Cluster analysis of conditional extreme value models

Patrick O'Toole[1], Christian Rohrbeck[1], **Jordan Richards**[2]

[1]Department of Mathematical Sciences, University of Bath, [2]School of Mathematics, University of Edinburgh

SCAN ME

Motivation

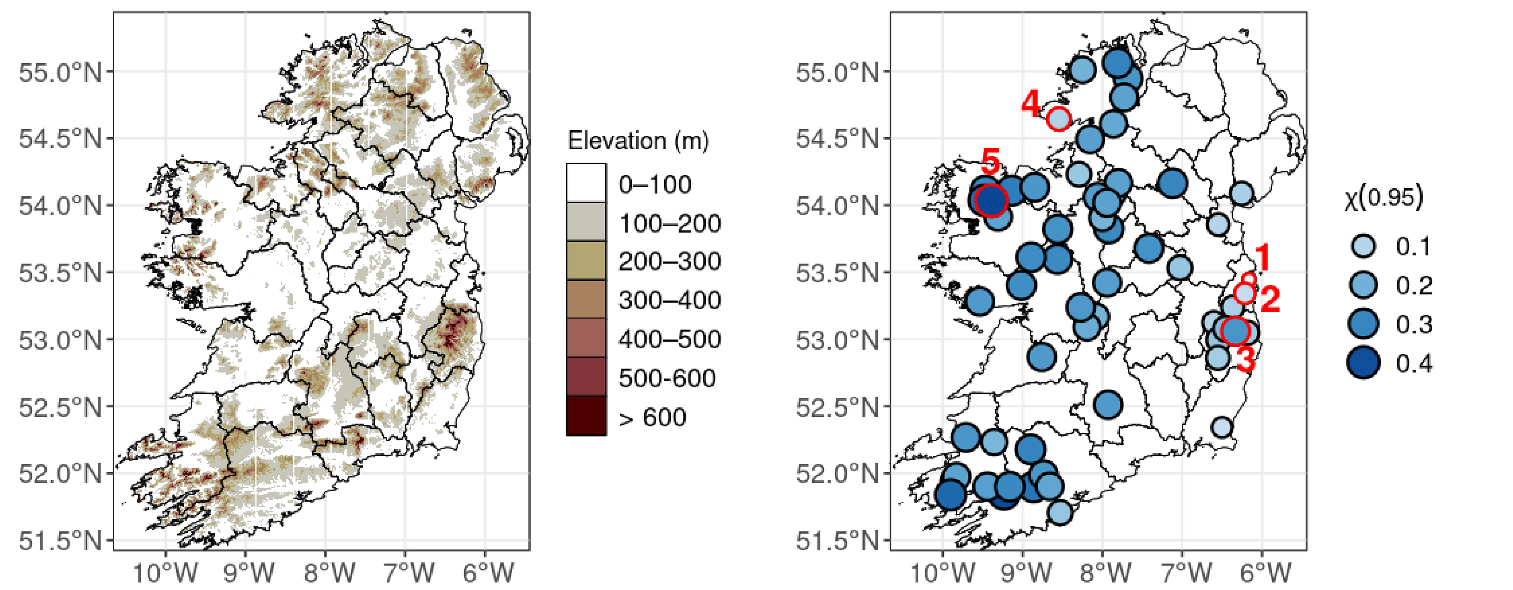
Environmental extremes often occur simultaneously across multiple variables (e.g., precipitation and wind speed). Understanding multivariate tail dependence is essential for risk assessment, but becomes difficult when data are observed across many locations.

Our goal is to identify groups of locations exhibiting similar extremal dependence, quantified by

$$\chi := \lim_{u \to 1} \chi(u) = \lim_{u \to 1} \Pr\left(F_1(X_1) > u \mid F_2(X_2) > u\right). \tag{1}$$

Although χ distinguishes asymptotic dependence ($\chi > 0$) from asymptotic independence ($\chi = 0$), it provides only a pairwise summary and cannot characterise the full multivariate tail.

Multivariate conditional extremes (CE; [1]) models provide a flexible description of tail dependence, but interpreting hundreds of fitted models across many locations is challenging.



Locations in red are (1) Malahide Castle, Dublin, (2) Ringsend, Dublin, (3) Glenmacnass, Wicklow, (4) Kilcar, Donegal, and (5) Derryhillagh, Mayo.

Our contribution

- Closed-form measure of divergence between multivariate CE models.
- Efficient clustering of multivariate extremes.
- Applicable in arbitrary dimensions, d (vector dimension) or D (number of sites).

Multivariate conditional extremes

For vector $\mathbf{Y}$ with margins $Y_i \sim \text{Laplace}, i = 1, \ldots, d$:

- Pick a conditioning variable, Y_i, e.g., wind speed. Consider extremes $Y_i > u$ for large $u > 0$.
- Model remaining vector $\mathbf{Y}_{-i}$ conditional on extremes of Y_i:

$$(\mathbf{Y}_{-i} \mid Y_i = y) = \boldsymbol{\alpha}_i y + y^{\boldsymbol{\beta}_i} \mathbf{Z}_i, \qquad y > u.$$

- Parameter vectors $\boldsymbol{\alpha} \in [-1, 1]^{d-1}$ and $\boldsymbol{\beta} \in (-\infty, 1]^{d-1}$ describe extremal dependence.
- A Gaussian working model, $\mathbf{Z}_i \sim \text{MVN}_{d-1}$, used for the residuals provides efficient estimation.

Multivariate tail divergence

With CE models at each site, $s = 1, \ldots, D$, differences in extremal dependence (between two sites, say s and s^*) can be quantified using the skew-geometric Jensen-Shannon divergence (JS^{G_λ}; [2]):

$$\text{JS}^{G_\lambda}(H \| H^*) = (1 - \lambda)\,\text{KL}\left\{G_\lambda(H, H^*) \| H\right\} + \lambda\,\text{KL}\left\{G_\lambda(H, H^*) \| H^*\right\}, \tag{2}$$

for distribution functions H and H^*, and where KL is the Kullback–Leibler divergence. We use this to define the *expected extremal skew-geometric Jensen-Shannon divergence*:

- $$\text{eJS}^{G_\lambda}_{i,s,s^*} = \mathbb{E}\left(\text{cJS}^{G_\lambda}_{i,s,s^*}(Y_i) \mid Y_i > u\right) = \int_u^\infty \text{cJS}^{G_\lambda}_{i,s,s^*}(y) h(y \mid Y_i > u)\,\mathrm{d}y,$$

  where $\text{cJS}^{G_\lambda}_{i,s,s^*}(y)$ measures the JS^{G_λ} between two CE models (at sites s and s^*, and for conditioning variable Y_i) and $h(y \mid Y_i > u)$ is the truncated Laplace density.
- Under the Gaussian working assumption and mild conditions, we show that this divergence measure is non-negative, bounded, and symmetric.

Clustering of multivariate extremal dependence

We collect dissimilarity in (conditional) tail dependence using the matrix:

$$M^{(i)} := \left(\text{eJS}^{G_\lambda}_{i,s,s^*}\right)_{s,s^* = \{1,\ldots,D\}}. \tag{3}$$

This matrix, or one that aggregates over conditioning variables (e.g., $M = \sum_{i=1}^d M^{(i)}$), can then be used to determine clusters of sites which exhibit similar tail dependence.

- In particular, we use Partitioning Around Medoids (PAM), with optimal number of clusters, k, determined by elbow plots of the *total within-group divergence*:

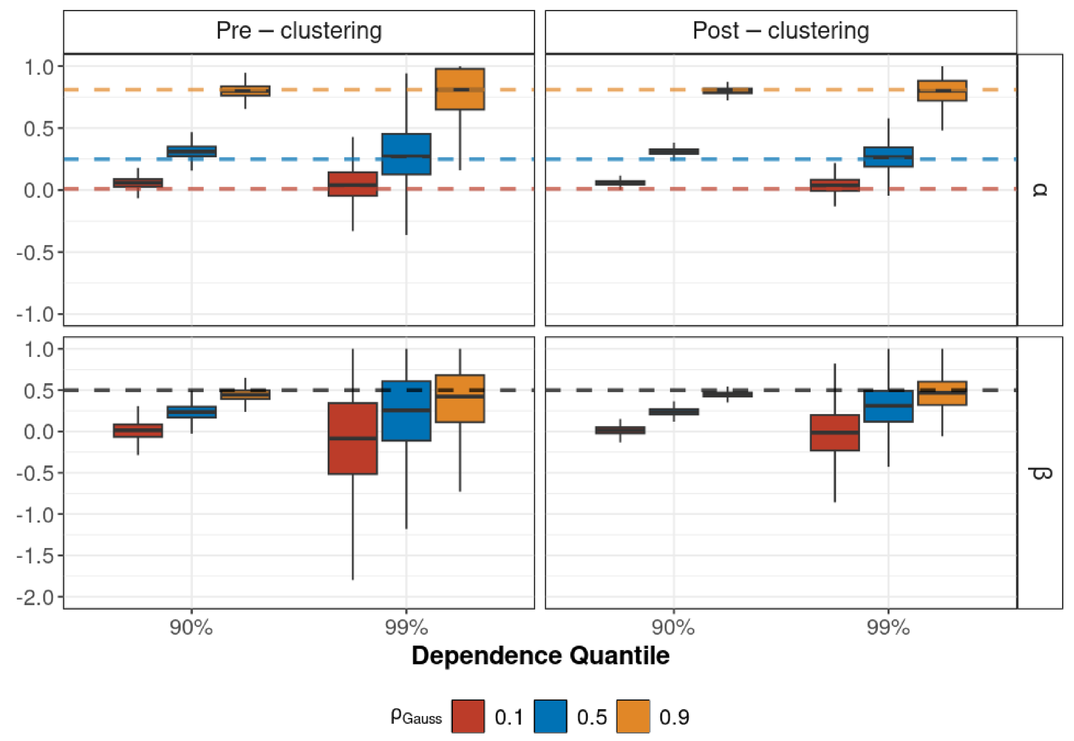
$$\text{TWD}(k) = \sum_{c=1}^{k} \sum_{s \in \mathcal{C}_c} M_{s,\kappa_c}, \tag{4}$$

where κ_c denotes the medoid site for the c-th distinct cluster, $\mathcal{C}_c \subseteq \{1, \ldots, D\}$.

Simulation study

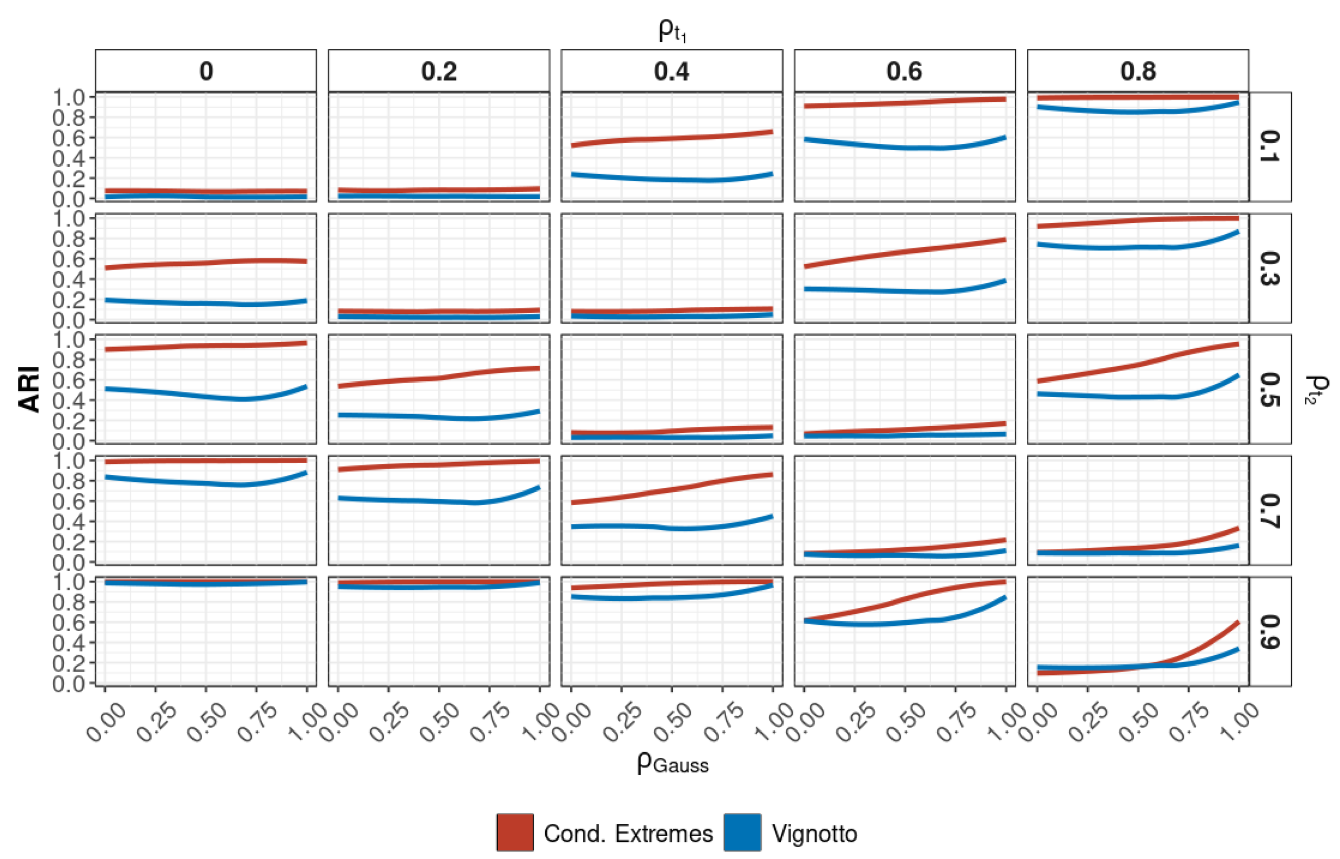
For simulation study 1 - CE parameter estimation uncertainty reduction:

- Simulated bivariate Gaussian copula, with correlation ρ_{Gauss} and $D = 12$, and $k = 3$ clusters.
- Known true values: $\alpha = \rho^2_{\text{Gauss}}$ and $\beta = 1/2$.
- Uncertainty in CE estimates compared *pre-clustering* - representing estimates obtained separately at each site, displayed according to the true cluster structure - and *post-clustering* - derived by pooling data across all sites according to their estimated clustering.



For simulation study 2 - Efficacy of extremal clustering, against competitor [3]:

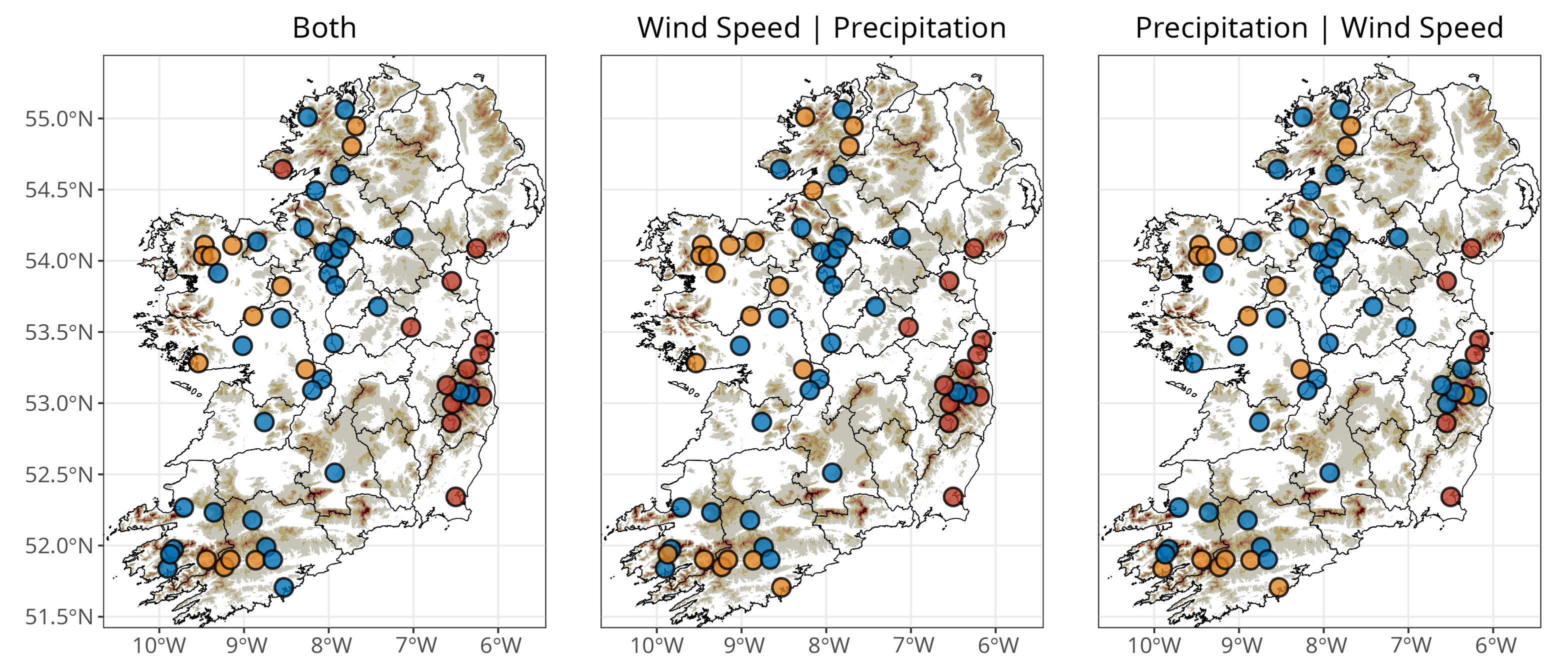
- Simulated bivariate mixture of 3 copulas: Gaussian, with correlation ρ_{Gauss} and two t-copulas, with correlations ρ_{t_1} and ρ_{t_2}. Mixture exhbits asymptotic dependence.
- Two latent clusters, $D = 12$; efficacy evaluated using the Adjusted Rand Index.



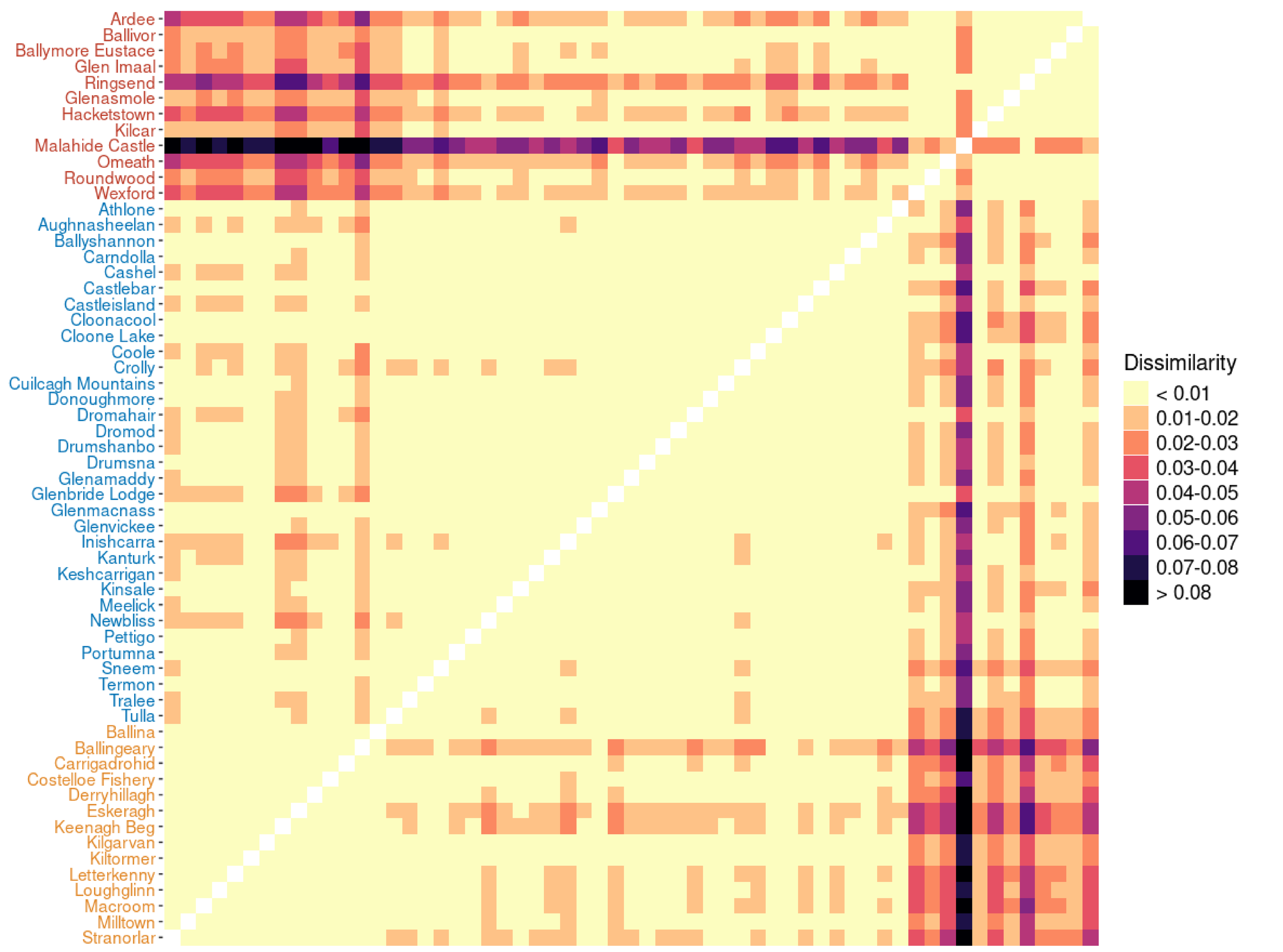
Irish precipitation and wind speed extremes

We analyse the joint extremal behaviour of precipitation and wind speed at $D = 59$ sites across the Republic of Ireland from 1990 to 2020:

- October–March; weekly total precipitation (Met Éireann) and average daily wind speeds (ERA5).
- Found $k = 3$ clusters to be optimal; clusters exhibit spatial structure, with a clear contrast between the east and west coasts, separated by a central cluster.



Extremal cluster estimates using the dissimilarity matrix for wind speed conditional on precipitation, precipitation conditional on wind speed, and the combined dissimilarity matrix.



Heat-map of the estimated aggregated dissimilarity matrix M.

References

[1] J. E. Heffernan and J. A. Tawn, "A conditional approach for multivariate extreme values (with discussion)," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 66, no. 3, pp. 497–546, 2004.

[2] F. Nielsen, "On the Jensen–Shannon symmetrization of distances relying on abstract means," *Entropy*, vol. 21, no. 5, p. 485, 2019.

[3] E. Vignotto, S. Engelke, and J. Zscheischler, "Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland," *Weather and Climate Extremes*, vol. 32, p. 100318, 2021.

[4] P. O'Toole, C. Rohrbeck, and J. Richards, "Clustering of multivariate tail dependence using conditional methods," *arXiv:2510.20424*, 2025.

LaTeX TikZposter