A deep learning approach to modelling joint environmental extremes

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

[1] University of Edinburgh, [2] Technische Universität Dresden, [3] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), [4] University of Arkansas

SCAN ME

Motivation

- Quantifying probabilities of joint occurrence of extremes is important in risk analysis and applications in, e.g., finance, climate, neuroscience.
- Many models for multivariate extremal dependence make restrictive assumptions about joint tail decay and do not scale well to high dimensions.
- *Geometric extremes* [1] is a flexible framework for modelling multivariate extremes that considers the deterministic limiting shape of scaled sample clouds.
- Estimation in this framework is i) limited to a low-dimensional setting, ii) requires restrictive parametric assumptions, or iii) provides estimates that do not satisfy theoretical properties.
- We provide new results for *Geometric extremes* that allow us to construct estimators that produce valid model estimates. We exploit these in a deep learning framework to build **flexible** and **scalable** models for multivariate extremal dependence.

Geometric representation of multivariate extremes

- For a d-dimensional random vector $\boldsymbol{X} \in \mathbb{R}^d$ on standard Laplace margins, with density $f(\cdot)$.
- Given n independent realisations of $\mathbf{X}$, consider the **scaled sample cloud**

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

as $n \to \infty$, where r_n is a suitably chosen normalising sequence.

- Under mild conditions C_n converges in probability onto the star-shaped and compact set

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

which intersects the boundary of $[-1,1]^d$ at least once in each component.

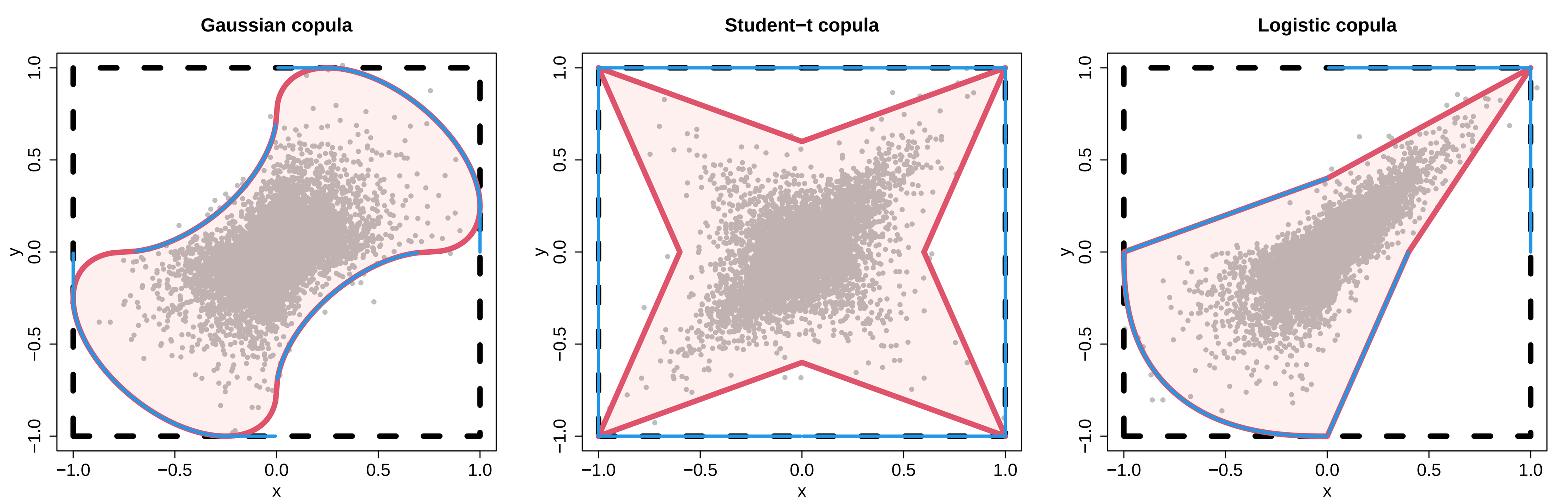

Figure 1. Scaled sample clouds of size $n = 10{,}000$. Shaded regions and solid red lines give the limit set $\mathcal{G}$ and its boundary $\partial\mathcal{G}$, while the blue lines denote the set $\{\boldsymbol{w}/\Lambda(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{S}^{d-1} \setminus \mathcal{A}\}$.

- A sufficient condition for convergence onto $\mathcal{G}$ [2] is that

$$-\log f(t\mathbf{x}) \sim t g(\mathbf{x}), \quad t \to \infty,\ \mathbf{x} \in \mathbb{R}^d,$$

where $g(\cdot)$ is a continuous 1-homogeneous **gauge function** on $\mathbb{R}^d$.

- [3] and [1] link $\partial\mathcal{G}$ to popular extremal dependence models. We can also use $\partial\mathcal{G}$ to get probabilities.
- For example, [4] assume that, as $u \to \infty$,

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

where $L(\cdot)$ is slowly-varying and $\eta \in (0,1]$ is the **coefficient of tail dependence**. Under asymptotic dependence, we have $\eta = 1$ and $\lim_{u\to\infty} L(u) > 0$.

- To connect this to $\partial\mathcal{G}$, we have that

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

- Define $(R, \boldsymbol{W})$ by $\boldsymbol{X} \mapsto (R, \boldsymbol{W}) := (\|\boldsymbol{X}\|, \boldsymbol{X}/\|\boldsymbol{X}\|)$ for $R > 0$ and $\boldsymbol{W} \in \mathcal{S}^{d-1}$, where $\mathcal{S}^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : ||\boldsymbol{x}|| = 1\}$ denotes the unit $(d-1)$-sphere and $||\cdot||$ is the Euclidean norm.
- Then

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

and so we only require a model for $g(\cdot)$ to estimate $\partial\mathcal{G}$.

Theoretical developments

- **Proposition**: For all $\boldsymbol{w} \in \mathcal{S}^{d-1}$, the gauge function $g(\cdot)$ satisfies the constraint that

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

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

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

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

where $\mathbf{0}_d := (0, \dots, 0)$. Then $\mathcal{H}$ is star-shaped, compact, and satisfies $\mathcal{H} \subseteq [-1,1]^d$, with unit-level set $\partial\mathcal{H} = \left\{\boldsymbol{w}h(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{S}^{d-1}\right\}$.

- To ensure $\partial\mathcal{H}$ also satisfies the blue property, we perform a numerical rescaling during the estimation procedure. This provides a new construction for valid (rescaled) gauge functions:

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

where $\kappa(\cdot)$ is a closed form bijective mapping and $b_i(\cdot), i = 1, \dots, d$, are scaling coefficients.

- We extend [5] and assume that, 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,

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

where $\Lambda(\boldsymbol{w})$ denotes the (extended) angular dependence function (ADF).

- To link Λ to $\partial\mathcal{G}$, we have $\Lambda(\boldsymbol{w}) = ||\boldsymbol{w}||_\infty \times \tilde{\mathfrak{r}}_{\boldsymbol{w}}^{-1}$, 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 $\tilde{\mathcal{R}}_{\boldsymbol{w}} := \bigotimes_{i=1,\dots,d} \mathcal{U}_{w_i}$, with $\mathcal{U}_{w_i} := [w_i/||\boldsymbol{w}||_\infty, \infty]$ for $w_i > 0$ and $[-\infty, w_i/||\boldsymbol{w}||_\infty]$ for $w_i < 0$.

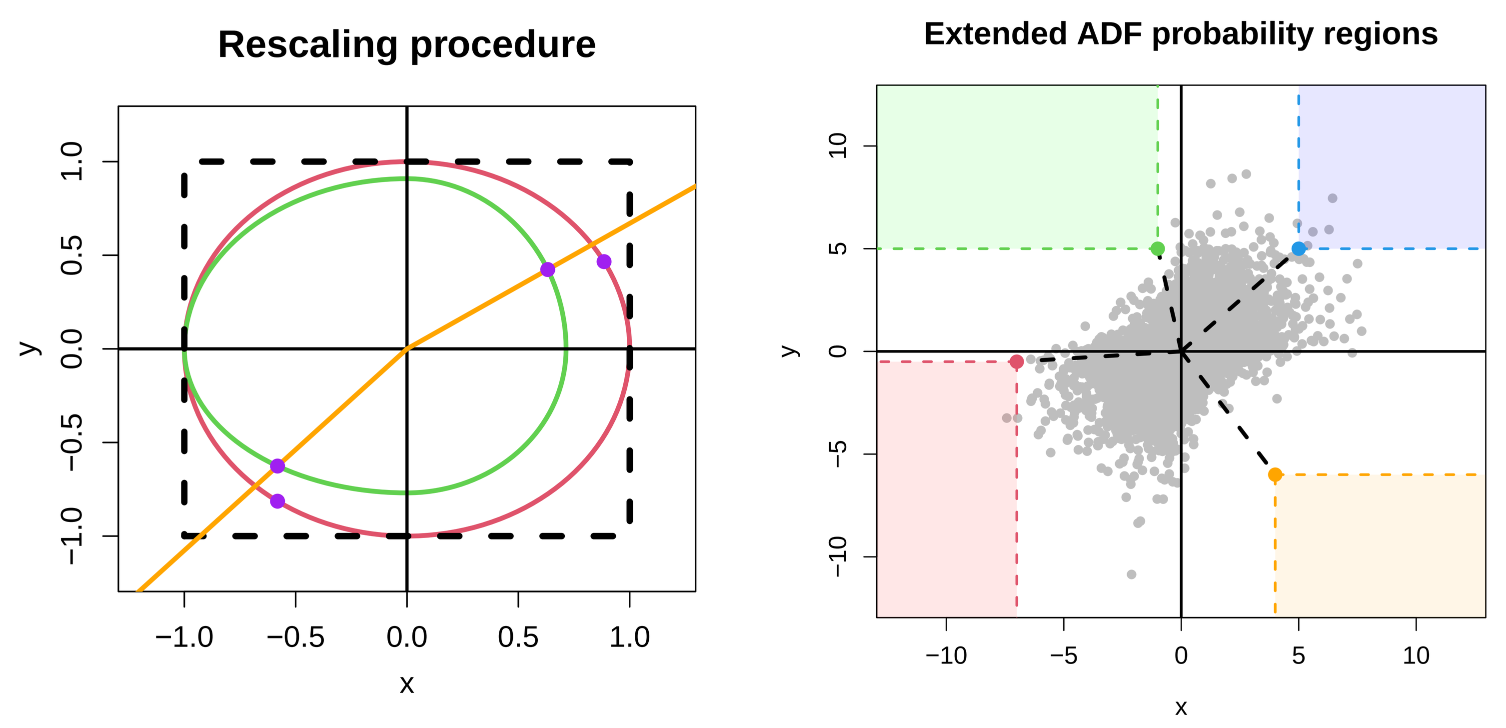

Figure 2. Left: rescaling. Right: example probability regions that can be evaluated using model (1).

Modelling and inference with neural networks

- For conditional radial distribution $R \mid (\boldsymbol{W} = \boldsymbol{w})$, we follow [6] and assume that

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

where $\alpha > 0$ and $r_\tau(\boldsymbol{w}) > 0$ satisfies $\Pr\{R \le r_\tau(\boldsymbol{w}) \mid \boldsymbol{W} = \boldsymbol{w}\} = \tau$ for $\tau \in (0,1)$ close to one.

- Then $\tilde{g}(\boldsymbol{w})$ is the rate parameter for the gamma distribution on $(R \mid \boldsymbol{W} = \boldsymbol{w}, R > r_\tau(\boldsymbol{w}))$ and can be found via standard maximum likelihood techniques.
- We model $r_\tau(\boldsymbol{w})$ and $1/h(\boldsymbol{w})$ as multi-layer perceptrons which take the angles $\boldsymbol{w}$ as their input. These are designed so that $r_\tau(\boldsymbol{w}) > 0$ and $1/h(\boldsymbol{w}) \ge ||\boldsymbol{w}||_\infty$ for all $\boldsymbol{w} \in \mathcal{S}^{d-1}$. The rescaling of $h(\cdot)$ is then performed to get the (rescaled) gauge function $\tilde{g}(\boldsymbol{w})$, $\boldsymbol{w} \in \mathcal{S}^{d-1}$.
- We estimate the conditional quantile $r_\tau(\boldsymbol{w})$ using standard quantile regression techniques. Inference for both this, and $\tilde{g}(\boldsymbol{w})$, is performed using the `R` interface to the deep learning library `Keras`.

Application to metocean extremes

- NORA10 hindcast dataset [NOrwegian ReAnalysis 10km, 7];
- Gridded, 3-hourly fields at 10 km resolution;
- September 1957 – December 2009;
- Wind speed (ws; m/s), significant wave height (hs; m), mean sea level pressure (mslp; hPa);
- Uncertainty through stationary bootstrap.

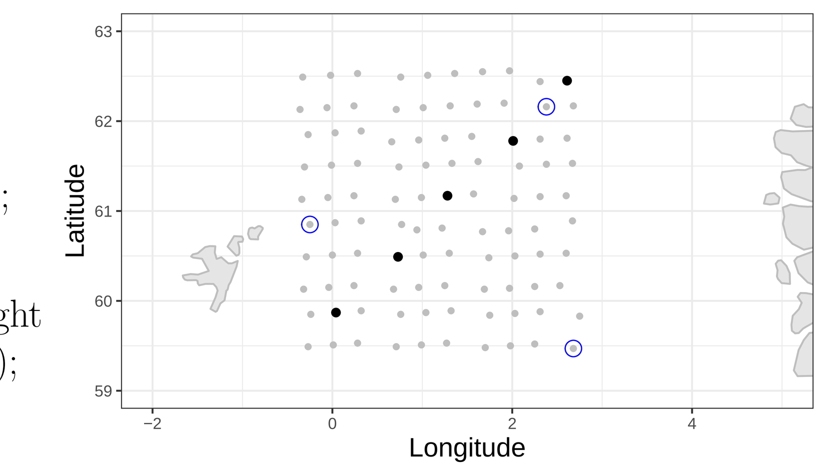

Figure 3. Study locations.

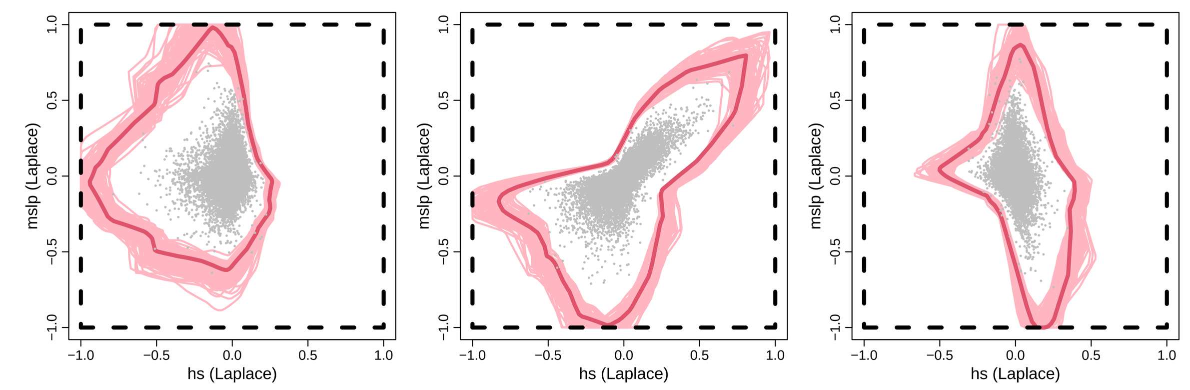

Figure 4. Estimated unit-level set slices for north-east location.

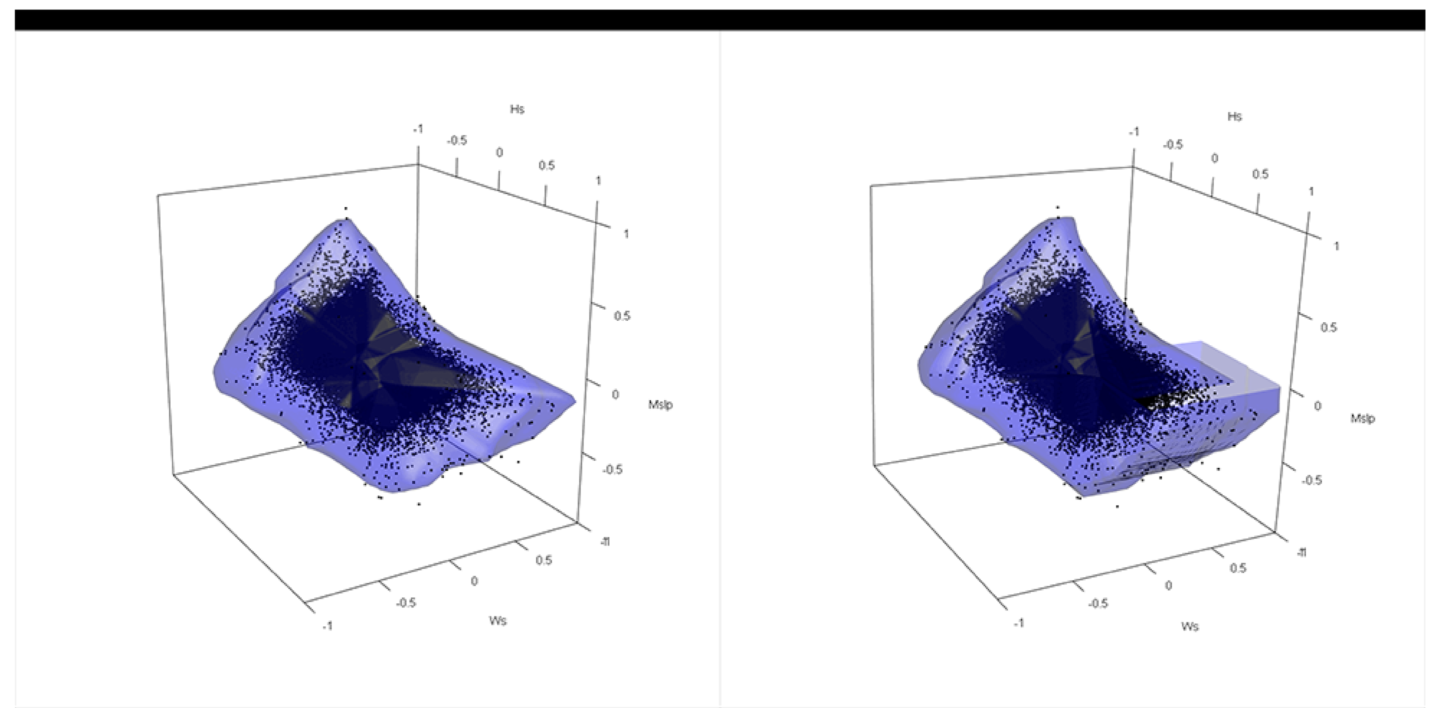

Figure 5. Estimated limit set (left) and $\Lambda(\cdot)$ (right) for north-east location.

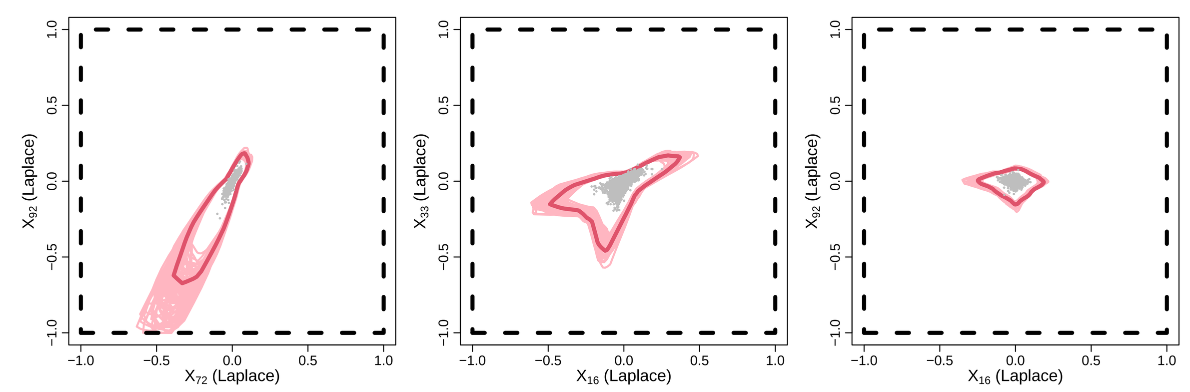

Figure 6. Estimated unit-level set slices for $d = 5$ transect (black locations).

References

[1] N. Nolde and J. L. Wadsworth, "Linking representations for multivariate extremes via a limit set," *Advances in Applied Probability*, vol. 54, pp. 688–717, 3 2022.
[2] G. Balkema and N. Nolde, "Asymptotic independence for unimodal densities," *Advances in Applied Probability*, vol. 42, pp. 411–432, 2 2010.
[3] N. Nolde, "Geometric interpretation of the residual dependence coefficient," *Journal of Multivariate Analysis*, vol. 123, pp. 85–95, 2014.
[4] A. W. Ledford and J. A. Tawn, "Statistics for near independence in multivariate extreme values," *Biometrika*, vol. 83, pp. 169–187, 1 1996.
[5] J. L. Wadsworth and J. A. Tawn, "A new representation for multivariate tail probabilities," *Bernoulli*, vol. 19, pp. 2689–2714, 5 B 2013.
[6] J. L. Wadsworth and R. Campbell, "Statistical inference for multivariate extremes via a geometric approach," *Journal of the Royal Statistical Society Series B: Methodology*, vol. 2208.14951, 2024.
[7] M. Reistad, Øyvind Breivik, H. Haakenstad, O. J. Aarnes, B. R. Furevik, and J.-R. Bidlot, "A high-resolution hindcast of wind and waves for the north sea, the norwegian sea, and the barents sea," *Journal of Geophysical Research*, vol. 116, p. C05019, C5 May 2011.
[8] C. J. Murphy-Barltrop, R. Majumder, and J. Richards, "Deep learning of multivariate extremes via a geometric representation," *arXiv:2406.19936*, 2024.

LaTeX TikZposter