

Clustering of multivariate tail dependence using conditional methods

Patrick O'Toole*

Department of Mathematical Sciences, University of Bath, UK

Christian Rohrbeck

Department of Mathematical Sciences, University of Bath, UK

Jordan Richards

School of Mathematics

and Maxwell Institute for Mathematical Sciences,
University of Edinburgh, UK

December 1, 2025

Abstract

The conditional extremes (CE) framework has proven useful for analysing the joint tail behaviour of random vectors. However, when applied across many locations or variables, it can be difficult to interpret or compare the resulting extremal dependence structures, particularly for high dimensional vectors. To address this, we propose a novel clustering method for multivariate extremes using the CE framework. Our approach introduces a closed-form, computationally efficient dissimilarity measure for multivariate tails, based on the skew-geometric Jensen-Shannon divergence, and is applicable in arbitrary dimensions. Applying standard clustering algorithms to a matrix of pairwise distances, we obtain interpretable groups of random vectors with homogeneous tail dependence. Simulation studies demonstrate that our method outperforms existing approaches for clustering bivariate extremes, and uniquely extends to the multivariate setting. In our application to Irish meteorological data, our clustering identifies spatially coherent regions with similar extremal dependence between precipitation and wind speeds.

Keywords: Conditional extremes; Extreme value analysis; Jensen-Shannon divergence; Multivariate extremes.

*Patrick O'Toole is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/S022945/1.

1 Introduction

Extreme Value Theory (EVT) provides asymptotically justified models for estimating the probability of events outside the range of observed values (Coles, 2001). EVT has found broad usage in numerous, disparate domains, including finance (Poon et al., 2004), public health (Vettori et al., 2019), neuroscience (Talento et al., 2025), and ecology (Koh and Opitz, 2025). Environmental applications are one area which benefits from the use of EVT methods, to estimate the severity of future extreme events and develop risk management strategies. Examples include the analysis of heatwaves (Tanarhte et al., 2015; French et al., 2019), extreme rainfall (Osei et al., 2021), wind speeds (Steinkohl et al., 2013; Soukissian and Tsalis, 2015), wildfire risks (Koh et al., 2023), and the impact of environmental extremes on insurance claims (Rohrbeck et al., 2018; Miralles and Davison, 2024).

In the multivariate EVT setting, one is interested in the *joint* tail behaviour of a random vector. Multivariate extremes frequently arise in spatial applications, where one or more variables are recorded at a discrete set of sampling locations. Within such settings, simultaneous or “concomitant” extremes, where multiple variables or locations concurrently exhibit extremal behaviour, can produce substantially more severe outcomes than single-variable extremes (Zscheischler et al., 2018; Bevacqua et al., 2021; Rohrbeck and Tawn, 2021; Sando et al., 2022). Consequently, joint risks may be underestimated unless tail *dependence* is explicitly modelled (Zhang et al., 2024). The strength of tail dependence between two random vectors is commonly summarised by the coefficient of asymptotic dependence $\chi \in [0, 1]$ (see e.g., Coles et al., 1999). For two random variables X_1 and X_2 with distribution functions F_1 and F_2 , respectively, χ is defined as

$$\chi = \lim_{u \rightarrow 1} \chi(u) = \lim_{u \rightarrow 1} \Pr \{F_1(X_1) > u \mid F_2(X_2) > u\}. \quad (1)$$

When $\chi > 0$, the variables X_1 and X_2 are *asymptotically dependent* (AD), while $\chi = 0$ corresponds to *asymptotic independence* (AI) between X_1 and X_2 . Although χ is a useful summary of pairwise extremal dependence, it does not fully characterise the multivariate tail dependence structure or support inference on joint exceedance probabilities, for which a full probabilistic model is required.

We consider a setting with D random vectors, $\mathbf{X}_1, \dots, \mathbf{X}_D$, of equal length d , but with potentially different joint tail behaviours. For instance, we may observe environmental variables (e.g., rainfall and wind speed) that exhibit non-stationary extremal dependence with respect to their sampling index or location: this could be time, or a fixed spatial location. Interest then lies in understanding the variation in tail behaviour across the D vectors and, where feasible, to pool information to reduce estimation uncertainty. However, we have limited tools for identifying random vectors which exhibit similar tail behaviour. To this end, we leverage the multivariate conditional extremes (CE) framework of [Heffernan and Tawn \(2004\)](#), which is a popular and flexible model for multivariate extremes, to propose a novel clustering algorithm for multivariate extremes. To illustrate the efficacy of our approach, we consider a motivating application with variables observed at D spatial locations (see [Section 4](#)), but our method is more broadly applicable to other fields, e.g. finance, where clustering multivariate extremes is of interest.

[Heffernan and Tawn \(2004\)](#) described the tail behaviour of a random vector *conditional* on one of its components being extreme, i.e. exceeding a suitably high threshold. This framework can flexibly capture both AD and AI, and is both conceptually simple and computationally efficient. Spatial and spatio-temporal extensions have been developed to handle high-dimensional data ([Simpson and Wadsworth, 2021](#); [Wadsworth and Tawn, 2022](#)), and adaptations that incorporate covariates, hierarchical structures, and mixtures have further increased the models' flexibility ([Richards et al., 2023](#); [Tendijck et al., 2023](#); [Talento et al., 2025](#)). As a result, CE models have seen widespread application in settings where AI

as well as AD may be prevalent, including river flow and flood risk (Gilleland et al., 2013; Jane et al., 2020), rainfall (Richards et al., 2022), sea-surface temperatures (Simpson et al., 2023; Kakampakou et al., 2024), heatwaves (Winter et al., 2016), neuroscience (Guerrero et al., 2023; Talento et al., 2025), and finance (Hilal et al., 2014). The broad applicability of CE models makes them well-suited to our motivating example. However, the scarcity of joint extreme observations, particularly in high dimensions, and the number of model parameters, often results in high uncertainty in estimates of the CE model.

Clustering is a popular tool in the extremes literature, where data sparsity and low sample sizes provide a natural motivation, and we distinguish three types of extremal clustering methods. *Univariate* methods group data by marginal tail behaviour, often through summaries such as return levels or parameter estimates (Coles, 2001), and have been used in environmental and financial applications (Alonso et al., 2014; de Carvalho et al., 2023; Zheng et al., 2023). *Spatial* methods cluster entire spatial fields or partition the spatial domain via pairwise tail dissimilarities, such as the F-madogram (Cooley et al., 2006; Bernard et al., 2013), or by fitting non-stationary extremal dependence models (Carreau et al., 2017; Rohrbeck and Tawn, 2021; Dupuis et al., 2023; Shao et al., 2025). Finally, *multivariate* methods cluster on joint tail summaries or fitted multivariate extremal dependence models. Vignotto et al. (2021) consider clusters of bivariate extremes of precipitation and wind speeds across spatial locations in Great Britain and Ireland. Their nonparametric approach estimated pairwise Kullback–Leibler (KL) (Kullback and Leibler, 1951) divergences for bivariate precipitation and wind extremes (see Zscheischler et al. (2021)) and then applied the k -medoids algorithm. However, their method is restricted to the bivariate case and, being non-parametric, lacks a generative model for tail dependence inference. Also, it is undefined where no joint extreme events are observed, which can occur frequently in environmental applications, where AI is prevalent.

We address the limitations of existing methods by developing a clustering framework for multivariate extremes that leverages the CE framework. This enables flexible, parametric exploration of tail dependence in arbitrary dimensions, in contrast to the existing non-parametric approach of [Vignotto et al. \(2021\)](#), and overcomes the data sparsity issue often associated with drawing interpretations from fitted CE models. Many classical models and clustering frameworks assume AD ([Vignotto et al., 2021](#); [Boulin et al., 2025](#)), but recent literature argues that AD models are inappropriate for environmental data ([Opitz, 2016](#); [Dawkins and Stephenson, 2018](#); [Huser et al., 2024](#)). Our clustering method can flexibly handle both AD and AI. Using the working Gaussian modelling assumption of [Heffernan and Tawn \(2004\)](#), we define a closed-form divergence measure for CE models which permits clustering using traditional similarity-based algorithms. Our clustering approach delivers a unified and statistically principled tool for exploratory analysis of multivariate extremes, facilitating the identification of groups with similar tail dependence.

The remainder of this paper is structured as follows. Section 2 introduces our novel dissimilarity measure for multivariate CE models, via the skew-geometric Jensen-Shannon divergence (JS^{G^λ}) of [Nielsen \(2019\)](#), and we further illustrate its use for clustering of multivariate extremes. Section 3 provides a simulation study where we demonstrate that our approach outperforms the clustering method of [Vignotto et al. \(2021\)](#). In Section 4, we apply our clustering framework to Irish meteorological data, and identify sites with similar tail dependence between precipitation and wind speeds. Section 5 concludes with a discussion of our findings, the limitations of our method, and potential avenues for future work.

2 Methodology

In this section, we introduce our clustering framework. Section 2.1 provides background on the conditional extremes (CE) model of [Heffernan and Tawn \(2004\)](#) and [Keef et al.](#)

(2013). Section 2.2 defines the setup for our clustering model and describes how inference is performed. In Section 2.3, we describe our novel dissimilarity measure for multivariate extremes, which uses the skew-geometric Jensen-Shannon divergence to efficiently quantify dissimilarity between fitted CE models. Finally, in Section 2.4, we outline how to perform clustering using this dissimilarity measure, and provide practical advice on choosing the number of clusters.

2.1 Background to multivariate conditional extremes

Like many other models for extremal dependence (see, e.g., Davison and Huser, 2015), the conditional extremes framework of Heffernan and Tawn (2004) is defined for data on standardised margins, which ensures that differences in marginal scale do not affect estimation of the dependence structure. As introduced in Keef et al. (2013), a pragmatic choice of margins for CE models is standard Laplace, as it allows for modelling of both positive and negative dependence. Let $\mathbf{X} := (X_1, \dots, X_d) \in \mathbb{R}^d$ be a d -dimensional random vector representing observed data with marginal distribution functions F_{X_1}, \dots, F_{X_d} . We apply the probability integral transformation to transform \mathbf{X} to a random vector with standard Laplace margins, denoted by $\mathbf{Y} := (Y_1, \dots, Y_d) \in \mathbb{R}^d$. Formally, for $i = 1, \dots, d$,

$$Y_i := \begin{cases} \log \{2F_{X_i}(X_i)\}, & \text{for } X_i < F_{X_i}^{-1}(0.5), \\ -\log [2\{1 - F_{X_i}(X_i)\}], & \text{for } X_i \geq F_{X_i}^{-1}(0.5). \end{cases} \quad (2)$$

Let $\mathbf{Y}_{-i} := \mathbf{Y} \setminus Y_i \in \mathbb{R}^{d-1}$ denote \mathbf{Y} with its i^{th} component removed, for $i \in \{1, \dots, d\}$. With all operations below taken component-wise, Heffernan and Tawn (2004) assumed that there exist parameter vectors $\boldsymbol{\alpha}_{|i} := \{\alpha_{j|i} : j \in \{1, \dots, d\} \setminus i\} \in [-1, 1]^{d-1}$ and $\boldsymbol{\beta}_{|i} := \{\beta_{j|i} : j \in \{1, \dots, d\} \setminus i\} \in (-\infty, 1]^{d-1}$ such that, for fixed $\mathbf{z} \in \mathbb{R}^{d-1}$,

$$\mathbb{P} \left(\frac{\mathbf{Y}_{-i} - \boldsymbol{\alpha}_{|i} Y_i}{Y_i^{\boldsymbol{\beta}_{|i}}} \leq \mathbf{z}, Y_i - u > y \mid Y_i > u \right) \rightarrow G_{|i}(\mathbf{z}) \exp(-y) \quad \text{as } u \rightarrow \infty, \quad (3)$$

where the distribution function $G_{|i}(\mathbf{z})$ is non-degenerate on $\mathbf{z} \in [-\infty, \infty)^{d-1}$. Equation (3) implies that, in the limit as $u \rightarrow \infty$, the excesses $Y_i - u$ and the standardised residuals $\mathbf{Z}_{|i} := Y_i^{-\beta_{|i}} (\mathbf{Y}_{-i} - \boldsymbol{\alpha}_{|i} Y_i)$ are independent.

Assuming that the limit in Equation (3) holds exactly for a large threshold $u_i > 0$ gives the heteroscedastic regression model

$$(\mathbf{Y}_{-i} | Y_i = y) = \boldsymbol{\alpha}_{|i} y + y^{\beta_{|i}} \mathbf{Z}_{|i}, \quad \text{for } y > u_i, \quad (4)$$

with residual vector $\mathbf{Z}_{|i} \sim G_{|i}$. The parameter vector $\boldsymbol{\alpha}_{|i}$ describes the strength of extremal association between components of \mathbf{Y}_{-i} and the conditioning variable Y_i , while $\beta_{|i}$ determines the spread induced by large values of Y_i . Some special cases (for $j \in \{1, \dots, d\} \setminus i$) of the CE model are: $\alpha_{j|i} = 0$ and $\beta_{j|i} = 0$, which imply asymptotic independence between Y_j and Y_i , corresponding to $\chi = 0$; $\alpha_{j|i} = 1$ (-1) and $\beta_{j|i} = 0$, which corresponds to positive (negative) asymptotic dependence, and $-1 < \alpha_{j|i} < 1$, which suggests asymptotic independence between Y_j and Y_i , with the strength of extremal dependence weakening with $\alpha_{j|i}$.

2.2 Framework and model inference

In our motivating spatial application, we consider a CE model that varies across sites and across conditioning variables, $i = 1, \dots, d$. To denote spatial variation, we include the additional subscript $s \in \{1, \dots, D\}$, where D is the number of spatial locations, so that $Y_{i,s}$ denotes the variable Y_i observed at site s . This leads to a spatial extension of Equation (4), with site-specific random vector $\mathbf{Y}_s := (Y_{1,s}, \dots, Y_{d,s}) \in \mathbb{R}^d$, and, for $i = 1, \dots, d$,

$$(\mathbf{Y}_{-i,s} | Y_{i,s} = y) = \boldsymbol{\alpha}_{|i,s} y + y^{\beta_{|i,s}} \mathbf{Z}_{|i,s}, \quad y > u_{i,s}, \quad (5)$$

where $\mathbf{Y}_{-i,s} := \mathbf{Y}_s \setminus Y_{i,s}$, and $\boldsymbol{\alpha}_{|i,s}, \boldsymbol{\beta}_{|i,s}, \mathbf{Z}_{|i,s} \sim G_{|i,s}$, and $u_{i,s}$ are the site-specific analogues of the regression parameters, residuals, and exceedance threshold in Equations (3) and (4). Note that we do not assume spatial smoothness in this regression model.

As a working assumption that permits inference on $\boldsymbol{\alpha}_{|i,s}$ and $\boldsymbol{\beta}_{|i,s}$, we follow [Heffernan and Tawn \(2004\)](#) and assume that the residual vector $\mathbf{Z}_{|i,s}$ in Equation (5) is multivariate Gaussian with mean vector $\boldsymbol{\mu}_{|i,s} \in \mathbb{R}^{d-1}$ and covariance matrix $\Sigma_{|i,s} \in \mathbb{R}^{(d-1) \times (d-1)}$ for $i = 1, \dots, d$ and $s = 1, \dots, D$. This working assumption is the key to deriving a closed-form divergence measure for computationally efficient clustering; see Section 2.3. It follows that

$$(\mathbf{Y}_{-i,s} \mid Y_{i,s} = y) \sim \text{MVN}_{d-1} \left(\boldsymbol{\alpha}_{|i,s} y + y^{\boldsymbol{\beta}_{|i,s}} \boldsymbol{\mu}_{|i,s}, (y^{\boldsymbol{\beta}_{|i,s}})^{\text{T}} \Sigma_{|i,s} y^{\boldsymbol{\beta}_{|i,s}} \right), \quad y > u_{i,s}, \quad (6)$$

where $\boldsymbol{\alpha}_{|i,s}, \boldsymbol{\beta}_{|i,s}, \boldsymbol{\mu}_{|i,s}, \Sigma_{|i,s}$ are the site-specific parameters for the i^{th} conditioning variable, $i = 1, \dots, d, s = 1, \dots, D$. To compare tail dependence across different random vectors \mathbf{Y}_s at different sites s , we compare the conditional distributions in Equation (6), i.e., the assumed multivariate Gaussian distributions parameterised by $\boldsymbol{\alpha}_{|i,s}, \boldsymbol{\beta}_{|i,s}, \boldsymbol{\mu}_{|i,s}$ and $\Sigma_{|i,s}$.

We estimate parameters $\boldsymbol{\alpha}_{|i,s}, \boldsymbol{\beta}_{|i,s}, \boldsymbol{\mu}_{|i,s}, \Sigma_{|i,s}$, for each conditioning variable $i = 1, \dots, d$ and site $s = 1, \dots, D$, using maximum likelihood estimation. In practice, we set the threshold $u_{i,s} := u_i$ to the q^{th} standard Laplace quantile for all s , as this is required to build a symmetric divergence measure for clustering (see Section 2.3). Choosing q involves the classic bias-variance tradeoff: q must be sufficiently low as to leave enough excesses above u_i for reliable estimation, but high enough to reduce bias induced when the asymptotic arguments in Equation (3) are not satisfied. To pick q , we use threshold stability plots and assess the stability of parameter estimates at different thresholds ([Southworth and Heffernan, 2012](#)). Using the bootstrap procedure of [Heffernan and Tawn \(2004\)](#), we obtain standard errors for parameter estimates at each site and choose the highest value of q such that the estimates are stable and the standard errors are reasonably small across sites.

2.3 A divergence measure for multivariate conditional extremes

There are several qualities we desire from a divergence measure based on the conditional distribution in Equation (6): (i) the measure should be non-negative and (ii) symmetric so that it can be interpreted as a distance (Vijendra and Laxman, 2015); (iii) the measure should have an upper bound, as large divergences can lead to numerical instability during clustering (Thiagarajan and Ghosh, 2025); (iv) it should be cheap to compute, which we provide through a closed-form solution for our divergence.

Nielsen (2019) introduced the skew-geometric Jensen-Shannon divergence ($\text{JS}^{\text{G}\lambda}$), which is a generalisation of the Jensen-Shannon (JS) divergence (Lin, 1991). Specifically, we use the dual form of the $\text{JS}^{\text{G}\lambda}$ divergence (Deasy et al., 2020), which is defined between two continuous $(d - 1)$ -variate distribution functions, say H and H^* , as

$$\text{JS}^{\text{G}\lambda}(H\|H^*) = (1 - \lambda) \text{KL}\{G_\lambda(H, H^*)\|H\} + \lambda \text{KL}\{G_\lambda(H, H^*)\|H^*\}, \quad (7)$$

where $G_\lambda(H, H^*) = H^{1-\lambda}(H^*)^\lambda$ is the weighted geometric mean of H and H^* , with $\lambda \in [0, 1]$, and $\text{KL}(H\|H^*)$ is the Kullback-Leibler (KL) divergence between H and H^* , given by

$$\text{KL}(H\|H^*) = \int_{\mathbb{R}^{d-1}} h(\mathbf{y}) \log\left(\frac{h(\mathbf{y})}{h^*(\mathbf{y})}\right) d\mathbf{y}, \quad (8)$$

where h and h^* are the continuous density functions of H and H^* , respectively. In our context, we replace H and H^* by the conditional multivariate Gaussian distributions induced by the CE framework at two different sites, say s and s^* ; see Equation (6).

The weight λ skews the geometric mean $G_\lambda(H, H^*)$ towards H as $\lambda \rightarrow 0$ (and towards H^* as $\lambda \rightarrow 1$). One possible advantage of specifying λ is that we can design our divergence measure to place more weight at sites s with more reliable CE model fits, as determined by, for example, the effective sample size. In this paper, we set $\lambda = 0.5$ throughout, as all sites have the same number of observations in both our simulations and case study.

For two multivariate Gaussian distributions, the KL divergence in Equation (8) has a closed-form solution (Duchi, 2007). The $\text{JS}^{\text{G}\lambda}$ divergence exploits the property that the (weighted) product of two distributions from the same exponential family is proportional to another member of that family, and hence, after normalisation, is itself an exponential-family distribution (Nielsen, 2019). It follows that $G_\lambda(H, H^*)$ is multivariate Gaussian, and so the KL divergences in Equation (7) are available in closed-form. Hence, the principle advantage of the $\text{JS}^{\text{G}\lambda}$ divergence is that it has a closed-form solution for multivariate Gaussian distributions. Its computational efficiency (Nielsen, 2019) has led to its use in several applications, including clustering, deep learning, variational inference, and remote sensing (Nielsen, 2019; Deasy et al., 2020; Ni et al., 2023; Thiagarajan and Ghosh, 2025). For two multivariate Gaussian distributions $H \sim \text{MVN}(\mathbf{m}, \Omega)$ and $H^* \sim \text{MVN}(\mathbf{m}^*, \Omega^*)$ in \mathbb{R}^d with mean vectors $\mathbf{m}, \mathbf{m}^* \in \mathbb{R}^d$ and covariance matrices $\Omega, \Omega^* \in \mathbb{R}^{d \times d}$, Nielsen (2019) showed that the $\text{JS}^{\text{G}\lambda}$ divergence introduced in Equation (7) has the closed-form solution

$$\begin{aligned} \text{JS}^{\text{G}\lambda}(H \| H^*) &= \frac{1}{2} \left\{ (1 - \lambda) \mathbf{m}^T \Omega^{-1} \mathbf{m} + \lambda (\mathbf{m}^*)^T (\Omega^*)^{-1} \mathbf{m}^* - \mathbf{m}_\lambda^T \Omega_\lambda^{-1} \mathbf{m}_\lambda + \log \left(\frac{|\Omega|^{1-\lambda} |\Omega^*|^\lambda}{|\Omega_\lambda|} \right) \right\}, \\ \mathbf{m}_\lambda &= \Omega_\lambda \{ (1 - \lambda) \Omega^{-1} \mathbf{m} + \lambda (\Omega^*)^{-1} \mathbf{m}^* \}, \text{ and} \\ \Omega_\lambda &= \{ (1 - \lambda) \Omega^{-1} + \lambda (\Omega^*)^{-1} \}^{-1}. \end{aligned} \tag{9}$$

Finally, the $\text{JS}^{\text{G}\lambda}$ is bounded above by $\text{JS}^{\text{G}\lambda}(H \| H^*) \leq \frac{1}{2} \max\{\lambda, 1 - \lambda\} \text{J}(H \| H^*)$, where the Jeffreys' divergence $\text{J}(H \| H^*) = \text{KL}(H \| H^*) + \text{KL}(H^* \| H)$ is simply the sum of the forward and backward KL divergences (Nielsen, 2025).

Now, suppose that for a specified conditioning variable $i \in \{1, \dots, d\}$, we have fitted the CE model at two different sites, s and s^* , and we want to compare their induced extremal dependence (i.e. the joint tails of $Y_{i,s}$ and Y_{i,s^*} ; see Equation (6)). Using the $\text{JS}^{\text{G}\lambda}$ divergence in Equation (9), we define the *conditional skew-geometric Jensen-Shannon divergence*

($\text{cJS}_{i,s,s^*}^{\text{G}\lambda}$) between two sites s and s^* , conditional on the i -th variable being extreme, as

$$\text{cJS}_{i,s,s^*}^{\text{G}\lambda}(y) = \text{JS}^{\text{G}\lambda}(H_{i,s}(y) \| H_{i,s^*}(y)), \quad y > u_i, \quad (10)$$

where $H_{i,s}(y)$ and $H_{i,s^*}(y)$ are the conditional $(d-1)$ -variate Gaussian distributions at sites s and s^* , respectively; see Equation (6). Equation (10) has a closed-form expression: from Equation (6), this can be derived by substituting, into Equation (9), the conditional parameters $\mathbf{m}_s(y) = \boldsymbol{\alpha}_{|i,s}y + y^{\beta_{|i,s}}\boldsymbol{\mu}_{|i,s}$, $\mathbf{m}_{s^*}(y) = \boldsymbol{\alpha}_{|i,s^*}y + y^{\beta_{|i,s^*}}\boldsymbol{\mu}_{|i,s^*}$, $\Omega_s(y) = (y^{\beta_{|i,s}})^{\text{T}}\Sigma_{|i,s}y^{\beta_{|i,s}}$, and $\Omega_{s^*}(y) = (y^{\beta_{|i,s^*}})^{\text{T}}\Sigma_{|i,s^*}y^{\beta_{|i,s^*}}$, and

$$\begin{aligned} \Omega_\lambda(y) &= \left[\lambda \left\{ (y^{\beta_{|i,s}})^{\text{T}}\Sigma_{|i,s}y^{\beta_{|i,s}} \right\}^{-1} + (1-\lambda) \left\{ (y^{\beta_{|i,s^*}})^{\text{T}}\Sigma_{|i,s^*}y^{\beta_{|i,s^*}} \right\}^{-1} \right]^{-1}, \\ \mathbf{m}_\lambda(y) &= \Omega_\lambda(y) \left[\lambda \left\{ (y^{\beta_{|i,s}})^{\text{T}}\Sigma_{|i,s}y^{\beta_{|i,s}} \right\}^{-1} \boldsymbol{\alpha}_{|i,s}y + y^{\beta_{|i,s}}\boldsymbol{\mu}_{|i,s} \right. \\ &\quad \left. + (1-\lambda) \left\{ (y^{\beta_{|i,s^*}})^{\text{T}}\Sigma_{|i,s^*}y^{\beta_{|i,s^*}} \right\}^{-1} \boldsymbol{\alpha}_{|i,s^*}y + y^{\beta_{|i,s^*}}\boldsymbol{\mu}_{|i,s^*} \right]. \end{aligned}$$

The $\text{cJS}_{i,s,s^*}^{\text{G}\lambda}(y)$ depends on $y > u_i$, the value of the corresponding conditioning variable. Therefore, to obtain a single measure of dissimilarity between the CE models at the two sites s and s^* for the i -th conditioning variable, we define the expected divergence

$$\begin{aligned} \text{eJS}_{i,s,s^*}^{\text{G}\lambda} &= \mathbb{E} \left(\text{cJS}_{i,s,s^*}^{\text{G}\lambda}(Y_{i,s}) \mid Y_{i,s} > u_i \right) \\ &= \mathbb{E} \left(\text{cJS}_{i,s^*,s}^{\text{G}\lambda}(Y_{i,s^*}) \mid Y_{i,s^*} > u_i \right) \\ &= \int_{u_i}^{\infty} \text{cJS}_{i,s,s^*}^{\text{G}\lambda}(y) h(y \mid Y_{i,s} > u_i) dy, \end{aligned} \quad (11)$$

where $h(y \mid Y_{i,s} > u_i)$ is the conditional density of $Y_{i,s} \mid (Y_{i,s} > u_i)$. We here exploit our standardised Laplace margins and note that $h(y \mid Y_{i,s} > u_i) = h(y \mid Y_{i,s^*} > u_i)$ for all s, s^* , ensuring symmetry of our dissimilarity measure.

We estimate the integral in Equation (11) using Monte Carlo methods. In practice, very large values of $Y_{i,s}$ can lead to unstable estimates, and we want to avoid extrapolating

beyond the range of the observations of $Y_{i,s}$ used to fit the CE model. Hence, we truncate the limits in integral (11), so that the upper limit is the 99th percentile of the empirical distribution of $Y_{i,s}$ pooled across all sites.

We evaluate the pairwise $\text{eJS}_{i,s,s^*}^{\text{G}\lambda}$ for all pairs of sites (s, s^*) and for each conditioning variable $i \in \{1, \dots, d\}$. These are then collected in a dissimilarity matrix:

$$M^{(i)} := \left(\text{eJS}_{i,s,s^*}^{\text{G}\lambda} \right)_{s,s^*=\{1,\dots,D\}}, \quad (12)$$

where D is the total number of sites, and $i = 1, \dots, d$.

If desired, we can aggregate the d dissimilarity matrices $M^{(1)}, \dots, M^{(d)}$ into a single dissimilarity matrix M by taking, for example, their element-wise average, i.e.

$$M = \frac{1}{d} \sum_{i=1}^d M^{(i)}. \quad (13)$$

This aggregated matrix M , or the individual matrices $M^{(i)}$, are then used as input to a clustering algorithm, which we describe in Section 2.4. Aggregating allows us to pool information across multiple comparisons and forms of information, and is supported in the extremal clustering literature (Mornet et al., 2017).

2.4 Clustering of multivariate extremes

We perform clustering on the dissimilarity matrices $M^{(i)}$, $i = 1, \dots, d$, or on the aggregated matrix M , defined in Equations (12) and (13), respectively. As in Bernard et al. (2013) and Vignotto et al. (2021), we use the Partitioning Around Medoids (PAM) algorithm, which is a variant of the popular k-medoids algorithm of Rousseeuw and Kaufman (1987). For each cluster $c = 1, \dots, k$, $\kappa_c \in \{1, \dots, D\}$ denotes its medoid index, and $\mathcal{C}_c \subseteq \{1, \dots, D\}$ the set of site indices assigned to cluster c ; the medoid site is the location with index κ_c . For the dissimilarity matrix $M^{(i)}$, a brief description of the algorithm is as follows:

Algorithm 1 Partitioning Around Medoids (PAM) algorithm using the $\text{eJS}^{\text{G}\lambda}$ divergence.

- 1: **Input:** dissimilarity matrix $M^{(i)} \in \mathbb{R}^{D \times D}$, number of clusters k
 - 2: **Output:** Cluster assignments $\mathcal{C}_1, \dots, \mathcal{C}_k$, medoid indices $\kappa = (\kappa_1, \dots, \kappa_k)$
 - 3: **Initialise:** choose k distinct medoid indices $\{\kappa_1, \dots, \kappa_k\} \subseteq \{1, \dots, D\}$ uniformly at random.
 - 4: **repeat**
 - 5: **Assignment:** Assign each vector index $s \in \{1, \dots, D\}$ to its closest medoid in terms of $\text{eJS}^{\text{G}\lambda}$ (see Equation (11)), i.e.
 assign s to \mathcal{C}_c , where $c = \arg \min_{\ell \in \{1, \dots, k\}} \text{eJS}_{i,s,\kappa_\ell}^{\text{G}\lambda}$.
 - 6: **for** $c = 1, \dots, k$ **do**
 - 7: **Update:** Update medoid index to minimise intra-cluster dissimilarity, i.e.
 $\kappa_c = \arg \min_{p \in \mathcal{C}_c} \sum_{q \in \mathcal{C}_c} \text{eJS}_{i,p,q}^{\text{G}\lambda}$.
 - 8: \triangleright Break ties by choosing the smallest index.
 - 9: **end for**
 - 10: **until** no observation changes cluster assignment
 - 11: **return** $\mathcal{C}_1, \dots, \mathcal{C}_k, \kappa$
-

The PAM algorithm takes a single hyperparameter, the number of clusters, k . We follow [Thorndike \(1953\)](#) and use the total within-group sum of squares (TWGSS) to select k . We find in our simulation study (Section 3) that this method correctly identifies k . Specifically, we define the TWGSS as

$$\text{TWGSS}(k) = \sum_{c=1}^k \sum_{s \in \mathcal{C}_c} \text{eJS}_{i,s,\kappa_c}^{\text{G}\lambda}, \quad (14)$$

where $\text{eJS}_{i,s,\kappa_c}^{\text{G}\lambda}$ is the dissimilarity in Equation (11) evaluated for site s and the medoid site κ_c of cluster c , $c = 1, \dots, k$. We plot $\text{TWGSS}(k)$ against k and select the optimal number of clusters as the value of k beyond which additional clusters yield only negligible reductions in TWGSS, corresponding to the ‘‘elbow’’ of the curve ([Hastie et al., 2009](#)).

3 Simulation study

We illustrate the performance of our multivariate extremal clustering approach across multiple simulation studies. Clustering accuracy is evaluated using the adjusted Rand index (ARI) of [Hubert and Arabie \(1985\)](#). Unreported experiments showed that the TWGSS method correctly identified k , and so we proceed with k fixed to the true number of clus-

ters. In Section 3.1, we generate data from a Gaussian copula for which the parameters of the CE model are known (Keef et al., 2013). This allows us to verify that our clustering approach reduces estimation uncertainty and to investigate the sensitivity of the results to the choice of conditioning threshold. Section 3.2 extends the study to mixtures of Gaussian and t -copulas. For the bivariate case, we compare our clustering approach to that of Vignotto et al. (2021), while we also show how our method can be applied in higher dimensions and for more complex dependence structures.

3.1 Gaussian copula setting

We consider a setting with $d = 2$ variables observed at $D = 12$ sites. Data are generated independently across sites from a Gaussian copula. For sites $s = 1, \dots, 12$, we sample 1000 observations of $(X_{1,s}, X_{2,s})$ from a bivariate Gaussian distribution with mean zero and covariance matrix whose off-diagonal elements $\rho_{\text{Gauss}}^s \in [0, 1]$ serve as site-specific correlation parameters. We then apply the transformation in Equation (2), with $F_X(\cdot)$ being the standard normal distribution function, to obtain $(Y_{1,s}, Y_{2,s})$ with standard Laplace margins.

The sites are assigned to one of three latent clusters of equal size, based on ρ_{Gauss}^s . As a proof of concept, we choose well-separated values to ensure clear clustering. Specifically, the first cluster, comprising sites 1 to 4, has correlation $\rho_{\text{Gauss}}^s = 0.1, s = 1, \dots, 4$; the second cluster has $\rho_{\text{Gauss}}^s = 0.5$, for $s = 5, \dots, 8$; and the third cluster has $\rho_{\text{Gauss}}^s = 0.9$, for $s = 9, \dots, 12$. Keef et al. (2013) showed that the corresponding parameters of the CE framework are $\alpha_{|i} = (\rho_{\text{Gauss}}^s)^2$ and $\beta_{|i} = 1/2, i = 1, 2$, implying asymptotic independence between $Y_{1,s}$ and $Y_{2,s}$.

At each site $s = 1, \dots, 12$, we estimate the CE model in Equation (4), with conditioning threshold u_i as the standard Laplace q -quantile for $q = 0.9$ and $q = 0.99$. Pairwise dissimilarities between each site are computed using the expected divergence in Equation (11), and

we average over both conditioning directions, $i = 1, 2$, to obtain a single dissimilarity measure (see M in Equation (13)). Clusters are identified using Algorithm 1. We repeat this procedure 500 times, with performance assessed by comparing finite-sample CE estimates to the asymptotic truths.

Figure 1 shows boxplots of the estimated model parameters stratified by the true values of ρ_{Gauss}^s . The results are shown pre- and post-clustering, with the latter derived by pooling data across all sites according to their estimated clustering (which was always correct in this setting). By pooling data across sites, we reduce estimation uncertainty for both $\alpha_{|i}$ and $\beta_{|i}$. Bias, as defined as the difference between the finite-sample estimate and the true theoretical parameter, and variance decrease with increasing ρ_{Gauss}^s , indicating faster convergence to the asymptotic model under stronger dependence. We observe the expected bias-variance tradeoff in the choice of quantile for the CE model, with the higher quantile ($q = 0.99$) yielding less bias, but with higher variance.

3.2 Mixture model setting

We extend our simulation design to mixtures of Gaussian and Student t -copulas (Demarta and McNeil, 2007), which induce asymptotic independence and asymptotic dependence, respectively, and whose dependence strength is determined by their respective correlation parameters ρ_{Gauss}^s and $\rho_t^s \in [0, 1]$.

To obtain a sample of size $n = 1000$ from the mixture copula, at each site $s = 1, \dots, 12$, we combine $n/2 = 500$ samples from the Gaussian copula described in Section 3.1 (with correlation parameter ρ_{Gauss}^s), and $n/2 = 500$ observations from a Student t -copula with 3 degrees of freedom and correlation ρ_t^s , before applying the Laplace transformation in Equation (2) to obtain the pairs $(Y_{1,s}, Y_{2,s})$ at site s .

In Section 3.2.1, we compare our method to the leading extremal competitor in the bivariate case (Vignotto et al., 2021) and, in Section 3.2.2, we show how our approach uniquely

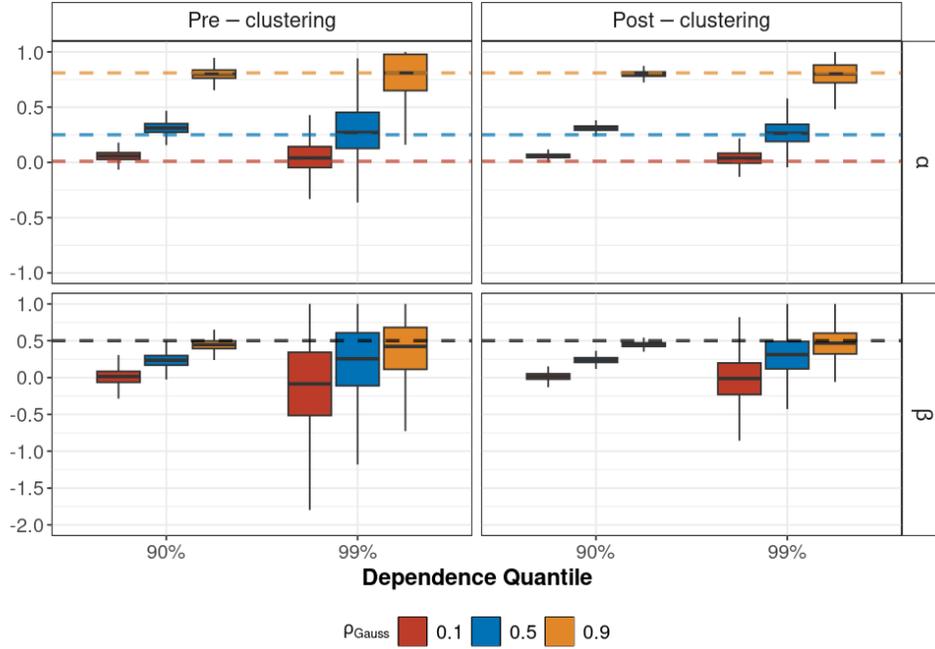


Figure 1: *Boxplots of estimates of $\alpha_{|i}$ (top) and $\beta_{|i}$ (bottom) for the conditional extremes model pre- and post-clustering with bivariate Gaussian data simulated at 12 sites, belonging equally to three known clusters. The x-axis gives q , the level of the conditioning quantile used to fit the model, and the boxes are coloured by the true Gaussian copula correlation parameter, ρ_{Gauss}^s . The horizontal lines indicate the true theoretical values of $\alpha_{|i}$ and $\beta_{|i}$, with the colours for $\alpha_{|i}$ corresponding with the true ρ_{Gauss}^s values.*

handles clustering of multivariate extremes in higher dimensions ($d > 2$). Finally, Section 3.2.3 considers a setting with $D = 60$ locations across three unequally sized clusters, and locations in the same cluster exhibiting perturbations in their correlation parameters. Performance is evaluated using the ARI, with 500 replicated experiments per setting.

3.2.1 Comparison to Vignotto et al. (2021)

When $d = 2$, we compare our extremal clustering method to that of Vignotto et al. (2021). We generate $n = 1000$ observations at $D = 12$ sites, using the mixture of Gaussian and t-copulas (with Laplace margins), as described above. Two clusters of sites $s = 1, \dots, 6$ and $s = 7, \dots, 12$ are defined with different values of ρ_t^s . We set $\rho_t^s = \rho_{t_1}$ for $s = 1, \dots, 6$ and $\rho_t^s = \rho_{t_2}$ for $s = 7, \dots, 12$, while ρ_{Gauss}^s is the same for both clusters. We consider a range of extremal dependence scenarios defined over a grid of values ($\rho_{Gauss}^s \in \{0.1, 0.2, \dots, 1.0\}, \rho_{t_1}^s \in \{0.1, 0.3, \dots, 0.9\}, \rho_{t_2}^s \in \{0, 0.2, \dots, 0.8\}$). We es-

timate the CE model using $q = 0.9$, and clusters were obtained using the aggregated dissimilarity matrix M in Equation (13).

The results are shown in Figure 2. Across the range of considered bivariate extremal dependence structures, our method consistently outperforms that of Vignotto et al. (2021). Both methods have high ARI when the cluster-wise difference in ρ_t^s is larger, since greater disparity naturally makes clusters more distinct and clustering easier. Both methods perform better when the Gaussian and t-copula correlations are higher. This may be because the Gaussian copula becomes more similar to the t-copula for higher values of ρ_{Gauss} , thus allowing the clustering approach to focus on differences in the t-copula.

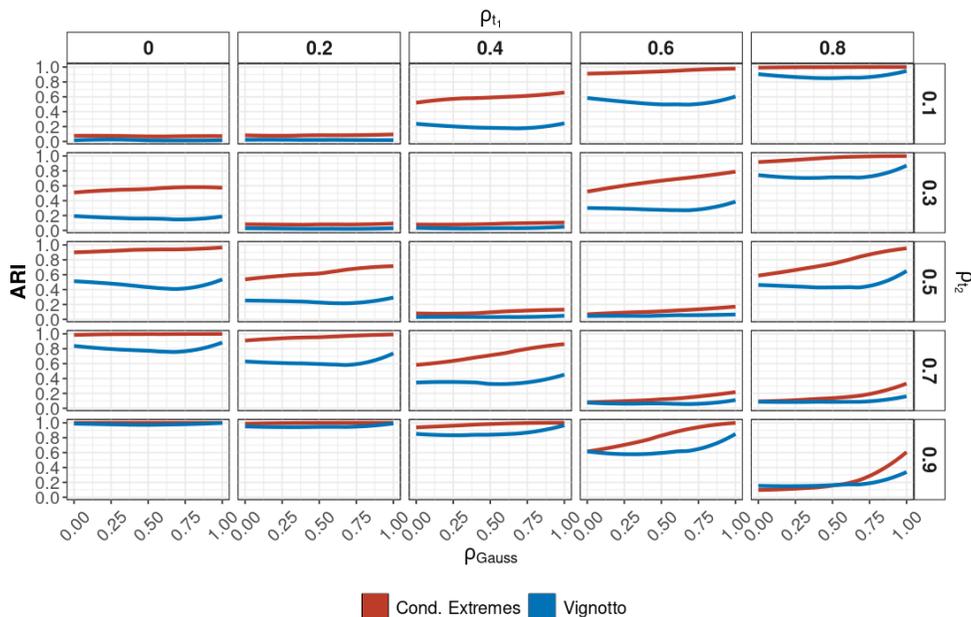


Figure 2: Comparison of our conditional extremes-based approach and the method of Vignotto et al. (2021). The x-axis represents the Gaussian correlation parameter ρ_{Gauss} , with facet labels indicating the t-copula correlation parameters, ρ_{t_1} and ρ_{t_2} , for the two true clusters. The smoothed lines show the average ARI across bootstrap samples, for both methods, with smoothing performed using a generalised additive model.

3.2.2 Extension to higher dimensions

While the method of Vignotto et al. (2021) is restricted to $d = 2$ dimensions, our approach can be extended to $d \geq 3$ variables. To illustrate this, we extend the simulation study of Section 3.2.1 by simulating at $D = 12$ locations, with two equally-sized clusters of sites

each with $n = 1000$ observations per site, but now for $d = 3$ variables, using the same grid of $(\rho_{\text{Gauss}}, \rho_{t_1}, \rho_{t_2})$ values as in Section 3.2.1. Our mixture copula now has three dimensions (with all pairwise correlations equal), and we fit three CE models at each site. Clustering is performed using the aggregated divergence matrix M from Equation (13). The results show that our proposed method performs well for $d = 3$, with the same patterns in ARI emerging as in the bivariate case in Section 3.2.1. In fact, performance is markedly stronger in three dimensions than in two. This is not surprising, as the CE fits in this setup are similar in both two and three dimensions, with the same ρ_t^s being used across all variable pairs. Therefore, the aggregated matrix M in Equation (13) is computed similarly, just averaged over more models with more data, which naturally improves clustering performance.

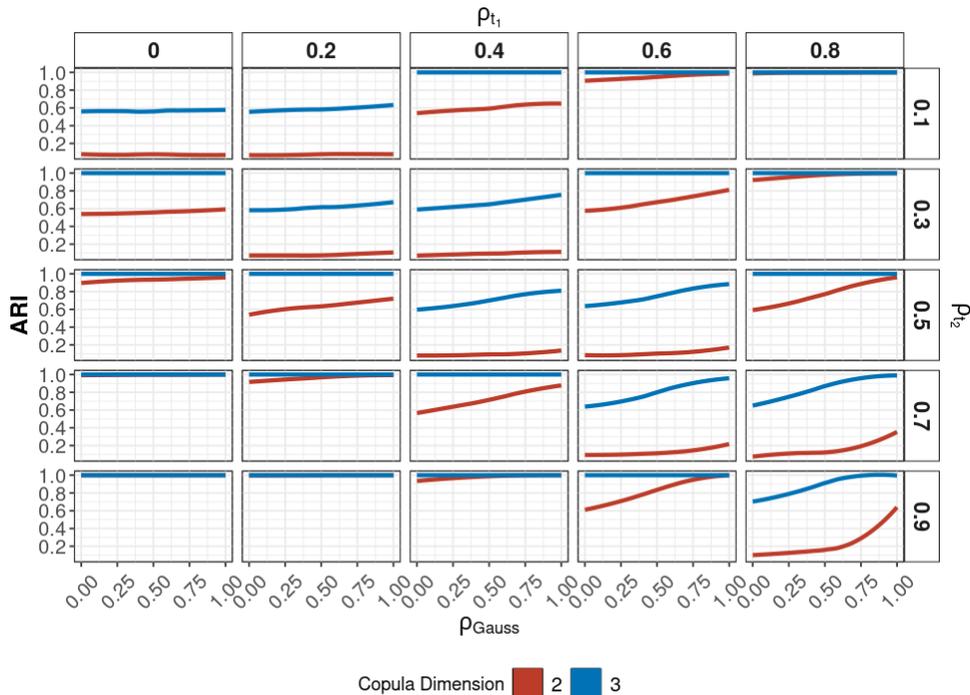


Figure 3: Comparison of clustering performance for $d = 2$ and $d = 3$ variables and two equally-sized clusters. The x-axis represents the Gaussian correlation parameter ρ_{Gauss} for both clusters, and the facet labels show the t-copula correlation parameters, ρ_{t_1} and ρ_{t_2} , for the two true clusters. The smoothed lines show the average ARI across bootstrap samples, with smoothing performed using a generalised additive model.

To further illustrate the performance of our method in higher dimensions, we consider a specific case from the grid where $\rho_{t_1} = 0.4$ and $\rho_{t_2} = 0.5$. This setting presents a challenging

clustering scenario in both two and three dimensions, as the cluster-wise difference in the t -copula correlation parameters is small. Figure 4 displays the results when the dimensionality of the copula mixture is increased from $d = 2$ to $d = 5$, based on 500 simulations across all considered ρ_{Gauss} values. Cluster accuracy improves steadily with dimension, reaching values close to 1 across all ρ_{Gauss} settings when $d = 5$. These gains should be interpreted cautiously, since as mentioned above we use the same ρ_t^s across all variable pairs, thus not altering the pairwise extremal dependence structure and instead increasing the amount of data as d increases. Having separate ρ_t^s values for each variable pair would increase the complexity and signal-to-noise ratio of the clustering task, and likely reduce performance. Nevertheless, the results demonstrate that the proposed method scales effectively to higher dimensions and performs well in multivariate settings.

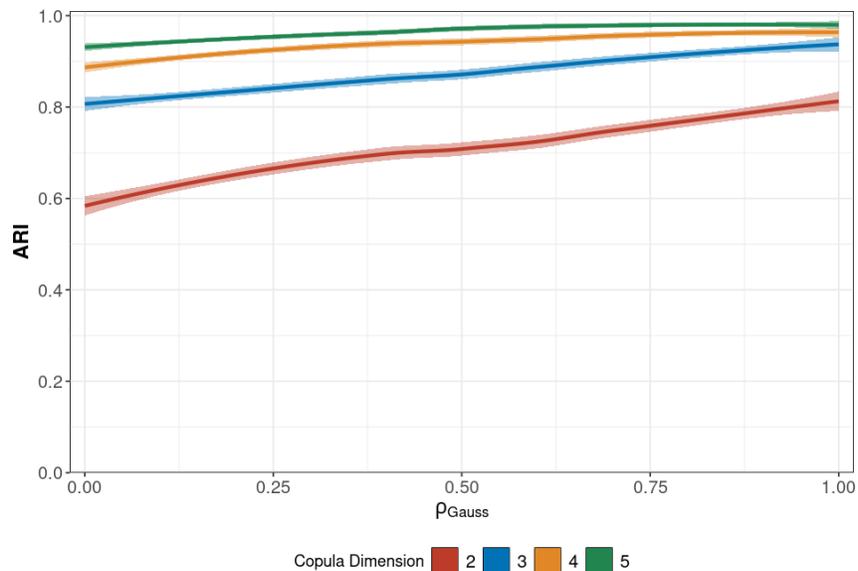


Figure 4: *Evaluation of clustering for $d = 2$ to $d = 5$ variables and two equally-sized clusters for simulations at 12 sites from a mixture of Gaussian and t -copulas. The t -copula correlation parameters for each true cluster were set to $\rho_{t_1} = 0.4$ and $\rho_{t_2} = 0.5$. A sequence of values are considered for the Gaussian correlation parameter ρ_{Gauss} , as represented on the x -axis. The smoothed lines show the average ARI and associated uncertainty across bootstrap samples, with smoothing performed using a generalised additive model.*

3.2.3 Extension to more sites

We now design another simulation study to better mimic the structure of the Irish meteorological dataset, to be analysed in Section 4. Data are generated at $D = 60$ sites, each with $n = 1000$ observations of $d = 2$ variables, using the mixture of Gaussian and t-copulas. The 60 sites are partitioned into three unequally-sized clusters of 10, 20, and 30 sites. At each location, the true ρ_t^s is perturbed by adding $\text{Unif}(-0.05, 0.05)$ noise, making the clustering task more challenging. We now have a separate t-copula correlation parameter for each of the three clusters, denoted by ρ_{t_1}, ρ_{t_2} , and ρ_{t_3} . To reduce the dimensionality of the grid of considered copula parameter values, a constant $\rho_{\text{Gauss}} = 0.5$ is used across all sites and experiments.

The results of 500 repetitions of this simulation study are shown in Figure 5. Similarly to the previous simulation studies, our clustering method performs well, particularly when the differences in t-copula correlation parameters between clusters are larger. For example, the best performance is observed where $\rho_{t_1} = 0.9, \rho_{t_2} = 0.1$, and $\rho_{t_3} = 0.5$, as the clusters are most distinct, while conversely, performance is worst for $\rho_{t_1} = 0.6, \rho_{t_2} = 0.4$, and $\rho_{t_3} = 0.5$.

4 Application to Irish meteorological data

We now apply our extremal clustering method to meteorological observations from the Republic of Ireland. In Section 4.1, we describe the precipitation and wind speed measurements used in our analysis and present exploratory analysis (using the χ statistic from Equation (1)) on the extremal dependence between the two variables. In Section 4.2, we discuss fits of the conditional extremes (CE) model, and, in Section 4.3, we apply our extremal clustering approach.

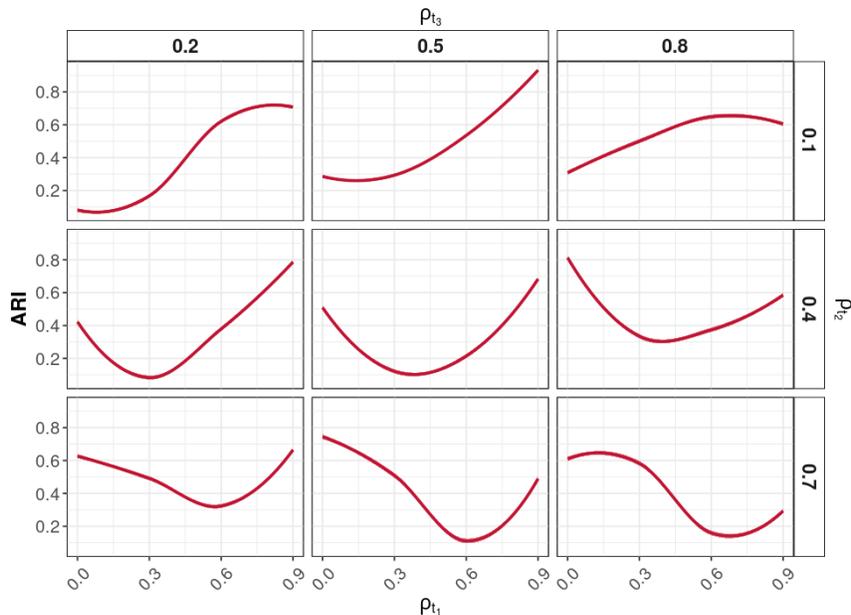


Figure 5: *Evaluation of clustering performance for $d = 2$ variables and three clusters of 60 locations for simulations from a mixture of Gaussian and t -copulas. The true x -axis and facet labels show the t -copula correlation parameters, ρ_{t_1} , ρ_{t_2} , and ρ_{t_3} , for each of the three true clusters, for a Gaussian copula correlation of $\rho_{Gauss} = 0.5$. The smoothed lines show the average ARI across bootstrap samples, with smoothing performed using a generalised additive model.*

4.1 Ireland meteorological data

We analyse the joint extremal behaviour of precipitation and wind speed at $D = 59$ sites across the Republic of Ireland from 1990 to 2020, inclusive. Precipitation and wind speed data are obtained from the Met Éireann archive (Met Éireann, 2025), and ERA5 Reanalysis (Copernicus Climate Change Service (C3S), 2024), respectively. Met Éireann data are irregularly spaced over space, while ERA5 provides spatial averages on a regular $0.25^\circ \times 0.25^\circ$ grid. We match each spatial site to the grid cell in which it lies. The dataset spans 11,323 consecutive daily observations.

We focus on the winter months, October to March, following Vignotto et al. (2021). Precipitation are aggregated into weekly sums, and daily wind speed maxima are aggregated into weekly averages. This weekly temporal scale reflects the typical lag between extreme precipitation and wind speed during storm events (Bengtsson et al., 2009), and reduces

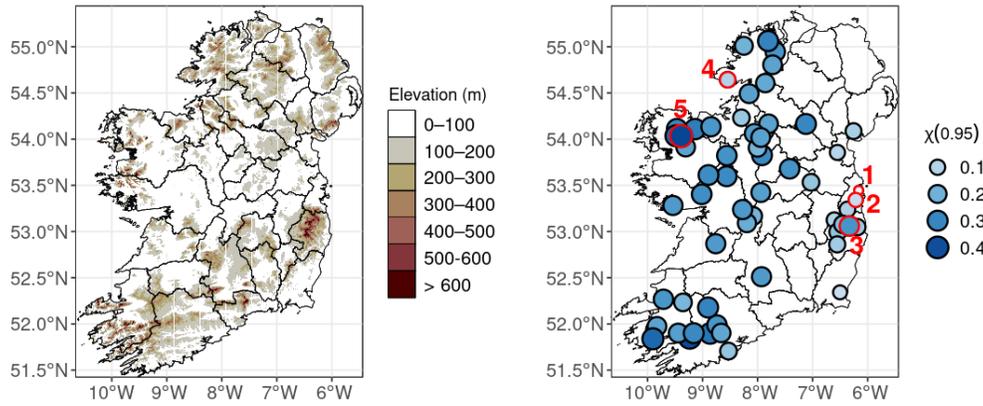


Figure 6: *Left: elevation (m) profile of Ireland. Right: Estimated $\chi(0.95)$ between precipitation and wind speed at all 59 locations. Darker colours and larger points indicate stronger asymptotic dependence. Locations in red are (1) Malahide Castle, Dublin, (2) Ringsend, Dublin, (3) Glenmacnass, Wicklow, (4) Kilcar, Donegal, and (5) Derryhillagh, Mayo.*

short-term temporal dependence not accounted for in our model. Weeks with no precipitation are excluded, resulting in 5,022 observations across all $D = 59$ sites.

We explore the extremal dependence structure of the data using the χ statistic introduced in Equation (1). Figure 6 shows empirical estimates of $\chi(0.95)$ at all 59 sites. An elevation profile of Ireland is also shown in Figure 6, which may help further explain some of this spatial structure. Non-zero values are observed at almost all sites, indicating positive extremal dependence between precipitation and wind speed. In practical terms, this suggests that extreme wind speeds tend to be accompanied by high precipitation, and vice versa. A clear spatial pattern emerges: extremal dependence is strongest in the south west, with the highest values along the Atlantic coast, notably in counties Kerry, Galway, Mayo, and Donegal, and weaker along the east coast, in Dublin, Wexford, and Louth. The highest $\chi(0.95)$ value (0.414) is observed at Derryhillagh in Mayo (Figure 6, location (5)), while the lowest (0.0005) is at Malahide Castle in Dublin (Figure 6, location (1); hereafter ‘Malahide’).

Some spatial outliers are also apparent. Kilcar in Donegal (Figure 6, location (4)) has a relatively low $\chi(0.95)$ value (0.109), despite its west-coast location, while some sites in county Wicklow show unusually high values for the east, particularly Glenmacnass (Figure 6, location (3); 0.268). The latter may be related to the influence of the Wicklow

Mountains, as reflected in the elevation profile (Figure 6). It may be of interest to investigate if these outliers in $\chi(0.95)$ are reflected in the CE model parameter estimates, and whether clustering can identify them. These results indicate distinct regional patterns of extremal dependence, as well as notable outliers, motivating the use of our extremal clustering framework to characterise and interpret this spatial structure more formally.

4.2 Conditional extremes estimates

We transform the precipitation and wind speed data, separately at each location, to Laplace margins, with the empirical rank transform replacing F_{X_i} in Equation (2). To select an appropriate quantile level q for fitting the CE model (see Section 2.1), we estimate the parameters $(\alpha_{|i}, \beta_{|i})$ at a sequence of levels q , to identify the lowest q beyond which parameter estimates appear stable. Estimates are generally stable for both models, i.e., precipitation conditional on wind speed and vice versa, down to $q = 0.85$, with similar maximum likelihood estimates across 500 bootstrap samples for $(\alpha_{|i}, \beta_{|i})$. Threshold stability plots for the CE fits at the five locations highlighted in Figure 6 are included in Figures S3 – S7 of the supplementary materials. Therefore, we fix $q = 0.85$ for the subsequent analysis. In Figure 10, we show that our estimated extremal clustering is robust to the choice of q , with estimates shown for $q = 0.85, 0.88$, and 0.90 .

The estimated site-wise CE parameters are shown in Figure 7. Although interpretation is challenging due to variability in the estimates, as with the $\chi(0.95)$ estimates (Figure 6), we observe some evidence of increasing $\alpha_{|i}$ values for both variables towards the west and south. To mitigate uncertainty and aid interpretation, we cluster the locations using our extremal clustering approach.

4.3 Clustering

We cluster the locations using the aggregated dissimilarity matrix M as defined in Equation (13), and separately for each individual matrix $M^{(i)}, i = 1, 2$, as in Equation (12). The

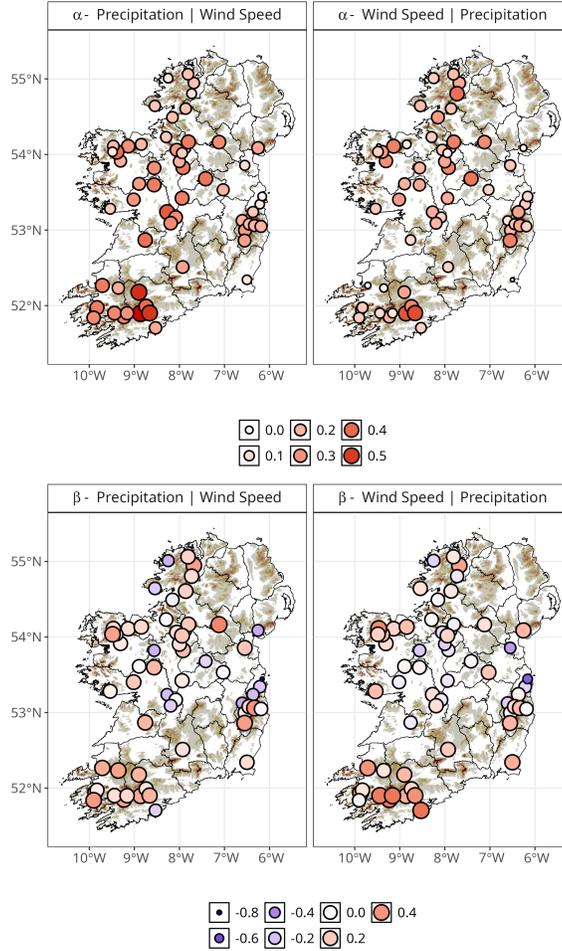


Figure 7: Maps of site-wise estimated model parameters $\alpha_{i,s}$ (top) and $\beta_{i,s}$ (bottom) for precipitation conditional on wind speed (left) and wind speed conditional on precipitation (right). The colour scale is the same for both variables, with darker red and larger points indicating larger positive values of α and β , and darker blue and smaller points indicating smaller negative values.

number of clusters k is chosen via the method in Section 2.4, using the elbow plot of the TWGSS. For all three clustering scenarios, clear elbows are observed at $k = 3$. The elbow plots are provided in Figure S1 of the supplementary materials.

The resulting cluster estimates are displayed in Figure 8. Across all three choices of dissimilarity matrix, the estimated clusters are broadly consistent, with three spatially distinct groups emerging. This spatial coherence is notable given that no explicit spatial information was used in model fitting or in the clustering, suggesting that the CE model parameters capture meaningful spatial structure in the extremal dependence of the data.

When interpreting the clusters, we observe a clear contrast between the east and west coasts, separated by a central cluster. For precipitation conditional on wind speed, the clustering is somewhat less distinct, with the central cluster extending more extensively into parts of Galway, Mayo, and Donegal on the west coast. There are also notable outliers in the clusters, which we further discuss below.

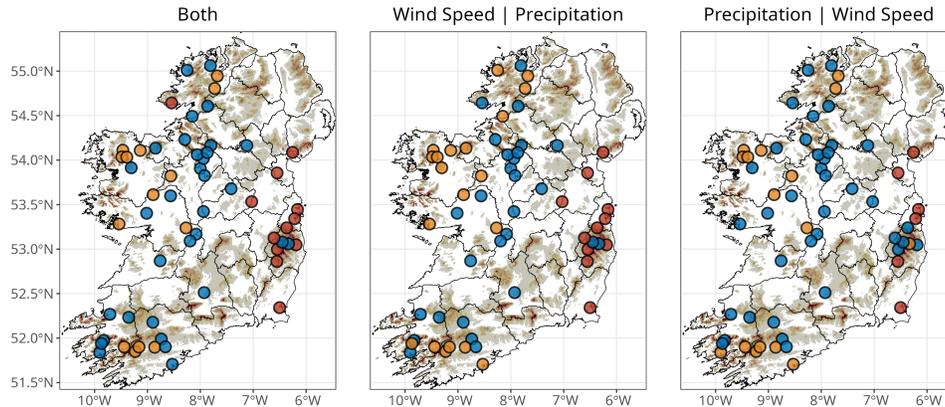


Figure 8: *Extremal cluster estimates using the dissimilarity matrix for wind speed conditional on precipitation (centre), precipitation conditional on wind speed (right), and the combined dissimilarity matrix (left). Sites are coloured by cluster label.*

In Figure 9, we visualise the dissimilarity matrix for the aggregated clusters (see left panel of Figure 8). The clusters are quite distinct, with the dissimilarity between clusters being much larger than within-cluster dissimilarity, which rarely exceeds 0.01. In particular, the difference between the eastern and western clusters is quite pronounced, with the intermediate central cluster acting as somewhat of a ‘buffer’ between the two. In contrast, the western and central clusters seem to be quite close, in that we observe many small inter-cluster pairwise dissimilarities between them. This indicates that allocation between the two is not always clear-cut and for several sites is nearly interchangeable, as observed in the variable-specific estimated clusters in Figure 8 and in the quantile sensitivity analysis in Figure 10, where membership between the two often switches. Also, if we only fit two clusters, we find that most of the central cluster members are assigned to the western cluster, reflecting this less obvious separation between the two compared to the more markedly different eastern cluster. A noticeable outlier in the dissimilarity matrix is Malahide, which

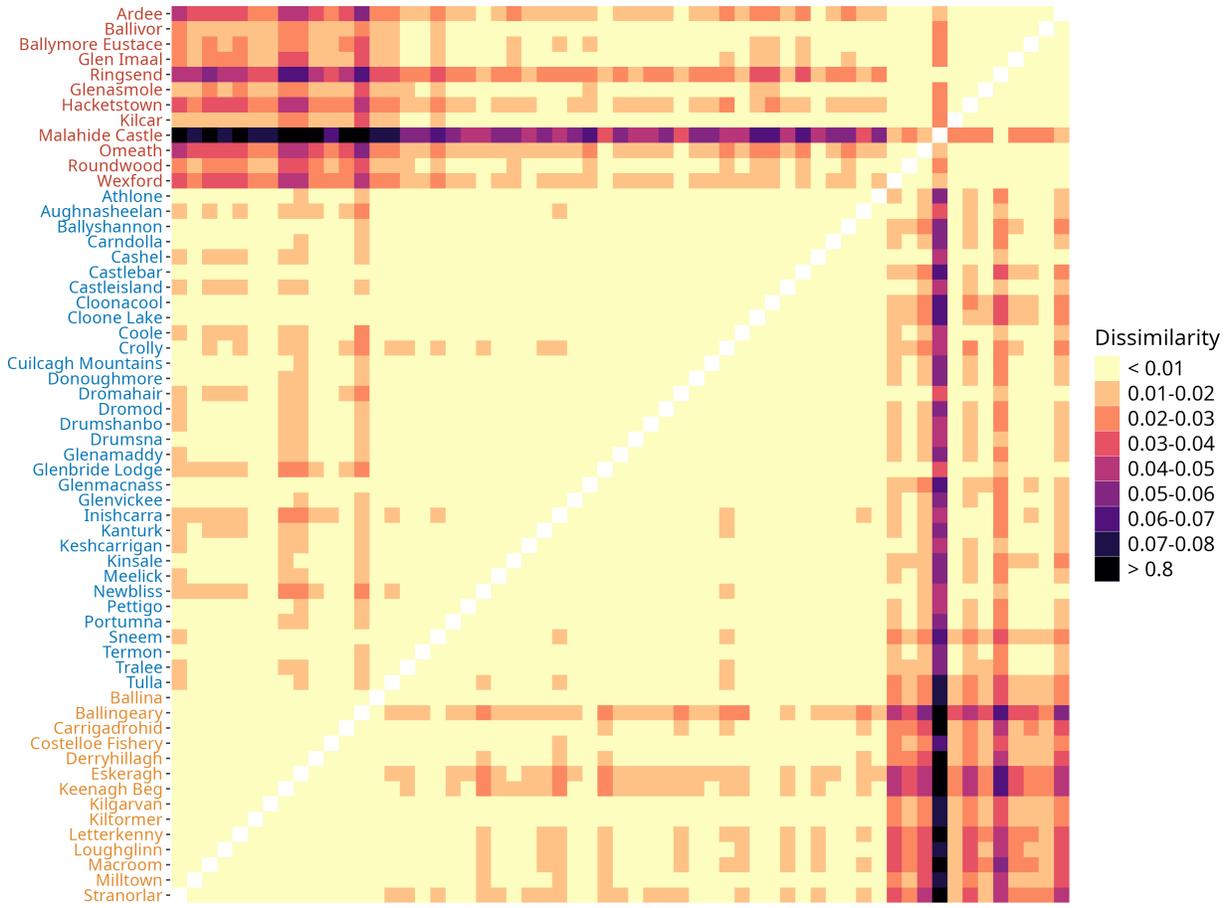


Figure 9: Heatmap of the estimated aggregated dissimilarity matrix M . The names of each site are coloured and ordered according to their estimated cluster labels, matching that of Figure 8. The dissimilarity colour scale is shown on the right, with darker colours indicating larger values.

we previously identified as having the lowest $\chi(0.95)$ value in Figure 6. Even within its own cluster, Malahide shows a relatively high dissimilarity to the other sites. Indeed, if we set $k = 4$, Malahide becomes its own cluster (clustering results for $k = 2$ and $k = 4$ are provided in Figure S2 of the supplementary materials). Malahide also has the lowest mean weekly precipitation of any site; the second lowest is Ringsend, also in county Dublin (Figure 6 location (2)), which is the site closest to Malahide in terms of our dissimilarity measure. Finally, Malahide has very low $\chi(0.95)$ (Figure 6) and α_i (Figure 7) estimates. Together, this may explain why we observe such outlying behaviour for Malahide. Finally, Kilcar in County Donegal is the member of the eastern cluster with the smallest dissimilarity to the

western cluster members. This may reflect its geographic location in northwest Ireland, despite its low $\chi(0.95)$ value (Figure 6) relative to other west-coast sites.

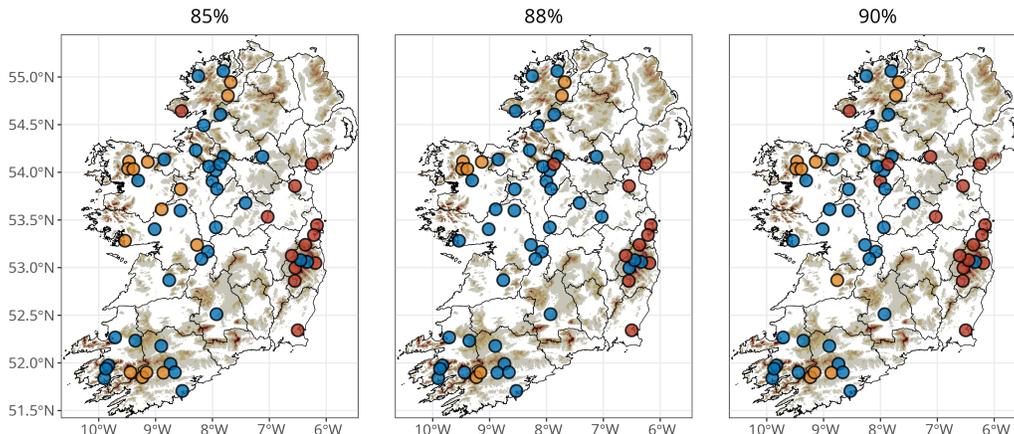


Figure 10: *Estimated extremal clusters with the conditional extremes model parameters estimated at $q = 0.85, q = 0.88$ and $q = 0.9$ quantiles, using the aggregated dissimilarity matrix. Sites are coloured by cluster label.*

We perform a sensitivity analysis of the estimated clusters (with the aggregated dissimilarity matrix) to the choice of quantile level $q \in \{0.85, 0.88, 0.90\}$, in Figure 10. The estimated clusters are similar across the three values of q , with some small differences. Noticeably, the earlier identified outlier at Kilcar in county Donegal is assigned to the eastern cluster when $q = 0.85$ and $q = 0.90$, but to the central cluster when $q = 0.88$. Also, there are more sites in the western cluster when $q = 0.85$, with some sites moving to the central cluster for $q = 0.88$. Interestingly, some of these sites then move back to the western cluster for $q = 0.90$. Apart from these observations, the estimated clusters are generally quite similar across the three quantile levels, and so we conclude that our clustering method is relatively robust to the choice of quantile level.

5 Discussion

In this paper, we introduced a novel method for clustering multivariate extremes. We proposed the expected conditional JS^{G_λ} divergence (Nielsen, 2019) as a measure of dissimilarity between multivariate conditional extremes (CE) models (Heffernan and Tawn, 2004).

We leveraged the commonly-applied working Gaussian assumption to derive a closed-form expression for the JS^{G_λ} divergence. Efficient k-medoids clustering was then performed on the dissimilarity matrix derived from this measure. In our simulation study, a bivariate Gaussian copula example illustrated that our clustering method reduces parameter uncertainty and bias when the CE model is re-fitted with data pooled post-clustering. Using mixtures of Gaussian and t -copulas, we showed that our method outperforms the method of [Vignotto et al. \(2021\)](#) across a range of bivariate extremal dependence scenarios. Our method was also shown to be uniquely extendable to $d > 2$ dimensions. We applied our extremal clustering method to Irish meteorological data, consisting of 59 locations with weekly rainfall and wind speed observations. Without explicitly incorporating spatial information, three clusters with distinct spatial patterns were identified, with well separated clusters representing the east and west coasts and the centre of Ireland.

While we focused on a spatial application in this work, our method is more generally applicable to any setting where clustering of multivariate extremal dependence is desired. For instance, it could be applied to clinical trial or public health data to cluster patients based on the tail dependence structure of their symptoms or biomarker measurements. A neuroscience application was explored in [Talento et al. \(2025\)](#), where the CE framework was used to highlight differences in the tail dependence of electroencephalogram (EEG) signals for seizure-prone and seizure-absent subjects. However, they used pre-defined clusters. Hence, we may expect our approach to be useful in similar applications, to identify subjects who have undergone extreme brain events, e.g., seizures. In financial applications, we could cluster time periods according to the extremal dependence of log-returns of different assets, offering potential insights for portfolio management and risk assessment.

While the residual Gaussian assumption on which our method relies may not hold for all datasets, our clustering approach showed good performance for all settings in Section 3. A possible alternative is to employ the generalised and asymmetric multivariate Gaussian dis-

tribution used in [Farrell et al. \(2025\)](#). While this does not permit a closed-form divergence, our extremal clustering method could still be used, with costly numerical techniques used to estimate the dissimilarity matrix. Indeed, if choosing to forgo this Gaussian assumption, it may be preferable to use the standard Jensen-Shannon divergence in place of the $\text{JS}^{\text{G}\lambda}$ divergence, since the former is bounded ([Thiagarajan and Ghosh, 2025](#)).

Another important limitation of our method is in its treatment of uncertainty. We do not propagate uncertainty in the CE model estimates through to the clustering. To account for this, one possible approach is to use the bootstrap method of [Heffernan and Tawn \(2004\)](#) to generate samples of CE model estimates, and compute our dissimilarity matrix for each sample. We could then produce a distribution of dissimilarity matrices, and use these to compute a distribution of estimated clusters. However, this would be computationally expensive, and so would compromise one of the main advantages of our method, namely its computational efficiency.

In this work, we implicitly assume spatial stationarity in the extremal dependence structure, as we fit a separate CE model at each location. Extensions of the CE model have been proposed, such as those incorporating covariates ([Richards et al., 2023](#)) and spatio- and spatio-temporal representations of the CE parameters ([Simpson and Wadsworth, 2021](#); [Wadsworth and Tawn, 2022](#)). These could be incorporated into our clustering framework, but care must be taken to ensure that the residual Gaussian assumption still holds, or that an appropriate alternative dissimilarity is used, as discussed above. Another approach to dealing with the uncertainty in the clustering solution would be to adopt a Bayesian perspective. By placing priors on the CE model parameters, we could obtain posterior distributions for these parameters, which could be used to compute a distribution of dissimilarity matrices and clustering solutions. Approaches such as that of [Rohrbeck and Tawn \(2021\)](#) have used a Reversible-Jump MCMC scheme to perform Bayesian cluster-

ing for spatial extremes, which does not require the number of clusters to be specified in advance.

Acknowledgements

Patrick O’Toole is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/S022945/1. We acknowledge the use of the ERA5 reanalysis dataset, available from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (<https://cds.climate.copernicus.eu/>). We also acknowledge the use of the Irish meteorological dataset obtained from Met Éireann (<https://www.met.ie/climate/available-data/historical-data>).

SUPPLEMENTARY MATERIAL

PDF Supplement: This supplement contains additional plots for the motivating data application in Section 4.3 of the main paper, including elbow plots for choosing the number of clusters, clustering solutions for $k = 2$ and $k = 4$ clusters, and threshold stability plots for the conditional extremes model fits for the five locations highlighted in Figure 6.

Data and R code: The proposed methods are made available in the R package `CeCl` at <https://github.com/potoole7/CeCl>. The analysis code used in the paper and the data are available upon request from the corresponding author.

References

Alonso, A. M., de Zea Bermudez, P., and Scotto, M. G. (2014). Comparing generalized Pareto models fitted to extreme observations: an application to the largest temperatures in Spain. *Stochastic Environmental Research and Risk Assessment*, 28(5):1221–1233.

- Bengtsson, L., Hodges, K. I., and Keenlyside, N. (2009). Will extratropical storms intensify in a warmer climate? *Journal of Climate*, 22(9):2276–2301.
- Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20):7929–7937.
- Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A. F., Ramos, A. M., Vignotto, E., Bastos, A., Blesić, S., Durante, F., et al. (2021). Guidelines for studying diverse types of compound weather and climate events. *Earth’s Future*, 9(11):e2021EF002340.
- Boulin, A., Di Bernardino, E., Laloë, T., and Toulemonde, G. (2025). Identifying regions of concomitant compound precipitation and wind speed extremes over Europe. *Journal of the Royal Statistical Society Series C: Applied Statistics*. qlaf014.
- Carreau, J., Naveau, P., and Neppel, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research*, 53(5):4407–4426.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer.
- Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.
- Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for Spatial Max–Stable Random Fields. In *Dependence in Probability and Statistics*, pages 373–390. Springer, New York.
- Copernicus Climate Change Service (C3S) (2024). ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available at <https://cds.climate.copernicus.eu/>, last access: 31 January 2025.
- Davison, A. C. and Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2(1):203–235.

- Dawkins, L. C. and Stephenson, D. B. (2018). Quantification of extremal dependence in spatial natural hazard footprints: independence of windstorm gust speeds and its impact on aggregate losses. *Natural Hazards and Earth System Sciences*, 18(11):2933–2949.
- de Carvalho, M., Huser, R., and Rubio, R. (2023). Similarity-based clustering for patterns of extreme values. *Stat*, 12(1):e560.
- Deasy, J., Simidjievski, N., and Liò, P. (2020). Constraining Variational Inference with Geometric Jensen–Shannon Divergence. *Advances in Neural Information Processing Systems*, 33:10647–10658.
- Demarta, S. and McNeil, A. J. (2007). The t Copula and Related Copulas. *International Statistical Review*, 73(1):111–129.
- Duchi, J. (2007). Derivations for Linear Algebra and Optimization. http://ai.stanford.edu/~jduchi/projects/general_notes.pdf. Accessed: 2025-01-31.
- Dupuis, D. J., Engelke, S., and Trapin, L. (2023). Modeling panels of extremes. *The Annals of Applied Statistics*, 17(1):498–517.
- Farrell, A., Eastoe, E. F., and Lee, C. (2025). Conditional Extremes with Graphical Models. arXiv preprint arXiv:2411.17013.
- French, J., Kokoszka, P., Stoev, S., and Hall, L. (2019). Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data. *Computational Statistics & Data Analysis*, 131:176–193.
- Gilleland, E., Brown, B. G., and Ammann, C. M. (2013). Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics*, 24(6):418–432.
- Guerrero, M. B., Huser, R., and Ombao, H. (2023). Conex–connect: Learning patterns in extremal brain connectivity from multichannel eeg data. *The Annals of Applied Statistics*, 17(1):178–198.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York.

- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3):497–546.
- Hilal, S., Poon, S.-H., and Tawn, J. (2014). Portfolio risk assessment using multivariate extreme value methods. *Extremes*, 17(4):531–556.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.
- Huser, R., Opitz, T., and Wadsworth, J. (2024). Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes.
- Jane, R., Cadavid, L., Obeysekera, J., and Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south Florida. *Natural Hazards and Earth System Sciences*, 20(10):2681–2699.
- Kakampakou, L., Simpson, E., and Wadsworth, J. (2024). Spatial Extremal Modelling: A Case Study on the Interplay Between Margins and Dependence. *Stat*, 13(4).
- Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.
- Koh, J. and Opitz, T. (2025). Extreme-value modelling of migratory bird arrival dates: Insights from citizen science data (with discussion). *Journal of the Royal Statistical Society Series A: Statistics in Society*, 188(3):674–699.
- Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, 17(1):560–582.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

- Lin, J. (1991). Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory*, 37(1):145–151.
- Met Éireann (2025). Historical Data - Met Éireann - The Irish Meteorological Service — met.ie. <https://www.met.ie/climate/available-data/historical-data>. [Accessed 2025-08-01].
- Miralles, O. and Davison, A. C. (2024). Bayesian modeling of insurance claims for hail damage. *The Annals of Applied Statistics*, 18(4):3091–3108.
- Mornet, A., Opitz, T., Luzi, M., Loisel, S., and Bailleul, B. (2017). Wind storm risk management: sensitivity of return period calculations and spread on the territory. *Stochastic Environmental Research and Risk Assessment*, 31(8):1977–1995.
- Ni, S., Lin, C., Wang, H., Li, Y., Liao, Y., and Li, N. (2023). Learning geometric Jensen-Shannon divergence for tiny object detection in remote sensing images. *Frontiers in Neurorobotics*, 17:1273251.
- Nielsen, F. (2019). On the Jensen-Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy*, 21(5):485.
- Nielsen, F. (2025). Two tales for a geometric Jensen-Shannon divergence. arXiv preprint arXiv:2508.05066.
- Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. *Spatial Statistics*, 16:1–18.
- Osei, M. A., Amekudzi, L. K., Omari-Sasu, A. Y., Yamba, E. I., Quansah, E., Aryee, J. N., and Preko, K. (2021). Estimation of the return periods of maxima rainfall and floods at the Pra River Catchment, Ghana, West Africa using the Gumbel extreme value theory. *Heliyon*, 7(5).
- Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, 17(2):581–610.

- Richards, J., Alotaibi, N., Cisneros, D., Gong, Y., Guerrero, M. B., Redondo, P., and Shao, X. (2023). Modern extreme value statistics for Utopian extremes. arXiv preprint arXiv:2311.11054.
- Richards, J., Tawn, J. A., and Brown, S. (2022). Modelling extremes of spatial aggregates of precipitation using conditional methods. *The Annals of Applied Statistics*, 16(4):2693–2713.
- Rohrbeck, C., Eastoe, E. F., Frigessi, A., and Tawn, J. A. (2018). Extreme value modelling of water-related insurance claims. *The Annals of Applied Statistics*, 12(1):246–282.
- Rohrbeck, C. and Tawn, J. A. (2021). Bayesian spatial clustering of extremal behavior for hydrological variables. *Journal of Computational and Graphical Statistics*, 30(1):91–105.
- Rousseeuw, L. and Kaufman, P. (1987). Clustering by means of medoids. In *Statistical Data Analysis Based on the L1 Norm and Related Methods*, pages 405–416. Neuchâtel, Switzerland.
- Sando, K., Wada, R., Rohmer, J., Lecacheux, S., and Jonathan, P. (2022). Estimating Joint Extremes of Significant Wave Height and Wind Speed for Tropical Cyclones. In *Volume 5B: Ocean Engineering; Honoring Symposium for Professor Günther F. Clauss on Hydrodynamics and Ocean Engineering*, OMAE2022. American Society of Mechanical Engineers.
- Shao, X., Hazra, A., Richards, J., and Huser, R. (2025). Flexible modeling of nonstationary extremal dependence using spatially fused LASSO and ridge penalties. *Technometrics*, 67(1):97–111.
- Simpson, E. S., Opitz, T., and Wadsworth, J. L. (2023). High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. *Extremes*, 26(4):669–713.
- Simpson, E. S. and Wadsworth, J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, 41:100482.
- Soukissian, T. H. and Tsalis, C. (2015). The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction. *Natural Hazards*, 78:1777–1809.

- Southworth, H. and Heffernan, J. E. (2012). Multivariate extreme value modelling of laboratory safety data from clinical studies. *Pharmaceutical Statistics*, 11(5):367–372.
- Steinkohl, C., Davis, R. A., and Klüppelberg, C. (2013). Extreme value analysis of multivariate high-frequency wind speed data. *Journal of Statistical Theory and Practice*, 7:73–94.
- Talento, M. S. D., Richards, J., Pinto-Orellana, M., Huser, R., and Ombao, H. C. (2025). Spectral Extremal Connectivity of Two-State Seizure Brain Waves. *arXiv preprint arXiv:2503.04169*.
- Tanarhte, M., Hadjinicolaou, P., and Lelieveld, J. (2015). Heat wave characteristics in the eastern Mediterranean and Middle East using extreme value theory. *Climate Research*, 63(2):99–113.
- Tendijck, S., Eastoe, E., Tawn, J., Randell, D., and Jonathan, P. (2023). Modeling the Extremes of Bivariate Mixture Distributions With Application to Oceanographic Data. *Journal of the American Statistical Association*, 118(542):1373–1384.
- Thiagarajan, P. and Ghosh, S. (2025). Jensen-Shannon divergence based novel loss functions for Bayesian neural networks. *Neurocomputing*, 618(129115):129115.
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.
- Vettori, S., Huser, R., and Genton, M. G. (2019). Bayesian modeling of air pollution extremes using nested multivariate max-stable processes. *Biometrics*, 75(3):831–841.
- Vignotto, E., Engelke, S., and Zscheischler, J. (2021). Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland. *Weather and Climate Extremes*, 32:100318.
- Vijendra, S. and Laxman, S. (2015). Symmetry based automatic evolution of clusters: A new approach to data clustering. *Computational Intelligence and Neuroscience*, 2015:796276.
- Wadsworth, J. and Tawn, J. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.

- Winter, H. C., Tawn, J. A., and Brown, S. J. (2016). Modelling the effect of the El Niño-Southern Oscillation on extreme spatial temperature events over Australia. *The Annals of Applied Statistics*, 10(4):2075 – 2101.
- Zhang, L., Risser, M. D., Wehner, M. F., and O’Brien, T. A. (2024). Leveraging Extremal Dependence to Better Characterize the 2021 Pacific Northwest Heatwave. *Journal of Agricultural, Biological and Environmental Statistics*.
- Zheng, S., Fan, K., Hou, Y., Feng, J., and Fu, Y. (2023). Clustering by the probability distributions from extreme value theory. *IEEE Transactions on Artificial Intelligence*, 4(2):292–303.
- Zscheischler, J., Naveau, P., Martius, O., Engelke, S., and Raible, C. C. (2021). Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics*, 12(1):1–16.
- Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477.