A deep geometric approach to modelling multivariate extremes

Jordan Richards[1]
Callum J. R. Murphy-Barltrop[2,3] Reetam Majumder[4]

[1]School of Mathematics, University of Edinburgh

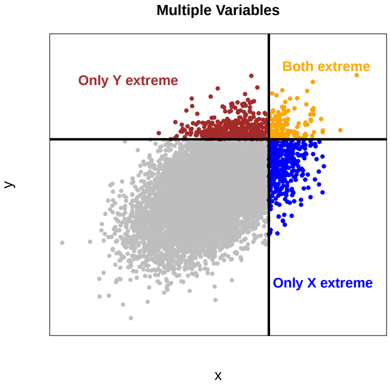
[2]Technische Universität Dresden, Institut Für Mathematische Stochastik, Dresden

[3]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig

[4]Southeast Climate Adaptation Science Center, North Carolina State University

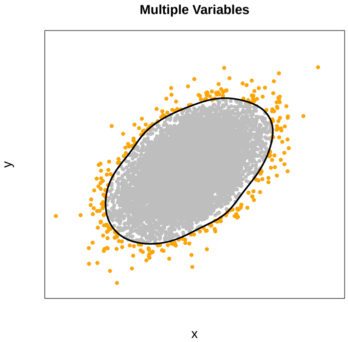
Multivariate extremes

- Quantifying extremal dependence is key for risk analysis
- Many classical approaches require restrictive assumptions on joint tail decay, e.g., regular variation, max-stability
- Focus on particular extremes only



Geometric extremes

- Recent works have shown that deterministic limit sets provide a useful tool for studying extremal dependence (Nolde and Wadsworth, 2022; Mackay and Jonathan, 2023; Papastathopoulos et al., 2024).
- The "geometric approach" does not require restrictive parametric assumptions about joint tail decay.



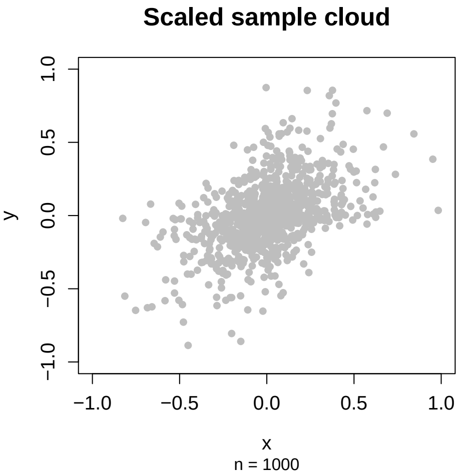
Limit sets

- All theory is applicable to d-dimensional random vectors $\boldsymbol{X} \in \mathbb{R}^d$ on standard margins, with density function $f(\cdot)$.
- Given n independent realisations of $\mathbf{X}$, consider **scaled sample cloud**

$$C_n := \{\mathbf{X}_i / r_n;\ i = 1, \ldots, n\},$$

  as $n \to \infty$, where r_n is a suitably chosen normalising sequence.
- For exponential (Laplace) margins, $r_n = \log(n)$ ($r_n = \log(n/2)$).

Limit sets



Gaussian copula $\rho = 0.5$ with Laplace margins

Limit sets

- Suppose
$$-\log f(t\mathbf{x}) \sim tg(\mathbf{x}), \quad t \to \infty, \ \mathbf{x} \in \mathbb{R}^d,$$
where $g(\cdot)$ is a continuous function on $\mathbb{R}^d$.
- $g(\mathbf{x})$ is termed the **gauge function**.
- $g(\mathbf{x})$ is 1-homogeneous, i.e., $g(c\mathbf{x}) = cg(\mathbf{x})$ for any $c > 0$.

Limit sets

- As $n \to \infty$, C_n converges in probability onto the set

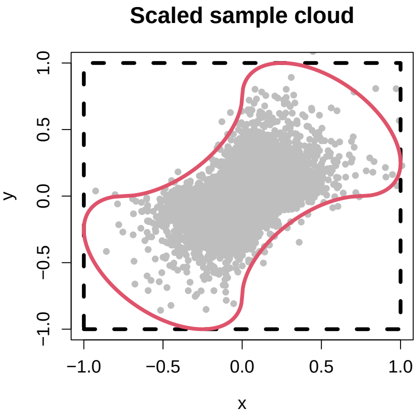
$$\mathcal{G} = \{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) \leq 1\} \subseteq [-1, 1]^d.$$

- $\mathcal{G}$ is star-shaped and compact (closed and bounded).
- For some centre $\boldsymbol{o}$, we have that the lines segment $\{\boldsymbol{o} + t\boldsymbol{x} : t \in [0, 1]\} \subset \mathcal{G}$ for all $\boldsymbol{x} \in \mathcal{G}$.
- Componentwise $\max \mathcal{G} = (1, \ldots, 1)^T$ and $\min \mathcal{G} = (-1, \ldots, -1)^T$.
- See Nolde and Wadsworth (2022) for further details.

Limit sets

Consider the unit level (boundary) set given by

$$\partial\mathcal{G} = \{\mathbf{x} : g(\mathbf{x}) = 1\} \subseteq [-1, 1]^d.$$



Limit sets

- There exist links between $\partial\mathcal{G}$ and several existing approaches for multivariate extremes: Ledford and Tawn (1996), Heffernan and Tawn (2004), Wadsworth and Tawn (2013) and Simpson et al. (2020).

Once we have $\partial\mathcal{G}$, we get the rest for free.

- Whereas the above models focus on specific parts of the distribution, knowing $\partial\mathcal{G}$ gives you the complete picture of extremal dependence.

Ledford and Tawn

- For example, the approach of Ledford and Tawn (1996) when $\mathbf{X}_E$ has exponential margins:

$$\Pr\left(\min_{i=1,\dots,d}\{X_{E,i}\} > u\right) \sim L(e^u)\exp(-u/\eta),$$

as $u \to \infty$, with $L(\cdot)$ slowly varying and $\eta \in (0,1]$.

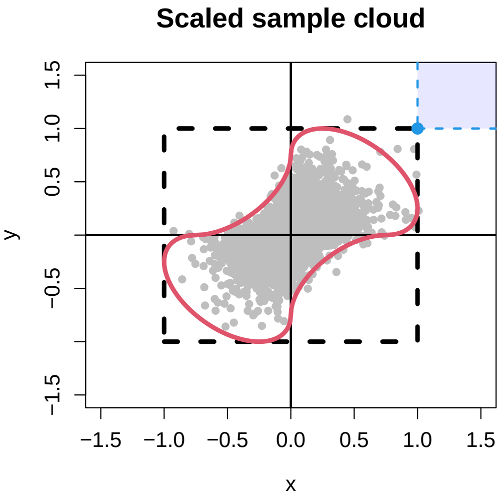
- η quantifies the form of extremal dependence, with asymptotic dependence in $\mathbf{X}_E$ implying $\eta = 1$.

Ledford and Tawn

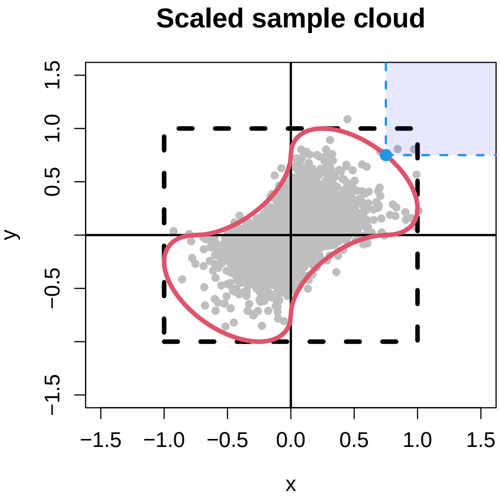
- We have that

$$\eta = \min \left\{ s \in (0,1] : [s,\infty]^d \cap \partial \mathcal{G} = \emptyset \right\}.$$

Ledford and Tawn



Limit sets



Limit sets

- $\widehat{\partial\mathcal{G}}$ gives risk metrics; **return level sets** (Papastathopoulos et al., 2024), **return curves** (Murphy-Barltrop et al., 2024) and **joint tail probabilities** (Wadsworth and Campbell, 2024).
- Knowing $\partial\mathcal{G}$ is invaluable for inference.
- How do we estimate $\partial\mathcal{G}$?

Limit sets

Many recent approaches have been proposed for estimating $\partial\mathcal{G}$.

- Simpson and Tawn (2022) used generalised additive models to approximate $\partial\mathcal{G}$ via scaled radii sets.
- Wadsworth and Campbell (2024) propose truncated parametric copula models for estimating $\partial\mathcal{G}$.
- Majumder et al. (2023) proposed a semi-parametric approach with Bézier polynomials.
- Papastathopoulos et al. (2024) provided a Bayesian inference approach using latent Gaussian variables.

Many more approaches (likely) to follow.

Limit sets

Most require parametric or semi-parametric assumptions about the gauge function $g(\cdot)$:

- Limited to generally low dimension $d \leq 3$;
- Inflexible;
- Do not necessarily guarantee that estimates of $\partial\mathcal{G}$ satisfy the red and blue properties

Talk outline

- Build valid estimators of $\partial\mathcal{G}$ using flexible neural networks
- Provide a means to estimate probabilities using an extension of the Wadsworth and Tawn (2013) ADF
- Application to metocean extremes

Radial-angular decomposition

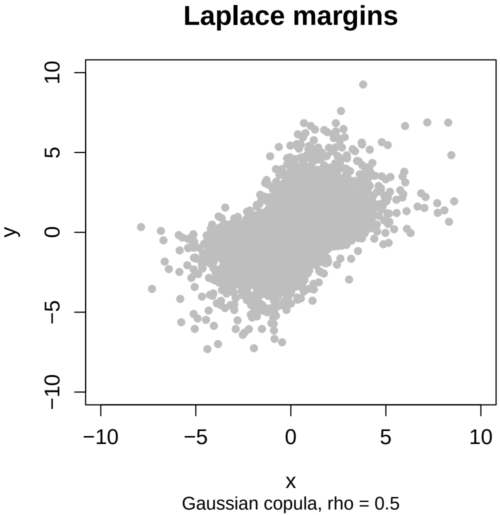
- Define the radial and angular components

$$R := \|\mathbf{X}\|_2, \quad \boldsymbol{W} := \frac{\mathbf{X}}{R},$$

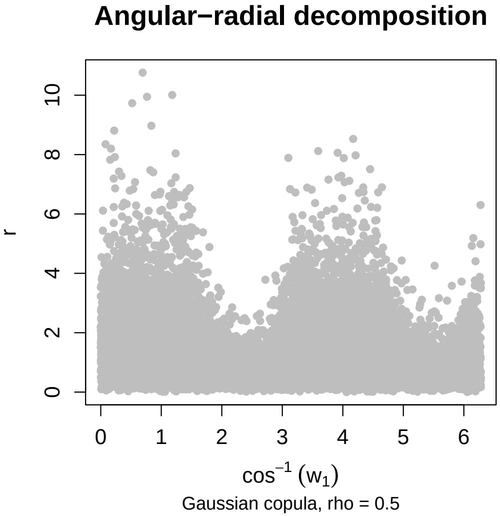
  so $\mathbf{X} = R\mathbf{W}$.

- Observe $\|\boldsymbol{W}\|_2 = 1$. Therefore, $\boldsymbol{W}$ will always exist on the unit $(d-1)$-sphere $\mathcal{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$.

Data



Polar coordinates



Limit set

- By star-shaped property of $\mathcal{G}$, we can re-characterise the unit-level set:

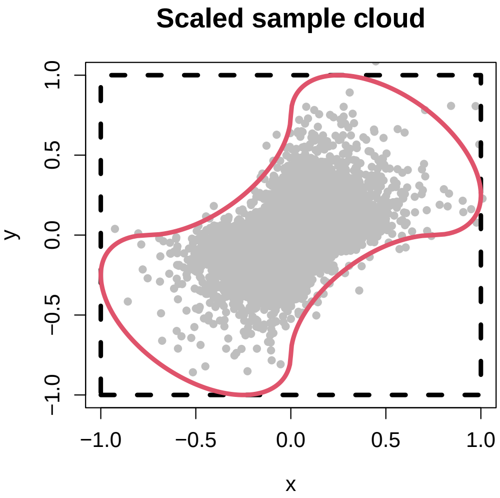
$$\partial\mathcal{G} = \{r\boldsymbol{w} : r > 0, \boldsymbol{w} \in \mathcal{S}^{d-1}, g(r\boldsymbol{w}) = 1\}.$$

- By homogeneity of $g(\cdot)$, we must have $r = 1/g(\boldsymbol{w})$, implying

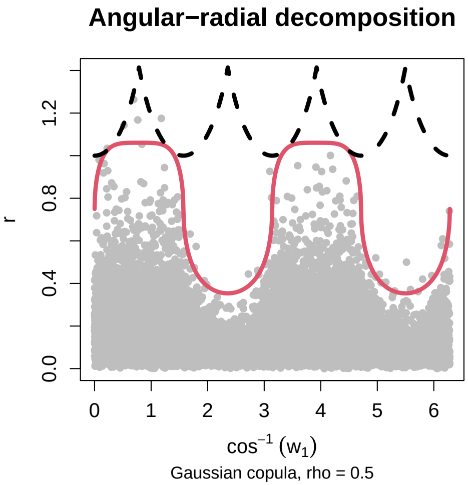
$$\partial\mathcal{G} = \{\boldsymbol{w}/g(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{S}^{d-1}\}.$$

- To evaluate $\partial\mathcal{G}$, we just need to evaluate $g(\cdot)$ over the sphere $\mathcal{S}^{d-1}$.

Limit set



Limit set: radially transformed



Red: $1/g(\mathbf{w})$. Black: upper-bound for $1/g(\mathbf{w})$.

Bound on g

Proposition

For all $\boldsymbol{w} \in \mathcal{S}^{d-1}$, *we have that* $g(\cdot)$ *satisfies*

$$g(\boldsymbol{w}) \geq ||\boldsymbol{w}||_\infty,$$

where $||\boldsymbol{x}||_\infty := \max\{|x_1|, \ldots, |x_d|\}$ *denotes the infinity norm.*

Bound on g : valid estimators

Suppose we have **any** $h(\cdot) : \mathcal{S}^{d-1} \mapsto \mathbb{R}_+$ satisfies $1/h(\boldsymbol{w}) \geq ||\boldsymbol{w}||_\infty$ for all $\boldsymbol{w} \in \mathcal{S}^{d-1}$. Then:

Proposition

Define the set

$$\mathcal{H} := \left\{ \boldsymbol{x} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\} \,\middle|\, ||\boldsymbol{x}|| \leq h(\boldsymbol{x}/||\boldsymbol{x}||) \right\} \bigcup \left\{ \mathbf{0}_d \right\},$$

where $\mathbf{0}_d := (0, \ldots, 0)$. *Then* $\mathcal{H}$ *is star-shaped, compact, and satisfies* $\mathcal{H} \subseteq [-1, 1]^d$.

Bound on g : valid estimators

The corresponding unit-level set $\partial\mathcal{H}$ is

$$\partial\mathcal{H} = \left\{ \boldsymbol{w}\, h(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{S}^{d-1} \right\}.$$

- Whilst $\partial\mathcal{H} \subseteq [-1, 1]^d$, it might not satisfy coordinate-wise max and min property of $\partial\mathcal{H}$, i.e., $\max \partial\mathcal{H} = \mathbf{1}_d$ and $\min \partial\mathcal{H} = -\mathbf{1}_d$
- Can be easily obtained using a straightforward rescaling
- **Implication**:starting with a general radial function $h(\cdot)$ satisfying $h(\mathbf{w}) \geq ||\boldsymbol{w}||_\infty$, we can construct **valid unit level sets**

Rescaling

For each $i = 1, \ldots, d$, we define
$b_i(w_i) := \mathbb{1}(w_i \geq 0) b_i^U - \mathbb{1}(w_i < 0) b_i^L > 0$, where

$$b_i^U := \max \left\{ w_i h(\boldsymbol{w}) \mid \boldsymbol{w} \in \mathcal{S}^{d-1} \right\} > 0, \quad \text{and}$$
$$b_i^L := \min \left\{ w_i h(\boldsymbol{w}) \mid \boldsymbol{w} \in \mathcal{S}^{d-1} \right\} < 0.$$

Using these scaling functions, we define the rescaled set

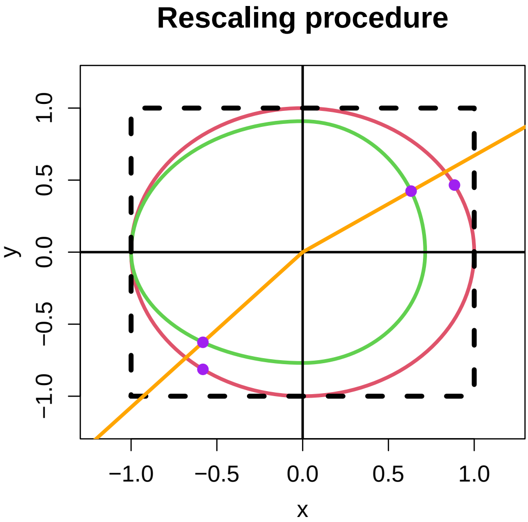
$$\widetilde{\partial \mathcal{H}} := \left\{ h(\boldsymbol{w}) \left(\frac{w_1}{b_1(w_1)}, \ldots, \frac{w_d}{b_d(w_d)} \right) \middle| \boldsymbol{w} \in \mathcal{S}^{d-1} \right\}.$$

Rescaling

Proposition

The rescaled set $\widetilde{\partial\mathcal{H}}$ is in one-to-one correspondence with $\partial\mathcal{H}$, satisfies $\widetilde{\partial\mathcal{H}} \subset [-1,1]^d$, and has componentwise maxima and minima $\mathbf{1}_d$ and $-\mathbf{1}_d$, respectively.

Rescaling



Rescaling

This provides a new construction for valid (rescaled) gauge functions

$$\tilde{g}(\boldsymbol{w}) := 1 \Big/ \left\| h(\kappa^{-1}(\boldsymbol{w})) \left(\frac{\kappa^{-1}(\boldsymbol{w})_1}{b_1(\kappa^{-1}(\boldsymbol{w})_1)}, \ldots, \frac{\kappa^{-1}(\boldsymbol{w})_d}{b_d(\kappa^{-1}(\boldsymbol{w})_d)} \right) \right\|$$

where $\kappa(\cdot)$ is a bjiective mapping.

- In practice, we model $h(\cdot)$ using a neural network;
- Compute scaling factors b_i numerically;
- By uniform sampling (many) points on $\mathcal{S}^{d-1}$.

With estimates of $\tilde{g}(\cdot)$, we can get tail probability estimates.

Ledford and Tawn

- Recall L+T (96):

$$\Pr\left(\min_{i=1,\ldots,d}\{X_{E,i}\} > u\right) \sim L(e^u)\exp(-u/\eta),$$

as $u \to \infty$, with $L(\cdot)$ slowly varying and $\eta \in (0, 1]$.

Angular dependence function (ADF)

Wadsworth and Tawn (2013) generalise L+T (96) with the angular dependence function $\lambda(\mathbf{w})$.

- For angle $\boldsymbol{w} = (w_1, \ldots, w_d)^T \in \mathcal{S}_+^{d-1}$:

$$\Pr\left(\min_{i=1,\ldots,d}\{X_{E,i}/w_i\} > u\right) \sim L(e^u; \boldsymbol{w}) \exp(-\lambda(\boldsymbol{w})u)$$

  as $u \to \infty$, with $\mathcal{S}_+^{d-1} := \{\boldsymbol{x} \in \mathbb{R}_+^d : ||\boldsymbol{x}|| = 1\}$.
- $\lambda(\boldsymbol{w})$ quantifies extremal dependence along different rays $\boldsymbol{w}$.
- With $\eta^{-1} = \{\sqrt{d}\lambda(d^{-1/2}, \ldots, d^{-1/2})\}$.

Extended ADF

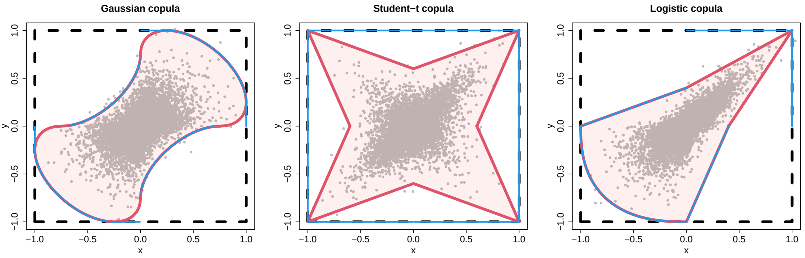
- Assume W + T (2013) model holds for any reflection $\boldsymbol{cX}$, where $\boldsymbol{c} \in \{-1, 1\}^d$ ($\boldsymbol{X}$ on Laplace margins).
- For any $\boldsymbol{w} \in \mathcal{S}^{d-1} \setminus \mathcal{A}$, where $\mathcal{A} := \bigcup_{i=1}^{d} \{\boldsymbol{w} \in \mathcal{S}^{d-1} : w_i = 0\}$ is the intersection of $\mathcal{S}^{d-1}$ with each axis:

$$\Pr\left(\min_{i=1,\ldots,d}\{X_i/w_i\} > u\right) \sim L(e^u; \boldsymbol{w}) \exp(-\Lambda(\boldsymbol{w})u), \quad u \to \infty,$$

  where $\Lambda(\boldsymbol{w})$ denotes the (extended) ADF.
- Quantifies strength of extremal dependence along any $\boldsymbol{w} \in \mathcal{S}^{d-1} \setminus \mathcal{A}$ (not $\mathcal{S}_{+}^{d-1}$) as moved from exponential to Laplace margins.
- See, also, Mackay and Jonathan (2023).

Extended ADF

Unit-level set $\partial\mathcal{G}$, alongside $\{\boldsymbol{w}/\Lambda(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{S}^{d-1} \setminus \mathcal{A}\}$.



Extended ADF

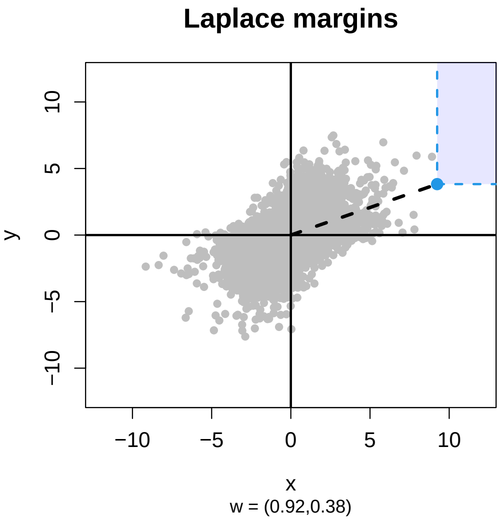
We prove the following links between Λ and $g/\partial\mathcal{G}$:

- $g(\boldsymbol{w}) \geq \Lambda(\boldsymbol{w}) \geq ||\mathbf{w}||_\infty$;
- $\Lambda(\boldsymbol{w}) = ||\boldsymbol{w}||_\infty \times \tilde{\mathfrak{r}}_{\boldsymbol{w}}^{-1}$, where $\tilde{\mathfrak{r}}_{\boldsymbol{w}} = \max\{\mathfrak{r} \in [0,1] : \mathfrak{r}\tilde{\mathcal{R}}_{\boldsymbol{w}} \cap \partial\mathcal{G} \neq \emptyset\}$ and $\tilde{\mathcal{R}}_{\boldsymbol{w}} := \bigotimes_{i=1,\ldots,d} \mathcal{U}_{w_i}$, with $\mathcal{U}_{w_i} := [w_i/||\boldsymbol{w}||_\infty, \infty]$ for $w_i > 0$ and $[-\infty, w_i/||\boldsymbol{w}||_\infty]$ for $w_i < 0$.

Can be used to estimate probabilities of the form

$$\Pr(\text{sgn}(x_i)X_i > \text{sgn}(x_i)x_i, i = 1, \ldots, d).$$

Extended ADF



Inference

To estimate $\tilde{g}(\cdot)$:

- Consider $R \mid (\boldsymbol{W} = \boldsymbol{w})$.
- Following Wadsworth and Campbell (2024), assume

$$R \mid (\boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w})) \sim \text{truncGamma}(\alpha, \tilde{g}(\boldsymbol{w})),$$

  where $\alpha > 0$ and $r_\tau(\boldsymbol{w}) > 0$ satisfies $\Pr\{R \leq r_\tau(\boldsymbol{w}) \mid \boldsymbol{W} = \boldsymbol{w}\} = \tau$ for $\tau \in (0,1)$ close to one.
- Then $\tilde{g}(\boldsymbol{w})$ is the rate parameter for the gamma distribution on $(R \mid \boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w}))$.

Inference

Our inference framework has two steps:

- Estimation of the threshold function $r_\tau(\boldsymbol{w})$, $\boldsymbol{w} \in \mathcal{S}^{d-1}$.
- Estimation of the (rescaled) gauge function $\tilde{g}(\boldsymbol{w})$, $\boldsymbol{w} \in \mathcal{S}^{d-1}$.

We perform both estimation steps using multilayer perceptrons, say $m(\cdot)$, which take input $\mathbf{w}$.

Neural net setup

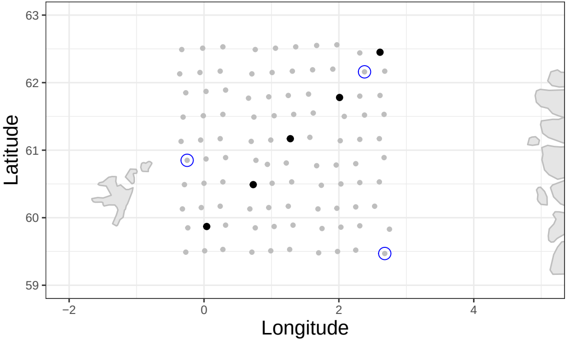
- The threshold $r_\tau(\cdot) > 0$ is represented as $\exp\{m(\mathbf{w})\}$ and estimated first using quantile regression.
- We model $h(\mathbf{w}) = \text{ReLU}(m(\mathbf{w})) + ||\mathbf{w}||_\infty$ to ensure $h(\mathbf{w}) \geq ||\mathbf{w}||_\infty$, and then rescaled to get $\tilde{g}(\cdot)$[1] .
- Inference is performed using Keras for R.

[1] $ReLU(\mathbf{x}) = \max\{x, 0\}$.

Case study

- NORA10 hindcast dataset (NOrwegian ReAnalysis 10km, Reistad et al., 2011)
- Gridded product, 3-hourly wave fields at 10 km res.
- covering the Norwegian Sea, the North Sea, and the Barents Sea
- September 1957 – December 2009
- wind speed (ws) measured in m/s, significant wave height (hs) measured in m, and mean sea level pressure (mslp) measured in hPa;

Case study



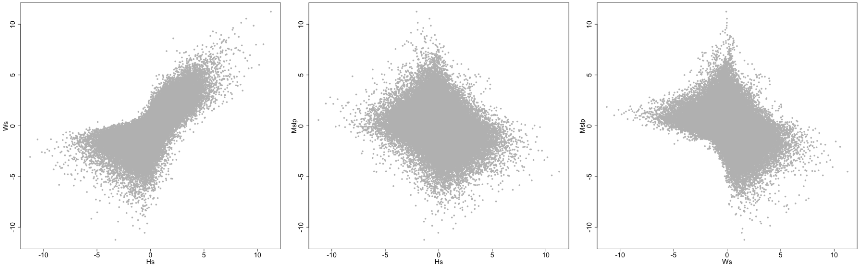


Figure: Pairwise ws, hs, and mslp at one location.

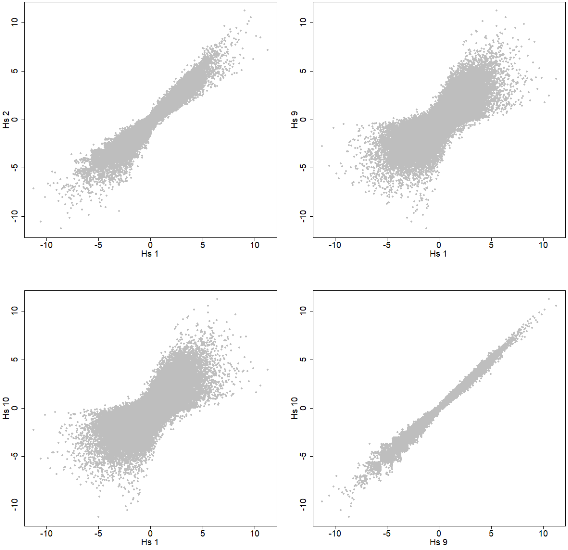


Figure: Pairwise distributions of hs across four locations in transect.

Diagnostics - $d = 3$

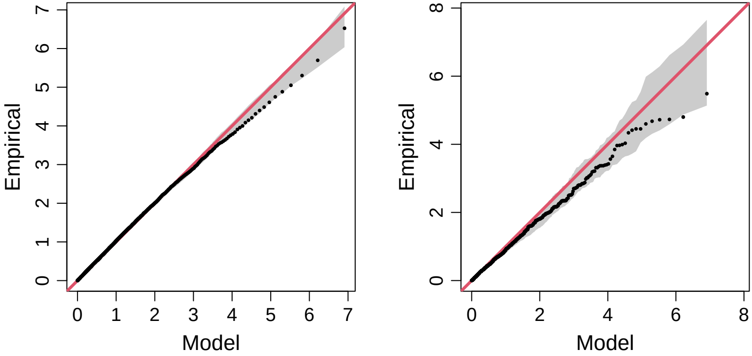


Figure: QQplots for fit of $\tilde{g}(\cdot)$ and $\Lambda(\cdot)$ for location 85.

Results - $d = 3$

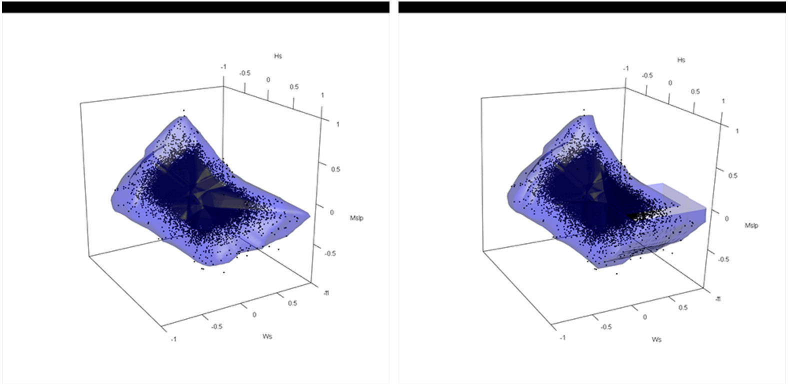


Figure: Limit set (left) and $\Lambda(\cdot)$ (right) for location 46.

Results - $d = 3$

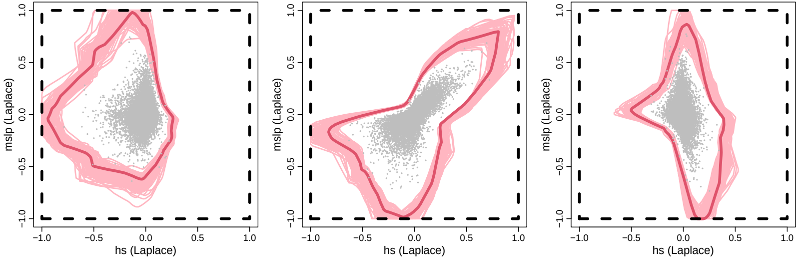


Figure: Unit level set slices for location 46.

Results - $d = 5$

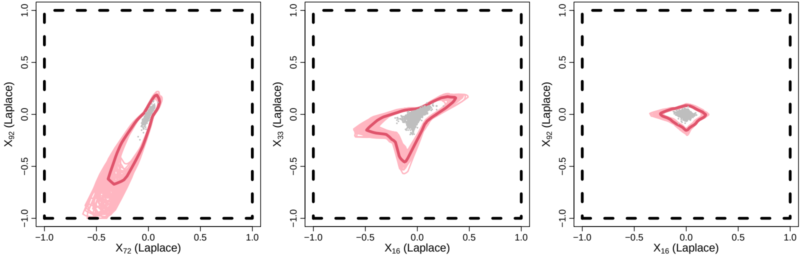


Figure: Unit level set slices for transect.

Conclusions

- "Deep Learning of Multivariate Extremes via a Geometric Representation" is available on arxiv (2406.19936)
- Further theoretical results for limit sets (with Laplace margins)
- Includes simulation study and practical advice for validating model fits
- R code available soon!

Fin.

SCAN ME

Scan for full details of my research.

References I

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66:497–546.

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.

Majumder, R., Shaby, B. A., Reich, B. J., and Cooley, D. (2023). Semiparametric estimation of the shape of the limiting multivariate point cloud. *arXiv*, 2306.13257.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2024). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *arXiv*, 2303.13237.

References II

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

Papastathopoulos, I., de Monte, L., Campbell, R., and Rue, H. (2024). Statistical inference for radially-stable generalized pareto distributions and return level-sets in geometric extremes. *arXiv*, 2310.06130.

Reistad, M., Øyvind Breivik, Haakenstad, H., Aarnes, O. J., Furevik, B. R., and Bidlot, J.-R. (2011). A high-resolution hindcast of wind and waves for the north sea, the norwegian sea, and the barents sea. *Journal of Geophysical Research*, 116:C05019.

Simpson, E. S. and Tawn, J. A. (2022). Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties. *arXiv*, 2207.02626.

References III

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107:513–532.

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*, 2208.14951.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714.