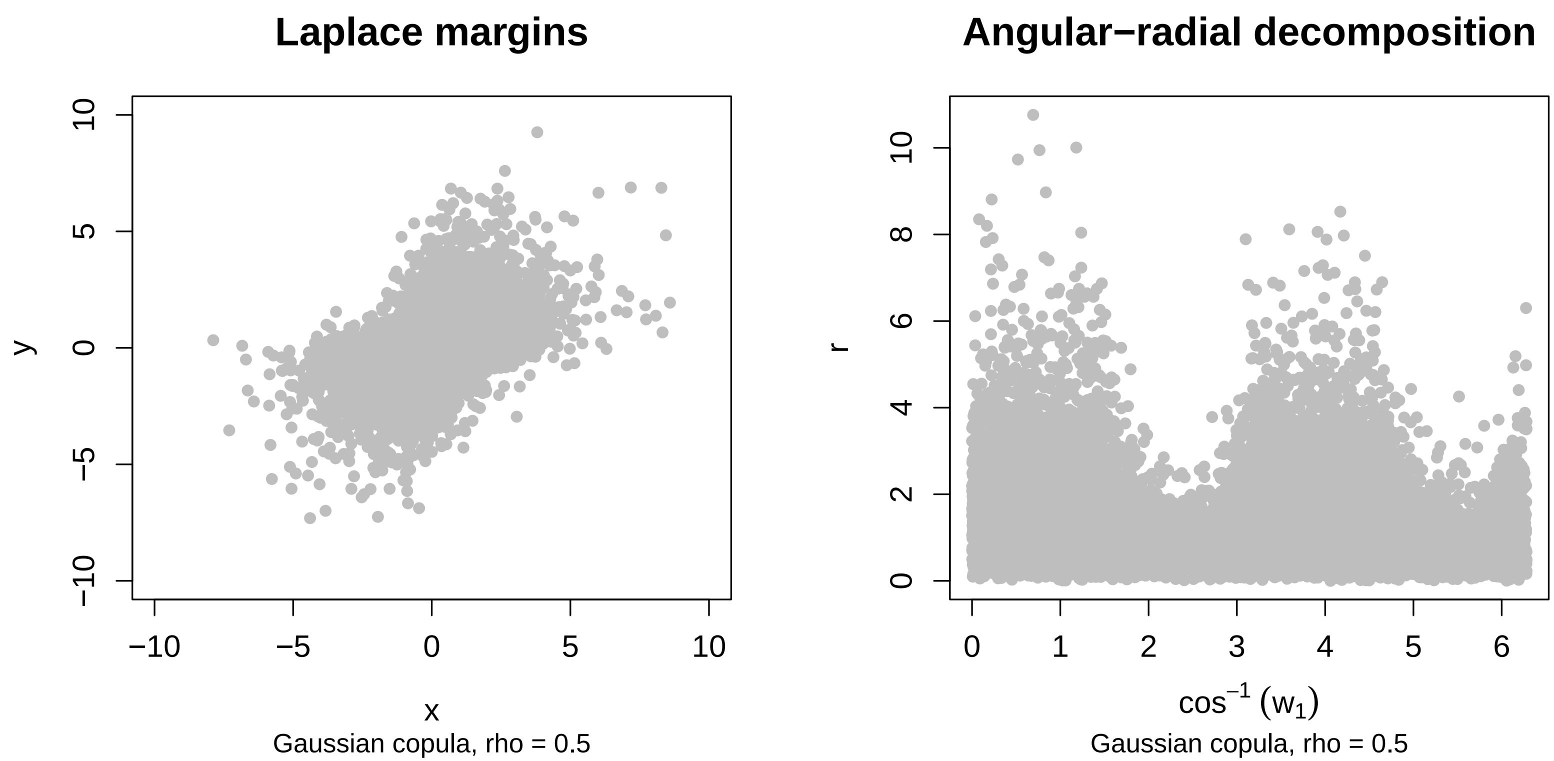


Papastathopoulos, I., De Monte, L., Campbell, R., & Rue, H. (2023). Statistical inference for radially-stable generalized Pareto distributions and return level-sets in geometric extremes. arXiv:2310.06130.

Angular-radial representations

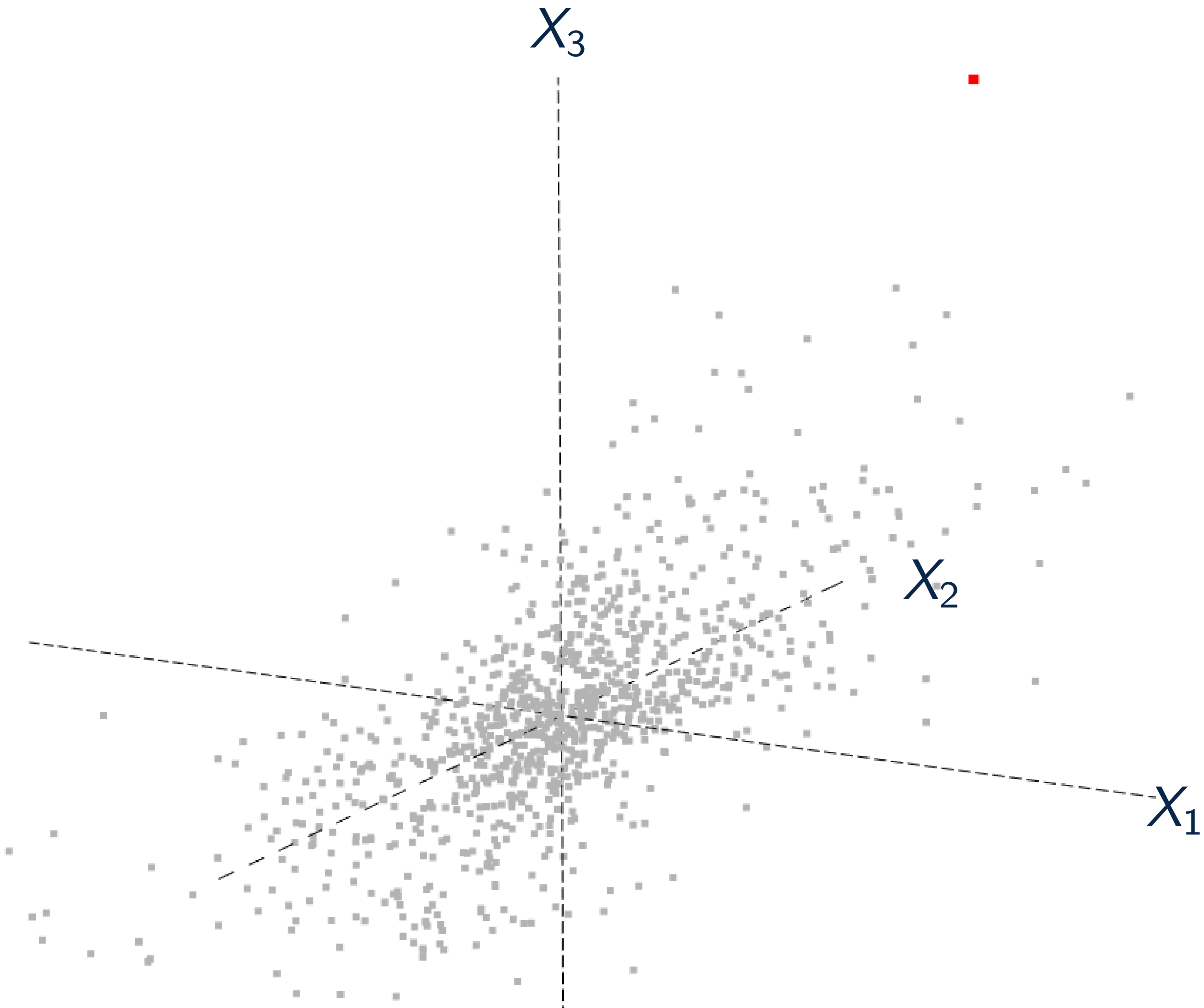
Let $\mathbf{X} \in \mathbb{R}^d$ and

$$R := \|\mathbf{X}\| > 0, \qquad \mathbf{W} := \frac{\mathbf{X}}{\|\mathbf{X}\|} \in \mathcal{S}^{d-1}.$$

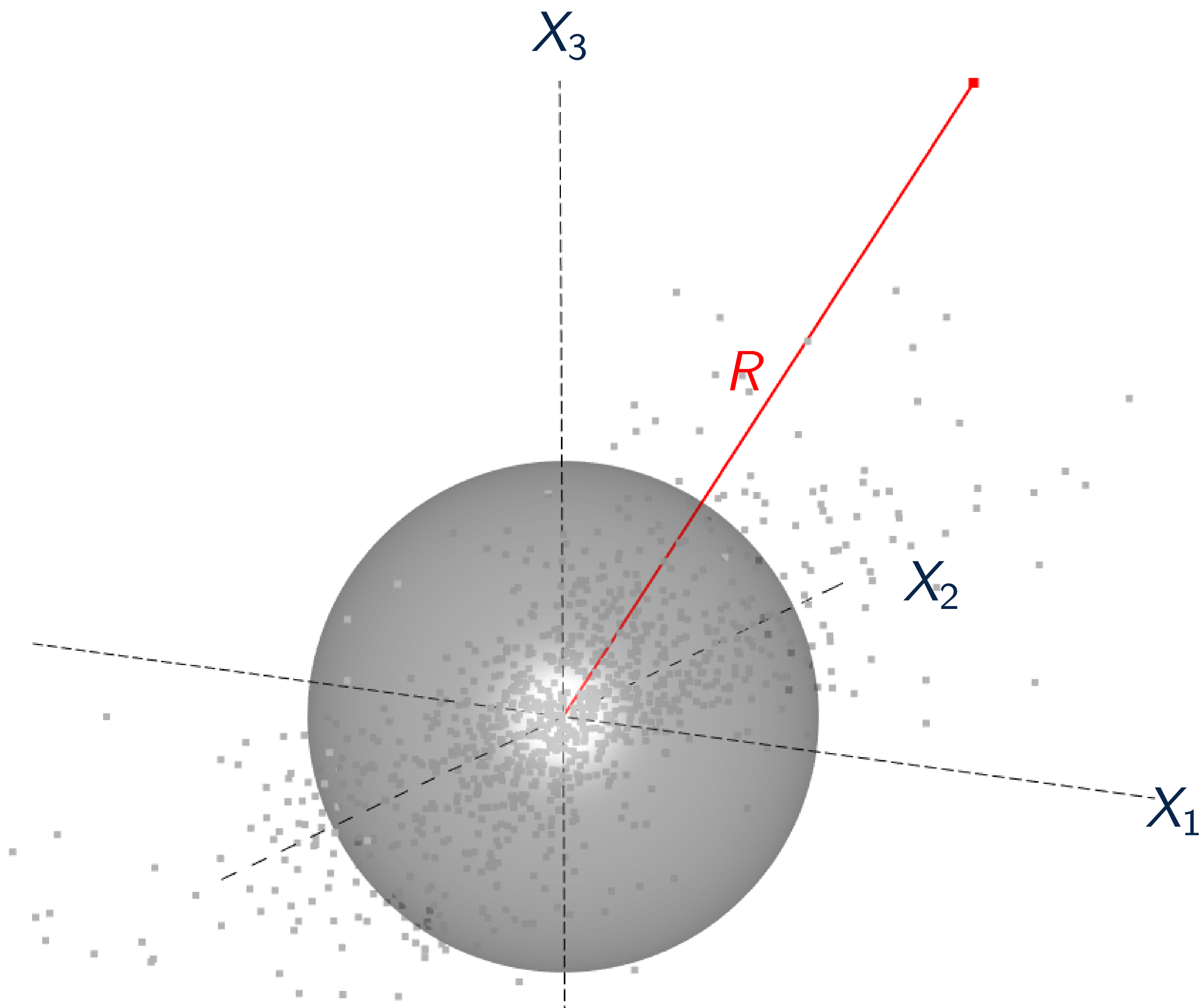


Murphy-Barltrop, C. J., Majumder, R., & Richards, J. (2024). Deep learning of multivariate extremes via a geometric representation. arXiv:2406.19936.

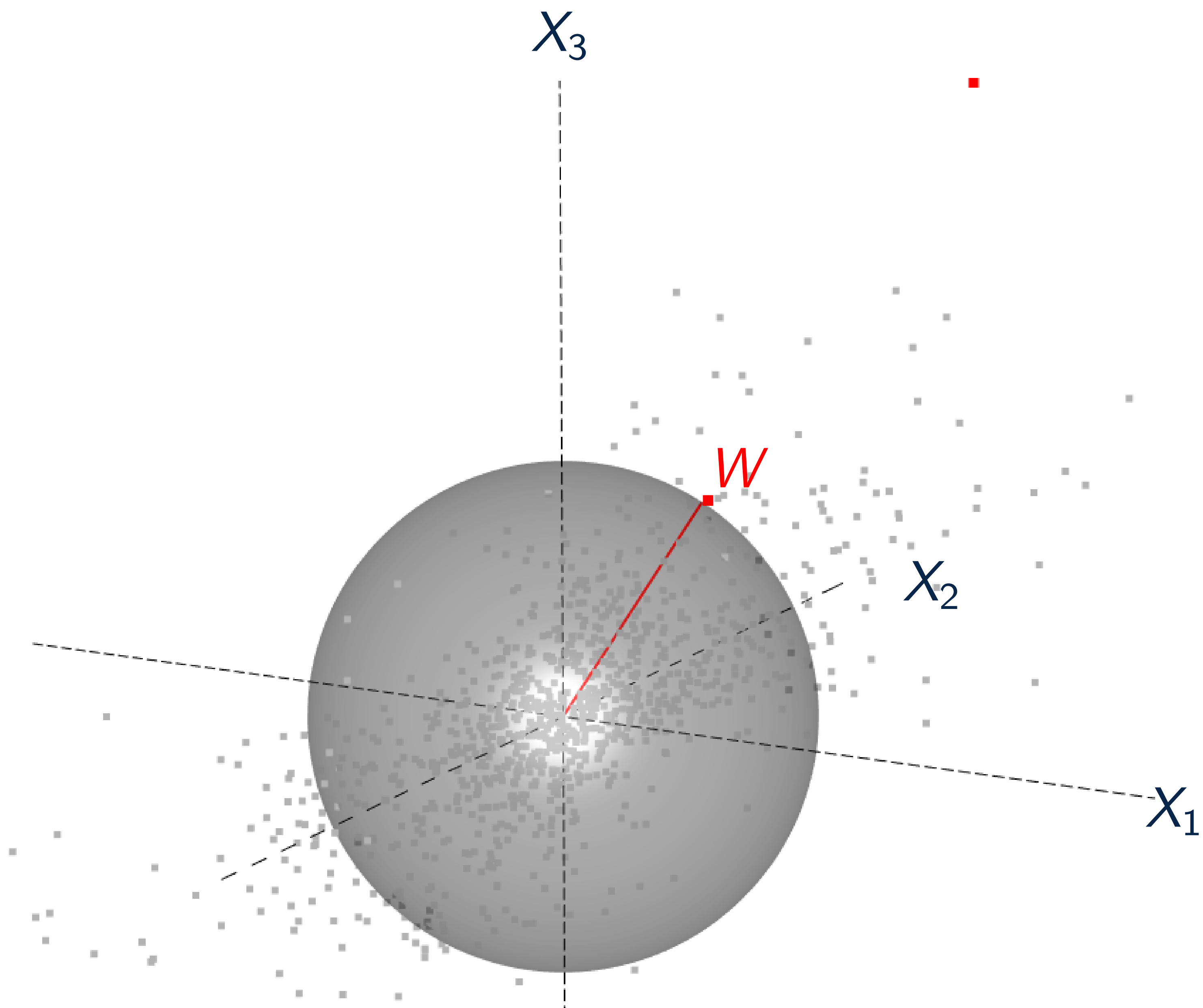
Angular-radial representations



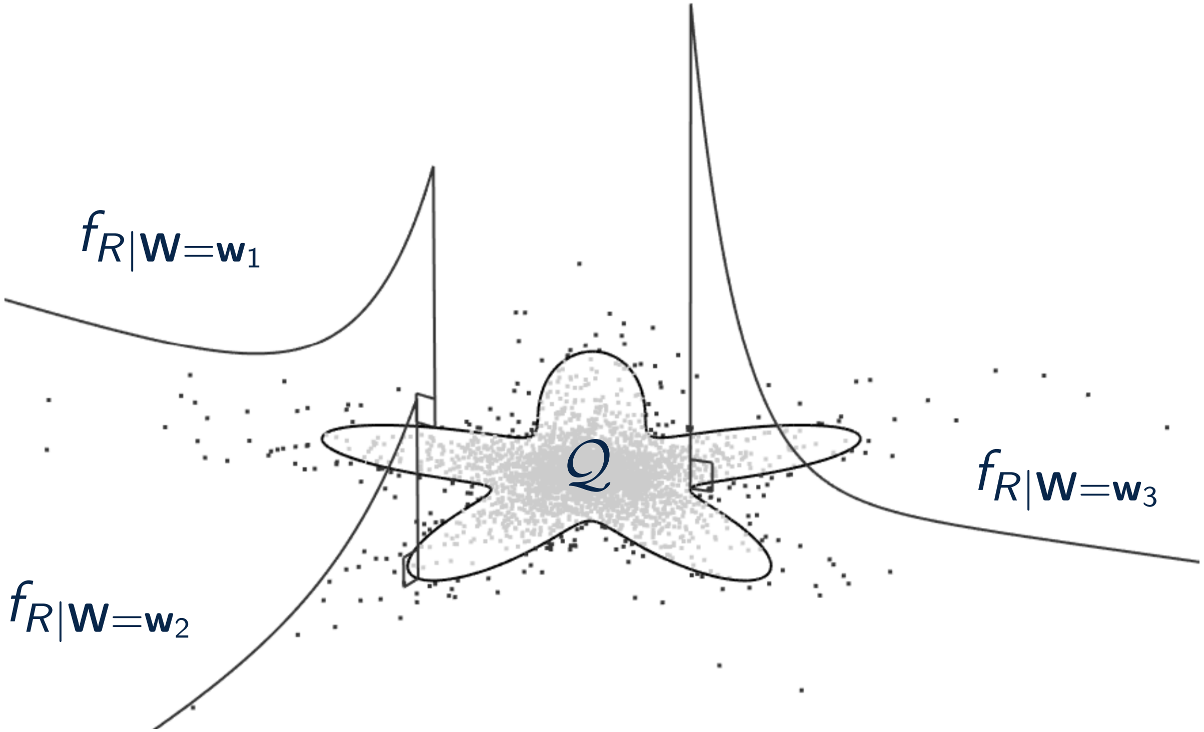
Angular-radial representations



Angular-radial representations



Angular-radial representations



De Monte, L., Huser, R., Papastathopoulos, I., & Richards, J. (2025). Generative modelling of multivariate geometric extremes using normalising flows. arXiv:2505.02957.

Recent developments

- Limit sets and scaled sample clouds:
  - Theory - Balkema and Nolde (2010); Nolde (2014); Nolde and Wadsworth (2022)
  - Inference - Campbell and Wadsworth (2024); Simpson and Tawn (2024a,b); Wadsworth and Campbell (2024); Murphy-Barltrop et al. (2024); Majumder et al. (2025)
- Semi-Parametric Angular-Radial model: Mackay and Jonathan (2023); Mackay et al. (2025a,b); Murphy-Barltrop et al. (2025)
- Radially-stable distributions: Papastathopoulos et al. (2023); De Monte et al. (2025)
- Angular modelling via generative AI - Lhaut et al. (2025); Wessel et al. (2025)

Session outline

- Jenny Wadsworth (Lancaster): *Methodologies for angular-radial models of multivariate extremes*
- Emma Simpson (UCL): *Application of angular-radial models*
- Ed Mackay (Exeter): *Deep learning for multivariate extremes*

Organisers: Callum Murphy-Barltrop (TU Dresden), Ed Mackay (Exeter), Jordan Richards (UoE), Phil Jonathan (Lancaster)

References I

Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.

Campbell, R. and Wadsworth, J. (2024). Piecewise-linear modeling of multivariate geometric extremes. *preprint arXiv:2412.05195*.

De Monte, L., Huser, R., Papastathopoulos, I., and Richards, J. (2025). Generative modelling of multivariate geometric extremes using normalising flows. *preprint arXiv:2505.02957*.

Lhaut, S., Rootzén, H., and Segers, J. (2025). Wasserstein-aitchison GAN for angular measures of multivariate extremes. *arXiv preprint arXiv:2504.21438*.

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

References II

Mackay, E., Murphy-Barltrop, C., and Jonathan, P. (2025a). The SPAR model: A new paradigm for multivariate extremes: Application to joint distributions of metocean variables. *Journal of Offshore Mechanics and Arctic Engineering*, 147(1):011205.

Mackay, E., Murphy-Barltrop, C. J., Richards, J., and Jonathan, P. (2025b). Deep learning of joint extremes of metocean variables using the SPAR model. *preprint arXiv:2412.15808*.

Majumder, R., Shaby, B. A., Reich, B. J., and Cooley, D. S. (2025). Semiparametric estimation of the shape of the limiting bivariate point cloud. *Bayesian Analysis*, 1(1):1–27.

Murphy-Barltrop, C. J., Majumder, R., and Richards, J. (2024). Deep learning of multivariate extremes via a geometric representation. *preprint arXiv:2406.19936*.

References III

Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2025). Inference for bivariate extremes via a semi-parametric angular-radial model. *Extremes*, 28(2):209–238.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717.

Papastathopoulos, I., De Monte, L., Campbell, R., and Rue, H. (2023). Statistical inference for radially-stable generalized Pareto distributions and return level-sets in geometric extremes. *preprint arXiv:2310.06130*.

Simpson, E. S. and Tawn, J. A. (2024a). Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties. *Electronic Journal of Statistics*, 18(2):4582–4611.

References IV

Simpson, E. S. and Tawn, J. A. (2024b). Inference for new environmental contours using extreme value analysis. *Journal of Agricultural, Biological and Environmental Statistics*. To appear.

Wadsworth, J. L. and Campbell, R. (2024). Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1243–1265.

Wessel, J. B., Murphy-Barltrop, C. J., and Simpson, E. S. (2025). A comparison of generative deep learning methods for multivariate angular simulation. *arXiv preprint arXiv:2504.21505*.

Analysing multivariate extremes through a geometric approach

Jenny Wadsworth
Lancaster University, UK

RSS conference, Edinburgh, 2025

Background: statistics of multivariate extremes

Main premise of extreme value statistics:

Link probabilities of being in **more extreme regions** to probabilities of being in **less extreme regions**

Challenges in $\mathbb{R}^d$:

- No unique direction of extrapolation
- Different methodologies useful for different dependence structures
- No established methodology fully general for $d \geq 3$

Brief overview of difficulties in multivariate extremes

Main question for finding appropriate multivariate extremes methodology:

Which variables can be large together?

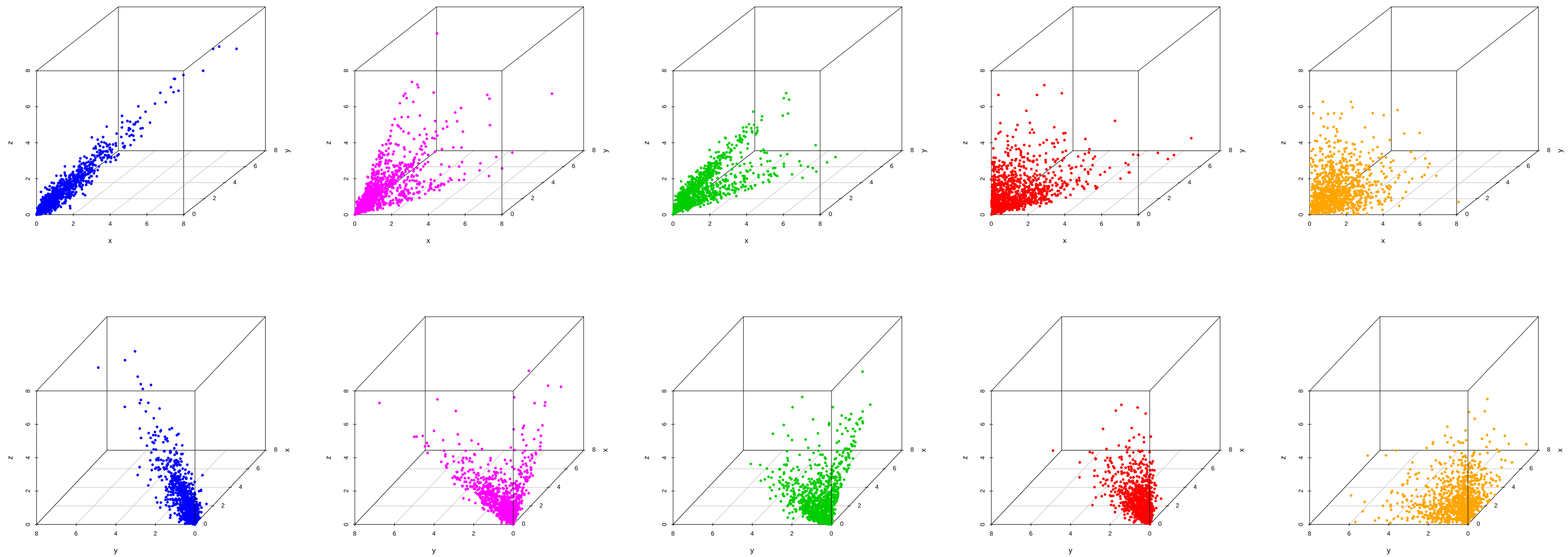
That is: for which groups of variables do **joint extremes** occur with similar frequency to **marginal extremes**.

Brief overview of difficulties in multivariate extremes

Main question for finding appropriate multivariate extremes methodology:

Which variables can be large together?

That is: for which groups of variables do **joint extremes** occur with similar frequency to **marginal extremes**.



Brief overview of difficulties in multivariate extremes

Main question for finding appropriate multivariate extremes methodology:

Which variables can be large together?

That is: for which groups of variables do **joint extremes** occur with similar frequency to **marginal extremes**.

Can be summarized by a mapping to the faces of a $(d-1)$ dimensional simplex:

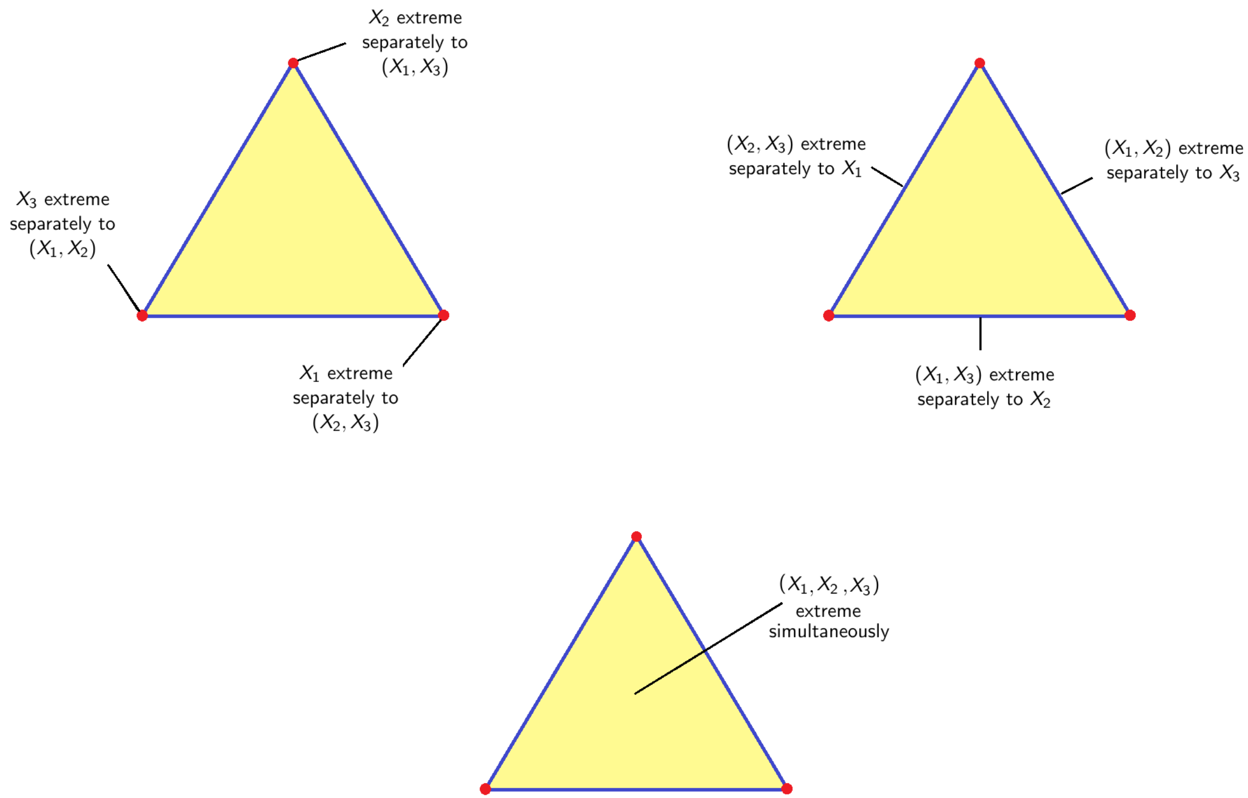
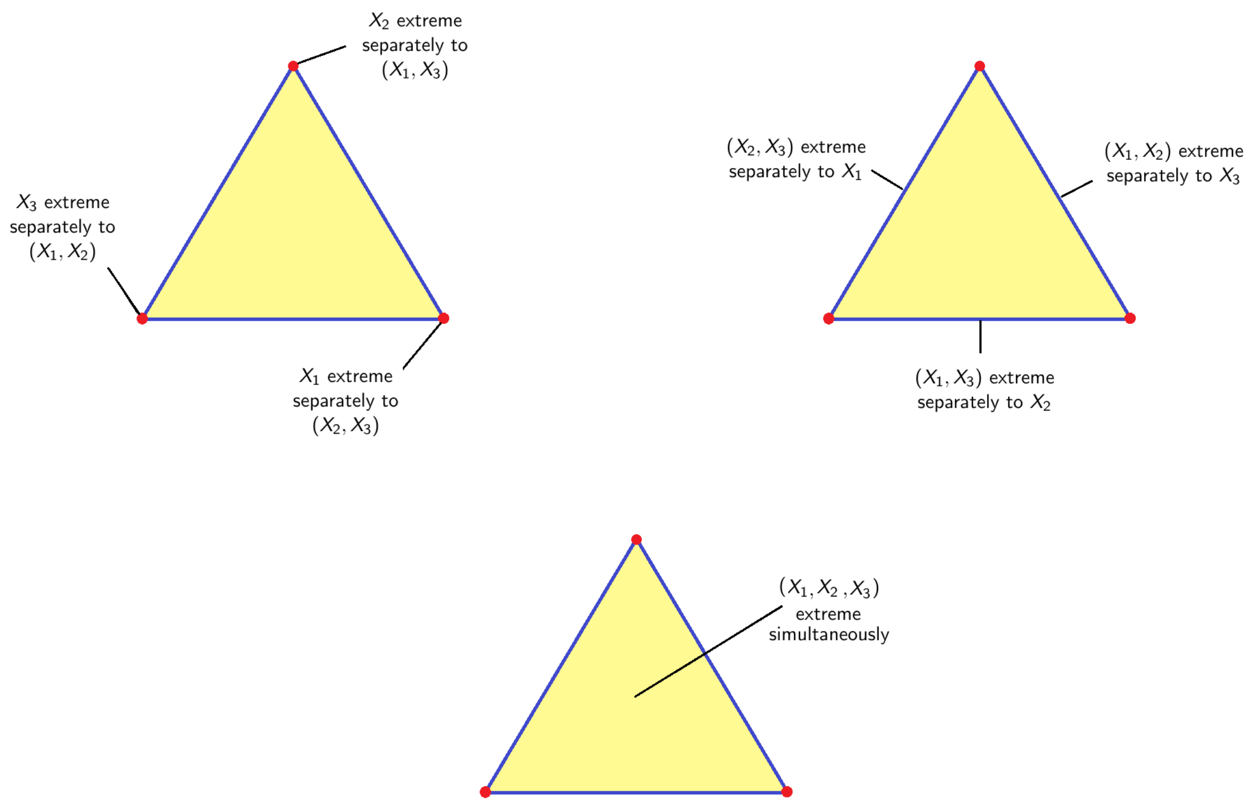


Image credit: Emma Simpson

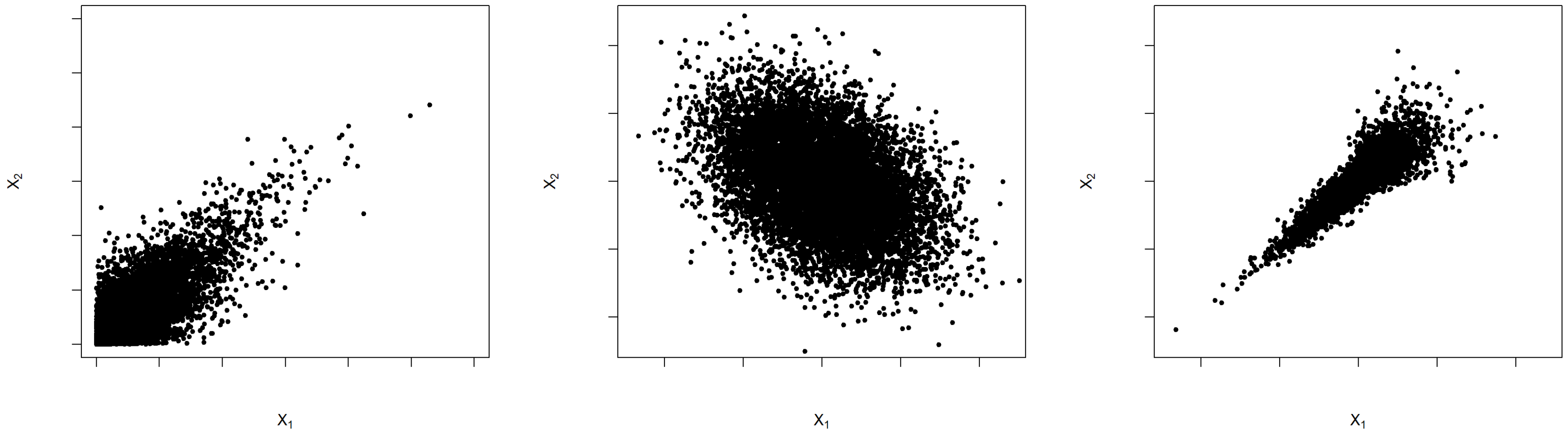
Brief overview of difficulties in multivariate extremes



- $2^d - 1 = 7$ possibilities for $d = 3$, **but** many can occur in combination
- Out of $2^7 = 128$ possibilities for whether a certain combination occurs or not, there are only 19 that cannot occur, leaving a mere 109...
- No general methodology to date to allow for all these structures

Geometric approach to multivariate extremes

When we view a scatterplot of datapoints in light-tailed margins, can often see a "shape" in this sample cloud.



We can formalize this using the notion of the **scaled sample cloud** and its **limit set**. (Davis et al., 1988; Kinoshita & Resnick, 1991; Balkema et al., 2010)

The shape of the limit set contains a lot of information on the extremal dependence structure (Nolde & Wadsworth, 2022).

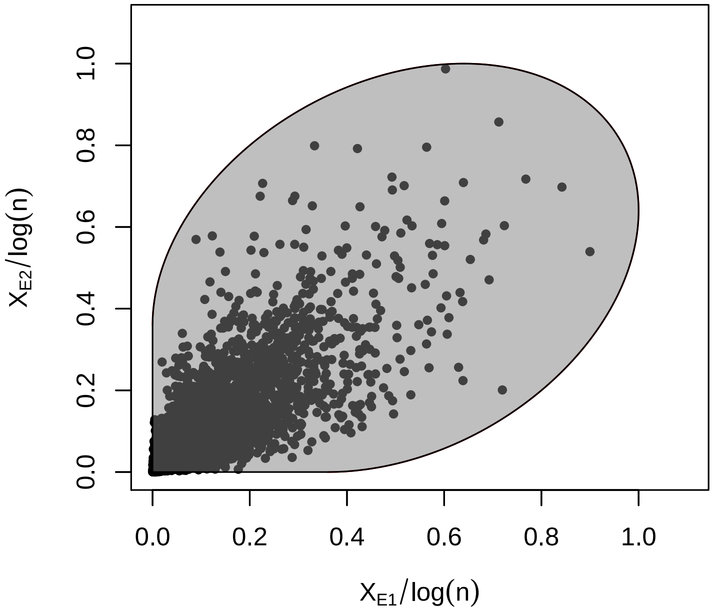
Asymptotic behaviour of scaled sample cloud

Sample clouds and limit sets

The shape of the sample cloud and limit set is impacted by both margins and dependence. To focus on dependence only, we **standardize to exponential margins**, $\boldsymbol{X}_{E,i}$, or **Laplace margins** $\boldsymbol{X}_{L,i}$, $i = 1, \ldots, n$.

Define the **scaled sample cloud** to be

$$N_n = \left(\frac{\boldsymbol{X}_{E,1}}{\log n}, \ldots, \frac{\boldsymbol{X}_{E,n}}{\log n} \right).$$

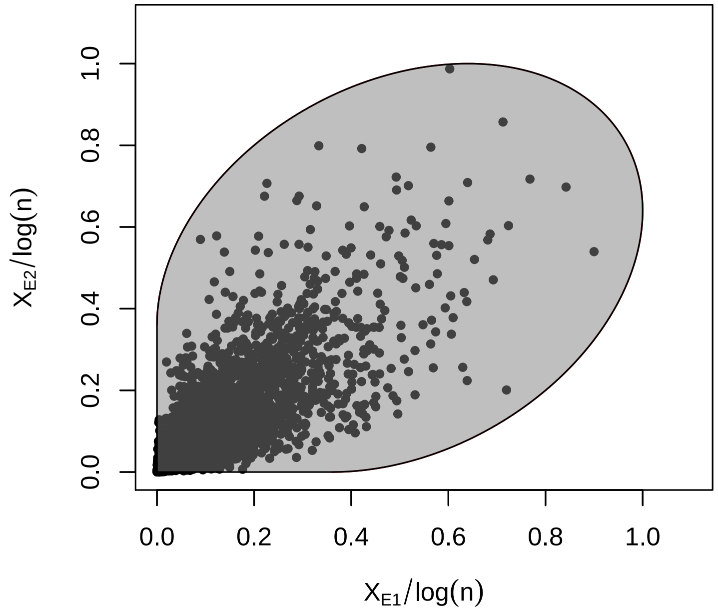


Sample clouds and limit sets

The shape of the sample cloud and limit set is impacted by both margins and dependence. To focus on dependence only, we **standardize to exponential margins**, $\boldsymbol{X}_{E,i}$, or **Laplace margins** $\boldsymbol{X}_{L,i}$, $i = 1, \ldots, n$.

Define the **scaled sample cloud** to be

$$N_n = \left(\frac{\boldsymbol{X}_{E,1}}{\log n}, \ldots, \frac{\boldsymbol{X}_{E,n}}{\log n} \right).$$



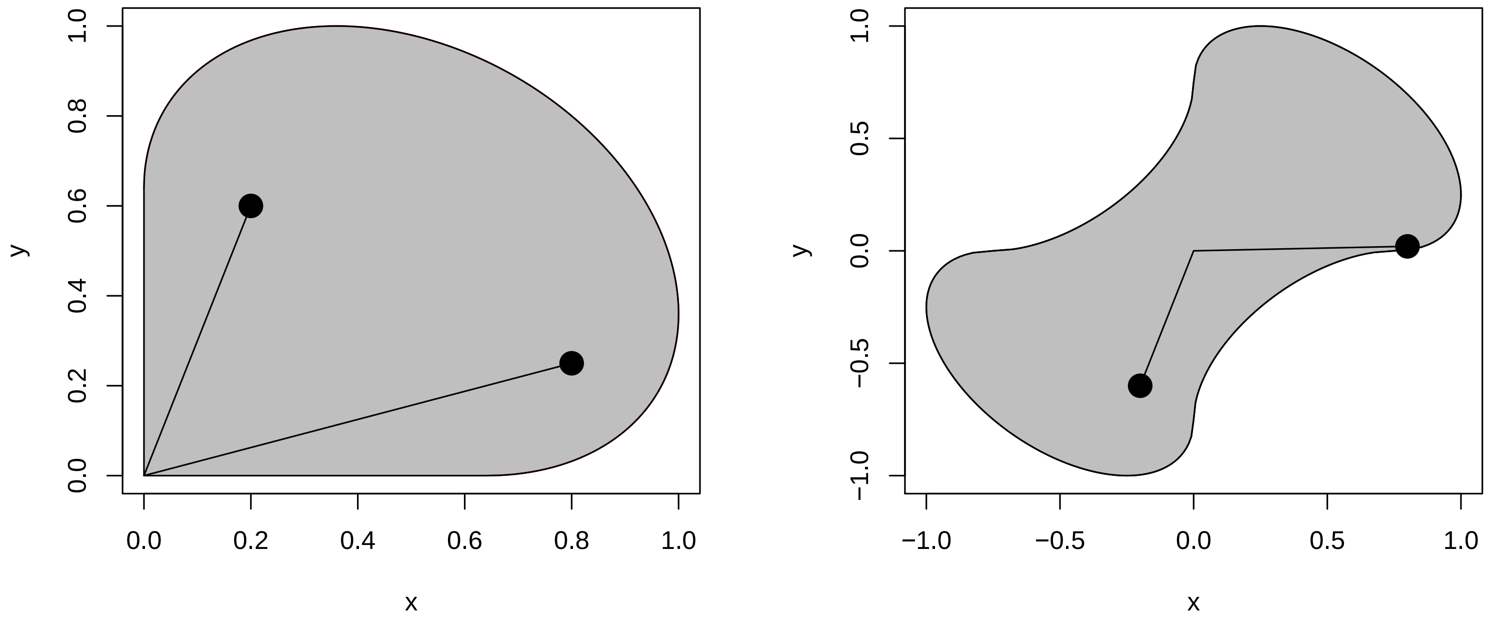
This converges (in probability) onto a **limit set** G if

1. $n \Pr(\frac{\boldsymbol{X}_E}{\log n} \in U^c) \to 0$ for any open set U containing G
2. $n \Pr(\frac{\boldsymbol{X}_E}{\log n} \in \{\boldsymbol{x} + \epsilon B\}) \to \infty$ for all $\boldsymbol{x} \in G$, any $\epsilon > 0$, B Euclidean unit ball

$\max_{1 \le i \le n}(X_{E,i}) / \log n \xrightarrow{p} 1$, so the coordinatewise supremum of G is $(1, \ldots, 1)$.

The limit set, G

The limit set G is **star-shaped**: $\boldsymbol{x} \in G \Rightarrow t\boldsymbol{x} \in G$ for $t \in (0, 1)$.



If for every $\boldsymbol{x} \in G$, the line segment $\mathbf{0} + t\boldsymbol{x}$, $t \in [0, 1)$, lies in the interior of G, then G can be characterized by a continuous **gauge function**, g.

$$G = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq 1\}$$

The gauge function g is homogeneous: $g(c\boldsymbol{x}) = cg(\boldsymbol{x}), c > 0$.

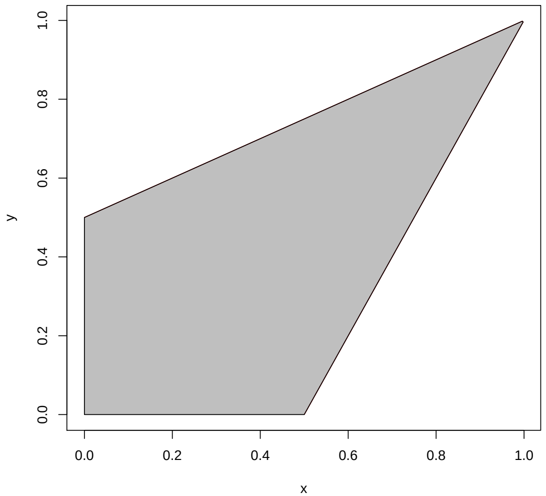
Calculating the gauge function

Suppose that $\boldsymbol{X}_E$ has d-variate density $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then the continuous g can be found through

$$\frac{-\log f_{\boldsymbol{X}}(t\boldsymbol{x})}{t} \to g(\boldsymbol{x}), \qquad t \to \infty.$$

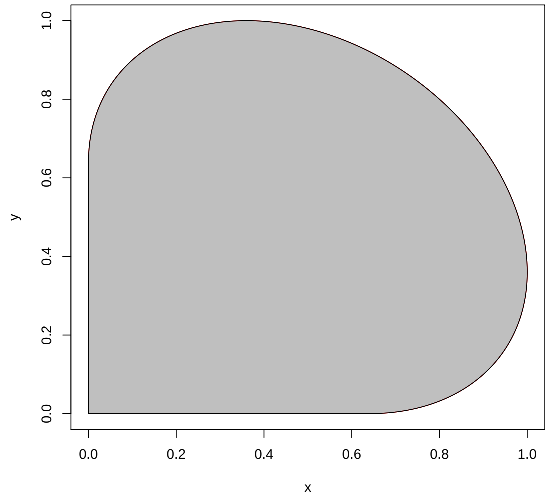
(Balkema and Nolde, 2010; Nolde, 2014).

Bivariate logistic



$$g(x, y) = \frac{\max(x, y)}{\theta} + \left(1 - \frac{1}{\theta}\right)\min(x, y)$$

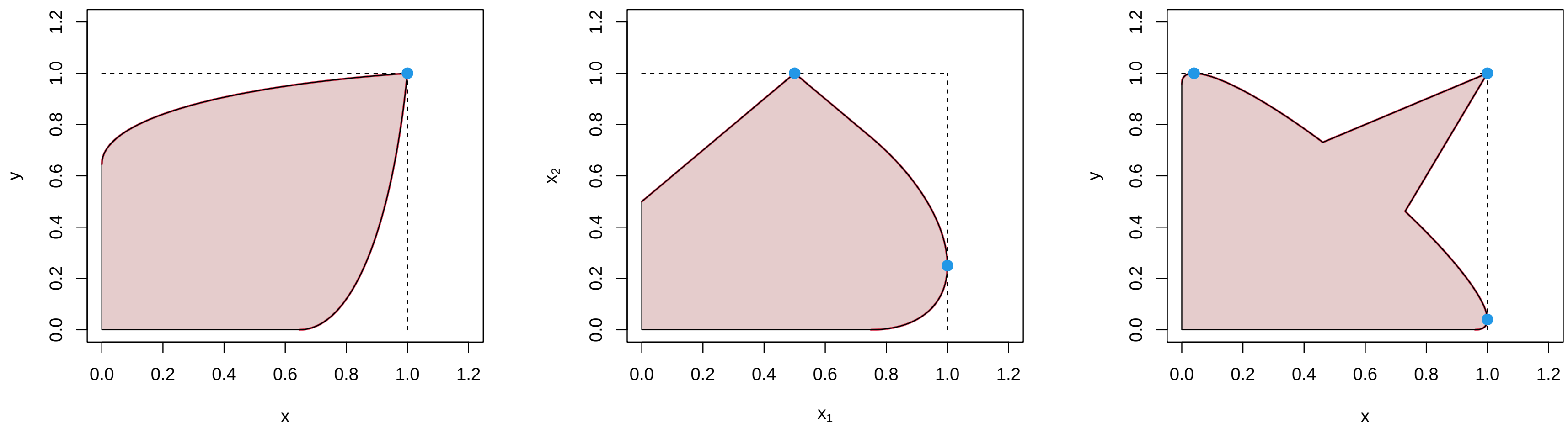
Bivariate Gaussian



$$g(x, y) = \frac{x + y - 2\rho(xy)^{1/2}}{1 - \rho^2}$$

Limit set shapes and extremal dependence

- In exponential margins, $G \subseteq [0,1]^d$ with $\sup(G) = (1, \ldots, 1)$.
- Points of intersection with the bounding box give directions in which most extremes lie
- Summarize collections of variables that are simultaneously extreme via these intersection points: **geometric extreme directions** (Papastathopoulos & Wadsworth, 2025+)



$\{\{1,2\}\}$ $\{\{1\},\{2\}\}$ $\{\{1\},\{2\},\{1,2\}\}$

Limit set shapes and extremal dependence

- Many existing extremal dependence coefficients and properties link to the shape of G (Nolde, 2014; Nolde & Wadsworth, 2022)
- Residual tail dependence coefficient (Ledford & Tawn, 1996); conditional extreme value model coefficients (Heffernan & Tawn, 2004); angular dependence function (Wadsworth & Tawn, 2013)

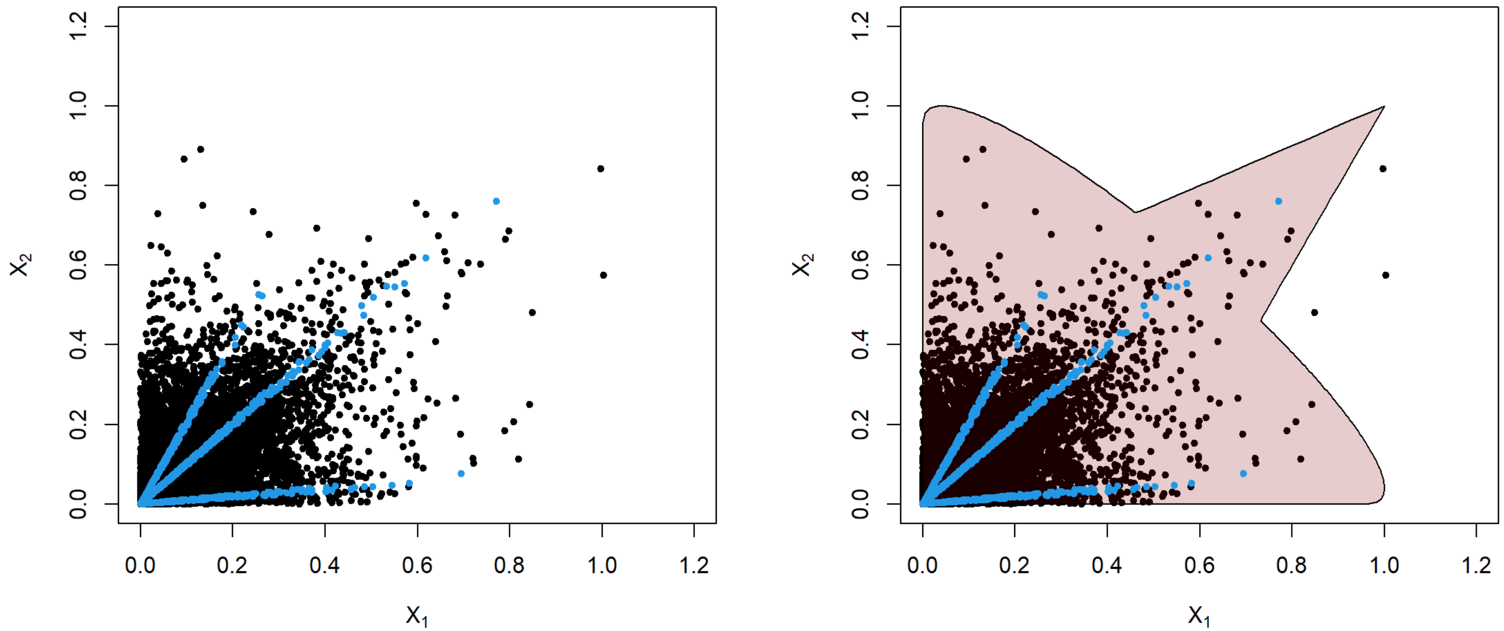
Limit set framework

- Provides a **comprehensive** description of extremal dependence allowing non-trivial treatment of different extremal dependence scenarios
- Offers **connections between existing representations** of the multivariate tail
- Represents a promising approach to **statistical inference for complex dependence structures** (and simpler ones!)

Inference on G

Star-shaped set suggests **radial-angular** decomposition of $\boldsymbol{X}$:

$$R = \|\boldsymbol{X}\|, \qquad \boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|.$$



- The **tail** of the distribution of $R|\boldsymbol{W}$ is informative about the boundary of G
- The gauge function $g(\boldsymbol{w})$ arises as a rate parameter in this asymptotic distribution

Ingredients for tail inference via the geometric limit set framework:

- A radial model for tail of $R|\boldsymbol{W}$ which depends on the gauge function $g(\boldsymbol{w})$
  - GPD (Simpson & Tawn, 2024; Papastathopoulos et al, 2023)
  - Truncated gamma (Wadsworth & Campbell, 2024)
- An angular model for $\boldsymbol{W}$

Other approaches

Semi-parametric angular-radial (SPAR) model (Mackay & Jonathan, 2023; Murphy-Barltrop et al, 2024; Mackay et al, 2025)

- Does not (necessarily) assume particular margins or limit set convergence – but can encompass limit set approach in standard margins
- Focus on angular-radial systems for extrapolation, including estimating environmental contours and other quantities
- Radial model GPD (not necessarily linked to limit set)
- Angular model also required

Thanks for your attention!

Balkema, A., Embrechts, P. and Nolde, N. (2010)

Meta densities and the shape of their sample clouds. *JMVA*

Balkema, A. and Nolde, N. (2010)

Asymptotic independence for unimodal densities. *Adv. Appl. Prob.*

Davis, R., Mulrow, E. Resnick, S. (1988)

Almost Sure Limit Sets of Random Samples in R^d. *Adv. Appl. Prob.*

Heffernan, J. and Tawn, J. (2004)

A conditional approach for multivariate extreme values. *JRSSB*

Kinoshita, K., Resnick, S. (1991)

Convergence of Scaled Random Samples in R^d. *Ann. Prob.*

Ledford, A. and Tawn, J. (1996)

Statistics for near independence in multivariate extreme values. *Biometrika*

Mackay, E. and Jonathan, P. (2023)

Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv*

Mackay, E., Murphy-Barltrop, C., Richards, J. and Jonathan, P. (2024)

Deep learning joint extremes of metocean variables using the SPAR model. *arXiv*

Murphy-Barltrop, C., Mackay, E., and Jonathan, P. (2024)

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

Thanks for your attention!

Nolde, N. (2014)

Geometric interpretation of the residual tail dependence coefficient. *JMVA*

Nolde, N. and Wadsworth, J. (2022)

Linking representations for multivariate extremes via a limit set. *Adv. Appl. Prob.*

Simpson, E. and Tawn, J. (2024)

Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties. *Electronic Journal of Statistics*

Papastathopoulos, I., de Monte, L., Campbell, R., Rue, H. (2023)

Statistical inference for radially-stable generalized Pareto distributions and return level-sets in geometric extremes. *arXiv*

Papastathopoulos, I., and Wadsworth, J. (2025+)

Geometric extremal graphical models and coefficients of extremal dependence on block graphs.

Wadsworth, J. and Campbell, R. (2024)

Statistical inference for multivariate extremes via a geometric approach. *JRSSB*

Wadsworth, J. and Tawn, J. (2013)

A new representation for multivariate tail probabilities. *Bernoulli*

Strategies for simulating a multivariate angular vector

Emma Simpson

University College London

RSS conference, Edinburgh
4 September 2025

Motivation

Angular-radial representations can be useful in the study of multivariate extremes.

For a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$, we can consider

$$R = \|\boldsymbol{X}\|_A, \qquad \boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|_B,$$

i.e., a *radial* component and a *(pseudo-)angular* vector.

Motivation

Angular-radial representations can be useful in the study of multivariate extremes.

For a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$, we can consider

$$R = \|\boldsymbol{X}\|_A, \qquad \boldsymbol{W} = \boldsymbol{X}/\|\boldsymbol{X}\|_B,$$

i.e., a *radial* component and a *(pseudo-)angular* vector.

We focus on $\boldsymbol{W}$, taking values on the $(d-1)$-sphere, or hypersphere,

$$\mathbb{S}^{d-1} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_B = 1\right\}.$$

Motivation - radial-angular multivariate extremes

For statistical inference, angular-radial approaches include:

- geometric extremes (e.g., Wadsworth and Campbell, 2024);
- SPAR models (e.g., Mackay and Jonathan, 2023).

Motivation - radial-angular multivariate extremes

For statistical inference, angular-radial approaches include:

- geometric extremes (e.g., Wadsworth and Campbell, 2024);
- SPAR models (e.g., Mackay and Jonathan, 2023).

For joint tail inference, including extrapolation, these require modelling or simulation of an angular component.

Possible strategies for generating angular data

- Empirical sampling.
  - Becomes infeasible in higher dimensions.
- Parametric models.
  - May lack the flexibility required for complex structures.

Possible strategies for generating angular data

- Empirical sampling.
  - Becomes infeasible in higher dimensions.
- Parametric models.
  - May lack the flexibility required for complex structures.

- Generative deep learning approaches?
  - Less restrictive than relying on exact observations.
  - Can estimate the angular distribution without specifying a density.

Wessel, J. B., Murphy-Barltrop, C. J. R. and Simpson, E. S. (2025).

A comparison of generative deep learning methods for multivariate angular simulation.

Spherical coordinates

Spherical coordinates $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_{d-1})$ can be useful:

$$
\begin{aligned}
\Theta_1 &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2 + \cdots + {X_2}^2}, X_1\right), \\
\Theta_2 &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2 + \cdots + {X_3}^2}, X_2\right), \\
&\;\vdots \\
\Theta_{d-2} &= \operatorname{atan2}\left(\sqrt{{X_d}^2 + {X_{d-1}}^2}, X_{d-2}\right), \\
\Theta_{d-1} &= \operatorname{atan2}\left(X_d, X_{d-1}\right),
\end{aligned}
$$

where $\Theta_1, \ldots \Theta_{d-2} \in [0, \pi]$ and $\Theta_{d-1} \in (-\pi, \pi]$. (Blumenson, 1960)

Deep learning approaches

- Generative adversarial networks (GANs) - Goodfellow et al. (2014)
- Normalizing flows
  - Neural spline flows (NFNSFs) - Durkan et al. (2019)
  - Masked autoregressive flows (NFMAFs) - Papamakarios et al. (2017)
- Flow matching (FM) - Lipman et al. (2022)

Adaptations for angular variables - GANs

- Composed of a generator and a discriminator.

Adaptations for angular variables - GANs

- Composed of a generator and a discriminator.
- To ensure the spherical coordinates lie in the correct range, we specify activation functions after the last generator layer as

$$\theta_i \leftarrow \pi \cdot \frac{1}{1+e^{-x_i}} \text{ for } i = 1, \dots, d-2,$$
$$\theta_{d-1} \leftarrow \pi \cdot \tanh x_{d-1}.$$

- Note: circular wrapping resulted in some training instabilities.

Adaptations for angular variables - NFs

- Involves a base distribution and a series of transformations.
- For both the NFNSF and NFMAF approaches, we use a transformation of the form

$$\theta_i \leftarrow \pi \cdot \Phi(x_i) \text{ for } i = 1, \ldots, d-2,$$
$$\theta_{d-1} \leftarrow (x_{d-1} \bmod 2\pi) - \pi.$$

- In these cases, the sigmoid has slower convergence at 0 and 1 than the Gaussian CDF, which led to poor results near the endpoints of $[0, \pi]$.
- Circular wrapping for the final component worked well in this case.

Adaptations for angular variables - FM

- For flow matching, we don't need to map onto the angular space, since flows can be defined directly on the sphere.

Parametric approach - mixture of vMF distributions

- The von Mises-Fisher (vMF) distribution has density

$$f_{vMF}(\boldsymbol{w} \mid \boldsymbol{\mu}, \kappa) = c_d(\kappa) e^{\kappa \boldsymbol{\mu}^T \boldsymbol{w}}, \quad \boldsymbol{w} \in \mathbb{S}^{d-1},$$

  for some concentration $\kappa \geq 0$, mean direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and normalising constant $c_d(\kappa)$.

- This is an angular extension of a multivariate Gaussian distribution.

Parametric approach - mixture of vMF distributions

- The von Mises-Fisher (vMF) distribution has density

$$f_{vMF}(\boldsymbol{w} \mid \boldsymbol{\mu}, \kappa) = c_d(\kappa) e^{\kappa \boldsymbol{\mu}^T \boldsymbol{w}}, \quad \boldsymbol{w} \in \mathbb{S}^{d-1},$$

  for some concentration $\kappa \geq 0$, mean direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and normalising constant $c_d(\kappa)$.

- This is an angular extension of a multivariate Gaussian distribution.

- A finite mixture of vMF distributions provides a flexible, parametric approach to modelling angular data.

Simulation study - overview

To cover a wide range of examples, we consider the following options:

- dimensions $d = 5$, $d = 10$ and $d = 50$;
- datasets of size $n = 1,000$, $n = 10,000$ and $n = 100,000$;
- light and heavy-tailed marginal distributions;
- a range of different dependence (copula) structures.

Simulation study - evaluation strategies

We consider a range of evaluation metrics to assess both marginal and joint structure.

- Numerical metrics - angular energy score (aES)
- Histograms and QQ-plots for spherical margins
- Scatterplots between pairs of spherical angles
- Orthant probability plots to assess full structure

Simulation study - some key findings

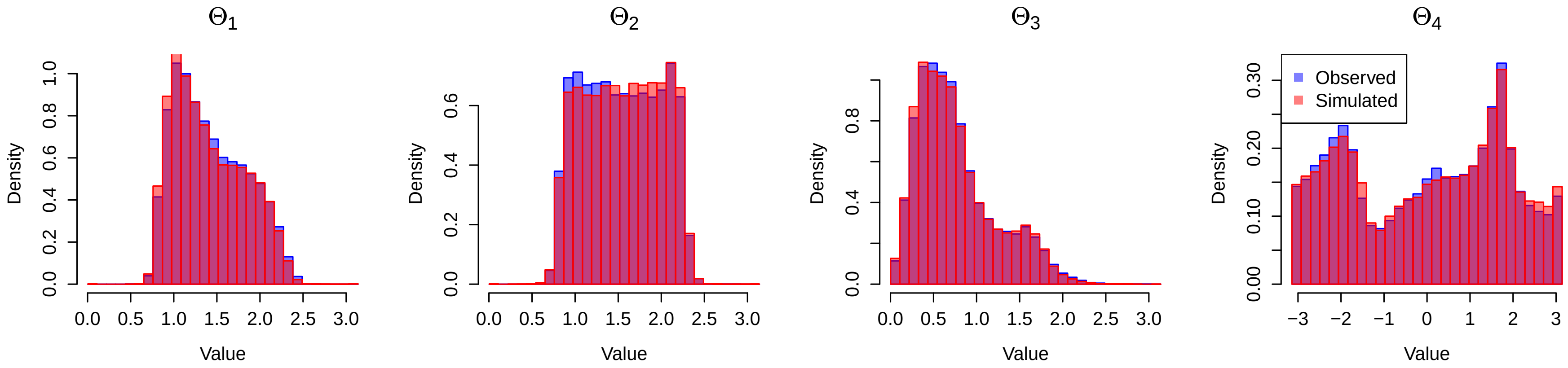
- Difficult to distinguish between methods using numeric scores alone.
- No method should be completely discounted.
- NFNSF, NFMAF and FM all provide good results across a range of examples.

Metocean data example

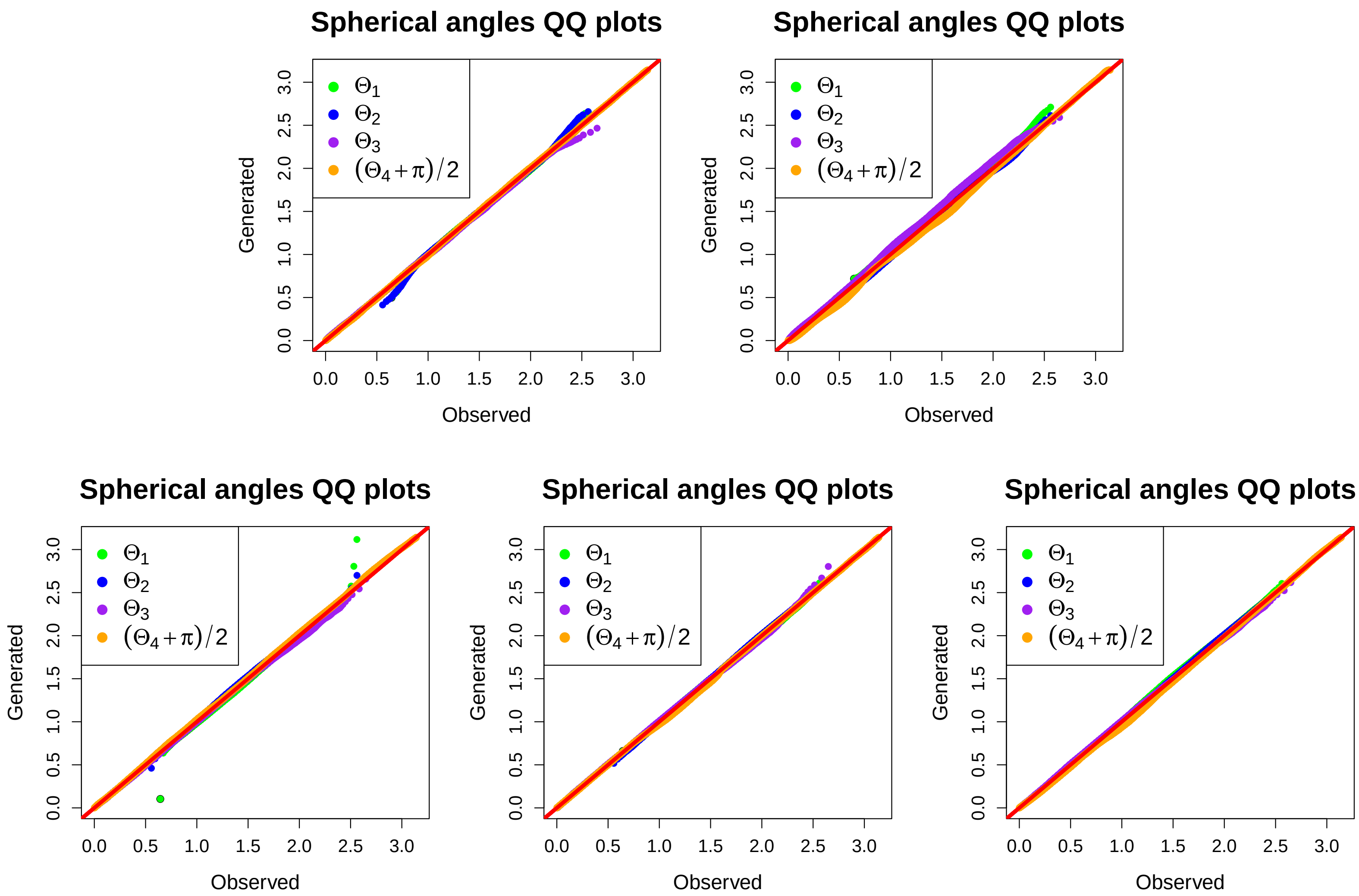
Five-dimensional dataset consisting of:

- significant wave height,
- mean wave period,
- mean wave direction,
- hourly mean wind speed at 10 m above sea level,
- wind direction.

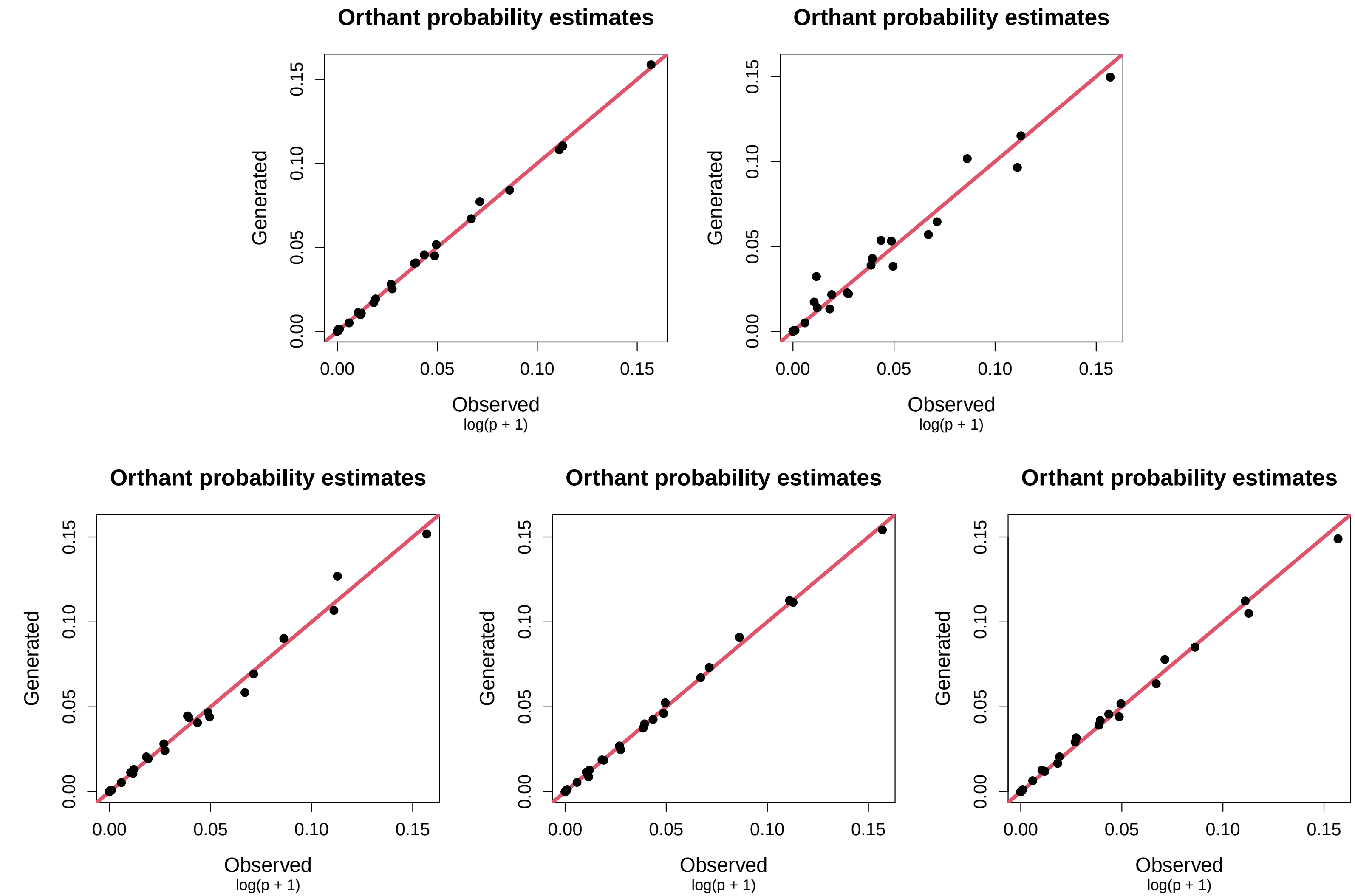
Metocean data - some example results



Metocean data - some example results



Metocean data - some example results



Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.

Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.
- The parametric approach was generally good at capturing marginal distributions, but struggled with some complex dependence structures.

Conclusions

- The flexibility and scalability offered by generative deep learning approaches make them suitable candidates for multivariate angular simulation.
- The parametric approach was generally good at capturing marginal distributions, but struggled with some complex dependence structures.
- No methods can be completely discounted, and there is no individual ‘best’ approach. So model checks and validation are key...
  - ... and it would be useful to have more metrics for evaluation in angular/spherical contexts!

Thank you!

Based on the paper:

- Wessel, J. B., Murphy-Barltrop, C. J. R. and Simpson, E. S. (2025). A comparison of generative deep learning methods for multivariate angular simulation. *arXiv:2504.21505.*

- Accompanying code available from `https://github.com/callumbarltrop/DeGeMoH`

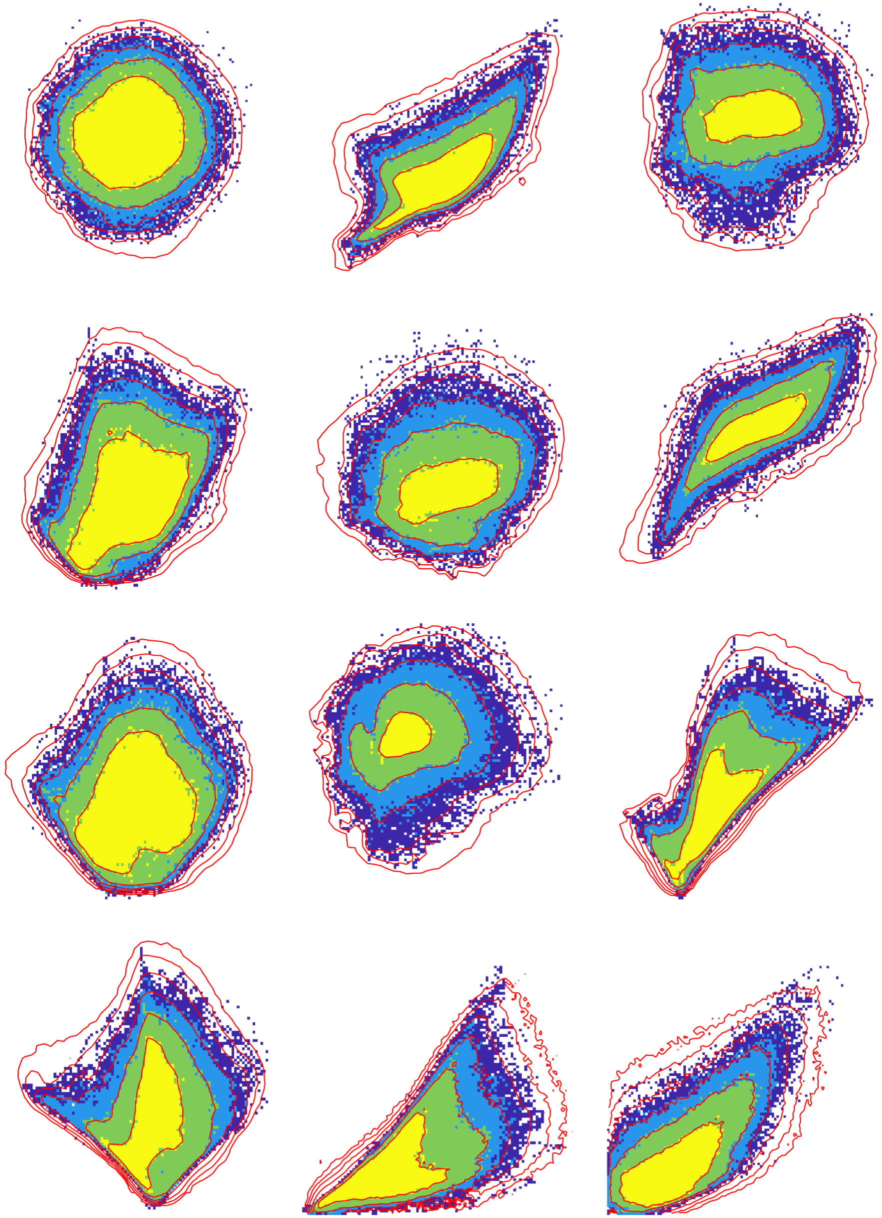
Other references:

- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2019). Neural spline flows. *NeurIPS.*
- Goodfellow, I. J., et al. (2014). Generative adversarial nets. *NeurIPS.*
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. (2022). Flow matching for generative modeling. *ICLR.*
- Mackay, E. and Jonathan, P. (2023). Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv:2310.12711.*
- Papamakarios, G., Pavlakou, T., and Murray, I. (2017). Masked autoregressive flow for density estimation. *NeurIPS.*
- Wadsworth, J. L. and Campbell, R. (2024). Statistical inference for multivariate extremes via a geometric approach. *JRSSB.*

Contact: emma.simpson@ucl.ac.uk

University of Exeter

Deep Learning Multivariate Extremes using the SPAR Model



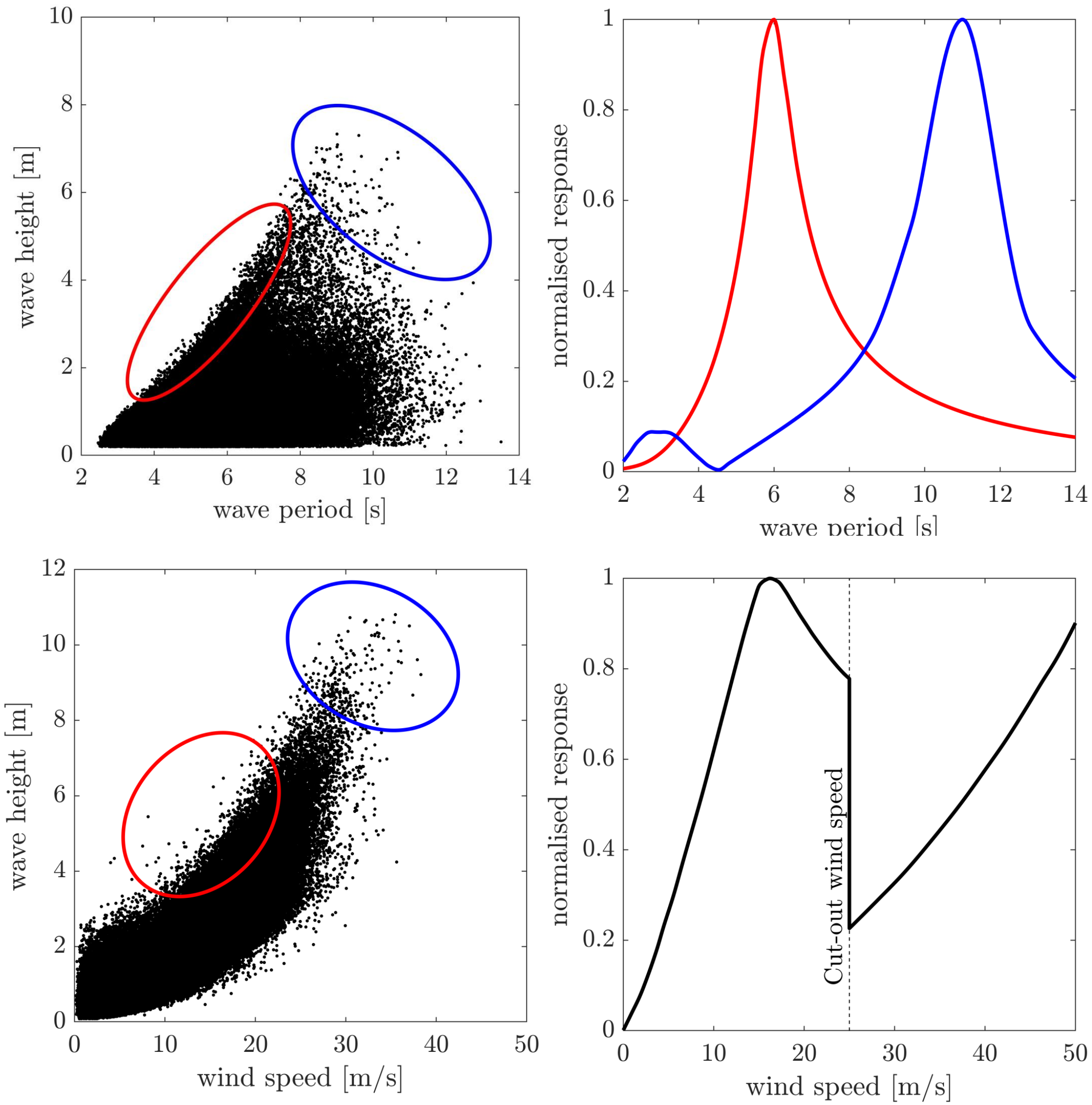
Ed Mackay

RRS International Conference, Edinburgh

4th September 2025

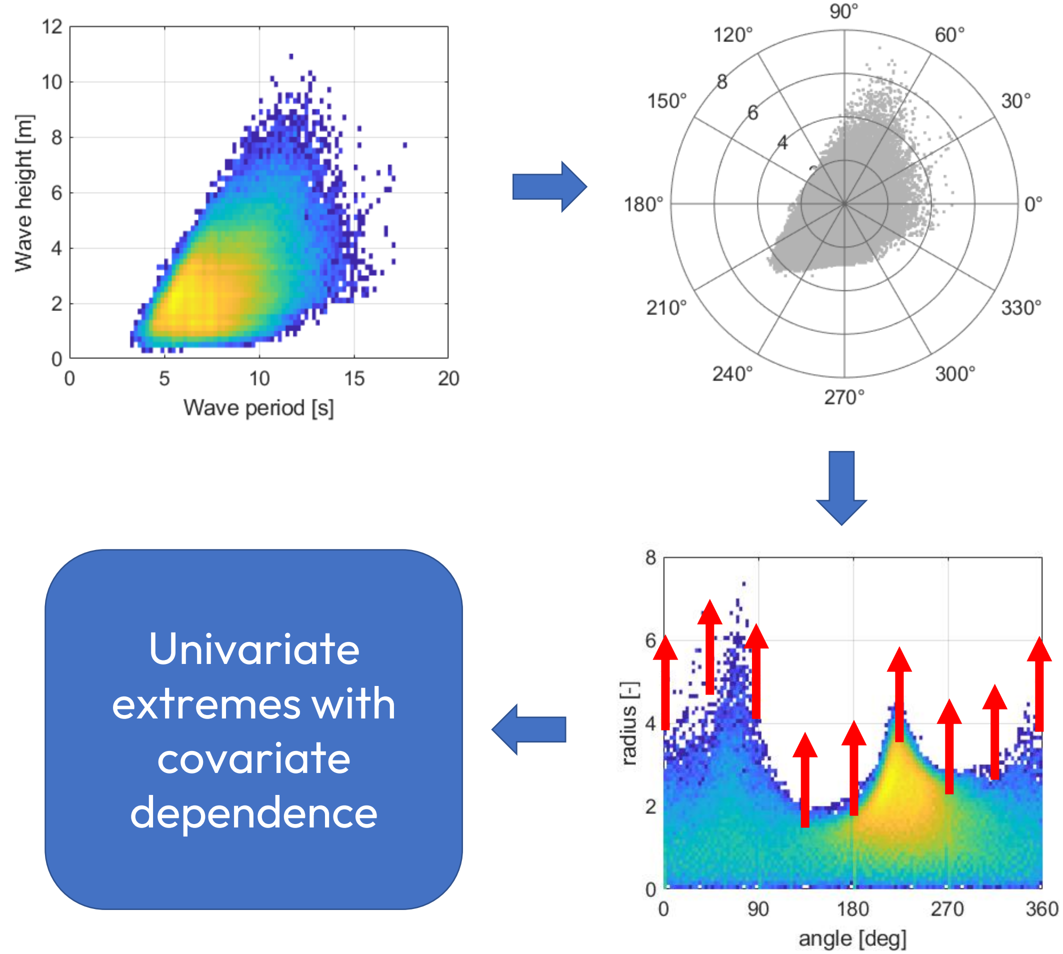
Motivating problems

- Environmental loads on offshore wind turbines dependent on winds, waves, tides etc.
- Largest response does not necessarily occur in largest conditions
- Blue regions: extreme at least one variable
- Red regions: not extreme in either variable, but are "locally extreme"
- Could use various types of existing model to characterise each region
- **Can we capture all "extreme regions" in a single inference?**



The SPAR model

- Semi-Parametric Angular-Radial (SPAR) model
- Reframes multivariate extremes as an intuitive extension of univariate theory, with angular dependence
- Inference is standard univariate peaks-over-threshold (POT) problem with covariate dependence
- Many existing methods for multivariate extremes are special cases of SPAR
- SPAR is more flexible than current methods



Transformation to angular-radial coordinates

- Let $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ have density $f_{\mathbf{X}}$

- Define radial and angular variables:

$$R = \|\mathbf{X}\|_2$$
$$\mathbf{W} = \mathbf{X}/R$$

where $\|\mathbf{X}\|_2 = \sqrt{X_1^2 + \cdots + X_d^2}$ is the Euclidean (L2) norm

- Note that
  - R is radial variable
  - $\mathbf{W}$ is a pseudo-angle corresponding to a point on the unit hypersphere

- Then $(R, \mathbf{W})$ has density

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

SPAR model statement

- Write joint density in conditional form:

$$f_{R,\mathbf{W}}(r,\mathbf{w}) = f_{\mathbf{W}}(\mathbf{w}) f_{R|\mathbf{W}}(r|\mathbf{w})$$

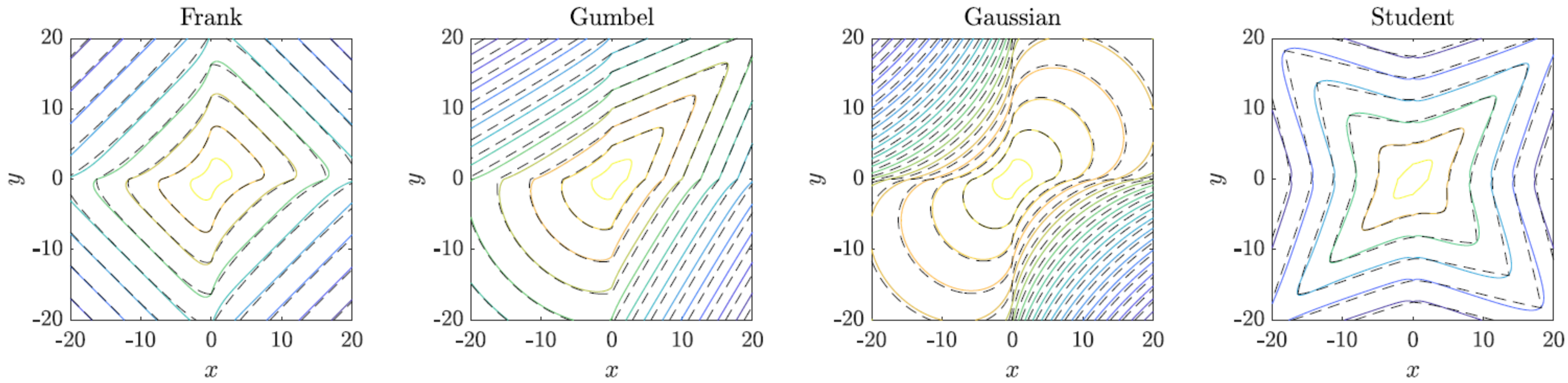
- Assume standard univariate POT model for tail of $f_{R|\mathbf{W}}(r|\mathbf{w})$
- Define threshold function $u(\mathbf{w})$ as conditional quantile of $R|(\mathbf{W}=\mathbf{w})$ at exceedance probability ζ
- SPAR model for the joint density is

$$f_{R,\mathbf{W}}(r,\mathbf{w}) = \zeta\ f_{\mathbf{W}}(\mathbf{w})\ f_{\mathrm{GP}}\big(r - u(\mathbf{w}); \xi(\mathbf{w}), \sigma(\mathbf{w})\big), \qquad r > u(\mathbf{w})$$

$f_{\mathrm{GP}}(r; \xi, \sigma)$ is a generalised Pareto density function with shape parameter ξ and scale parameter σ

SPAR representations of parametric copulas

- SPAR parameters can be derived analytically for parametric families of copulas
- SPAR can accurately capture distributions with different dependence properties



- Coloured lines: contours of true density
- Dashed lines: contours of SPAR density at the same levels

Marginal distributions

- SPAR formulation does not make any assumption about the marginal distributions

- SPAR on exponential / Laplace margins is consistent with limit set theory
  - Additional assumption: constant GP shape $\xi(\mathbf{w}) = 0$

- SPAR can be applied on original marginal scale
  - Pros: removes uncertainty with estimating marginal model
  - Cons: need to be confident that marginal effects will not mask dependence effects

- For many environmental variables it is reasonable to assume that distributions are bounded, due to physical limits
  - Rules out possibility of e.g. one margin heavy-tailed and another light-tailed

Limit sets

- **Definition:** Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a sequence of IID random vectors in $\mathbb{R}^d$ with either standard exponential or Laplace margins, such that

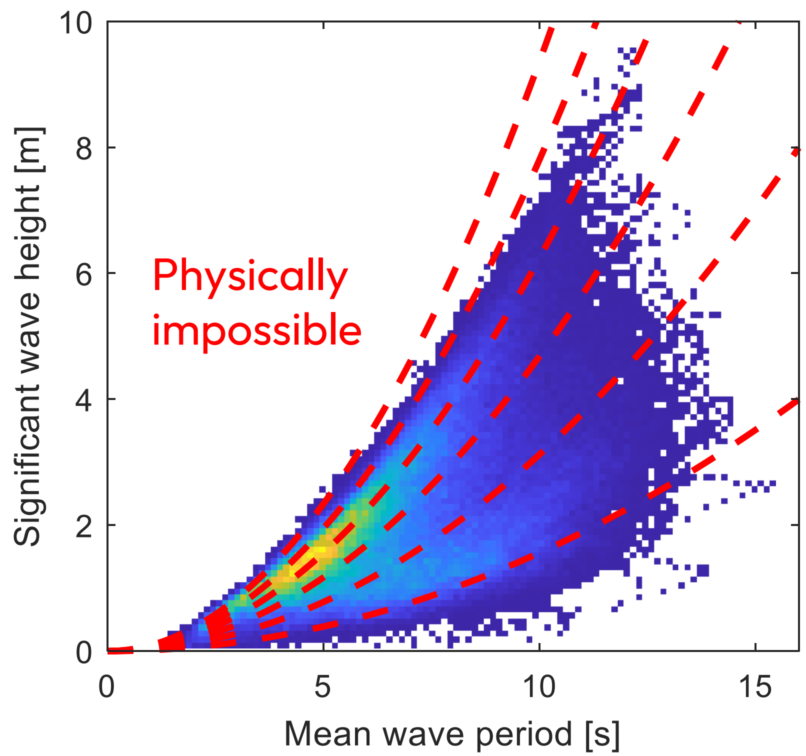
$$\left\{\frac{\mathbf{X}_1}{\log n}, \ldots, \frac{\mathbf{X}_n}{\log n}\right\} \to G, \qquad n \to \infty,$$

then we say G is the limit set of the scaled sample cloud.

- Assumes that extreme value index of the radial variable is constant at all angles
  - All regions of variable space have non-zero probability

- Cannot be true for certain environmental variables, since physical limits result in regions of variable space with zero probability

- Allowing the GP shape parameter to vary with angle gives more flexibility to model these effects



Inference

- Model:

$$f_{R,\mathbf{W}}(r,\mathbf{w}) = \zeta\, f_{\mathbf{W}}(\mathbf{w})\, f_{\mathrm{GP}}\big(r - u(\mathbf{w}); \xi(\mathbf{w}), \sigma(\mathbf{w})\big)$$

- Need to estimate:
  - Angular density: $f_{\mathbf{W}}(\mathbf{w})$
    - ➢ **Directional statistics**
  - Threshold function: $u(\mathbf{w})$
    - ➢ **Quantile regression**
  - GP shape and scale parameter functions: $\xi(\mathbf{w}), \sigma(\mathbf{w})$
    - ➢ **Non-stationary POT**

- Problems are separable:
  - Angular density can be estimated independently of GP threshold and parameter functions

- Many existing methods for these problems

- Key consideration is choosing suitable semi-parametric form for threshold and GP parameter functions

Covariate representations

- Various ways of representing covariate functions: e.g.
  - Generalised additive models (GAMs), e.g. splines
  - Gaussian processes
  - Artificial neural networks (ANNs)

- All provide flexible function approximation

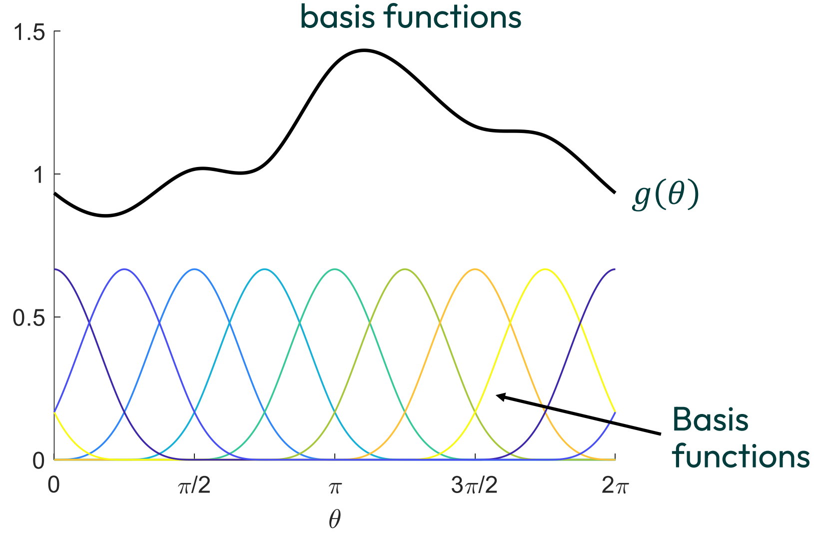
- In all cases, parameters estimated by minimising the same loss functions
  - Tilted loss for quantile regression
  - Likelihood function for GP parameters

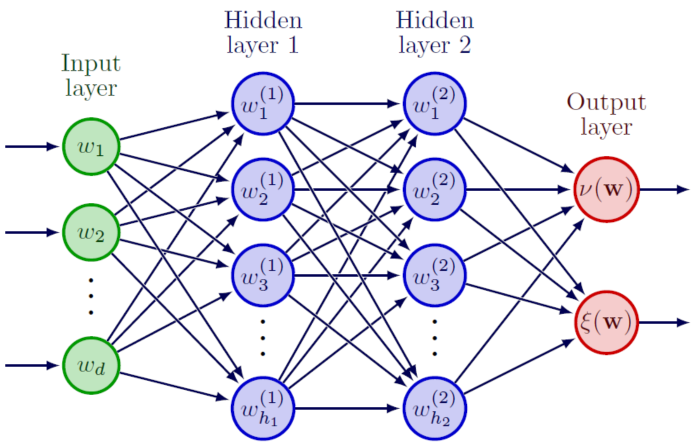
- Neural networks have computational advantages over GAMs in higher dimensions

- Inference with GAMs:
  - Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2025). Inference for bivariate extremes via a semi-parametric angular-radial model. Extremes, 28(2):209–238.
  - https://github.com/callumbarltrop/SPAR

- Inference with ANNs:
  - Mackay, E., Murphy-Barltrop, C. J., Richards, J., and Jonathan, P. (2025). Deep learning of joint extremes of metocean variables using the SPAR model. arXiv:2412.15808.
  - https://github.com/callumbarltrop/DeepSPAR

GAM representation of a function as a sum of basis functions





Schematic of ANN with two hidden layers

ANN setup

- Two ANNs:
  - ANN for threshold trained first
  - ANN for GP parameters trained afterwards

- Activation functions:
  - Hidden layer activation functions all set as rectified linear units
  - Final layers use exponential transformation to ensure GP scale and threshold are strictly positive

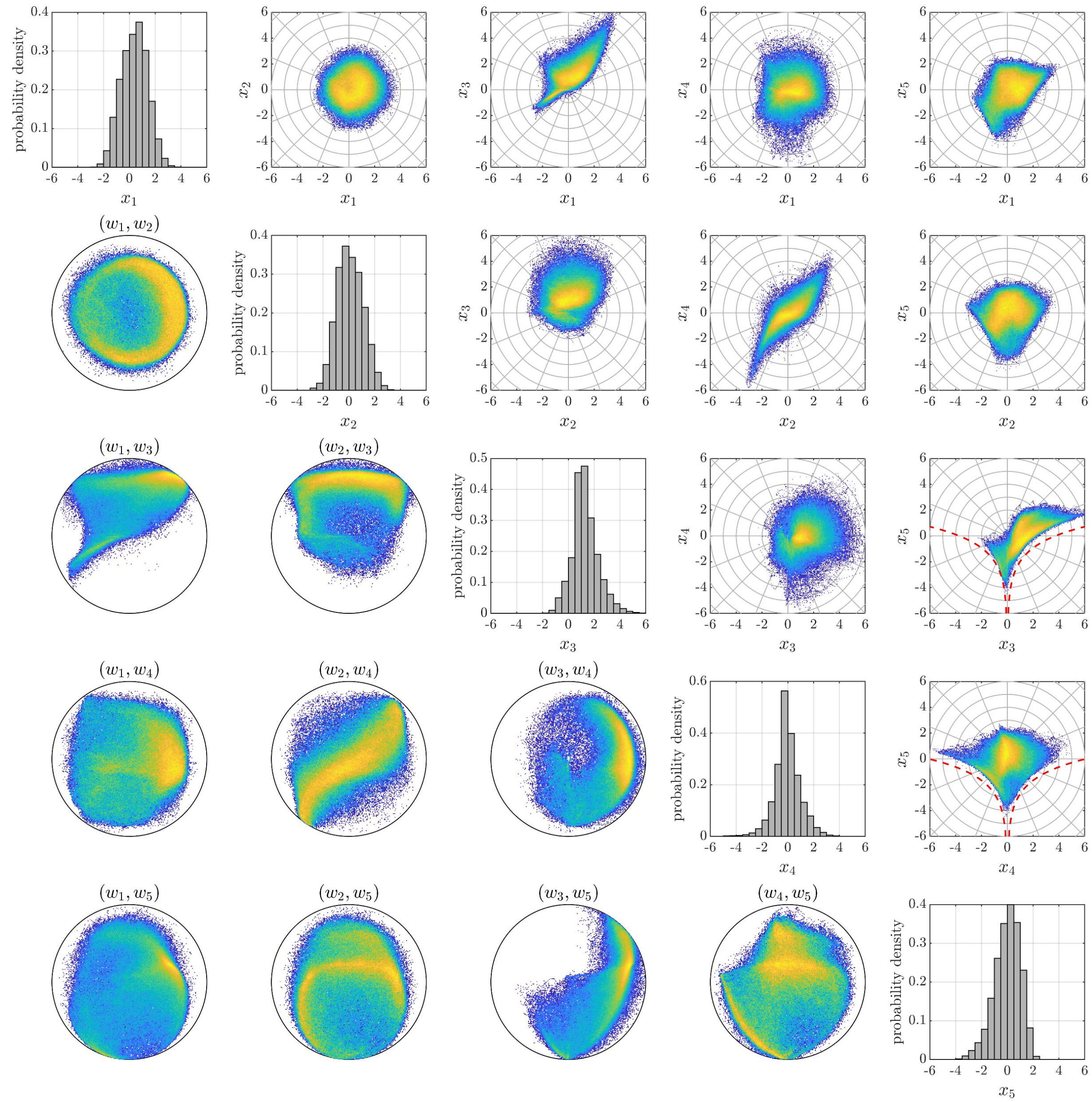
- Network depth (number of layers) and width (neurons per layer) determine model flexibility
  - All layers fully connected
  - Architecture 'optimised' using a grid search
  - Relatively simple architecture selected for both networks: (3 layers, 16 neurons per layer) → 650 parameters

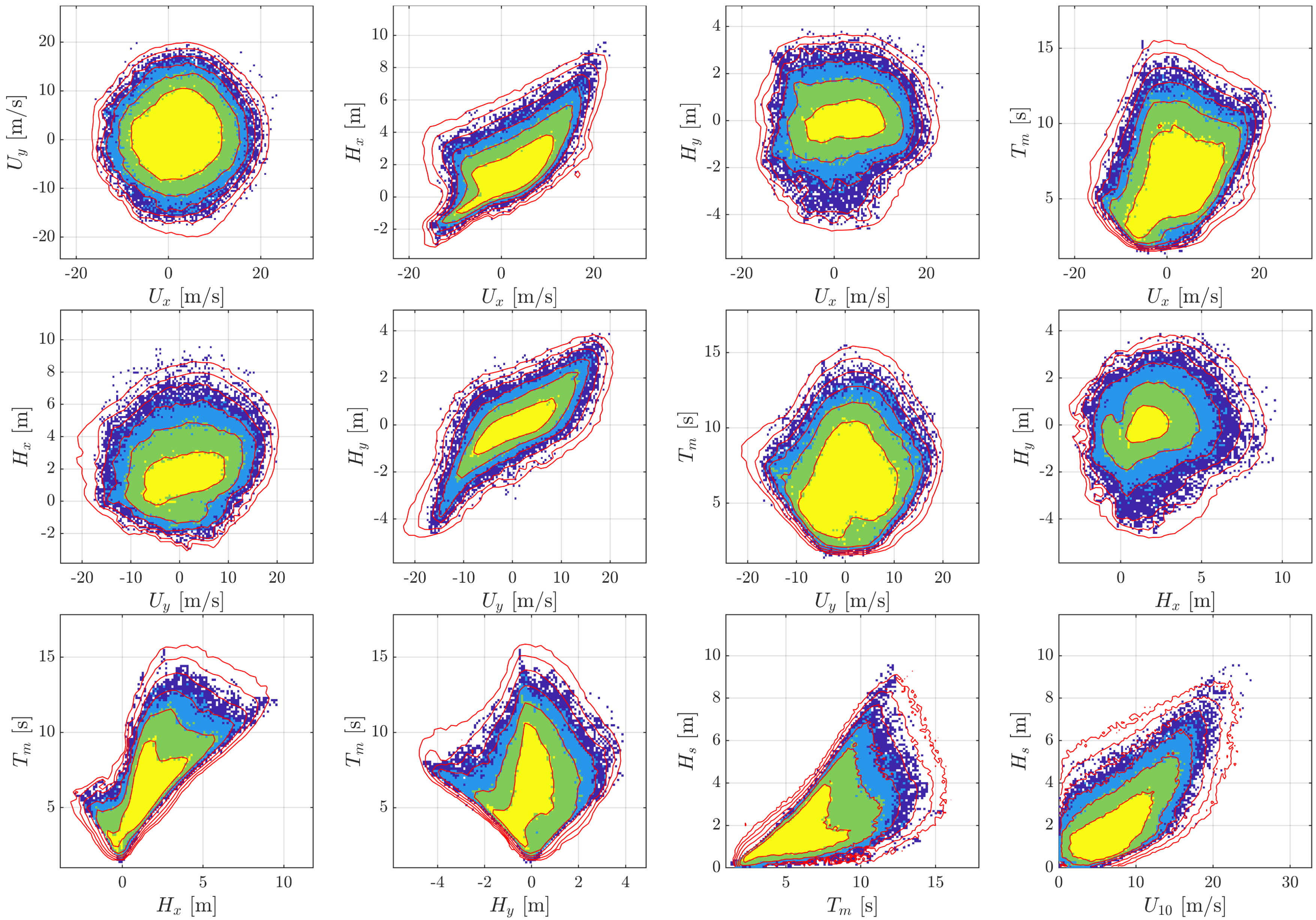
- Minimisation of loss function
  - Stochastic gradient descent (ADAM)
  - Data split into 80% training, 20% validation
  - Early stopping: training terminated when validation loss plateauxs

- Special care needed to deal with GP upper end point

Five-dimensional application

- 31-year hindcast of hourly values for a site in Celtic Sea
- Variables
  - Wind speed & direction
  - Wave height, period & direction
- Transformed to:
  - East-North components of winds and waves
  - Logarithm of wave period
  - All normalised to have unit variance
- Kernel density method used for angular model in this example

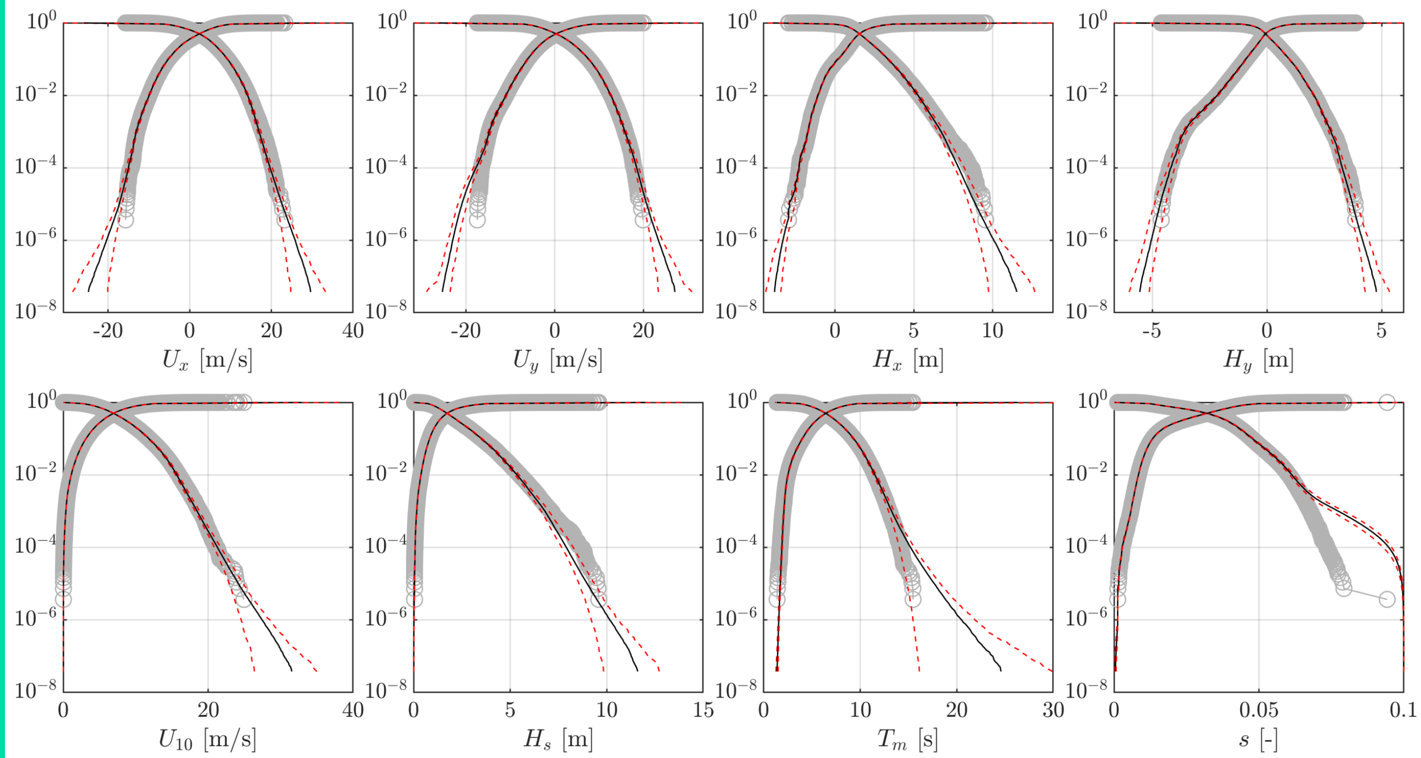




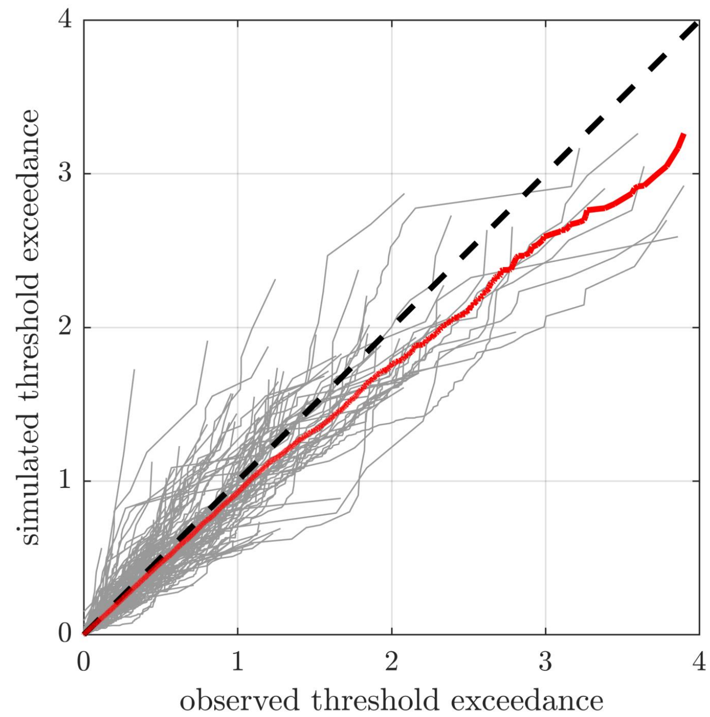
U_y [m/s]
U_x [m/s]
H_x [m]
U_x [m/s]
H_y [m]
U_x [m/s]
T_m [s]
U_x [m/s]
H_x [m]
U_y [m/s]
H_y [m]
U_y [m/s]
T_m [s]
U_y [m/s]
H_y [m]
H_x [m]
T_m [s]
H_x [m]
T_m [s]
H_y [m]
H_s [m]
T_m [s]
H_s [m]
U_{10} [m/s]

Radial model diagnostics

University of Exeter



Marginal lower & upper tail exceedance plots



Local tail QQ plots

Summary

University of Exeter

- Existing ‘engineering’ methods for modelling multivariate extremes have strong limitations
  - Rely on ad-hoc choices
  - Poor fit in practice

- SPAR model
  - Provides model for joint density in all ‘extreme regions’
  - Requires no strong assumptions about form of margins or dependence structure
  - Inference built on existing practices
  - Shown to provide good fit to metocean datasets

- Future work
  - Detailed comparison with existing inference methods
  - Effect of choice of margins
  - Effect of choice of origin
  - Optimisation of neural network inference

https://arxiv.org/abs/2412.15808

Questions?

e.mackay@exeter.ac.uk