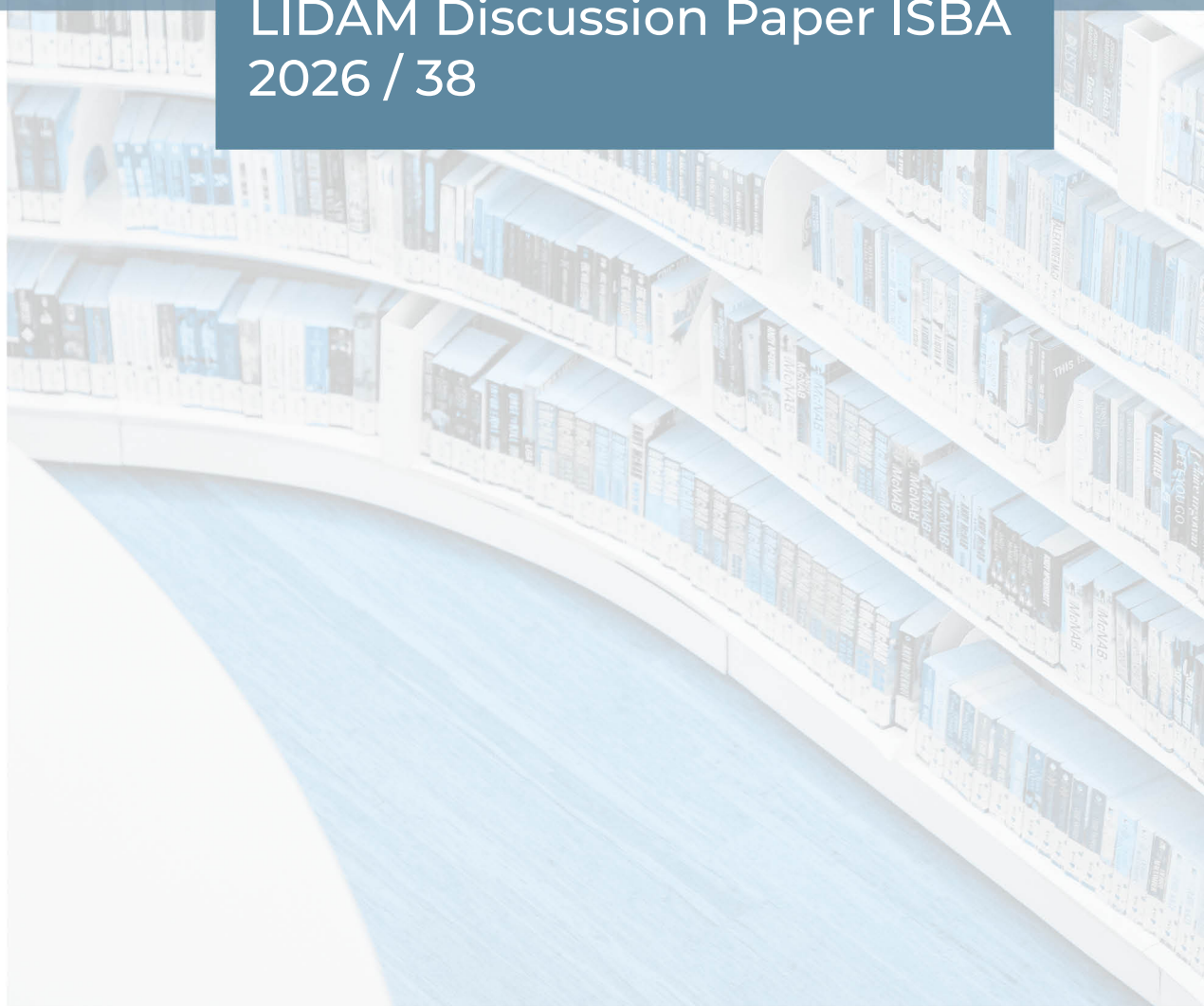
CLUSTERING EXTREMAL DEPENDENCE STRUCTURES VIA WASSERSTEIN DISTANCES BETWEEN ANGULAR MEASURES

Kamal Gasser, Johan Segers, Francesco Ragone

LIDAM Discussion Paper ISBA
2026 / 38



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

arXiv:2610.21931v2 [stat.ME] 2 Nov 2026

Clustering extremal dependence structures via Wasserstein distances between angular measures

K. Gasser[1*], J. Segers[2,1] and F. Ragone[3]

[1]Institute of Statistics, Biostatistics and Actuarial Sciences, LIDAM, UCLouvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.
[2]Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium.
[3]School of Computing and Mathematical Sciences, University of Leicester, University Road, Leicester LE1 7RH, U.K.

*Corresponding author(s). E-mail(s): kamal.gasser@uclouvain.be;
Contributing authors: jjjsegers@kuleuven.be; fr120@le.ac.uk;

Abstract

Grouping random vectors by their joint extremal behaviour is a recurring task in climate science, hydrology, and finance. Classical clustering is driven by typical observations, while existing extremal procedures either summarize dependence by a single coefficient or discretize the simplex in an ad hoc way. We propose a dissimilarity that compares complete angular measures by optimal transport: each angular measure is estimated above a high radial threshold, mapped from the simplex into a Euclidean space of Aitchison coordinates, and compared by the 1-Wasserstein distance. The methodology is fully nonparametric and requires no fitted model. The dissimilarities feed any distance-based clustering algorithm. Under intermediate-threshold asymptotics and mild moment and separation conditions, hierarchical clustering and partitioning around medoids based on them recovers the latent partition with probability tending to one. In simulations, the proposed approach matches established competitors when clusters differ mainly in the strength of extremal dependence and performs substantially better when equal-strength clusters differ in the shape of their angular measures, especially when asymmetry drives the separation. Applications to rain–snowmelt–runoff extremes in Swiss catchments and to extremal co-movement between international equity markets and the US market reveal hydrologically and economically interpretable groups.

Keywords: Aitchison coordinate, Angular measure, Compound extreme, Extreme value theory, Optimal transport, Wasserstein distance

1 Introduction

Clustering is one of the fundamental tasks in unsupervised statistical learning, aiming to partition vector of variables into homogeneous groups while maximizing their separation from other groups. As such, it has become a tool across a wide range of scientific disciplines, including biology, economics, finance, environmental sciences and climatology, where it is routinely used to uncover hidden structures. In many applications, the primary scientific interest instead lies in understanding the behaviour of rare, high-impact events. In climate science, for example, identifying regions that exhibit similar patterns of joint heatwaves and droughts is of direct interest (Zscheischler et al. 2020); in hydrology, clustering river basins according to the dependence structure of flood characteristics can support hazard mapping; and in finance, grouping markets according to their tendency to experience simultaneous crashes provides valuable information for systemic risk management and portfolio diversification. These applications call for clustering methods that focus specifically on the extremal behaviour of multivariate random vectors.

The development of clustering methods specifically designed for extreme events has attracted increasing attention over the last decade. Early contributions focused on the regionalization of spatial extremes by adapting classical clustering algorithms to dependence measures arising from extreme-value theory. Bernard et al. (2013) proposed a partitioning around medoids algorithm in which the classical Euclidean distance is replaced by the F-madogram, yielding clusters of sites exhibiting similar extremal dependence while preserving a meaningful representative within each cluster. More recently, Kiriliouk and Naveau (2020) extended this line of research by proposing a regionalization approach based on the tail dependence coefficient χ, providing an alternative characterization of extremal dependence between locations. In parallel, Janßen and Wan (2020) introduced an extremal version of spherical k-means, where clustering is performed directly on extremal observations to estimate representative directions of extremal dependence, together with consistency guarantees for the resulting cluster centres. Beyond these approaches based on extreme-value theory, Benevento et al. (2025, 2024) developed several copula-based clustering procedures combining tail dependence measures and spatial information to construct dissimilarity matrices, with applications to environmental and climatological data.

More recently, the focus has shifted from clustering spatial locations to clustering multivariate extremal dependence structures themselves. Fuchs et al. (2021) proposed a dependence-based clustering framework relying on the notion of π-comonotonicity, allowing multivariate random vectors to be grouped according to their copula. Although broadly applicable, this framework is designed for general dependence structures rather than specifically for extremal dependence, and therefore does not fully exploit the asymptotic theory of multivariate extremes. From an extreme-value perspective, Boulin et al. (2025) introduced the asymptotic-independence block model, where variables are partitioned into groups that are asymptotically independent. Under this model, clusters are estimated through a statistically consistent procedure with polynomial computational complexity. While elegant from a theoretical viewpoint, the asymptotic-independence block assumption may be restrictive in practice, as extremal dependence

structures often exhibit more complex interactions than strict block-wise asymptotic independence.

Vignotto et al. (2021) proposed a fully nonparametric clustering approach based on partitioning the extreme region and comparing the resulting empirical probabilities through a Kullback–Leibler dissimilarity. More recently, Engelke et al. (2026) established asymptotic theory for the underlying Kullback–Leibler statistic in a two-sample testing framework. Although this construction avoids restrictive parametric assumptions, its implementation requires a predefined partition, whose choice may substantially influence the resulting dissimilarities and clustering. Fomichov and Ivanovs (2023) developed spherical k-principal-components clustering to detect groups of concomitant extremes. More recently, O'Toole et al. (2025) introduced a conditional extremes clustering framework, which compares extremal dependence through fitted conditional extremes models. By relying on the conditional extremes framework, the method provides a flexible characterization of extremal dependence and has demonstrated excellent empirical performance. Nevertheless, its implementation requires fitting a conditional model for each random vector, which can become computationally demanding for large collections of datasets. Moreover, the method typically relies on Gaussian residual assumptions, which may be inadequate.

We propose an optimal-transport framework for clustering multivariate extremal dependence structures. The approach builds on the Wasserstein–Aitchison construction of Lhaut et al. (2026), who map angular observations from the simplex to a linear space using Aitchison coordinates and train a Wasserstein generative adversarial network to learn the angular measure and generate multivariate extreme observations. We use this construction for a different purpose: Wasserstein distances are computed directly between empirical angular measures in Aitchison coordinates, producing a dissimilarity metric that can be supplied to any distance-based clustering algorithm. The resulting procedure compares complete extremal dependence structures without summary coefficients, ad hoc discretization of the simplex, or fitted parametric or generative models.

We establish convergence of the empirical dissimilarities and consistency of the resulting clustering procedures. An extensive simulation study compares the proposed approach with existing methods across a wide range of extremal dependence structures. Applications to climatological and financial data illustrate its practical usefulness.

2 Methods

2.1 Statistical framework

We consider n random vectors of dimension p, each observed through T independent and identically distributed realizations. For $i = 1, \dots, n$ and $t = 1, \dots, T$, let

$$X_{i,t} = \left(X_{i,t}^{(1)}, \dots, X_{i,t}^{(p)}\right) \in \mathbb{R}^p.$$

We assume that $X_{i,1}, \dots, X_{i,T}$ have the same distribution function F_i. Moreover, we assume that the marginal distributions are continuous, and that the n random vectors are partitioned into K latent groups. Let $g(i) \in \{1, \dots, K\}$ denote the group

membership of vector i. Then random vectors belonging to the same group are assumed to share a common dependence structure. To lighten the notation, we write $X_i = \big(X_i^{(1)}, \ldots, X_i^{(p)}\big)$ for a generic random vector with distribution function F_i, that is, for a random vector distributed as any of $X_{i,1}, \ldots, X_{i,T}$. For each i, define the probability integral transforms

$$U_i^{(j)} = F_i^{(j)}\big(X_i^{(j)}\big), \qquad j = 1, \ldots, p,$$

where $F_i^{(j)}$ is the marginal distribution function of the jth component of random vector i.

By Sklar's theorem, the joint distribution of the p-variate random vector can be represented through a copula (Sklar 1959). Specifically,

$$\mathrm{pr}\big(X_i^{(1)} \le x_1, \ldots, X_i^{(p)} \le x_p\big) = C_i\big(F_i^{(1)}(x_1), \ldots, F_i^{(p)}(x_p)\big),$$

where C_i denotes the copula associated with random vector i. We assume that each C_i is in the domain of attraction of an extreme-value copula, which guarantees the existence of an angular measure. Background on the extreme-value copulas used in the simulation study is given in Section S1 of the Supplementary Material.

We assume that random vectors belonging to the same latent group share a common angular measure $H_{g(i)}$ concentrated on the open simplex Δ_{p-1}°, as described in Section 2.2, where $g : \{1, \ldots, n\} \to \{1, \ldots, K\}$ denotes the group-assignment function. The objective is to construct a dissimilarity measure between the n angular measures, which can then be used as input to a clustering algorithm in order to partition the n random vectors into the K latent groups according to their extremal dependence structure.

2.2 Angular measure

To characterize extremal dependence, we rely on the classical radial–angular decomposition from multivariate extreme-value theory (Beirlant et al. 2004; Resnick 2007). Following the marginal standardization introduced in the previous section, each component is transformed to a standard Pareto scale. For $j = 1, \ldots, p$, define

$$Z_i^{(j)} = \frac{1}{1 - U_i^{(j)}} = \frac{1}{1 - F_i^{(j)}\big(X_i^{(j)}\big)}. \tag{1}$$

By construction, $\mathrm{pr}\big(Z_i^{(j)} > z\big) = z^{-1}$ for $z \ge 1$, so that each transformed margin follows a standard Pareto distribution. We therefore obtain the transformed random vector $Z_i = \big(Z_i^{(1)}, \ldots, Z_i^{(p)}\big)$.

Using the L_1-norm, the radial and angular components associated with Z_i are defined by

$$R_i = \|Z_i\|_1 = \sum_{j=1}^{p} Z_i^{(j)}, \qquad \Theta_i = \frac{Z_i}{R_i}. \tag{2}$$

By construction, $\sum_{j=1}^{p} \Theta_i^{(j)} = 1$, so that Θ_i belongs to the unit simplex

$$\Delta_{p-1} = \{\Theta \in [0,1]^p : \sum_{j=1}^{p} \Theta^{(j)} = 1\}. \tag{3}$$

The radial variable R_i measures the overall magnitude of an observation, whereas the angular component Θ_i describes how this magnitude is distributed across the different variables. Under standard multivariate regular variation assumptions, the conditional distribution satisfies

$$\text{pr}(\Theta_i \in A \mid R_i > r) \longrightarrow H_i(A), \qquad r \to \infty, \tag{4}$$

for Borel sets $A \subset \Delta_{p-1}$ such that $H_i(\partial A) = 0$, where ∂A denotes the boundary of A relative to Δ_{p-1}. The limiting probability measure H_i is called the angular measure and characterizes the extremal dependence structure of random vector i (Resnick 2007; de Haan and Ferreira 2006). The angular measure H_i may assign positive probability to the boundary of the simplex Δ_{p-1}, namely to subsets where at least one coordinate is equal to zero. From the perspective of multivariate extremes, this represents situations in which some components of X are extreme while others remain non-extreme. Such configurations, often referred to as extreme directions (Mourahib et al. 2025), are excluded from our methodology. We assume that the angular measure concentrates on the open simplex

$$\Delta_{p-1}^{\circ} = \{\Theta \in (0,1)^p : \sum_{j=1}^{p} \Theta^{(j)} = 1\}. \tag{5}$$

The geometry of the angular measure reflects the strength and structure of extremal dependence. When extremes tend to occur in only a few components, the angular mass concentrates near the vertices of the simplex. In contrast, when large values occur simultaneously across several components, the angular mass tends to concentrate in the interior of the simplex (Coles 2001). Figure S1 of the Supplementary Material illustrates this behaviour for a bivariate random vector under two extremal dependence regimes. In the low-dependence regime, the angular observations concentrate near the vertices of the simplex, indicating that extremes mainly occur in one component at a time. In contrast, under strong extremal dependence, the angular observations concentrate near the centre of the simplex, reflecting the simultaneous occurrence of extremes.

In practice, the marginal distribution functions are unknown and the angular measure must be estimated from the observed sample. For $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $j = 1, \ldots, p$, define the empirical probability integral transform

$$\widehat{U}_{i,t}^{(j)} = \frac{\text{rank}\big(X_{i,t}^{(j)}\big)}{T+1}, \tag{6}$$

where, the rank of $X_{i,t}^{(j)}$ is computed among $\{X_{i,s}^{(j)} : s = 1, \ldots, T\}$. The corresponding empirical standard Pareto transformation is

$$\widehat{Z}_{i,t}^{(j)} = \frac{1}{1 - \widehat{U}_{i,t}^{(j)}}. \tag{7}$$

Write $\widehat{Z}_{i,t} = \big(\widehat{Z}_{i,t}^{(1)}, \ldots, \widehat{Z}_{i,t}^{(p)}\big)$. The empirical radial and angular components are then defined by

$$\widehat{R}_{i,t} = \left\|\widehat{Z}_{i,t}\right\|_1 = \sum_{j=1}^{p} \widehat{Z}_{i,t}^{(j)}, \qquad \widehat{\Theta}_{i,t} = \frac{\widehat{Z}_{i,t}}{\widehat{R}_{i,t}}. \tag{8}$$

For a sufficiently high threshold r_i, let

$$\mathcal{T}_i = \big\{t \in \{1, \ldots, T\} : \widehat{R}_{i,t} > r_i\big\}$$

denote the set of radial exceedances, and let $k_{i,T} = |\mathcal{T}_i|$ be the corresponding number of exceedances. The empirical angular measure associated with random vector i is then defined, for every Borel set $A \subset \Delta_{p-1}$, by

$$\widehat{H}_{i,T}(A) = \frac{1}{k_{i,T}} \sum_{t \in \mathcal{T}_i} \mathbb{1}\big\{\widehat{\Theta}_{i,t} \in A\big\}. \tag{9}$$

Under standard multivariate regular variation assumptions, the empirical angular measure consistently estimates the underlying angular measure when the radial threshold is chosen according to an intermediate sequence. More precisely, let

$$r_{i,T} = \frac{T}{\ell_{i,T}}, \qquad \ell_{i,T} \to \infty, \qquad \frac{\ell_{i,T}}{T} \to 0, \qquad T \to \infty.$$

Then $r_{i,T} \to \infty$, while the number of observations exceeding the threshold increases but remains asymptotically negligible relative to the sample size. Consequently, $\widehat{H}_i$ consistently estimates the limiting angular measure H_i. Consistency and asymptotic normality results for nonparametric estimators of spectral and angular measures have been established by Einmahl et al. (2001), Einmahl and Segers (2009), Janßen and Wan (2020) and more recently, Lhaut and Segers (2024).

2.3 Aitchison coordinates for angular measures

The angular measure of a multivariate regularly varying random vector is supported on the unit simplex Δ_{p-1}. Consequently, angular observations are compositional data, that is, vectors whose components are constrained to be positive and to sum to one. This simplex structure induces a non-Euclidean geometry and makes direct statistical analysis challenging (Aitchison 1982; Egozcue et al. 2003). A central idea of compositional data analysis is to represent simplex-valued observations through

coordinates in a linear Euclidean space. This is achieved using the centred log-ratio transformation

$$\mathrm{clr}(\theta) = \left(\log \frac{\theta^{(1)}}{f(\theta)}, \ldots, \log \frac{\theta^{(p)}}{f(\theta)} \right), \qquad \theta \in \Delta^{\circ}_{p-1}, \tag{10}$$

where $f(\theta) = \left(\prod_{j=1}^{p} \theta^{(j)} \right)^{1/p}$ denotes the geometric mean of the components of θ. The transformed vector belongs to the hyperplane

$$\mathcal{H} = \left\{ w \in \mathbb{R}^p : \sum_{j=1}^{p} w^{(j)} = 0 \right\},$$

which is a $(p-1)$-dimensional linear subspace of $\mathbb{R}^p$. The centred log-ratio transformation removes the simplex constraint and maps angular observations to an unconstrained Euclidean space while preserving their relative information. In particular, distances between compositions are naturally measured through the Aitchison distance

$$d_A(\theta_1, \theta_2) = \left\| \mathrm{clr}(\theta_1) - \mathrm{clr}(\theta_2) \right\|_2, \tag{11}$$

which corresponds to the Euclidean distance after the centred log-ratio transformation.

Following Egozcue et al. (2003), we use the Helmert orthonormal basis of $\mathcal{H}$. For $j = 1, \ldots, p-1$, define

$$e_j = \left\{ \frac{j}{j+1} \right\}^{1/2} \left(\underbrace{j^{-1}, \ldots, j^{-1}}_{j \text{ entries}}, -1, 0, \ldots, 0 \right).$$

This basis is orthonormal in $\mathbb{R}^p$ and spans $\mathcal{H}$. We then represent each angular observation through its Aitchison coordinates. Let $\mathcal{A} : \Delta^{\circ}_{p-1} \to \mathbb{R}^{p-1}$ denote the Aitchison coordinate map, defined for every $\theta \in \Delta^{\circ}_{p-1}$ by

$$\mathcal{A}(\theta) = \left(\langle \mathrm{clr}(\theta), e_1 \rangle, \ldots, \langle \mathrm{clr}(\theta), e_{p-1} \rangle \right). \tag{12}$$

Applying this map to the empirical angular observations in equation (8) yields the Aitchison coordinates

$$a_{i,t} = \mathcal{A}\left(\widehat{\Theta}_{i,t} \right) \in \mathbb{R}^{p-1}, \qquad t \in \mathcal{T}_i. \tag{13}$$

The use of Aitchison coordinates is particularly advantageous for clustering angular measures. Directly comparing distributions on the simplex may be difficult because the simplex is a constrained space where variability is compressed by the positivity and unit-sum constraints. The centred log-ratio transformation effectively decompresses the angular information by mapping the simplex onto a linear Euclidean space. As a consequence, differences between angular distributions become easier to detect and

quantify. This idea was leveraged by Lhaut et al. (2026) to simulate multivariate extremes through generative adversarial networks operating in Aitchison coordinates.

In the proposed methodology, the empirical angular observations associated with each random vector, as defined in Eq. (9), are first transformed into Aitchison coordinates, leading to an empirical probability measure on $\mathbb{R}^{p-1}$ defined, for every Borel set $B \subset \mathbb{R}^{p-1}$, in terms of the vectors $a_{i,t}$ in equation (13) by

$$\widehat{G}_{i,T}(B) = \frac{1}{k_{i,T}} \sum_{t \in \mathcal{T}_i} \mathbb{1}\{a_{i,t} \in B\}\,. \tag{14}$$

Distances between random vectors are then computed by applying the Wasserstein distance to the corresponding distributions in Aitchison coordinates.

2.4 Wasserstein distance between angular measures

To compare the extremal dependence structures associated with different random vectors, we compare the empirical measures either directly on the simplex or after transformation to Aitchison coordinates using Wasserstein distances from optimal transport theory. Intuitively, the Wasserstein distance quantifies the minimal cost required to transport one probability distribution onto another. Let P_i and P_l denote two probability measures of random vectors i and l, respectively. The 1-Wasserstein distance between P_i and P_l is (Villani 2009)

$$W_1(P_i, P_l) = \inf_{\pi \in \Pi(P_i, P_l)} \int_{\mathbb{R}^{p-1} \times \mathbb{R}^{p-1}} \|x - y\|_2 \, \mathrm{d}\pi(x, y),$$

where $\Pi(P_i, P_l)$ denotes the collection of all probability measures on $\mathbb{R}^{p-1} \times \mathbb{R}^{p-1}$ with marginals P_i and P_l. The quantity $\|x - y\|_2$ represents the transportation cost between the points x and y.

The Wasserstein distance therefore provides a comparison between the distributions by taking into account the global organization of probability mass. In contrast with the partition-based approach of Vignotto et al. (2021), this methodology does not require the specification of a partition parameter and avoids the information loss induced by discretizing the simplex.

For arbitrary $p \geq 2$, the empirical Wasserstein distance can be evaluated as a finite-dimensional discrete optimal-transport problem. Let

$$k_i = |\mathcal{T}_i|, \qquad k_l = |\mathcal{T}_l|,$$

be the numbers of retained observations. Re-index the observations in the chosen representation as

$$y_{i,1}, \ldots, y_{i,k_i} \qquad \text{and} \qquad y_{l,1}, \ldots, y_{l,k_l},$$

where $y_{i,r}$ denotes either an angular observation on the simplex or its Aitchison coordinates. The corresponding empirical measures are

$$\widehat{P}_i = \frac{1}{k_i}\sum_{r=1}^{k_i}\delta_{y_{i,r}}, \qquad \widehat{P}_l = \frac{1}{k_l}\sum_{s=1}^{k_l}\delta_{y_{l,s}}.$$

Define the set of admissible transport plans by

$$\Gamma_{il} = \left\{\gamma \in \mathbb{R}_+^{k_i\times k_l} : \sum_{s=1}^{k_l}\gamma_{rs} = \frac{1}{k_i}, \quad r = 1,\dots,k_i, \qquad \sum_{r=1}^{k_i}\gamma_{rs} = \frac{1}{k_l}, \quad s = 1,\dots,k_l\right\}.$$

The empirical 1-Wasserstein distance is then

$$\widehat{d}_{il}^{W} = W_1(\widehat{P}_i, \widehat{P}_l) = \min_{\gamma\in\Gamma_{il}}\sum_{r=1}^{k_i}\sum_{s=1}^{k_l}\gamma_{rs}d\left(y_{i,r}, y_{l,s}\right), \tag{15}$$

where d is the Euclidean distance in the chosen representation and γ_{rs} is the amount of probability mass transported from $y_{i,r}$ to $y_{l,s}$. The marginal constraints ensure that γ has $\widehat{P}_i$ and $\widehat{P}_l$ as its two marginals. We evaluate Equation (15) numerically using the network-flow implementation of the discrete Kantorovich problem provided by the `transport` package in `R` (Peyré and Cuturi 2019; Schuhmacher et al. 2024). Finally, the dissimilarity matrix,

$$\widehat{D}_T^{W} = \left(\widehat{d}_{il}^{W}\right)_{i,l=1,\dots,n} \in \mathbb{R}^{n\times n}, \tag{16}$$

serves as input to any distance-based clustering algorithm.

3 Consistency of Clustering Procedures

We establish the consistency of the proposed Wasserstein clustering methodology for hierarchical clustering with single, complete, or average linkage and for the global K-medoids solution in a projected Aitchison-coordinate space. Recall that Δ_{p-1}° is the open unit simplex defined in equation (5), and that $\mathcal{A} : \Delta_{p-1}^{\circ} \to \mathbb{R}^{p-1}$ is the Aitchison-coordinate map in equation (12). We primarily advocate the Aitchison transformation in the bivariate setting, where it substantially improves clustering performance. In higher dimensions, working directly on the simplex can also be effective, making the transformation optional.

The consistency argument for the direct-simplex approach is immediate. Because the simplex has finite Euclidean diameter, weak convergence of the empirical angular measures implies convergence in W_1. The clustering-consistency result then follows from the same separation argument used below, without any projection. The Aitchison case requires additional care because $\mathcal{A}$ maps the bounded simplex onto the unbounded space $\mathbb{R}^{p-1}$: angular observations approaching the simplex boundary are mapped

arbitrarily far from the origin, so weak convergence alone need not imply convergence in W_1 after transformation. To avoid imposing additional moment conditions, we project the Aitchison coordinates onto a closed ball of radius M. The projection is only needed for the proof. For any finite dataset, one can always choose M large enough that the projection does not change any observation.

For $M > 0$, let

$$\mathcal{B}_M = \{a \in \mathbb{R}^{p-1} : \|a\|_2 \le M\}$$

and define the Euclidean projection $\Pi_M : \mathbb{R}^{p-1} \to \mathcal{B}_M$ by

$$\Pi_M(a) = \begin{cases} a, & \|a\|_2 \le M, \\ Ma/\|a\|_2, & \|a\|_2 > M. \end{cases}$$

The map Π_M is continuous and satisfies $|\Pi_M(a)|_2 = \min\{|a|_2, M\}$.

Assumption 1 After marginal standardization to a common Pareto scale, each random vector X_i, $i \in \{1, \dots, n\}$, admits a limiting angular probability measure H_i on Δ_{p-1}°. Let

$$G_i = \mathcal{A}_{\#} H_i$$

be its image in Aitchison coordinates, where $\mathcal{A}_{\#} H_i$ denotes the pushforward of H_i under $\mathcal{A}$.

Assumption 2 The radial threshold used to construct $\widehat{H}_{i,T}$ satisfies

$$r_{i,T} = \frac{T}{\ell_{i,T}}, \qquad \ell_{i,T} \to \infty, \qquad \frac{\ell_{i,T}}{T} \to 0, \qquad T \to \infty.$$

Assumption 3 The latent groups are characterized by their limiting angular measures:

$$g(i) = g(l) \quad \Longleftrightarrow \quad H_i = H_l.$$

Assumption 4 The number of latent groups K is known and fixed.

Let $\widehat{G}_{i,T}$ be the empirical Aitchison-coordinate measure defined in equation (14). For $M > 0$, define

$$G_i^{(M)} = (\Pi_M)_{\#} G_i, \qquad \widehat{G}_{i,T}^{(M)} = (\Pi_M)_{\#} \widehat{G}_{i,T}.$$

These measures are supported on $\mathcal{B}_M$ and thus have first moments bounded by M.

Lemma 1 (Preservation of group identifiability) *Under Assumption 3, there exists $M_0 < \infty$ such that, for every $M \ge M_0$ and every $i, l \in 1, \dots, n$,*

$$g(i) = g(l) \quad \Longleftrightarrow \quad G_i^{(M)} = G_l^{(M)}.$$

Consequently,

$$\Delta_M = \min_{g(i) \ne g(l)} W_1\left(G_i^{(M)}, G_l^{(M)}\right) > 0.$$

Define

$$\widehat{d}_{il,T}^{W,M} = W_1\left(\widehat{G}_{i,T}^{(M)}, \widehat{G}_{l,T}^{(M)}\right), \qquad \widehat{D}_T^{W,M} = \left(\widehat{d}_{il,T}^{W,M}\right)_{1\le i,l\le n}.$$

Theorem 1 (Consistency of projected Wasserstein–Aitchison clustering) *Fix $M \ge M_0$ and let the partition $\widehat{\mathcal{P}}_T^{(M)}$ of $\{1,\dots,n\}$ be obtained from $\widehat{D}_T^{W,M}$ by either: (i) hierarchical clustering with single, complete, or average linkage, cut to produce K clusters; or (ii) assignment to the nearest medoid in a global solution*

$$\widehat{\mathcal{S}}_T^{(M)} \in \underset{\substack{\mathcal{S}\subseteq\{1,\dots,n\}\\ |\mathcal{S}|=K}}{\arg\min} \sum_{i=1}^{n} \min_{s\in\mathcal{S}} \widehat{d}_{is,T}^{W,M}.$$

Under Assumptions 1–4,

$$\mathrm{pr}\left\{\widehat{\mathcal{P}}_T^{(M)} = \mathcal{P}\right\} \longrightarrow 1, \qquad T\to\infty,$$

where $\mathcal{P}$ is the true latent-group partition.

The proofs of Lemma 1 and Theorem 1 are given in Section S3 of the Supplementary Material.

4 Simulation Study

The simulation study is designed to evaluate the clustering performance of the proposed Wasserstein–Aitchison methodology and to compare it with the approaches of Vignotto et al. (2021) and O'Toole et al. (2025) in a wide range of controlled scenarios of extremal dependence. Details on the two competitors can be found in Section S4.1 of the Supplementary Material. Clustering agreement is measured using the adjusted Rand index; details on this metric are provided in Section S4.2 of the Supplementary Material.

Because the simulation study involves only a finite collection of finite samples, a common finite projection radius M can be chosen larger than the maximum norm of all observed Aitchison coordinates. The projection then leaves every coordinate unchanged, so the projected and unprojected Wasserstein–Aitchison procedures coincide exactly for all results reported below.

In the simulation studies and applications, the K-medoids partition is computed using partitioning around medoids (Kaufman and Rousseeuw 1990), an iterative algorithm for approximating the global K-medoids solution considered in Theorem 1. Partitioning around medoids is used because it produced better empirical results than hierarchical clustering in our experiments and was also employed by Vignotto et al. (2021) and O'Toole et al. (2025).

The methodology of Section 2 and the consistency result of Section 3 are stated for a general dimension $p \ge 2$. Simulation studies I and II below are bivariate, whereas simulation study III is trivariate. Mean adjusted Rand indices, Monte Carlo standard errors and exact-recovery rates for all simulation are reported in Tables S1– S6 of the Supplementary Material.

4.1 Simulation study I: varying strength of extremal dependence

For every simulation setting, the latent number of groups is fixed to $K = 3$. The study varies the separation between dependence regimes, the sample size, the cluster-size configuration and the copula family. Extremal dependence is summarized by the upper tail dependence coefficient χ, defined for $p = 2$ by

$$\chi = 2 \int_0^1 \min(\theta, 1 - \theta)\, dH(\theta).$$

Three levels of separation between the latent groups are considered: easy, with $\chi = (0.1, 0.5, 0.9)$, yielding clearly separated extremal dependence structures; moderate, with $\chi = (0.2, 0.5, 0.8)$, producing a more challenging clustering problem; and hard, with $\chi = (0.3, 0.45, 0.6)$, corresponding to weakly separated extremal dependence structures.

We consider four different sample sizes, namely $T \in \{250, 500, 750, 1\,000\}$. Two cluster-size configurations are considered: a balanced design with $N = (10, 10, 10)$, in which each latent group contains the same number of random vectors, and an unbalanced design with $N = (5, 10, 20)$. The simulations are conducted under two bivariate extreme-value copula families, the Gumbel–Hougaard and the Hüsler–Reiss copula. The parameter W which controls the partitioning of the space, is set to $W \in \{2, 3, 4, 5, 8, 10, 15\}$. This allows us to assess the sensitivity of the method to the partition resolution. For both methods, extremal dependence is estimated using only the largest 10% of radial observations and the conditional extremes approach is fitted above the 90th percentile threshold. Finally, for each scenario, 500 independent replications are generated.

Figure 1 report the adjusted Rand indices for a subset of the unbalanced design. While, Figures S2 and S3 in Section S4 of the Supplementary Material report it for the full design. We observed that Wasserstein–Aitchison and conditional extremes clustering consistently outperform the partition-based method of Vignotto et al. (2021). Although the latter improves as the number of partitions increases, particularly from small values, it rarely attains the accuracy of the other two methods. Performance also improves substantially with sample size: in the easy and moderate scenarios, both leading methods approach perfect recovery for the Gumbel and Hüsler–Reiss models under both designs. Their performance is similar for balanced groups, whereas conditional extremes clustering is generally more accurate under unbalanced designs, especially in the moderate and hard scenarios. As expected, decreasing separation from the easy to the hard setting reduces accuracy and increases variability. Finally, Wasserstein–Aitchison outperforms Wasserstein–simplex, indicating that the Aitchison transformation provides a more informative geometry for distinguishing bivariate extremal dependence structures.

4.2 Simulation study II: equal tail dependence and different angular shapes

The first simulation study considers groups with different upper tail-dependence coefficients χ. Because χ does not fully characterize the angular measure, we now

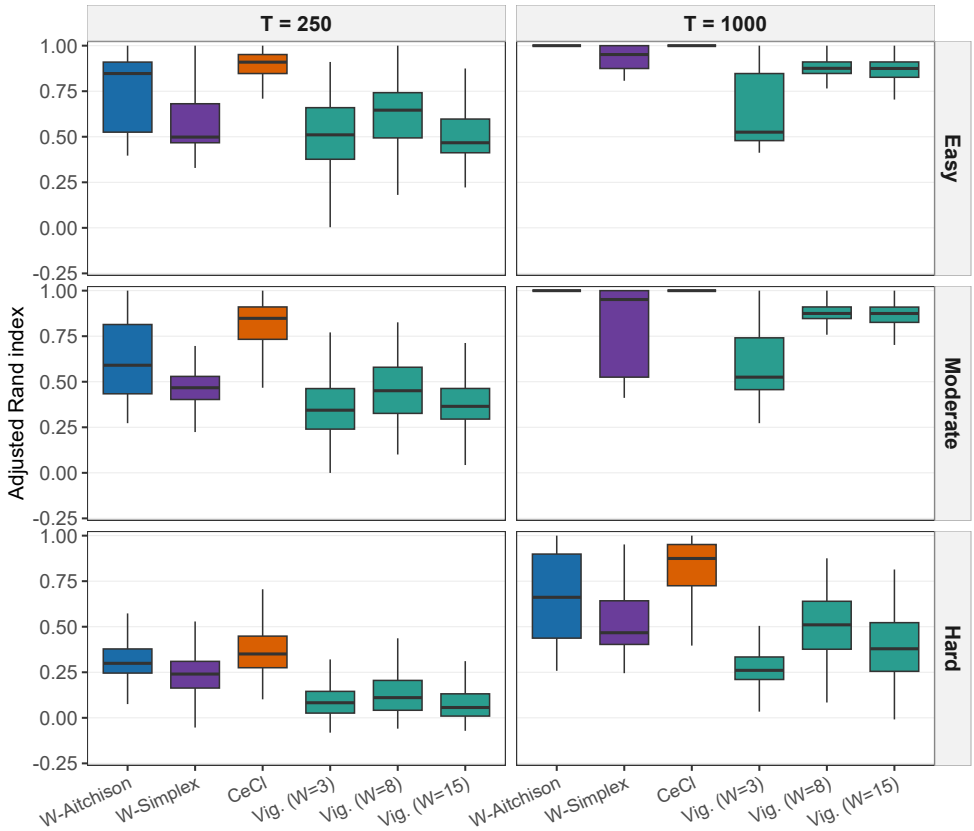


Fig. 1: Adjusted Rand index for a subset of the unbalanced design $N = (5, 10, 20)$, with $K = 3$ latent groups, for sample sizes $T \in \{250, 1\,000\}$ and the three separation levels easy, moderate and hard, corresponding to $\chi = (0.1, 0.5, 0.9)$, $(0.2, 0.5, 0.8)$ and $(0.3, 0.45, 0.6)$, under the Hüsler–Reiss copulas. The Wasserstein–Aitchison, Wasserstein–simplex and conditional extremes dissimilarities are compared with the partition-based dissimilarity of Vignotto et al. (2021), the latter for $W \in \{3, 8, 15\}$ partitions. Results are based on 500 Monte Carlo replications.

examine whether the procedures can distinguish groups having the same χ but different angular structures. The three groups are generated from Galambos, Hüsler–Reiss and extremal-t copulas, the latter with $\nu = 5$ degrees of freedom. For each $\chi \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, the copula parameters are calibrated to give the same value of χ across groups. Their recovery must therefore rely on differences between the complete angular measures rather than on the overall strength of tail dependence. Given the difficulty of this setting, we consider $T \in \{10\,000, 100\,000\}$.

We compare Wasserstein–Aitchison and conditional extremes clustering, both combined with K-medoids and the true value $K = 3$. The dissimilarity of Vignotto et al. (2021) is omitted because it was dominated by both methods in the first simulation

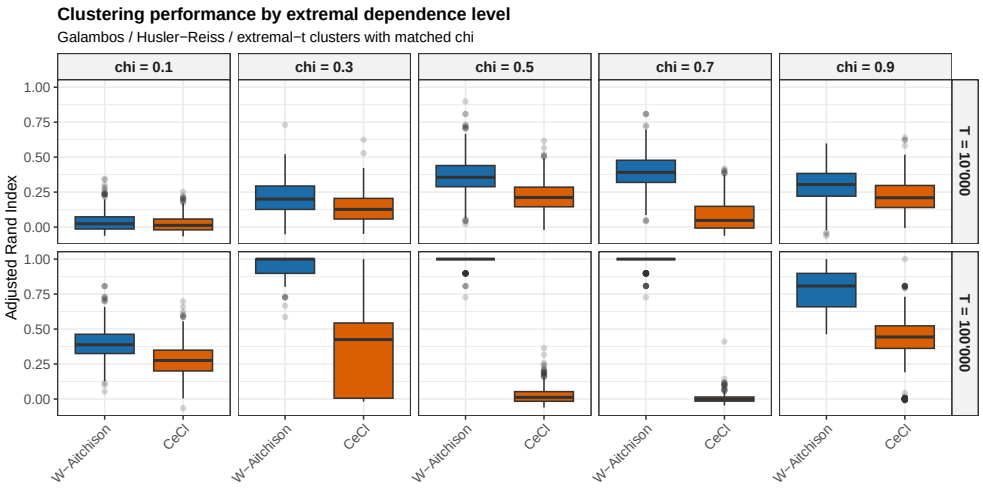


Fig. 2: Adjusted Rand index obtained by the Wasserstein–Aitchison and conditional extremes clustering procedures for the equal-tail-dependence simulation study. The three latent groups are generated from Galambos, Hüsler–Reiss and extremal-t copulas calibrated to share the same upper tail dependence coefficient χ. Results are shown for $\chi \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and sample sizes $T \in \{10\,000, 100\,000\}$ based on 500 Monte Carlo replications.

study. Wasserstein–Aitchison retains the largest 10% of radial observations, while the conditional extremes model is fitted above the 90th percentile. Each scenario is evaluated using 500 independent Monte Carlo replications.

Figure 2 reports the adjusted Rand index obtained by the two procedures. The Wasserstein–Aitchison method outperforms conditional extremes clustering in this setting as well, attaining a higher adjusted Rand index. The difference between the two procedures is most pronounced for the larger sample size $T = 100\,000$ and for intermediate levels of extremal dependence, that is, for χ around 0.5. Since all three latent groups share the same upper tail dependence coefficient by construction, this difference cannot be attributed to the strength of extremal dependence. It indicates instead that comparing the complete angular measures, rather than a fitted conditional model, retains more of the information contained in the shape of the angular measure.

4.3 Simulation study III: trivariate asymmetric angular measures

The third simulation study examines whether the Wasserstein–Aitchison and Wasserstein–simplex procedures remain effective in dimension $p = 3$, where angular measures are supported on the two-dimensional simplex embedded in $\mathbb{R}^3$. We consider three asymmetric angular measures that have identical pairwise tail-dependence coefficients, $\chi_{12} = \chi_{13} = \chi_{23} = 0.45$, but differ in their orientation and higher-order

dependence structure. Consequently, the latent groups cannot be distinguished using pairwise tail-dependence coefficients alone.

The three angular measures are denoted by $H_{+,\delta}$, H_0 and $H_{-,\delta}$, where $\delta \in \{0.20, 0.60, 1\}$ controls their separation. For small δ, the positive- and negative-orientation measures are close to H_0, whereas $\delta = 1$ yields their greatest separation, without changing their pairwise tail-dependence coefficients. The measures are constructed using constrained mixtures of Dirichlet densities; full details are provided in Section S5 of the Supplementary Material. Observations are generated by drawing an angle from the relevant measure and multiplying it by an independent Pareto radial component. Figure 3 displays 3 000 independent angular observations from each measure for every value of δ.

The three vertices of the simplex represent extreme events dominated by a single component. For example, observations close to the first vertex satisfy $\Theta_1 \approx 1$ and $\Theta_2, \Theta_3 \approx 0$. Observations close to the centre of the simplex correspond to events in which all three components contribute similar proportions to the overall magnitude.

The middle column does not change across rows because it always represents H_0. In the left and right columns, the angular measures move progressively farther from H_0 as δ increases, and they do so in opposite directions: their orientations, in the sense made precise in Section S5 of the Supplementary Material, have opposite signs. For small values of δ, the three columns are visually similar because $H_{+,\delta}$ and $H_{-,\delta}$ are both dominated by the common component H_0. When $\delta = 1$, the differences between the three angular measures are the most pronounced.

For each value of δ, one simulated dataset contains three latent groups, with four random vectors per group and $T = 1\,000$ independent observations per random vector. For each random vector, the largest 10% of the observations according to their radial component $\|Z_{i,t}\|_1$ are retained to compute the empirical angular measure.

The experiment is repeated independently 500 times for each of the five values of δ. For each dataset, the proposed Wasserstein–Aitchison and Wasserstein–simplex dissimilarities and the conditional extremes dissimilarity are computed. The resulting dissimilarity matrices are then clustered using partitioning around medoids with the true number of groups, $K = 3$.

Figure 4 compares the three dissimilarities as a function of δ. Performance improves monotonically with δ. For $\delta = 0.20$, the angular measures remain close to H_0, and the mean adjusted Rand indices are only 0.069, 0.085 and 0.055 for Wasserstein–Aitchison, Wasserstein–simplex and conditional extremes clustering, respectively. For $\delta = 0.40$, 0.60 and 0.80, the corresponding values are $(0.336, 0.351, 0.229)$, $(0.598, 0.624, 0.466)$ and $(0.867, 0.881, 0.745)$. Thus, both Wasserstein methods outperform conditional extremes clustering at intermediate separation levels. Unlike in the bivariate experiments, Wasserstein–simplex performs similarly to Wasserstein–Aitchison: their mean adjusted Rand indices differ by at most 0.026.

At $\delta = 1$, the mean adjusted Rand indices reach 0.977, 0.981 and 0.939, with perfect recovery in 91.4%, 93.0% and 78.8% of the replications. Overall, the two Wasserstein formulations perform nearly equivalently and consistently outperform conditional extremes clustering. This advantage arises from comparing the complete trivariate angular measures.

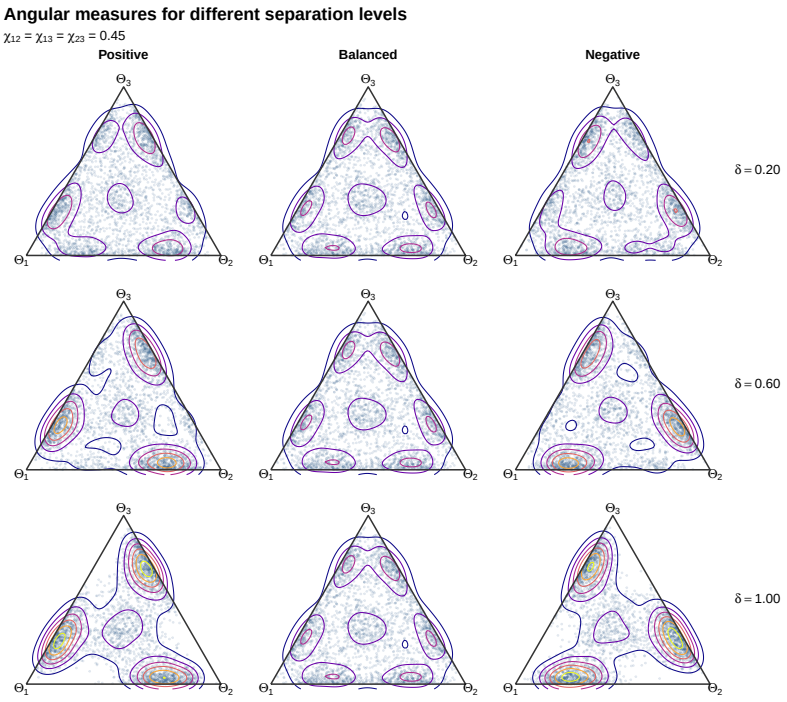


Fig. 3: Angular measures used in the trivariate simulation study. Rows correspond to the separation parameter $\delta \in \{0.20, 0.40, 0.60, 0.80, 1\}$. Columns correspond, respectively, to the positive-orientation, orientation-balanced and negative-orientation angular measures. Points represent independent draws from the angular measures, while contours indicate regions of greater angular density. All panels have identical pairwise tail-dependence coefficients, $\chi_{12} = \chi_{13} = \chi_{23} = 0.45$.

5 Data Application

5.1 Application to international equity markets

We now illustrate the proposed Wasserstein–Aitchison clustering procedure on international equity-market data. The objective is to cluster national stock-market indices according to their extremal co-movement with the United States market. For each country i, we consider the bivariate vector $(-R_{\mathrm{US},t}, -R_{i,t})$, where $R_{\mathrm{US},t}$ denotes the daily log-return of the S&P 500 index and $R_{i,t}$ denotes the daily log-return of the equity index associated with country i. Positive values therefore correspond to market losses. The analysis focuses on the extremal dependence between large United States market

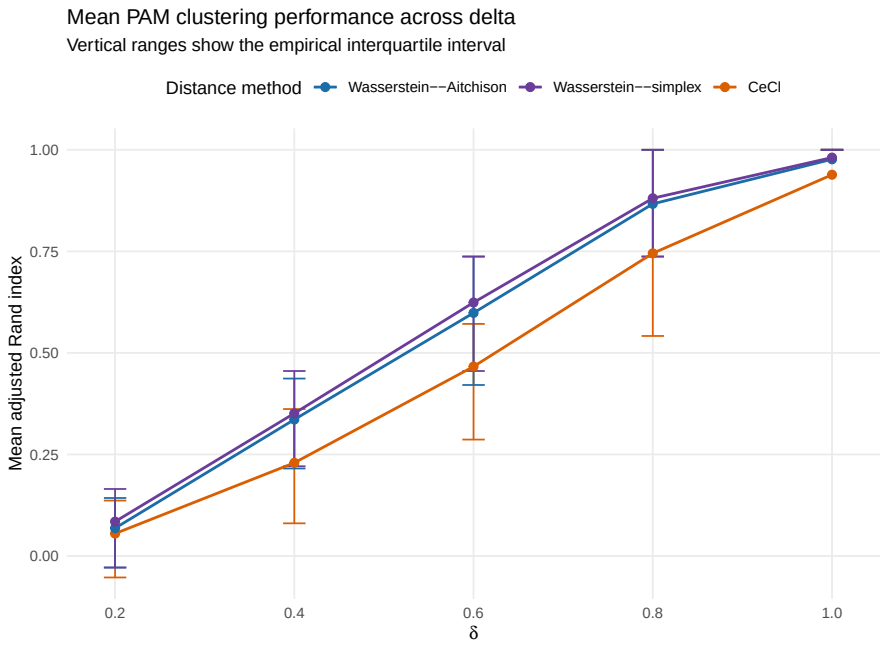


Fig. 4: Mean adjusted Rand index for partitioning around medoids using the Wasserstein–Aitchison, Wasserstein–simplex and conditional extremes dissimilarities. Results are based on 500 replications; vertical ranges are empirical interquartile intervals. For every δ, the three groups have identical first angular moments and pairwise coefficients $\chi_{12} = \chi_{13} = \chi_{23} = 0.45$.

losses and large country-specific losses. Daily closing prices are downloaded from Yahoo Finance for a broad set of national equity indices covering North America, South America, Europe, Asia–Pacific, the Middle East and Africa. The sample period extends from January 2000 until June 2025. Only indices with at least 1 000 valid daily observations are retained. Daily log-returns are then computed from the cleaned price series.

Since financial returns typically exhibit serial dependence and volatility clustering, the return series are filtered prior to the extremal analysis. For each country, both the United States and local-market returns are fitted separately using a first-order autoregressive model with generalized autoregressive conditional heteroscedasticity of order $(1, 1)$ and Student t innovations. The standardized residuals are then extracted and used as inputs for the clustering procedure. This filtering step aims to remove temporal dependence and conditional heteroscedasticity.

For each country, a bivariate sample is constructed by pairing the filtered United States residuals with the filtered residuals of the corresponding local market, $(\widehat{\varepsilon}_{\mathrm{US},t}, \widehat{\varepsilon}_{i,t})$. The methodology described in the previous sections is then applied. Observations are transformed to Pareto margins, radial–angular coordinates are computed, and only the largest 5% of radial observations are retained. The resulting angular observations are transformed into Aitchison coordinates and pairwise Wasserstein distances are

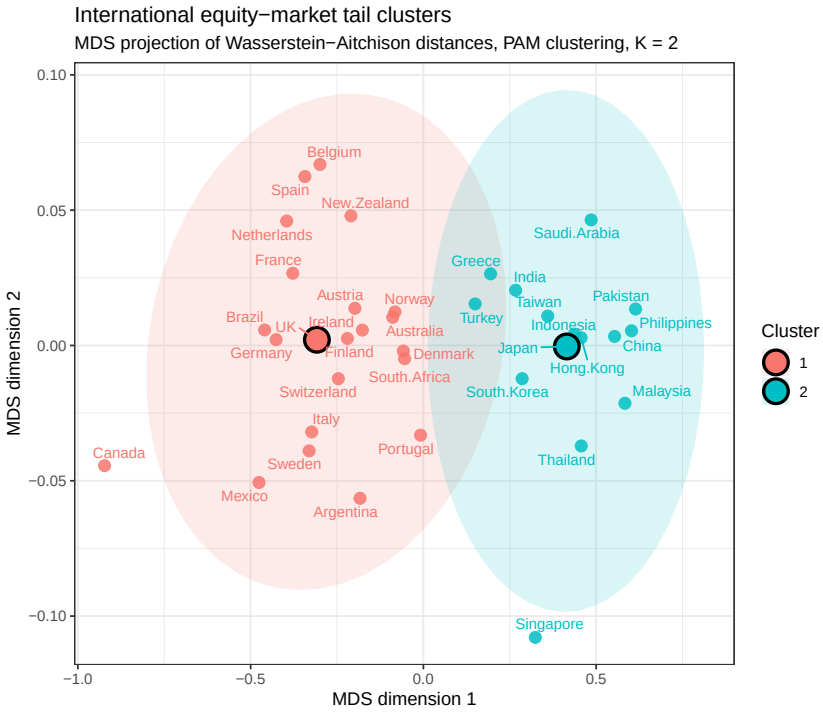


Fig. 5: International equity-market tail clusters obtained from the Wasserstein–Aitchison dissimilarity matrix. The figure shows a two-dimensional classical multidimensional scaling projection of the pairwise distances together with the partitioning around medoids result. Black-circled observations correspond to the cluster medoids.

computed between countries. The resulting dissimilarity matrix is finally clustered using partitioning around medoids.

The number of clusters is selected using the average silhouette width. Figure S4 in Section S6.1 of the Supplementary Material displays the silhouette criterion for values of K ranging from 2 to 10. The maximum silhouette value is obtained at $K = 2$, indicating that the dominant structure in the data is a two-cluster partition.

Figure 5 presents a two-dimensional classical multidimensional scaling representation of the Wasserstein–Aitchison distance matrix together with the partitioning around medoids results. The two clusters exhibit a clear geographical and economic interpretation. The first cluster contains most North American, South American and European markets together with Australia and New Zealand. The second cluster is composed primarily of Asian markets, including China, Hong Kong, India, South Korea, Taiwan, Singapore, Malaysia, Thailand, Indonesia, the Philippines and Pakistan. The cluster medoids are the United Kingdom and Japan.

The two medoids show distinct extremal-dependence regimes. The angular distribution of the United Kingdom is more concentrated near the centre of the simplex, indicating a relatively balanced sharing of extreme losses between the United States and

United Kingdom markets, whereas that of Japan has greater mass near the boundaries, indicating that extreme events are more frequently dominated by one market. Detailed angular-density estimates for the medoids and for countries near the cluster boundaries are provided in Figs. S5 and S6 in Section S6 of the Supplementary Material. Applying the conditional extremes clustering methodology of O'Toole et al. (2025) to the same data set yields exactly the same two-cluster partition; the corresponding multidimensional scaling representation is shown in Fig. S7 of the Supplementary Material. This agreement supports the robustness of the identified extremal-dependence structure.

5.2 Clustering rain–snowmelt–runoff extremal dependence in hydrologic Switzerland

This application investigates whether Swiss catchments can be grouped according to the extremal dependence structure linking liquid precipitation, snowpack depletion and runoff response. We use version 0.8 of the Catchment Attributes and Meteorology for Large-sample Studies–Switzerland data set (CAMELS-CH) (Höge et al. 2023a,b). For catchment i and day t, the analysis is based on the trivariate vector

$$X_{i,t} = \left(L_{i,t}, M_{i,t}, Q^{\star}_{i,t}\right),$$

whose components represent three aspects of a potential flood-generating event. First, $L_{i,t}$ is liquid precipitation accumulated over the current and two preceding days, obtained by partitioning daily precipitation into liquid and solid fractions using air temperature. Second, $M_{i,t}$ is the positive net depletion of snow water equivalent over three days and is used as a proxy for snowpack water release. Third, $Q^{\star}_{i,t}$ is the maximum observed specific discharge on day t or either of the following two days, allowing for a delay between meteorological forcing and runoff response. Three day accumulation was chosen based on the typical passage time of extratropical cyclones (Froidevaux et al. 2015). The analysis covers 1 September 1998–31 December 2020, determined by the availability of snow-water-equivalent data. Complete definitions and data-quality criteria are provided in Section S6.2 of the Supplementary Material.

Within each catchment, temporal dependence among extreme observations was handled using runs declustering with a run length of three days, following Sabourin and Renard (2015). An event was considered to have ended after three consecutive days without a radial exceedance, and the observation with the largest radial component within each event was retained. We retained only catchments with at least 50 events, a deliberately low minimum for estimating a trivariate angular measure. We applied the proposed Wasserstein–Aitchison methodology to the upper 5% of each catchment-specific radial distribution. The average silhouette width slightly favoured $K = 3$ over $K = 2$ (Fig. S10), but the sensitivity analysis in Section S6.2 of the Supplementary Material showed substantially greater stability for $K = 2$, which we therefore retained.

The resulting partition exhibits broad spatial organization without forming strictly contiguous regions, as shown in Fig. S9 of the Supplementary Material. Because geographical coordinates and catchment adjacency were not used in the clustering, this organization indicates an association with regional hydroclimatic conditions rather than a direct reproduction of geographical proximity. Figure 6 displays the empirical angular

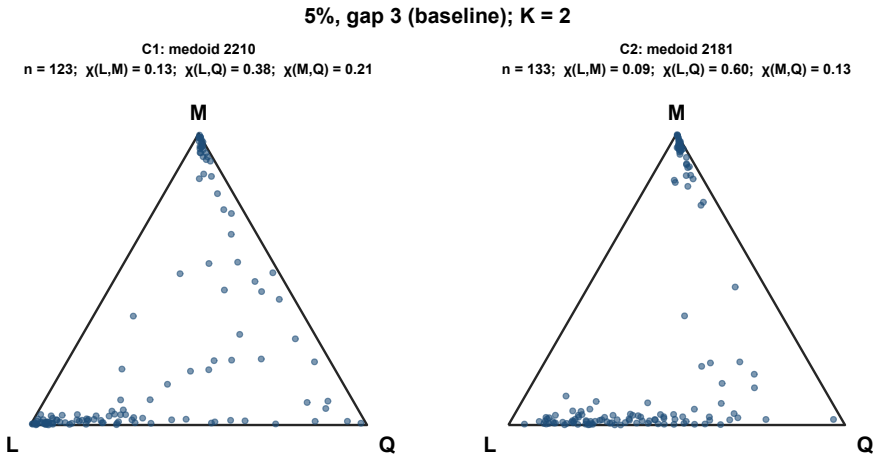


Fig. 6: Empirical angular measures of the $K = 2$ cluster medoids (C1, left; C2, right). Here L, M and Q denote standardized liquid precipitation, snow-water-equivalent depletion and runoff, respectively. Annotations report pairwise tail-dependence coefficients and numbers of declustered events.

measures of the two cluster medoids and reveals two distinct extremal-dependence regimes. The C1 medoid has the most broadly dispersed angular measure, with substantial mass in rainfall-related directions but also comparatively frequent snowmelt-related events. In contrast, the C2 medoid is strongly concentrated near the liquid-rain–runoff edge and has a substantially larger rain–runoff tail-dependence coefficient than its rain–melt and melt–runoff coefficients. It represents the clearest rainfall–runoff regime.

The hydrological interpretation is supported by four catchment attributes that were excluded from the clustering: mean elevation, fraction of precipitation falling as snow, mean slope and mean half-flow date. As shown in Fig. 7, C1 contains higher and more snow-influenced catchments with later runoff timing; C2 contains lower catchments with smaller snow fractions.

6 Conclusion

We have introduced a dissimilarity between multivariate extremal dependence structures based on optimal transport. The angular measure of each random vector is estimated nonparametrically above a high radial threshold, mapped from the unit simplex to $\mathbb{R}^{p-1}$ through the centred log-ratio transformation of Aitchison geometry, and two random vectors are compared through the 1-Wasserstein distance between the resulting empirical measures. The construction requires neither an ad hoc partition of the simplex, as in Vignotto et al. (2021), nor the fitting of a conditional extremes model, as in O'Toole et al. (2025). Theorem 1 shows that hierarchical clustering and global

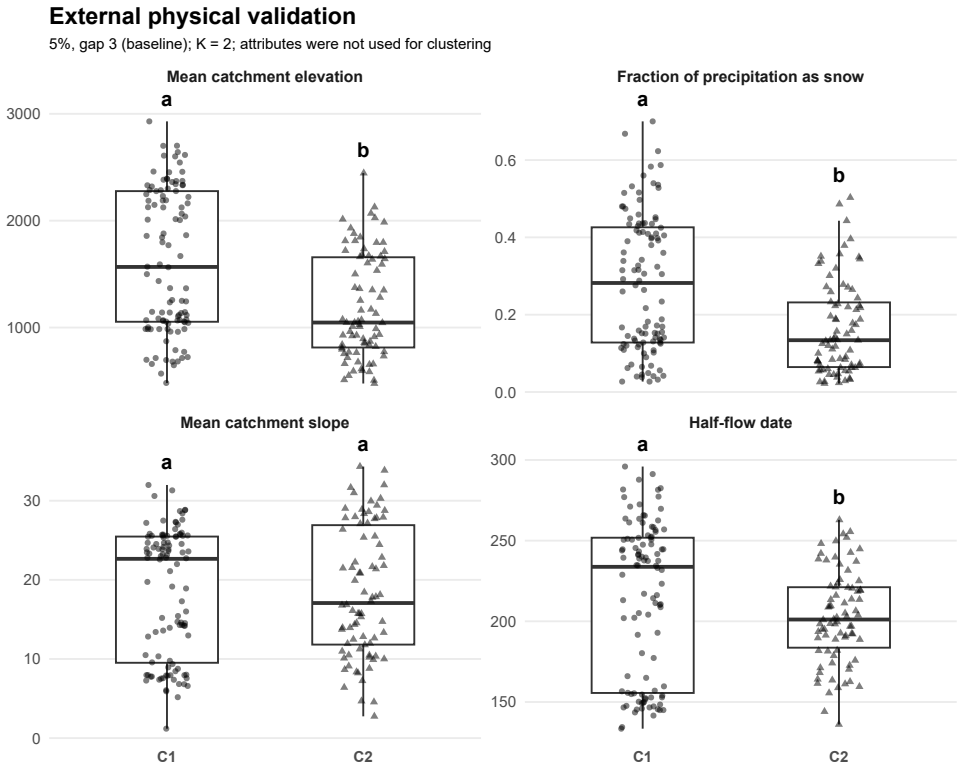


Fig. 7: External physical validation of the two extremal-dependence clusters. The panels show mean catchment elevation, fraction of precipitation falling as snow, mean catchment slope and mean half-flow date. Individual symbols represent catchments, with circles, triangles corresponding to C1, C2. Lowercase letters denote significance groups based on pairwise Wilcoxon rank-sum tests at the 5% level.

K-medoid based on the resulting dissimilarity matrix recovers the latent partition with probability tending to one. The simulation study of Section 4 confirms this behaviour in finite samples and shows that the gain over existing procedures is largest precisely in the situations that motivate the approach, namely when the latent groups cannot be separated by their pairwise tail dependence coefficients alone. The two applications of Section 5 yield interpretable partitions: a separation of Swiss catchments into more snowmelt-influenced and predominantly rainfall–runoff regimes, and a separation between Western developed markets and Asian and emerging markets according to their extremal co-movement with the US market.

Several directions deserve further attention. First, the radial threshold is chosen as a fixed proportion of the observations, and a data-driven choice, possibly differing across random vectors, would make the procedure more automatic. Second, the consistency result assumes that the number of latent groups K is known, whereas in the applications it is selected by maximizing the average silhouette width; a formal treatment of

this selection step is left for future work. Finally, the observations are assumed to be independent over time. In the applications, temporal dependence is removed by prefiltering the series or run declustering, but a direct treatment of serially dependent extremes would be a natural extension.

Possible future applications in climate science could include analysis of heatwave-drought compounding mechanisms and their changes under global warming.

Supplementary material. Supplementary material containing additional methodological details, proofs, simulation results and application results is provided after the main manuscript.

Declarations

Funding. This research is supported by the Projet d'Actions de Recherche Concertées (ARC) No. 24/29-146 (EXALT) of the Communauté française de Belgique.

Use of generative AI and AI-assisted technologies. During the preparation of this manuscript, the authors used ChatGPT for coding and debugging assistance, support in checking mathematical proofs, and concise language editing. All AI-assisted output was critically reviewed and verified by the authors, who take full responsibility for the manuscript.

Supplementary Material

S1 Extreme-value copulas

Since the objective is to model dependence between extreme realizations of the innovation process, we restrict attention to the class of extreme-value copulas. Extreme-value copulas arise naturally as limiting dependence structures of componentwise maxima and constitute a fundamental class of models in multivariate extreme-value theory (Gudendorf and Segers 2010). They are particularly well suited for describing extremal dependence between variables. In the multivariate setting, every p-variate extreme-value copula admits the representation

$$C(u_1, \ldots, u_p) = \exp\{-\ell(-\log u_1, \ldots, -\log u_p)\}, \qquad (u_1, \ldots, u_p) \in (0,1]^p,$$

where $\ell : [0,\infty)^p \to [0,\infty)$ is the stable tail dependence function. The function ℓ is convex, homogeneous of order one, and satisfies

$$\max(x_1, \ldots, x_p) \le \ell(x_1, \ldots, x_p) \le x_1 + \cdots + x_p, \qquad (x_1, \ldots, x_p) \in [0,\infty)^p.$$

In the bivariate setting, several summary measures can be used to quantify the dependence structure induced by a copula. For a bivariate copula C, we consider the upper tail dependence coefficient

$$\chi = \lim_{u \uparrow 1} \mathrm{pr}(U_1 > u \mid U_2 > u), \tag{S1}$$

provided the limit exists. Equivalently,

$$\chi = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}.$$

The coefficient χ measures the tendency of the variables to become simultaneously extreme. Its values satisfy $0 \le \chi \le 1$. The case $\chi = 0$ corresponds to asymptotic independence, whereas $\chi > 0$ indicates asymptotic dependence (Joe 2014). For bivariate extreme-value copulas with Pickands dependence function A, the coefficient is given by $\chi = 2\{1 - A(1/2)\}$ (Gudendorf and Segers 2010).

We now present the parametric families of copulas that will be used in the simulation study: the Gumbel–Hougaard, Hüsler–Reiss and Galambos copulas, together with the t-copula. The first three belong to the class of extreme-value copulas but exhibit different forms of extremal dependence, whereas the t-copula does not.

The Gumbel–Hougaard copula (Gumbel 1960, 1961), also known as the logistic extreme-value copula, is defined in dimension p by

$$C_\phi(u_1, \ldots, u_p) = \exp\left\{ -\left(\sum_{j=1}^{p} (-\log u_j)^\phi \right)^{1/\phi} \right\}, \qquad \phi \ge 1. \tag{S2}$$

Its associated stable tail dependence function is

$$\ell_\phi(x_1, \ldots, x_p) = \left(\sum_{j=1}^{p} x_j^\phi \right)^{1/\phi},$$

and the corresponding Pickands dependence function is

$$A_\phi(w_1, \ldots, w_p) = \left(\sum_{j=1}^{p} w_j^\phi \right)^{1/\phi}, \qquad (w_1, \ldots, w_p) \in \Delta_{p-1}.$$

The parameter ϕ controls the strength of extremal dependence. The case $\phi = 1$ corresponds to independence, whereas $\phi \to \infty$ yields complete dependence. In the bivariate case, the upper tail dependence coefficient admits the closed-form expression

$$\chi = 2 - 2^{1/\phi}. \tag{S3}$$

Consequently, larger values of ϕ correspond to stronger extremal dependence.

Another important class of extreme-value copulas is provided by the Hüsler–Reiss family (Hüsler and Reiss 1989). In dimension p, the Hüsler–Reiss copula is parameterized by a variogram matrix

$$\Gamma = (\Gamma_{ij})_{1 \le i,j \le p},$$

where $\Gamma_{ii} = 0$, $\Gamma_{ij} = \Gamma_{ji} \ge 0$, and Γ is conditionally negative definite. The corresponding stable tail dependence function admits the representation

$$\ell(x_1, \ldots, x_p) = \sum_{j=1}^{p} x_j \Phi_{p-1}\left(\eta^{(j)}; \Sigma^{(j)}\right),$$

where Φ_{p-1} denotes the $(p-1)$-variate Gaussian distribution function and the vectors $\eta^{(j)}$ and covariance matrices $\Sigma^{(j)}$ are determined by the variogram matrix Γ (Engelke and Hitz 2020). The variogram controls the strength of extremal dependence between the components. Small values of Γ_{ij} correspond to strong extremal dependence, whereas large values of Γ_{ij} correspond to weak extremal dependence. In the bivariate case, the model reduces to a single variogram parameter $\Gamma_{12} = \Gamma$ and the Pickands dependence function becomes

$$A_\Gamma(t) = (1-t)\,\Phi\left(\frac{\sqrt{\Gamma}}{2} + \frac{1}{\sqrt{\Gamma}} \log\frac{1-t}{t}\right) + t\,\Phi\left(\frac{\sqrt{\Gamma}}{2} + \frac{1}{\sqrt{\Gamma}} \log\frac{t}{1-t}\right),$$

where Φ denotes the standard normal distribution function. Independence is approached as $\Gamma \to \infty$, whereas complete dependence corresponds to $\Gamma \to 0$. The upper tail

dependence coefficient is given by

$$\chi = 2\left(1 - \Phi\left(\frac{\sqrt{\Gamma}}{2}\right)\right). \tag{S4}$$

Consequently, increasing values of Γ lead to weaker extremal dependence.

Another family of extreme-value copulas is the Galambos copula (Galambos 1975). In dimension p, its stable tail dependence function is given by

$$\ell_\theta(x_1, \ldots, x_p) = \sum_{\varnothing \neq I \subseteq \{1,\ldots,p\}} (-1)^{|I|+1} \left(\sum_{i \in I} x_i^{-\theta}\right)^{-1/\theta}, \qquad \theta > 0.$$

Equivalently, the corresponding extreme-value copula is

$$C_\theta(u_1, \ldots, u_p) = \exp\left\{-\ell_\theta\left(-\log u_1, \ldots, -\log u_p\right)\right\}.$$

The parameter θ controls the strength of extremal dependence. Independence is recovered when $\theta \to 0$, whereas complete dependence is approached as $\theta \to \infty$. In the bivariate case, the stable tail dependence function simplifies to

$$\ell_\theta(x, y) = x + y - \left(x^{-\theta} + y^{-\theta}\right)^{-1/\theta},$$

and the corresponding Pickands dependence function is

$$A_\theta(t) = 1 - \left(t^{-\theta} + (1-t)^{-\theta}\right)^{-1/\theta}, \qquad t \in [0, 1].$$

The upper tail dependence coefficient admits the closed-form expression

$$\chi = 2^{-1/\theta}. \tag{S5}$$

Although not belonging to the class of extreme-value copulas, the t-copula provides a flexible model for multivariate dependence with non-trivial tail dependence and is widely used in finance and risk management (Demarta and McNeil 2005). Let $T = (T_1, \ldots, T_p)$ follow a multivariate Student-t distribution with correlation matrix R and $\nu > 0$ degrees of freedom. The associated t-copula is defined by

$$C_{\nu,R}(u_1, \ldots, u_p) = t_{\nu,R}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_p)\right),$$

where t_ν denotes the univariate Student-t distribution function with ν degrees of freedom and $t_{\nu,R}$ the corresponding multivariate distribution function. The dependence structure is governed jointly by the correlation matrix R and the degrees-of-freedom parameter ν. Smaller values of ν produce heavier tails and stronger extremal dependence. As $\nu \to \infty$, the t-copula converges to the Gaussian copula. In the bivariate case, with

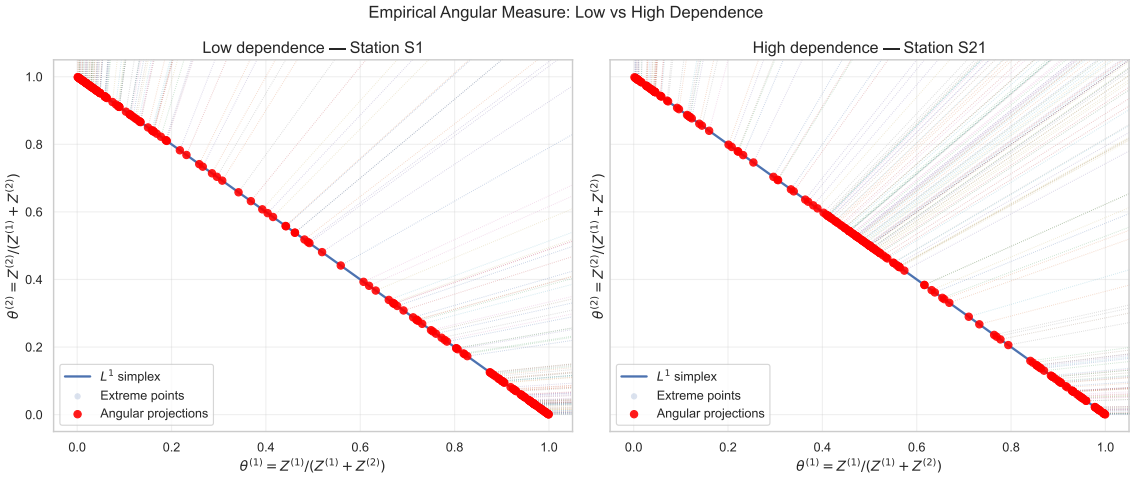


Fig. S1: Illustration of angular observations for a bivariate random vector under two extremal dependence regimes.

correlation parameter ρ, the upper tail dependence coefficient admits the closed-form expression

$$\chi = 2\, t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right), \tag{S6}$$

where $t_{\nu+1}$ denotes the univariate Student-t distribution function with $\nu+1$ degrees of freedom. Consequently, stronger correlation and smaller degrees of freedom lead to stronger extremal dependence.

Unlike the Gumbel–Hougaard, Hüsler–Reiss and Galambos families, the t-copula is not itself an extreme-value copula. Nevertheless, it belongs to the max-domain of attraction of the extremal t distribution (Opitz 2013). As a result, its asymptotic extremal dependence structure is characterized by an extreme-value copula.

S2 Supplementary material for the method section

Figure S1 illustrate the angular observations for a bivariate random vector on the simplex under two extremal dependence regimes.

S3 Consistency of Wasserstein clustering procedures

This appendix provides the proofs of Lemma 1 and Theorem 1.

Proof Lemma 1 If $g(i) = g(l)$, then $G_i = G_l$ by Assumption 3. Consequently,

$$G_i^{(M)} = (\Pi_M)_{\#} G_i = (\Pi_M)_{\#} G_l = G_l^{(M)}$$

for every $M > 0$. Conversely, suppose that $g(i) \neq g(l)$. By Assumption 3 and the one-to-one character of the Aitchison-coordinate map,

$$G_i \neq G_l.$$

Hence, there exists a Borel set $A \subset \mathbb{R}^{p-1}$ such that

$$G_i(A) \neq G_l(A).$$

For $q \geq 1$, define

$$A_q = A \cap \left\{ a \in \mathbb{R}^{p-1} : \|a\| < q \right\}.$$

Since $A_q \uparrow A$ as $q \to \infty$, continuity from below gives

$$G_i(A_q) \longrightarrow G_i(A), \qquad G_l(A_q) \longrightarrow G_l(A).$$

It follows that there exists a finite $q_{il} > 0$ such that

$$G_i(A_{q_{il}}) \neq G_l(A_{q_{il}}).$$

For every $M \geq q_{il}$, the set $A_{q_{il}}$ lies strictly inside $\mathcal{B}_M$. The projection leaves this set unchanged and maps every point outside $\mathcal{B}_M$ to its boundary. Therefore,

$$\Pi_M^{-1}(A_{q_{il}}) = A_{q_{il}},$$

and hence

$$G_i^{(M)}(A_{q_{il}}) = G_i(A_{q_{il}}) \neq G_l(A_{q_{il}}) = G_l^{(M)}(A_{q_{il}}).$$

Thus,

$$G_i^{(M)} \neq G_l^{(M)}$$

for every $M \geq q_{il}$. There are only finitely many between-group pairs. We may therefore choose

$$M_0 = \max_{\substack{1 \leq i,l \leq n \\ g(i) \neq g(l)}} q_{il} < \infty.$$

For every $M \geq M_0$, we then have

$$g(i) = g(l) \quad \Longleftrightarrow \quad G_i^{(M)} = G_l^{(M)}.$$

If $K \geq 2$, the set of between-group pairs is finite and nonempty. Moreover, the projected measures are supported on $\mathcal{B}_M$ and therefore possess finite first moments. Since W_1 is a metric on this space,

$$W_1\left(G_i^{(M)}, G_l^{(M)}\right) > 0$$

for every between-group pair. Consequently,

$$\Delta_M = \min_{g(i) \neq g(l)} W_1\left(G_i^{(M)}, G_l^{(M)}\right) > 0.$$

□

Proof Theorem 1 If $K = 1$, both clustering procedures return the unique partition containing all random vectors, and the result is immediate. We therefore assume throughout the proof that $K \geq 2$ and fix $M \geq M_0$.

Step 1: Wasserstein consistency of the projected empirical measures.

Under Assumptions 1 and 2, together with the sampling conditions stated in the main text, the empirical angular measures satisfy

$$\widehat{H}_{i,T} \rightsquigarrow H_i$$

in probability for every $i \in \{1, \ldots, n\}$; see the proof of Proposition 3.3 of Janßen and Wan (2020). The composition

$$\Pi_M \circ \mathcal{A} : \Delta_{p-1}^{\circ} \longrightarrow \mathcal{B}_M$$

is continuous. The Continuous Mapping Theorem therefore gives

$$\begin{aligned}\widehat{G}_{i,T}^{(M)} &= (\Pi_M)_{\#}\widehat{G}_{i,T} \\ &= (\Pi_M \circ \mathcal{A})_{\#}\widehat{H}_{i,T} \rightsquigarrow (\Pi_M \circ \mathcal{A})_{\#}H_i \\ &= (\Pi_M)_{\#}G_i = G_i^{(M)}\end{aligned}$$

in probability. To verify convergence of the first moments, define

$$f_M(a) = \min\{\|a\|, M\}, \qquad a \in \mathbb{R}^{p-1}.$$

The function f_M is bounded and continuous. Since both projected measures are supported on $\mathcal{B}_M$,

$$\begin{aligned}\int_{\mathbb{R}^{p-1}} \|a\| \, \mathrm{d}\widehat{G}_{i,T}^{(M)}(a) &= \int_{\mathbb{R}^{p-1}} f_M(a) \, \mathrm{d}\widehat{G}_{i,T}^{(M)}(a) \\ &\longrightarrow \int_{\mathbb{R}^{p-1}} f_M(a) \, \mathrm{d}G_i^{(M)}(a) \\ &= \int_{\mathbb{R}^{p-1}} \|a\| \, \mathrm{d}G_i^{(M)}(a)\end{aligned}$$

in probability. Thus, weak convergence and convergence of the first moments hold. By Theorem 6.9 of Villani (2009),

$$W_1\left(\widehat{G}_{i,T}^{(M)}, G_i^{(M)}\right) \longrightarrow 0$$

in probability for every i.

Step 2: Uniform consistency of the empirical dissimilarity matrix.

Define the population projected dissimilarities by

$$d_{il}^{W,M} = W_1\left(G_i^{(M)}, G_l^{(M)}\right).$$

For every $i, l \in \{1, \ldots, n\}$, the triangle inequality gives

$$|\widehat{d}_{il,T}^{W,M} - d_{il}^{W,M}| \leq W_1\left(\widehat{G}_{i,T}^{(M)}, G_i^{(M)}\right) + W_1\left(\widehat{G}_{l,T}^{(M)}, G_l^{(M)}\right).$$

The two terms on the right-hand side converge to zero in probability. Since n is fixed, it follows that

$$\varepsilon_T := \max_{1 \leq i,l \leq n} |\widehat{d}_{il,T}^{W,M} - d_{il}^{W,M}| \longrightarrow 0$$

in probability. By Lemma 1,

$$d_{il}^{W,M} = 0 \quad \text{if } g(i) = g(l),$$

whereas

$$d_{il}^{W,M} \geq \Delta_M \quad \text{if } g(i) \neq g(l).$$

Define

$$A_T = \left\{\varepsilon_T < \frac{\Delta_M}{3n}\right\}.$$

The preceding uniform convergence implies that

$$\mathrm{pr}(A_T) \longrightarrow 1.$$

On A_T, every within-group empirical dissimilarity satisfies

$$\widehat{d}_{il,T}^{W,M} \leq \varepsilon_T < \frac{\Delta_M}{3n}, \qquad g(i) = g(l),$$

whereas every between-group empirical dissimilarity satisfies

$$\widehat{d}_{il,T}^{W,M} \geq \Delta_M - \varepsilon_T, \qquad g(i) \neq g(l).$$

Since $\varepsilon_T < \Delta_M/2$ on A_T, every within-group empirical dissimilarity is strictly smaller than every between-group empirical dissimilarity.

Step 3: Consistency of hierarchical clustering.

Initially, every cluster is a singleton and is therefore contained in one true latent group. Suppose that, at some stage of the hierarchical procedure, every cluster formed so far remains contained in a single true group.

Let C and C' be two current clusters. If they are contained in the same true group, all pairwise dissimilarities between their elements are bounded above by ε_T. Their single-, complete-, and average-linkage dissimilarities are therefore also bounded above by ε_T.

If C and C' are contained in different true groups, all pairwise dissimilarities between their elements are bounded below by $\Delta_M - \varepsilon_T$. Their single-, complete-, and average-linkage dissimilarities are therefore also bounded below by $\Delta_M - \varepsilon_T$.

As long as more than K clusters remain, at least one true group is split across two or more current clusters. Hence, at least one within-group merge is available. On A_T,

$$\varepsilon_T < \Delta_M - \varepsilon_T,$$

so every available within-group merge has a smaller linkage dissimilarity than every between-group merge. The next merge must therefore occur within a true group.

By induction, no between-group merge occurs before every true group has been merged into a single cluster. Consequently, when the dendrogram is cut at exactly K clusters, the resulting partition is $\mathcal{P}$.

Step 4: Consistency of the global K-medoids solution.

Write the true partition as

$$\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_K\}, \qquad \mathcal{P}_k = \{i \in \{1, \ldots, n\} : g(i) = k\}.$$

For any set $\mathcal{S} \subseteq \{1, \ldots, n\}$ containing exactly K medoid indices, define the population and empirical K-medoids objectives by

$$\Phi^{(M)}(\mathcal{S}) = \sum_{i=1}^{n} \min_{s \in \mathcal{S}} d_{is}^{W,M}$$

and

$$\widehat{\Phi}_T^{(M)}(\mathcal{S}) = \sum_{i=1}^{n} \min_{s \in \mathcal{S}} \widehat{d}_{is,T}^{W,M}.$$

Suppose first that $\mathcal{S}$ contains exactly one index from each true group. Every $i \in \{1, \ldots, n\}$ then has a medoid $s \in \mathcal{S}$ belonging to the same true group. Consequently,

$$\Phi^{(M)}(\mathcal{S}) = 0.$$

Conversely, suppose that $\mathcal{S}$ does not contain an index from every true group. Since $\mathcal{S}$ contains K indices and there are K true groups, at least one true group, say $\mathcal{P}_k$, is not represented. For every $i \in \mathcal{P}_k$ and every $s \in \mathcal{S}$,

$$d_{is}^{W,M} \geq \Delta_M.$$

Since $\mathcal{P}_k$ is nonempty, this implies

$$\Phi^{(M)}(\mathcal{S}) \geq \Delta_M.$$

Thus, the population objective is zero precisely for those medoid sets containing one representative from each true group. Although the medoid set need not be unique, every population minimizer represents all K groups. For every admissible medoid set $\mathcal{S}$,

$$\begin{aligned}|\widehat{\Phi}_T^{(M)}(\mathcal{S}) - \Phi^{(M)}(\mathcal{S})| &\leq \sum_{i=1}^{n} |\min_{s\in\mathcal{S}} \widehat{d}_{is,T}^{W,M} - \min_{s\in\mathcal{S}} d_{is}^{W,M}| \\ &\leq n\varepsilon_T.\end{aligned}$$

On A_T, we have $n\varepsilon_T < \Delta_M/3$. Hence, if $\mathcal{S}$ contains one medoid from each true group,

$$\widehat{\Phi}_T^{(M)}(\mathcal{S}) < \frac{\Delta_M}{3}.$$

If $\mathcal{S}$ omits at least one true group, then

$$\widehat{\Phi}_T^{(M)}(\mathcal{S}) > \Delta_M - \frac{\Delta_M}{3} = \frac{2\Delta_M}{3}.$$

Therefore, on A_T, every global minimizer $\widehat{\mathcal{S}}_T^{(M)}$ of the empirical K-medoids objective contains exactly one medoid from each true group.

It remains to consider the assignment to the selected medoids. Let $i \in \mathcal{P}_k$, and let s_k denote the selected medoid belonging to $\mathcal{P}_k$. Then

$$\widehat{d}_{is_k,T}^{W,M} \leq \varepsilon_T.$$

For every selected medoid $s \notin \mathcal{P}_k$,

$$\widehat{d}_{is,T}^{W,M} \geq \Delta_M - \varepsilon_T.$$

On A_T, $\varepsilon_T < \Delta_M - \varepsilon_T$. Thus, every random vector is strictly closer to the selected medoid from its own true group than to any medoid from another group. The nearest-medoid assignment therefore recovers $\mathcal{P}$ exactly.

We have shown that, on A_T, both clustering procedures appearing in Theorem 1 recover the true partition:

$$\widehat{\mathcal{P}}_T^{(M)} = \mathcal{P}.$$

Consequently,

$$\text{pr}\left(\widehat{\mathcal{P}}_T^{(M)} = \mathcal{P}\right) \geq \text{pr}(A_T).$$

Since $\text{pr}(A_T) \to 1$, it follows that

$$\text{pr}\left(\widehat{\mathcal{P}}_T^{(M)} = \mathcal{P}\right) \longrightarrow 1, \qquad T \to \infty.$$

□

S4 Supplementary for the simulation studies

S4.1 Competing methods

Vignotto et al. (2021) proposed a clustering methodology for compound extremes based on a dissimilarity measure derived from the Kullback–Leibler divergence. The methodology first transforms each marginal distribution to the standard Pareto scale as in Equation (1). To identify extreme regions, they introduce a risk function $r : \mathbb{R}^2 \to \mathbb{R}$, applied to the Pareto-transformed observations. Typical choices include

$$r(z) = \max\left(z^{(1)}, z^{(2)}\right), \qquad r(z) = z^{(1)} + z^{(2)}.$$

For each random vector i, they define $Z_{i,t} = \left(Z_{i,t}^{(1)}, Z_{i,t}^{(2)}\right)$ and the associated univariate random variables $R_{i,t} = r(Z_{i,t})$. A point is considered extreme whenever $R_{i,t} > q_{u,i}$, where $q_{u,i}$ denotes a high quantile associated with an exceedance probability $u \in (0,1)$. The corresponding tail region is therefore $\{R_{i,t} > q_{u,i}\}$. They proposed to partition the extreme region into W disjoint subsets $A_1, \ldots, A_W$. The empirical proportion of extreme observations falling into region A_w is estimated by

$$\widehat{p}_{i,w} = \frac{\sum_{t=1}^{T} \mathbb{1}\{Z_{i,t} \in A_w\}}{\sum_{t=1}^{T} \mathbb{1}\{R_{i,t} > q_{u,i}\}}, \qquad w = 1, \ldots, W.$$

The vector $(\widehat{p}_{i,1}, \ldots, \widehat{p}_{i,W})$ can therefore be interpreted as an empirical multinomial distribution describing how extremes are distributed across the tail regions. The dissimilarity between the extremal dependence structures of two random vectors i and l is then quantified using the symmetrized Kullback–Leibler divergence (van Erven and Harremoës 2014),

$$\widehat{d}_{il}^{\mathrm{KL}} = \frac{1}{2} \sum_{w=1}^{W} (\widehat{p}_{i,w} - \widehat{p}_{l,w}) \log\left(\frac{\widehat{p}_{i,w}}{\widehat{p}_{l,w}}\right).$$

A small value of $\widehat{d}_{il}^{\mathrm{KL}}$ indicates that both random vectors exhibit similar extremal dependence structures, whereas large values indicate different tail behaviours. Finally, the pairwise divergences are assembled into a dissimilarity matrix

$$\widehat{D}_T^{\mathrm{KL}} = \left(\widehat{d}_{il}^{\mathrm{KL}}\right)_{i,l=1,\ldots,n} \in \mathbb{R}^{n \times n}, \tag{S7}$$

which is then used as input for a clustering algorithm. The parameter W controls the resolution of the tail partition. Large values of W allow finer differences in the extremal dependence structure to be detected, but may lead to unstable probability estimates when too few observations fall into each subset. The choice of W therefore induces a classical bias–variance trade-off.

We consider another competing approach, conditional extremes clustering, which was introduced more recently by O'Toole et al. (2025). The method is based on the conditional extremes framework (Heffernan and Tawn 2004), which models the distribution of a random vector conditional on one of its components being extreme. Separate conditional extremes models are fitted for each random vector, yielding parameters that characterize its extremal dependence structure. To compare extremal dependence across random vectors, O'Toole et al. (2025) introduce a dissimilarity measure based on the expected skew-geometric Jensen–Shannon divergence (Menéndez et al. 1997) between the fitted conditional distributions. Under a Gaussian working assumption for the conditional extremes residuals, this divergence admits a closed-form expression, allowing efficient computation of pairwise dissimilarities. For multivariate observations, the procedure is repeated for each conditioning component $j \in \{1, \ldots, p\}$. Let $\widehat{D}_T^{(j),\mathrm{CE}}$ denote the dissimilarity matrix obtained by fitting the conditional extremes

model with component j as the conditioning variable. The final dissimilarity matrix is then constructed by averaging over all conditioning components,

$$\widehat{D}_T^{\mathrm{CE}} = \frac{1}{p} \sum_{j=1}^{p} \widehat{D}_T^{(j),\mathrm{CE}}. \tag{S8}$$

This aggregation yields a single dissimilarity matrix that incorporates information from all possible conditioning variables.

S4.2 Evaluation metric

To evaluate the performance of the clustering in the simulation study, we use the Adjusted Rand Index (ARI), which measures the agreement between the true partition and the partition estimated by the clustering algorithm (Hubert and Arabie 1985). The ARI is a corrected version of the Rand Index that accounts for agreement occurring by chance. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ denote the true partition of the random vectors, and let $\widehat{\mathcal{P}} = \left\{\widehat{P}_1, \ldots, \widehat{P}_{\widehat{K}}\right\}$ be the partition estimated by the clustering procedure. The ARI is based on all pairwise comparisons between random vectors. More precisely, it compares whether two random vectors belong to the same cluster or to different clusters under both partitions. Let $n_{k\ell} = |P_k \cap \widehat{P}_\ell|$ denote the contingency table associated with the two partitions, and define

$$a_k = \sum_{\ell=1}^{\widehat{K}} n_{k\ell}, \qquad b_\ell = \sum_{k=1}^{K} n_{k\ell}.$$

The Adjusted Rand Index is then defined by (Hubert and Arabie 1985; Steinley 2004)

$$\mathrm{ARI} = \frac{\displaystyle\sum_{k,\ell} \binom{n_{k\ell}}{2} - \frac{\displaystyle\sum_k \binom{a_k}{2} \sum_\ell \binom{b_\ell}{2}}{\binom{n}{2}}}{\displaystyle\frac{1}{2}\left[\sum_k \binom{a_k}{2} + \sum_\ell \binom{b_\ell}{2}\right] - \frac{\displaystyle\sum_k \binom{a_k}{2} \sum_\ell \binom{b_\ell}{2}}{\binom{n}{2}}}. \tag{S9}$$

The ARI takes values in $[-1, 1]$. A value close to 1 indicates perfect agreement between the true and estimated partitions, whereas values close to 0 correspond to a clustering performance similar to random assignment. Negative values indicate agreement worse than random clustering.

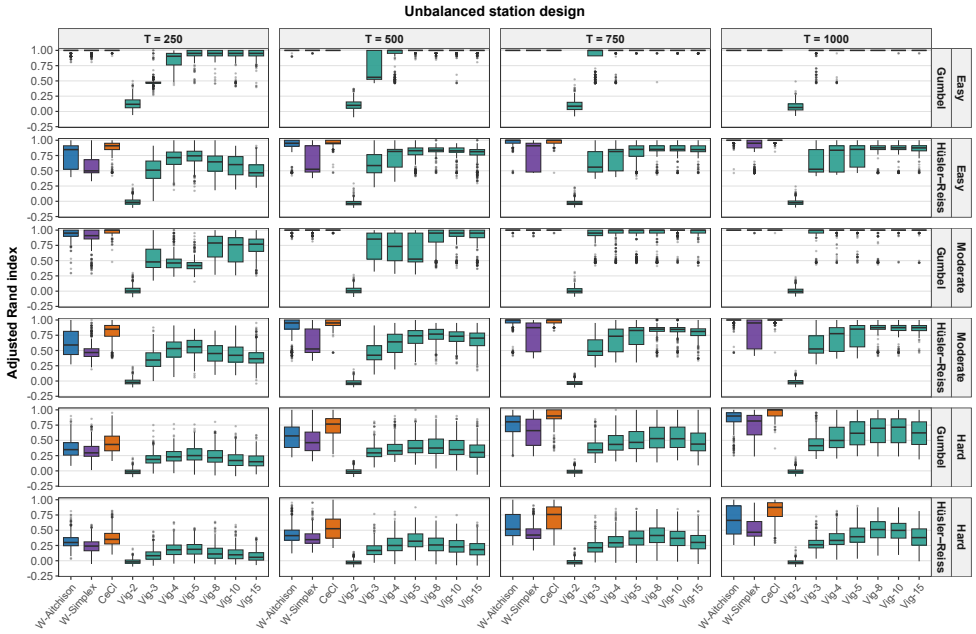


Fig. S2: Adjusted Rand index for the unbalanced design $N = (5, 10, 20)$, with $K = 3$ latent groups, for sample sizes $T \in \{250, 500, 750, 1\,000\}$ and the three separation levels easy, moderate and hard, corresponding to $\chi = (0.1, 0.5, 0.9)$, $(0.2, 0.5, 0.8)$ and $(0.3, 0.45, 0.6)$, under the Gumbel–Hougaard and Hüsler–Reiss copulas. The Wasserstein–Aitchison, Wasserstein–simplex and conditional extremes dissimilarities are compared with the partition-based dissimilarity of Vignotto et al. (2021), the latter for $W \in \{2, 3, 4, 5, 8, 10, 15\}$ partitions. Results are based on 500 Monte Carlo replications.

S4.3 Detailed Results for Simulation Study 1

Table S1 reports the corresponding mean values and Monte Carlo standard errors. Table S2 reports the percentage of replications in which the latent partition is recovered exactly.

Table S1: Mean adjusted Rand index, with Monte Carlo standard errors in parentheses, based on 500 replications. WA and WS denote Wasserstein–Aitchison and Wasserstein–simplex. The Vignotto column is an oracle comparison obtained by selecting the value of W with the largest mean adjusted Rand index separately within each simulation setting.

Design	Copula	Scenario	T	WA	WS	CeCl	Vignotto
Balanced	Gumbel	Easy	250	0.978 (0.002)	0.978 (0.002)	0.996 (0.001)	0.906 (0.005) [$W = 5$]
Balanced	Gumbel	Easy	500	0.999 (0.000)	0.998 (0.001)	1.000 (0.000)	0.991 (0.001) [$W = 5$]
Balanced	Gumbel	Easy	750	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.999 (0.001) [$W = 8$]
Balanced	Gumbel	Easy	1 000	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000) [$W = 5$]
Balanced	Gumbel	Moderate	250	0.880 (0.005)	0.880 (0.005)	0.950 (0.003)	0.667 (0.006) [$W = 8$]

Table S1: Mean adjusted Rand index, with Monte Carlo standard errors in parentheses, based on 500 replications. WA and WS denote Wasserstein–Aitchison and Wasserstein–simplex. The Vignotto column is an oracle comparison obtained by selecting the value of W with the largest mean adjusted Rand index separately within each simulation setting. *(continued)*

Design	Copula	Scenario	T	WA	WS	CeCl	Vignotto
Balanced	Gumbel	Moderate	500	0.983 (0.002)	0.980 (0.002)	0.996 (0.001)	0.941 (0.004) [$W = 8$]
Balanced	Gumbel	Moderate	750	0.997 (0.001)	0.997 (0.001)	0.999 (0.000)	0.985 (0.002) [$W = 10$]
Balanced	Gumbel	Moderate	1 000	1.000 (0.000)	0.999 (0.000)	1.000 (0.000)	0.999 (0.001) [$W = 10$]
Balanced	Gumbel	Hard	250	0.416 (0.006)	0.396 (0.006)	0.510 (0.006)	0.278 (0.005) [$W = 5$]
Balanced	Gumbel	Hard	500	0.635 (0.006)	0.597 (0.006)	0.744 (0.006)	0.490 (0.006) [$W = 5$]
Balanced	Gumbel	Hard	750	0.774 (0.006)	0.753 (0.006)	0.873 (0.005)	0.618 (0.007) [$W = 8$]
Balanced	Gumbel	Hard	1 000	0.862 (0.005)	0.835 (0.005)	0.934 (0.004)	0.743 (0.007) [$W = 8$]
Balanced	Hüsler–Reiss	Easy	250	0.751 (0.005)	0.660 (0.005)	0.755 (0.005)	0.494 (0.004) [$W = 5$]
Balanced	Hüsler–Reiss	Easy	500	0.896 (0.004)	0.809 (0.005)	0.898 (0.004)	0.604 (0.005) [$W = 5$]
Balanced	Hüsler–Reiss	Easy	750	0.954 (0.003)	0.884 (0.005)	0.957 (0.003)	0.678 (0.006) [$W = 5$]
Balanced	Hüsler–Reiss	Easy	1 000	0.978 (0.002)	0.938 (0.004)	0.982 (0.002)	0.748 (0.006) [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	250	0.705 (0.006)	0.601 (0.006)	0.734 (0.006)	0.379 (0.005) [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	500	0.890 (0.004)	0.786 (0.006)	0.896 (0.004)	0.544 (0.005) [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	750	0.944 (0.003)	0.888 (0.005)	0.954 (0.003)	0.658 (0.006) [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	1 000	0.983 (0.002)	0.932 (0.004)	0.982 (0.002)	0.743 (0.006) [$W = 5$]
Balanced	Hüsler–Reiss	Hard	250	0.365 (0.005)	0.282 (0.005)	0.395 (0.005)	0.117 (0.005) [$W = 8$]
Balanced	Hüsler–Reiss	Hard	500	0.551 (0.006)	0.442 (0.006)	0.583 (0.006)	0.226 (0.005) [$W = 5$]
Balanced	Hüsler–Reiss	Hard	750	0.696 (0.006)	0.574 (0.006)	0.728 (0.006)	0.326 (0.005) [$W = 5$]
Balanced	Hüsler–Reiss	Hard	1 000	0.785 (0.006)	0.663 (0.007)	0.816 (0.006)	0.414 (0.005) [$W = 5$]
Unbalanced	Gumbel	Easy	250	0.986 (0.002)	0.985 (0.001)	0.999 (0.000)	0.940 (0.003) [$W = 8$]
Unbalanced	Gumbel	Easy	500	0.999 (0.000)	0.999 (0.000)	1.000 (0.000)	0.996 (0.001) [$W = 5$]
Unbalanced	Gumbel	Easy	750	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.999 (0.000) [$W = 5$]
Unbalanced	Gumbel	Easy	1 000	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000) [$W = 5$]
Unbalanced	Gumbel	Moderate	250	0.924 (0.004)	0.881 (0.007)	0.974 (0.002)	0.734 (0.008) [$W = 8$]
Unbalanced	Gumbel	Moderate	500	0.991 (0.001)	0.990 (0.001)	0.999 (0.000)	0.896 (0.007) [$W = 10$]
Unbalanced	Gumbel	Moderate	750	0.999 (0.000)	0.999 (0.000)	1.000 (0.000)	0.974 (0.004) [$W = 10$]
Unbalanced	Gumbel	Moderate	1 000	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.995 (0.002) [$W = 10$]
Unbalanced	Gumbel	Hard	250	0.376 (0.007)	0.327 (0.006)	0.464 (0.007)	0.271 (0.006) [$W = 5$]
Unbalanced	Gumbel	Hard	500	0.566 (0.009)	0.493 (0.009)	0.730 (0.008)	0.411 (0.007) [$W = 8$]
Unbalanced	Gumbel	Hard	750	0.739 (0.009)	0.634 (0.010)	0.887 (0.006)	0.550 (0.009) [$W = 10$]
Unbalanced	Gumbel	Hard	1 000	0.848 (0.007)	0.752 (0.010)	0.946 (0.003)	0.666 (0.010) [$W = 8$]
Unbalanced	Hüsler–Reiss	Easy	250	0.770 (0.009)	0.581 (0.007)	0.875 (0.005)	0.725 (0.006) [$W = 5$]
Unbalanced	Hüsler–Reiss	Easy	500	0.899 (0.007)	0.659 (0.009)	0.954 (0.003)	0.834 (0.004) [$W = 8$]
Unbalanced	Hüsler–Reiss	Easy	750	0.964 (0.003)	0.759 (0.010)	0.979 (0.002)	0.850 (0.004) [$W = 10$]
Unbalanced	Hüsler–Reiss	Easy	1 000	0.986 (0.002)	0.860 (0.009)	0.992 (0.001)	0.866 (0.004) [$W = 15$]
Unbalanced	Hüsler–Reiss	Moderate	250	0.627 (0.009)	0.495 (0.006)	0.798 (0.008)	0.555 (0.007) [$W = 5$]
Unbalanced	Hüsler–Reiss	Moderate	500	0.858 (0.009)	0.633 (0.009)	0.945 (0.003)	0.749 (0.005) [$W = 8$]
Unbalanced	Hüsler–Reiss	Moderate	750	0.933 (0.006)	0.747 (0.010)	0.980 (0.002)	0.831 (0.005) [$W = 8$]
Unbalanced	Hüsler–Reiss	Moderate	1 000	0.981 (0.003)	0.831 (0.010)	0.991 (0.001)	0.865 (0.003) [$W = 15$]
Unbalanced	Hüsler–Reiss	Hard	250	0.321 (0.005)	0.240 (0.005)	0.377 (0.006)	0.193 (0.005) [$W = 5$]
Unbalanced	Hüsler–Reiss	Hard	500	0.447 (0.007)	0.377 (0.006)	0.542 (0.008)	0.335 (0.007) [$W = 5$]
Unbalanced	Hüsler–Reiss	Hard	750	0.574 (0.009)	0.456 (0.006)	0.706 (0.009)	0.424 (0.007) [$W = 8$]
Unbalanced	Hüsler–Reiss	Hard	1 000	0.662 (0.010)	0.524 (0.008)	0.802 (0.009)	0.507 (0.008) [$W = 8$]

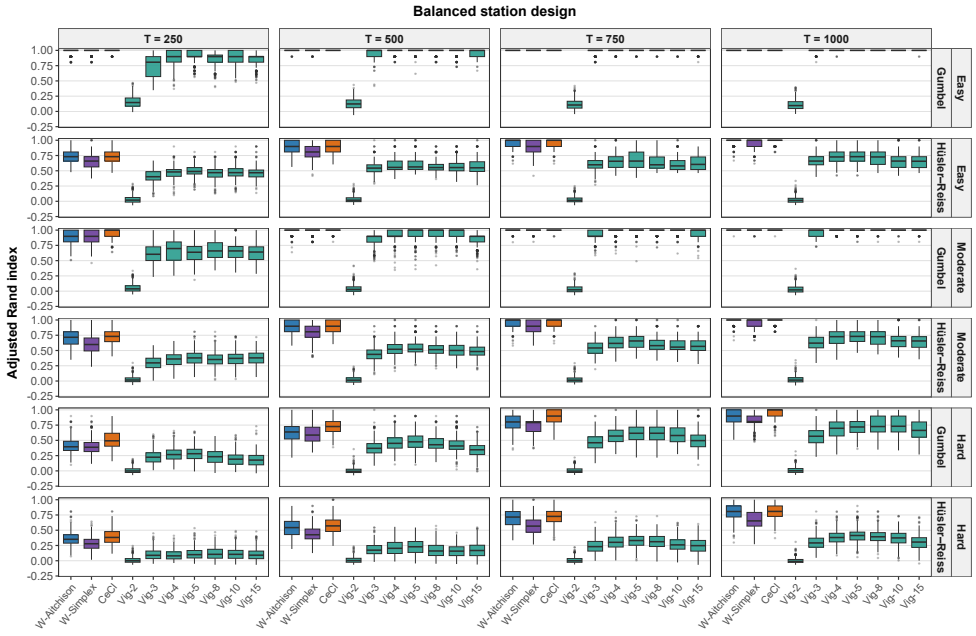


Fig. S3: Adjusted Rand Index (ARI) for the balanced design $N = (10, 10, 10)$, with $K = 3$ latent groups, for sample sizes $T \in \{250, 500, 750, 1\,000\}$ and the three separation levels easy, moderate and hard ($\chi = (0.1, 0.5, 0.9)$, $(0.2, 0.5, 0.8)$ and $(0.3, 0.45, 0.6)$), under the Gumbel–Hougaard and Hüsler–Reiss copulas. The Wasserstein–Aitchison, Wasserstein–simplex and CeCl dissimilarities are compared with the partition-based dissimilarity of Vignotto et al. (2021), the latter for $W \in \{2, 3, 4, 5, 8, 10, 15\}$ partitions. Results are based on 500 Monte Carlo replications.

Table S2: Percentage of the 500 replications achieving exact recovery, ARI = 1. WA and WS denote Wasserstein–Aitchison and Wasserstein–simplex. For Vignotto, W is selected using the mean adjusted Rand index reported in Table S1.

Design	Copula	Scenario	T	WA	WS	CeCl	Vignotto
Balanced	Gumbel	Easy	250	80.0%	79.6%	96.0%	41.8% [$W = 5$]
Balanced	Gumbel	Easy	500	98.8%	97.8%	100.0%	92.2% [$W = 5$]
Balanced	Gumbel	Easy	750	100.0%	100.0%	100.0%	98.6% [$W = 8$]
Balanced	Gumbel	Easy	1 000	100.0%	100.0%	100.0%	100.0% [$W = 5$]
Balanced	Gumbel	Moderate	250	32.2%	30.2%	61.8%	3.4% [$W = 8$]
Balanced	Gumbel	Moderate	500	85.2%	83.0%	96.0%	58.4% [$W = 8$]
Balanced	Gumbel	Moderate	750	97.2%	96.8%	99.4%	87.2% [$W = 10$]
Balanced	Gumbel	Moderate	1 000	99.8%	99.0%	99.8%	99.0% [$W = 10$]
Balanced	Gumbel	Hard	250	0.0%	0.0%	0.0%	0.0% [$W = 5$]
Balanced	Gumbel	Hard	500	0.6%	1.0%	4.4%	0.0% [$W = 5$]
Balanced	Gumbel	Hard	750	7.8%	6.2%	28.4%	1.4% [$W = 8$]
Balanced	Gumbel	Hard	1 000	26.2%	18.0%	52.6%	7.8% [$W = 8$]
Balanced	Hüsler–Reiss	Easy	250	3.6%	0.8%	4.4%	0.0% [$W = 5$]
Balanced	Hüsler–Reiss	Easy	500	34.6%	12.2%	36.8%	1.0% [$W = 5$]

Table S2: Percentage of the 500 replications achieving exact recovery, ARI = 1. WA and WS denote Wasserstein–Aitchison and Wasserstein–simplex. For Vignotto, W is selected using the mean adjusted Rand index reported in Table S1. *(continued)*

Design	Copula	Scenario	T	WA	WS	CeCl	Vignotto
Balanced	Hüsler–Reiss	Easy	750	64.4%	31.4%	67.2%	2.0% [$W = 5$]
Balanced	Hüsler–Reiss	Easy	1 000	81.4%	54.6%	82.8%	6.8% [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	250	2.6%	0.2%	4.0%	0.0% [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	500	31.4%	9.4%	36.2%	0.4% [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	750	57.6%	34.4%	64.8%	1.2% [$W = 5$]
Balanced	Hüsler–Reiss	Moderate	1 000	85.6%	52.2%	83.8%	6.6% [$W = 5$]
Balanced	Hüsler–Reiss	Hard	250	0.0%	0.0%	0.0%	0.0% [$W = 8$]
Balanced	Hüsler–Reiss	Hard	500	0.0%	0.0%	0.8%	0.0% [$W = 5$]
Balanced	Hüsler–Reiss	Hard	750	3.0%	0.6%	5.0%	0.0% [$W = 5$]
Balanced	Hüsler–Reiss	Hard	1 000	10.0%	2.4%	15.2%	0.0% [$W = 5$]
Unbalanced	Gumbel	Easy	250	83.2%	77.0%	97.2%	32.0% [$W = 8$]
Unbalanced	Gumbel	Easy	500	98.8%	98.8%	100.0%	92.6% [$W = 5$]
Unbalanced	Gumbel	Easy	750	100.0%	100.0%	100.0%	97.6% [$W = 5$]
Unbalanced	Gumbel	Easy	1 000	100.0%	100.0%	100.0%	100.0% [$W = 5$]
Unbalanced	Gumbel	Moderate	250	36.8%	27.4%	67.6%	3.0% [$W = 8$]
Unbalanced	Gumbel	Moderate	500	88.2%	85.0%	97.4%	40.6% [$W = 10$]
Unbalanced	Gumbel	Moderate	750	98.0%	97.8%	99.4%	80.2% [$W = 10$]
Unbalanced	Gumbel	Moderate	1 000	100.0%	99.2%	99.8%	96.8% [$W = 10$]
Unbalanced	Gumbel	Hard	250	0.0%	0.0%	0.0%	0.0% [$W = 5$]
Unbalanced	Gumbel	Hard	500	1.0%	0.4%	5.6%	0.0% [$W = 8$]
Unbalanced	Gumbel	Hard	750	7.6%	2.8%	27.8%	0.4% [$W = 10$]
Unbalanced	Gumbel	Hard	1 000	22.2%	15.4%	50.8%	4.6% [$W = 8$]
Unbalanced	Hüsler–Reiss	Easy	250	10.0%	0.4%	13.2%	0.2% [$W = 5$]
Unbalanced	Hüsler–Reiss	Easy	500	33.4%	6.2%	43.8%	0.6% [$W = 8$]
Unbalanced	Hüsler–Reiss	Easy	750	59.4%	22.4%	69.0%	2.2% [$W = 10$]
Unbalanced	Hüsler–Reiss	Easy	1 000	80.4%	44.4%	84.6%	5.4% [$W = 15$]
Unbalanced	Hüsler–Reiss	Moderate	250	4.4%	0.2%	8.0%	0.0% [$W = 5$]
Unbalanced	Hüsler–Reiss	Moderate	500	31.0%	5.2%	40.6%	0.0% [$W = 8$]
Unbalanced	Hüsler–Reiss	Moderate	750	52.8%	20.8%	69.0%	2.2% [$W = 8$]
Unbalanced	Hüsler–Reiss	Moderate	1 000	80.2%	41.4%	85.4%	4.8% [$W = 15$]
Unbalanced	Hüsler–Reiss	Hard	250	0.0%	0.0%	0.0%	0.0% [$W = 5$]
Unbalanced	Hüsler–Reiss	Hard	500	0.0%	0.0%	0.8%	0.0% [$W = 5$]
Unbalanced	Hüsler–Reiss	Hard	750	1.6%	0.0%	6.4%	0.0% [$W = 8$]
Unbalanced	Hüsler–Reiss	Hard	1 000	7.8%	0.0%	15.4%	0.0% [$W = 8$]

S4.4 Detailed Results for Simulation Study 2

Table S3 reports the mean adjusted Rand indices and their Monte Carlo standard errors. Table S4 reports the percentage of replications in which the latent partition is recovered exactly.

S4.5 Detailed Results for Simulation Study 3

Table S5 reports the mean adjusted Rand indices and their Monte Carlo standard errors. Table S6 reports the percentage of replications in which the latent partition is recovered exactly.

Table S3: Mean adjusted Rand index for Simulation Study 2, with Monte Carlo standard errors in parentheses, based on up to 500 replications per setting.

χ	T	Wasserstein–Aitchison	Conditional extremes
0.1	10 000	0.038 (0.003)	0.024 (0.003)
0.1	100 000	0.397 (0.005)	0.279 (0.005)
0.3	10 000	0.210 (0.005)	0.140 (0.005)
0.3	100 000	0.957 (0.003)	0.384 (0.013)
0.5	10 000	0.367 (0.006)	0.217 (0.005)
0.5	100 000	0.994 (0.001)	0.027 (0.003)
0.7	10 000	0.394 (0.006)	0.084 (0.005)
0.7	100 000	0.990 (0.002)	0.004 (0.001)
0.9	10 000	0.299 (0.005)	0.223 (0.005)
0.9	100 000	0.790 (0.007)	0.411 (0.009)

Table S4: Percentage of replications achieving exact recovery, ARI = 1, in Simulation Study 2.

χ	T	Wasserstein–Aitchison	Conditional extremes
0.1	10 000	0.0%	0.0%
0.1	100 000	0.0%	0.0%
0.3	10 000	0.0%	0.0%
0.3	100 000	66.8%	2.2%
0.5	10 000	0.0%	0.0%
0.5	100 000	95.2%	0.0%
0.7	10 000	0.0%	0.0%
0.7	100 000	92.0%	0.0%
0.9	10 000	0.0%	0.0%
0.9	100 000	21.6%	0.2%

Table S5: Mean adjusted Rand index for Simulation Study 3, with Monte Carlo standard errors in parentheses, based on 500 independent replications for each value of δ.

δ	Wasserstein–Aitchison	Wasserstein–simplex	Conditional extremes
0.20	0.069 (0.006)	0.085 (0.007)	0.055 (0.006)
0.40	0.336 (0.009)	0.351 (0.009)	0.229 (0.009)
0.60	0.598 (0.011)	0.624 (0.011)	0.466 (0.011)
0.80	0.867 (0.008)	0.881 (0.007)	0.745 (0.010)
1.00	0.977 (0.003)	0.981 (0.003)	0.939 (0.005)

Table S6: Percentage of the 500 replications achieving exact recovery, ARI = 1, in Simulation Study 3.

δ	Wasserstein–Aitchison	Wasserstein–simplex	Conditional extremes
0.20	0.0%	0.0%	0.0%
0.40	0.8%	1.0%	0.0%
0.60	15.6%	18.2%	6.2%
0.80	59.2%	62.8%	35.0%
1.00	91.4%	93.0%	78.8%

S5 Construction of the trivariate angular measures

This Section describes in detail the construction of the three trivariate angular measures used in the trivariate simulation study, together with the generation of the corresponding observations.

Bernstein–Dirichlet representation

Let $\Delta_2^\circ = \{\theta = (\theta_1, \theta_2, \theta_3) \in (0,1)^3 : \theta_1 + \theta_2 + \theta_3 = 1\}$ denote the interior of the two-dimensional simplex.

Finite mixtures of Dirichlet densities provide a flexible representation of angular measures in arbitrary dimensions (Boldi and Davison 2007; Sabourin and Naveau 2014). Our construction uses the connection between integer-valued Dirichlet densities and multivariate Bernstein polynomial bases (Marcon et al. 2016, 2017). We fix an integer order $J = 18$ and define

$$\mathcal{I}_J = \{\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}^3 : \alpha_j \geq 1 \ (j = 1, 2, 3),\ \alpha_1 + \alpha_2 + \alpha_3 = J\}.$$

Since every component of α is strictly positive, all Dirichlet components are supported on the open simplex. The number of basis components is $|\mathcal{I}_J| = \binom{J-1}{2} = 136$. For each $\alpha \in \mathcal{I}_J$, define

$$b_\alpha(\theta) = \frac{\Gamma(J)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \prod_{j=1}^{3} \theta_j^{\alpha_j - 1} \qquad (\theta \in \Delta_2^\circ),$$

which is the density of a Dirichlet distribution with parameter α. Given non-negative weights $q = (q_\alpha)_{\alpha \in \mathcal{I}_J}$ with $\sum_{\alpha \in \mathcal{I}_J} q_\alpha = 1$, we define the angular density

$$h_q(\theta) = \sum_{\alpha \in \mathcal{I}_J} q_\alpha b_\alpha(\theta).$$

The corresponding angular probability measure is denoted by H_q.

Angular moment constraints

Under the probability normalization of the angular measure, a valid three-dimensional angular measure must satisfy

$$\int_{\Delta_2^\circ} \theta_j \, \mathrm{d}H_q(\theta) = \frac{1}{3} \qquad (j = 1, 2, 3).$$

For a Dirichlet random vector with parameter α, $\mathbb{E}_\alpha(\Theta_j) = \alpha_j/J$. The moment constraints can therefore be written as linear constraints on the mixture weights:

$$\sum_{\alpha \in \mathcal{I}_J} q_\alpha \frac{\alpha_j}{J} = \frac{1}{3} \qquad (j = 1, 2, 3).$$

Since $\Theta_1 + \Theta_2 + \Theta_3 = 1$, the third constraint is implied by the first two together with the unit-mass constraint. Consequently, only the constraints for $j = 1, 2$ are supplied to the numerical linear program.

Fixing the pairwise tail-dependence coefficients

Let $\Theta = (\Theta_1, \Theta_2, \Theta_3) \sim H_q$. Under the probability normalization adopted here, H_q has total mass one and satisfies $\mathbb{E}_{H_q}(\Theta_j) = 1/3$ $(j = 1, 2, 3)$. The corresponding spectral measure is therefore $3H_q$. It follows that the pairwise tail-dependence coefficient is

$$\chi_{ij} = 3 \int_{\Delta_2^\circ} \min(\theta_i, \theta_j) \, \mathrm{d}H_q(\theta) = 3\mathbb{E}_{H_q}\{\min(\Theta_i, \Theta_j)\}.$$

Recall that H_q is a mixture of Dirichlet distributions with density $h_q(\theta) = \sum_{\alpha \in \mathcal{I}_J} q_\alpha b_\alpha(\theta)$.To calculate the expectation under this mixture, we first calculate it under each Dirichlet component. For every $\alpha \in \mathcal{I}_J$, let $\Theta^{(\alpha)} = (\Theta_1^{(\alpha)}, \Theta_2^{(\alpha)}, \Theta_3^{(\alpha)}) \sim$ Dirichlet(α), and define

$$m_{ij}(\alpha) = \mathbb{E}\Big[\min\Big(\Theta_i^{(\alpha)}, \Theta_j^{(\alpha)}\Big)\Big].$$

This expectation is available analytically from standard properties of the Dirichlet distribution (Kotz et al., 2000, Ch. 49). Let $I_x(a, b) = \mathrm{B}_x(a, b)/\mathrm{B}(a, b)$ denote the regularized incomplete beta function, where $\mathrm{B}_x(a, b) = \int_0^x u^{a-1}(1-u)^{b-1} \, \mathrm{d}u$ and $\mathrm{B}(a, b) = \mathrm{B}_1(a, b)$. If $U \sim \mathrm{Beta}(a, b)$, then $I_x(a, b) = \mathrm{pr}(U \le x)$.

For a fixed component α, define $S_{ij}^{(\alpha)} = \Theta_i^{(\alpha)} + \Theta_j^{(\alpha)}$ and $U_{ij}^{(\alpha)} = \Theta_i^{(\alpha)}/(\Theta_i^{(\alpha)} + \Theta_j^{(\alpha)})$. Aggregation properties of the Dirichlet distribution imply that $U_{ij}^{(\alpha)} \sim \mathrm{Beta}(\alpha_i, \alpha_j)$, $\mathbb{E}(S_{ij}^{(\alpha)}) = (\alpha_i + \alpha_j)/J$, and that $S_{ij}^{(\alpha)}$ and $U_{ij}^{(\alpha)}$ are independent. Moreover,

$$\min\Big(\Theta_i^{(\alpha)}, \Theta_j^{(\alpha)}\Big) = S_{ij}^{(\alpha)} \min\Big(U_{ij}^{(\alpha)}, 1 - U_{ij}^{(\alpha)}\Big).$$

Using independence, we obtain

$$m_{ij}(\alpha) = \mathbb{E}\left(S_{ij}^{(\alpha)}\right)\left[\mathbb{E}\left\{U_{ij}^{(\alpha)} I\left(U_{ij}^{(\alpha)} \le \frac{1}{2}\right)\right\} + \mathbb{E}\left\{\left(1 - U_{ij}^{(\alpha)}\right) I\left(U_{ij}^{(\alpha)} > \frac{1}{2}\right)\right\}\right],$$

where $I(\cdot)$ is the indicator function. The truncated expectations yield

$$m_{ij}(\alpha) = \frac{1}{J}\left[\alpha_i I_{1/2}(\alpha_i + 1, \alpha_j) + \alpha_j \left\{1 - I_{1/2}(\alpha_i, \alpha_j + 1)\right\}\right].$$

Linearity of expectation gives

$$\chi_{ij} = 3 \sum_{\alpha \in \mathcal{I}_J} q_\alpha m_{ij}(\alpha).$$

The mixture weights are constrained so that $3\sum_{\alpha\in\mathcal{I}_J} q_\alpha m_{ij}(\alpha) = 0.45$ for $(i,j) \in \{(1,2),(1,3),(2,3)\}$. Since $m_{ij}(\alpha)$ is fixed for every $\alpha \in \mathcal{I}_J$, these conditions are linear constraints in $\{q_\alpha : \alpha \in \mathcal{I}_J\}$.

Constructing angular measures with different orientations

To distinguish angular distributions satisfying the same moment and pairwise tail-dependence constraints, we introduce the alternating polynomial

$$\psi(\theta) = (\theta_1 - \theta_2)(\theta_2 - \theta_3)(\theta_3 - \theta_1) \qquad (\theta \in \Delta_2^\circ).$$

This polynomial is invariant under cyclic permutations of coordinates and changes sign when any two coordinates are exchanged. We use its expectation as a measure of orientation:

$$\mathrm{Or}(H) = \int_{\Delta_2^\circ} \psi(\theta)\, \mathrm{d}H(\theta).$$

For a Dirichlet component with parameter α, standard Dirichlet moments yield

$$o_\alpha = \mathbb{E}\left\{\psi\left(\Theta^{(\alpha)}\right)\right\} = \frac{(\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1)}{J(J+1)(J+2)}.$$

The positive-orientation angular measure H_+ is obtained by maximizing $\sum_{\alpha\in\mathcal{I}_J} q_\alpha o_\alpha$ subject to $q_\alpha \ge 0$, $\sum_{\alpha\in\mathcal{I}_J} q_\alpha = 1$, the first moment constraints, and the pairwise coefficient constraints. The negative-orientation measure H_- minimizes the same objective subject to the same constraints.

Let $\mathcal{O}_+ = \mathrm{Or}(H_+)$ and $\mathcal{O}_- = \mathrm{Or}(H_-)$. Define $\lambda = -\mathcal{O}_-/(\mathcal{O}_+ - \mathcal{O}_-)$ and construct the orientation-balanced measure $H_0 = \lambda H_+ + (1-\lambda)H_-$. By construction, $\mathrm{Or}(H_0) = 0$. Because all constraints are linear in the angular measure, H_0, H_+, and H_- share the same first angular moments and pairwise tail-dependence coefficients.

Controlling cluster separation with δ

For $\delta \in [0,1]$, define the three angular measures

$$H_{+,\delta} = (1-\delta)H_0 + \delta H_+, \qquad H_{-,\delta} = (1-\delta)H_0 + \delta H_-, \qquad H_{0,\delta} = H_0.$$

Because moment and pairwise constraints are linear, they are preserved under convex mixing for every $g \in \{+,0,-\}$ and $\delta \in [0,1]$. Furthermore, $\mathrm{Or}(H_{+,\delta}) = \delta\mathcal{O}_+$, $\mathrm{Or}(H_{-,\delta}) = \delta\mathcal{O}_-$, and $\mathrm{Or}(H_{0,\delta}) = 0$.

The parameter δ acts as an interpolation parameter controlling clustering difficulty without altering first moments or pairwise tail-dependence strength:

- when $\delta = 1$, $H_{+,\delta} = H_+$ and $H_{-,\delta} = H_-$, corresponding to the extreme LP solutions;
- as δ decreases, $H_{+,\delta}$ and $H_{-,\delta}$ move toward H_0;
- when $\delta = 0$, all three measures coincide with H_0.

Generation of the observations

Let $g(i) \in \{+,0,-\}$ denote the latent group of random vector i. For each δ, angular observations are generated independently as $\Theta_{i,t} \sim H_{g(i),\delta}$. Independently, standard Pareto radial variables $R_{i,t}$ are generated with $\mathrm{pr}(R_{i,t} > r) = 1/r$ $(r \geq 1)$. The trivariate observation is defined as $Z_{i,t} = 3R_{i,t}\Theta_{i,t}$.

Since $\mathbb{E}(\Theta_{i,t}^{(j)}) = 1/3$, the components of $Z_{i,t}$ have unit-Pareto upper tails. Specifically, for $z \geq 3$,

$$\mathrm{pr}\left(Z_{i,t}^{(j)} > z\right) = \mathbb{E}\left\{\mathrm{pr}\left(R_{i,t} > \frac{z}{3\Theta_{i,t}^{(j)}} \,\middle|\, \Theta_{i,t}\right)\right\} = \frac{3}{z}\mathbb{E}\left(\Theta_{i,t}^{(j)}\right) = \frac{1}{z}.$$

Furthermore, $\|Z_{i,t}\|_1 = 3R_{i,t}$ and $Z_{i,t}/\|Z_{i,t}\|_1 = \Theta_{i,t}$, so the angular component of every generated observation follows exactly $H_{g(i),\delta}$.

S6 Supplementary Results for the Data Applications

S6.1 Additional Results for the International Equity Markets

To examine the extremal-dependence regimes represented by the two cluster medoids, Fig. S5 displays kernel density estimates of their angular distributions. In the bivariate setting, the angular component is completely characterized by θ_1, which represents the proportion of an extreme event attributable to the United States market. Values close to one indicate events dominated by United States losses, whereas values close to zero indicate events dominated by local-market losses. The United Kingdom medoid exhibits more mass near the centre of the simplex, indicating a relatively balanced sharing of extreme losses between the United States and United Kingdom markets. In contrast, the Japan medoid displays a more pronounced concentration near the boundaries, suggesting that extreme events are more frequently dominated by one market or the other.

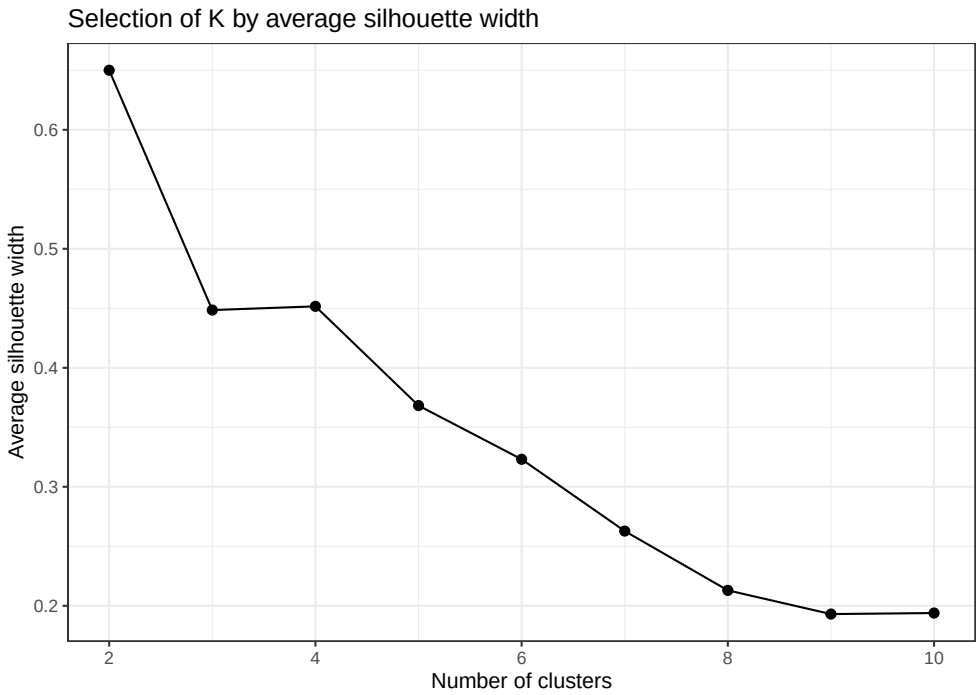


Fig. S4: Average silhouette width as a function of the number of clusters. The criterion is maximized at $K = 2$.

We next examine several countries located near the boundary of the two clusters. Figure S6 displays the estimated angular densities for Canada and Singapour. Canada exhibit substantially more mass near the centre of the simplex. This indicates a strong extremal dependence between United States and Canada losses during extreme events than that observed for the United Kingdom medoid. In contrast, Singapour shows much weaker dependence with US market even weaker than the Japan medoid and occupies a peripheral position in the multidimensional scaling representation shown in Fig. 5. These observations suggest that Saudi Arabia lies near the boundary between the two identified extremal-dependence regimes.

As a robustness comparison, we applied the conditional extremes clustering methodology of O'Toole et al. (2025) to the same data set. It yields exactly the same two-cluster partition as the Wasserstein–Aitchison procedure. Figure S7 shows the corresponding multidimensional scaling representation. This agreement indicates that the identified grouping is not an artefact of the particular dissimilarity measure employed, but reflects a structure in the extremal dependence of the international equity markets.

S6.2 Additional Details for the Swiss Catchment Application

We use the observation-based time-series collection from version 0.8 of CAMELS-CH (Höge et al. 2023a,b). The complete data set contains daily hydrometeorological

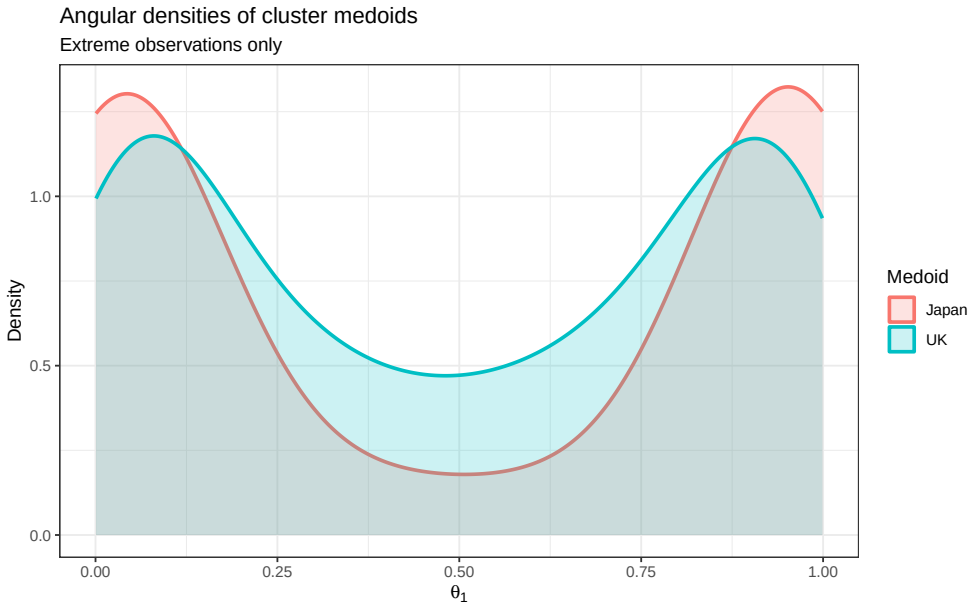


Fig. S5: Kernel density estimates of the angular distributions associated with the two cluster medoids. The variable θ_1 denotes the angular share corresponding to the United States market.

observations and catchment attributes for 331 hydrological units, comprising 298 river catchments and 33 lakes. The domain covers Switzerland and Liechtenstein together with transboundary catchments in Austria, France, Germany and Italy.

For catchment i and day t, the original variables are catchment-mean precipitation $p_{i,t}$ in $\mathrm{mm\,d^{-1}}$, catchment-mean daily air temperature $T_{i,t}$ in °C, snow water equivalent $s_{i,t}$ in mm, and observed specific discharge $q_{i,t}$ in $\mathrm{mm\,d^{-1}}$. Precipitation and temperature originate from the 2-km observation-based products of MeteoSwiss and were spatially averaged over each catchment. Discharge was recorded at the corresponding gauging stations and converted to specific discharge. The snow-water-equivalent data were obtained by aggregating a distributed 1-km snow-cover product constrained by observations from 345 snow-monitoring stations and MeteoSwiss precipitation data; the associated data-assimilation methodology is described by Magnusson et al. (2014). Snow water equivalent should therefore be regarded as a model–data-assimilation product rather than a direct catchment-scale measurement. Its availability from September 1998 determines the analysis period of 1 September 1998–31 December 2020. Daily precipitation is partitioned into liquid and solid fractions using the temperature-dependent liquid fraction

$$f(T) = \begin{cases} 0, & T \leq 0\,^{\circ}\mathrm{C}, \\ T/2, & 0\,^{\circ}\mathrm{C} < T < 2\,^{\circ}\mathrm{C}, \\ 1, & T \geq 2\,^{\circ}\mathrm{C}. \end{cases} \tag{S10}$$

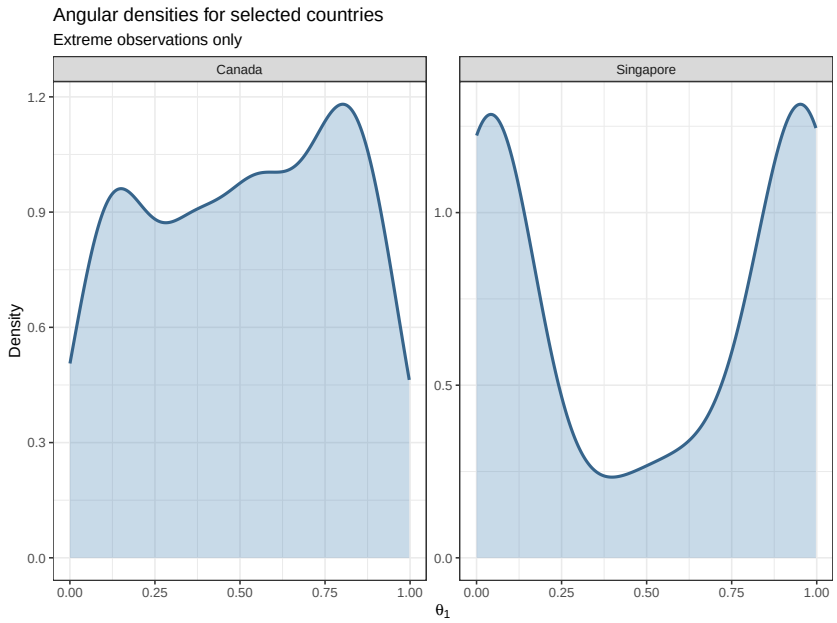


Fig. S6: Estimated angular densities for selected countries located near the boundary of the two clusters.

Following the temperature-range approach discussed by Jennings and Molotch (2019), we use a simple linear interpolation between 0 and 2 °C, centred on their Northern Hemisphere mean 50% rain–snow threshold of approximately 1 °C.

The three event-scale variables are defined by

$$L_{i,t} = \sum_{j=0}^{2} p_{i,t-j} f\left(T_{i,t-j}\right), \tag{S11}$$

$$M_{i,t} = \left(s_{i,t-3} - s_{i,t}\right)_+, \qquad (x)_+ = \max(x, 0), \tag{S12}$$

$$Q^{\star}_{i,t} = \max_{h \in \{0,1,2\}} q_{i,t+h}. \tag{S13}$$

Thus, $L_{i,t}$ is accumulated liquid precipitation over the current and two preceding days, $M_{i,t}$ is positive net snow-water-equivalent depletion over three days, and $Q^{\star}_{i,t}$ is the maximum specific discharge on day t or either of the following two days.

The three-day event scale is motivated by the rapid response of many Alpine catchments. Froidevaux et al. (2015) found that precipitation during the days immediately preceding a flood was generally the most relevant for Swiss Alpine catchments. Three-day precipitation totals and snow-water-equivalent decreases have also been used to distinguish rainfall-driven, snowmelt-driven and mixed rainfall–snowmelt runoff events (Davenport et al. 2020). The chosen window is nevertheless an approximation and

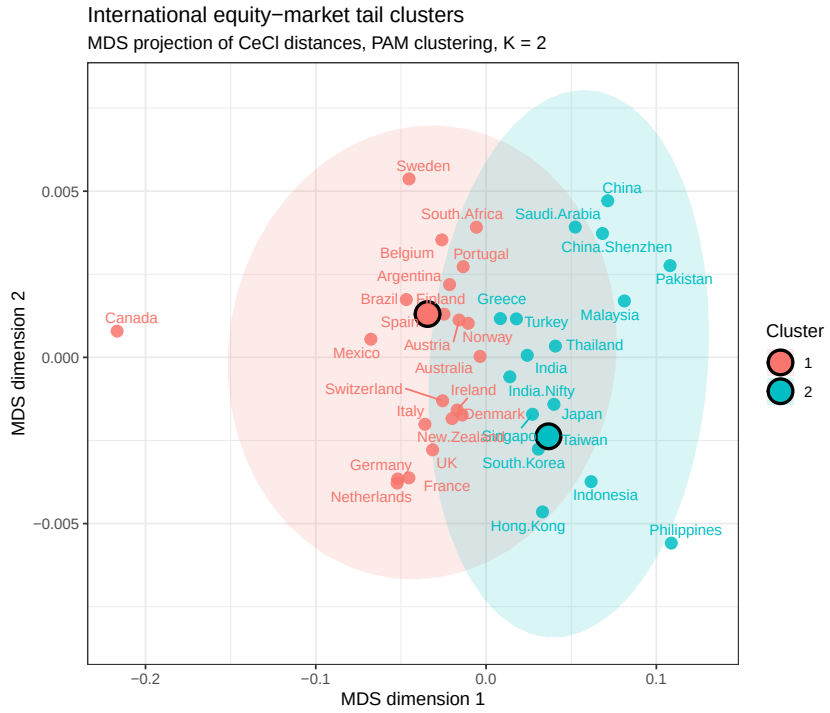


Fig. S7: International equity-market tail clusters obtained with the conditional extremes clustering dissimilarity. The figure shows a two-dimensional multidimensional scaling projection of the dissimilarities together with the PAM partition. Black-circled observations indicate the cluster medoids.

should not be interpreted as a universal catchment-response time. Net snow-water-equivalent depletion is a proxy for snowpack water release rather than gross snowmelt. Melt occurring simultaneously with new snowfall may be partly or completely masked by increasing snow water equivalent. Negative depletion is therefore set to zero rather than interpreted as negative melt.

Negative precipitation, snow water equivalent and discharge values were treated as missing, and missing snow-water-equivalent observations were not replaced by zero. A daily triplet

$$X_{i,t} = \left(L_{i,t}, M_{i,t}, Q^{\star}_{i,t}\right)$$

was retained only when all observations required by equations (S11)–(S13) were available. Catchments with fewer than 3 650 complete triplets, corresponding approximately to ten years of daily observations, were excluded. Selection was based exclusively on data availability.

Within each retained catchment, the three margins were transformed separately to standard Pareto scale. Because $L_{i,t}$ and $M_{i,t}$ contain substantial point masses at zero, tied observations were assigned reproducible random ranks within their tie blocks. For

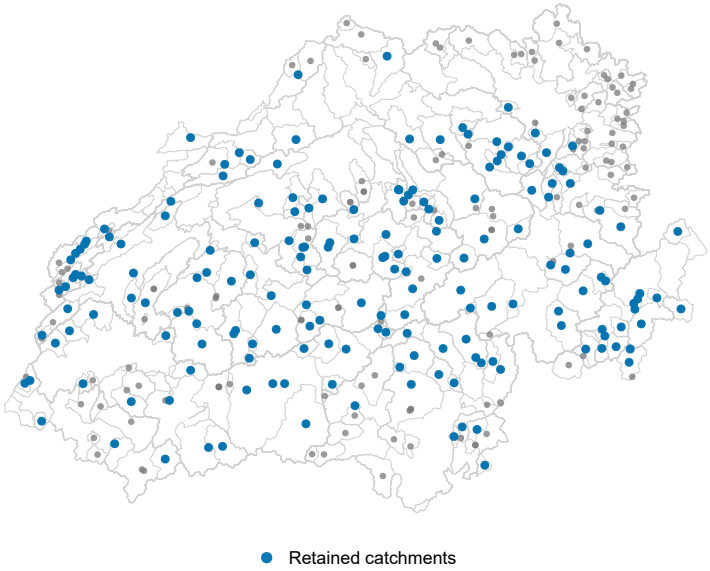


Fig. S8: Locations of the 331 catchments in the CAMELS-CH data set. Blue points indicate the 182 catchments retained for analysis. Grey points indicate catchments excluded because they had fewer than 10 years of available data or fewer than 50 declustered extreme events.

marginal sample size n_i, we used

$$U_{i,t,j} = \frac{\text{rank}(X_{i,t,j})}{n_i + 1}, \qquad Z_{i,t,j} = \frac{1}{1 - U_{i,t,j}}, \qquad j = 1, 2, 3.$$

The radial component was then

$$R_{i,t} = \sum_{j=1}^{3} Z_{i,t,j}.$$

Observations above the catchment-specific 95th percentile of $R_{i,t}$ were identified as multivariate extremes. Successive exceedances separated by at most three days were assigned to the same event, of which only the observation with the largest radial component was retained. This is our de-clustering method similarly as what was done by Sabourin and Renard (2015). Requiring at least 50 such events left 182 of the original 331 catchments, whose locations are shown in Fig. S8.

The corresponding angular observations were transformed to two-dimensional Aitchison coordinates. Pairwise 1-Wasserstein distances were computed between the empirical Aitchison-coordinate distributions and clustered using partitioning around medoids. We considered $K = 2, \ldots, 6$; the average silhouette width shown in Fig. S10

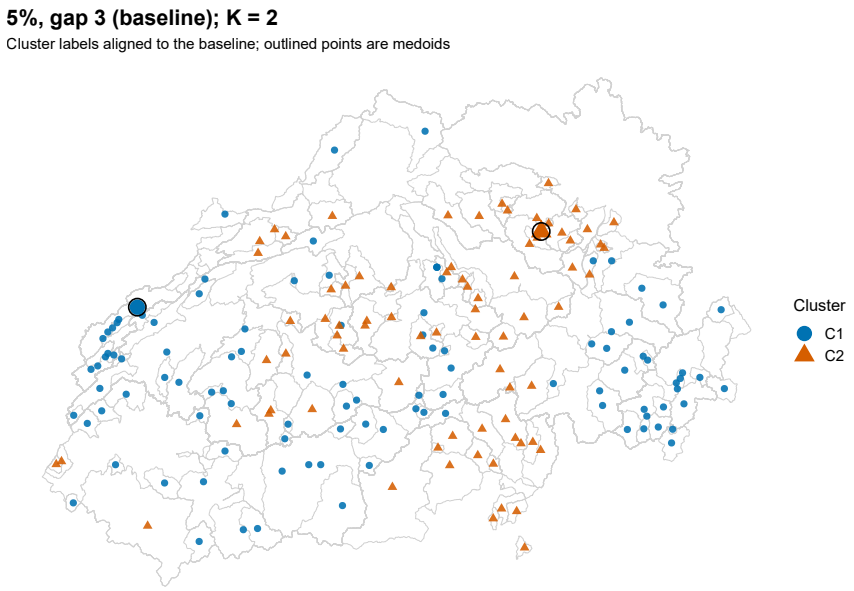


Fig. S9: Spatial distribution of the two extremal-dependence clusters. Blue circles, orange triangles denote C1, C2, respectively. Larger outlined symbols identify the cluster medoids. Light-grey lines delimit the catchments included in the analysis.

selected $K = 3$. The resulting partition exhibits broad spatial organization without forming strictly contiguous regions, as shown in Fig. S9

The interpretation of the two clusters was assessed using four CAMELS-CH attributes that were not used in the clustering: mean catchment elevation, fraction of precipitation falling as snow, mean catchment slope and mean half-flow date. The half-flow date is the number of days from the beginning of the hydrological year, defined as 1 October, required for cumulative discharge to reach one half of annual discharge (Höge et al. 2023a). Figure 7 shows that C1 catchments have the highest mean elevations and largest fractions of precipitation falling as snow. They also have steeper slopes and later half-flow dates, consistently with delayed runoff associated with seasonal snow storage and melt. C2 catchments have the lower elevations and snow fractions, supporting their interpretation as predominantly rainfall–runoff catchments. This is consistent with the angular regime identified in Fig. 6 of the main paper. We assessed sensitivity to the declustering window and radial threshold by comparing the baseline specification (upper 5% of radial observations and a three-day window) with windows of two and six days and retained fractions of 2.5% and 10%. The average silhouette widths in Fig. S10 generally support two clusters; although $K = 3$ is marginally favoured in some specifications, this preference is not consistent. Table S7 compares each two-cluster partition with the baseline. After aligning cluster labels, 89.0%–99.5% of catchments retain their assignment, with adjusted Rand indices between 0.607 and 0.978. The partition is nearly invariant to the declustering window and moderately more sensitive

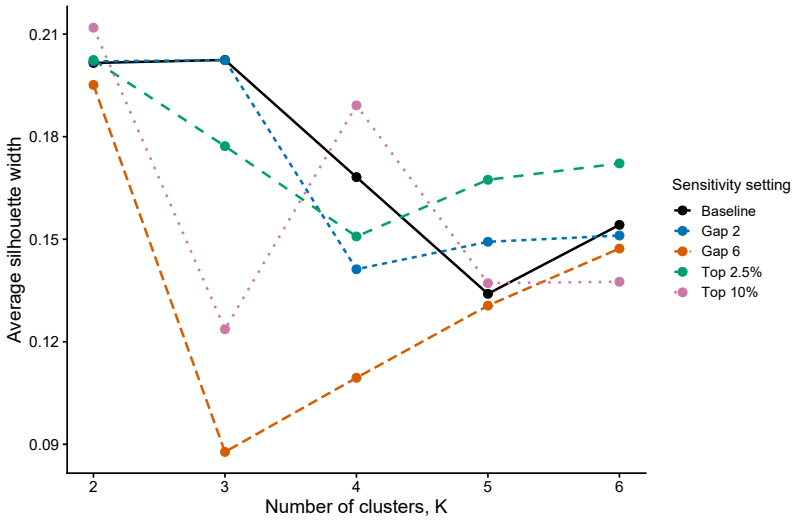


Fig. S10: Average silhouette width as a function of the number of clusters under the baseline and alternative radial-threshold and declustering-window specifications. Although some specifications marginally favour three clusters, the evidence is not consistent across settings and generally supports a two-cluster representation.

to the threshold, consistent with the usual bias–variance trade-off in extreme-value analysis. Overall, these results support $K = 2$ as the most stable choice.

Table S7: Sensitivity of the clustering results to the radial threshold and declustering window. Each partition is compared with the baseline specification (upper 5% and a three-day window) using the 182 catchments common to all specifications. Direct agreement is computed after optimal alignment of the cluster labels. ARI denotes the adjusted Rand index.

K	Sensitivity scenario	Common catchments	Same assignment	Changed assignment	Direct agreement	ARI
2	5%, gap 2	182	179	3	98.4%	0.935
2	5%, gap 6	182	181	1	99.5%	0.978
2	Top 2.5%, gap 3	182	162	20	89.0%	0.607
2	Top 10%, gap 3	182	167	15	91.8%	0.695
3	5%, gap 2	182	178	4	97.8%	0.927
3	5%, gap 6	182	99	83	54.4%	0.307
3	Top 2.5%, gap 3	182	126	56	69.2%	0.363
3	Top 10%, gap 3	182	101	81	55.5%	0.336
4	5%, gap 2	182	115	67	63.2%	0.442
4	5%, gap 6	182	114	68	62.6%	0.335
4	Top 2.5%, gap 3	182	119	63	65.4%	0.326
4	Top 10%, gap 3	182	161	21	88.5%	0.709

References

Aitchison, J.: The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Methodological) **44**(2), 139–160 (1982) https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Boldi, M.-O., Davison, A.C.: A mixture model for multivariate extremes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**(2), 217–229 (2007) https://doi.org/10.1111/j.1467-9868.2007.00585.x

Boulin, A., Di Bernardino, E., Laloë, T., Toulemonde, G.: High-dimensional variable clustering based on maxima of a weakly dependent random process. Journal of the American Statistical Association **120**(551), 1933–1944 (2025) https://doi.org/10.1080/01621459.2025.2459443

Benevento, A., Durante, F., Pappadà, R.: Tail-dependence clustering of time series with spatial constraints. Environmental and Ecological Statistics **31**(3), 801–817 (2024) https://doi.org/10.1007/s10651-024-00626-6

Benevento, A., Durante, F., Pappadà, R.: Comonotonic-based time series clustering with constraints: A review and a conceptual framework. Environmetrics **36**(8), 70047 (2025) https://doi.org/10.1002/env.70047

Beirlant, J., Goegebeur, Y., Teugels, J.L., Segers, J.: Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics. Wiley, Chichester (2004). https://doi.org/10.1002/0470012382

Bernard, E., Naveau, P., Vrac, M., Mestre, O.: Clustering of maxima: Spatial dependencies among heavy rainfall in France. Journal of Climate **26**(20), 7929–7937 (2013) https://doi.org/10.1175/JCLI-D-12-00836.1

Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. Springer, London (2001). https://doi.org/10.1007/978-1-4471-3675-0

Davenport, F.V., Herrera-Estrada, J.E., Burke, M., Diffenbaugh, N.S.: Flood size increases nonlinearly across the western United States in response to lower snow-precipitation ratios. Water Resources Research **56**(1), 2019–025571 (2020) https://doi.org/10.1029/2019WR025571

Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. Springer, New York (2006). https://doi.org/10.1007/0-387-34471-3

Demarta, S., McNeil, A.J.: The t copula and related copulas. International Statistical Review **73**(1), 111–129 (2005) https://doi.org/10.1111/j.1751-5823.2005.tb00254.x

Einmahl, J.H.J., Haan, L., Piterbarg, V.I.: Nonparametric estimation of the spectral

measure of an extreme value distribution. The Annals of Statistics **29**(5), 1401–1423 (2001) https://doi.org/10.1214/aos/1013203459

Engelke, S., Hitz, A.S.: Graphical models for extremes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **82**(4), 871–932 (2020) https://doi.org/10.1111/rssb.12355

Engelke, S., Naveau, P., Zhou, C.: A Kullback–Leibler divergence test for multivariate extremes: theory and practice. arXiv:2602.02316 (2026). https://doi.org/10.48550/arXiv.2602.02316 . https://arxiv.org/abs/2602.02316

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C.: Isometric logratio transformations for compositional data analysis. Mathematical Geology **35**(3), 279–300 (2003) https://doi.org/10.1023/A:1023818214614

Einmahl, J.H.J., Segers, J.: Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. The Annals of Statistics **37**(5B), 2953–2989 (2009) https://doi.org/10.1214/08-AOS677

Fuchs, S., Di Lascio, F.M.L., Durante, F.: Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Computational Statistics & Data Analysis **159**, 107201 (2021) https://doi.org/10.1016/j.csda.2021.107201

Fomichov, V., Ivanovs, J.: Spherical clustering in detection of groups of concomitant extremes. Biometrika **110**(1), 135–153 (2023) https://doi.org/10.1093/biomet/asac020

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., Martius, O.: Flood triggering in Switzerland: The role of daily to monthly preceding precipitation. Hydrology and Earth System Sciences **19**(9), 3903–3924 (2015) https://doi.org/10.5194/hess-19-3903-2015

Galambos, J.: Order statistics of samples from multivariate distributions. Journal of the American Statistical Association **70**(351a), 674–680 (1975) https://doi.org/10.1080/01621459.1975.10482493

Gudendorf, G., Segers, J.: Extreme-value copulas. In: Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics, vol. 198, pp. 127–145. Springer, Berlin and Heidelberg (2010). https://doi.org/10.1007/978-3-642-12465-5_6

Gumbel, E.J.: Bivariate exponential distributions. Journal of the American Statistical Association **55**(292), 698–707 (1960) https://doi.org/10.1080/01621459.1960.10483368

Gumbel, E.J.: Bivariate logistic distributions. Journal of the American Statistical Association **56**(294), 335–349 (1961) https://doi.org/10.1080/01621459.1961.

10482117

Hubert, L., Arabie, P.: Comparing partitions. Journal of Classification **2**(1), 193–218 (1985) https://doi.org/10.1007/BF01908075

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M.G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A.E., Addor, N., Brunner, M., Pool, S., Zappa, M., Fenicia, F.: CAMELS-CH: Hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland. Earth System Science Data **15**(12), 5755–5784 (2023) https://doi.org/10.5194/essd-15-5755-2023

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M.G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A.E., Addor, N., Brunner, M., Pool, S., Zappa, M., Fenicia, F.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH). Zenodo data set, version 0.8 (2023). https://doi.org/10.5281/zenodo.10354485 . https://doi.org/10.5281/zenodo.10354485

Hüsler, J., Reiss, R.-D.: Maxima of normal random vectors: Between independence and complete dependence. Statistics & Probability Letters **7**(4), 283–286 (1989) https://doi.org/10.1016/0167-7152(89)90106-5

Heffernan, J.E., Tawn, J.A.: A conditional approach for multivariate extreme values (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) **66**(3), 497–546 (2004) https://doi.org/10.1111/j.1467-9868.2004.02050.x

Jennings, K.S., Molotch, N.P.: The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient. Hydrology and Earth System Sciences **23**(9), 3765–3786 (2019) https://doi.org/10.5194/hess-23-3765-2019

Joe, H.: Dependence Modeling with Copulas, 1st edn. Chapman & Hall/CRC Monographs on Statistics and Applied Probability, vol. 134. Chapman & Hall/CRC, Boca Raton (2014)

Janßen, A., Wan, P.: k-means clustering of extremes. Electronic Journal of Statistics **14**(1), 1211–1233 (2020) https://doi.org/10.1214/20-EJS1689

Kiriliouk, A., Naveau, P.: Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. The Annals of Applied Statistics **14**(3), 1342–1358 (2020) https://doi.org/10.1214/20-AOAS1355

Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series in Probability and Statistics. Wiley, New York (1990). https://doi.org/10.1002/9780470316801

Lhaut, S., Rootzén, H., Segers, J.: Simulation of multivariate extremes: A Wasserstein–Aitchison GAN approach. Extremes **29**(2), 157–194 (2026) https://doi.org/10.1007/s10687-026-00530-1

Lhaut, S., Segers, J.: An asymptotic expansion of the empirical angular measure for bivariate extremal dependence. In: Barigozzi, M., Hörmann, S., Paindaveine, D. (eds.) Recent Advances in Econometrics and Statistics: Festschrift in Honour of Marc Hallin, pp. 187–208. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-61853-6_10

Magnusson, J., Gustafsson, D., Hüsler, F., Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. Water Resources Research **50**(10), 7816–7835 (2014) https://doi.org/10.1002/2014WR015302

Mourahib, A., Kiriliouk, A., Segers, J.: Multivariate generalized Pareto distributions along extreme directions. Extremes **28**(2), 239–272 (2025) https://doi.org/10.1007/s10687-024-00501-4

Marcon, G., Padoan, S.A., Antoniano-Villalobos, I.: Bayesian inference for the extremal dependence. Electronic Journal of Statistics **10**(2), 3310–3337 (2016) https://doi.org/10.1214/16-EJS1162

Marcon, G., Padoan, S.A., Naveau, P., Muliere, P., Segers, J.: Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. Journal of Statistical Planning and Inference **183**, 1–17 (2017) https://doi.org/10.1016/j.jspi.2016.10.004

Menéndez, M.L., Pardo, J.A., Pardo, L., Pardo, M.d.C.: The Jensen–Shannon divergence. Journal of the Franklin Institute **334**(2), 307–318 (1997) https://doi.org/10.1016/S0016-0032(96)00063-4

Opitz, T.: Extremal t processes: Elliptical domain of attraction and a spectral representation. Journal of Multivariate Analysis **122**, 409–413 (2013) https://doi.org/10.1016/j.jmva.2013.08.008

O'Toole, P., Rohrbeck, C., Richards, J.: Clustering of multivariate tail dependence using conditional methods. arXiv:2510.20424 (2025). https://doi.org/10.48550/arXiv.2510.20424 . https://arxiv.org/abs/2510.20424

Peyré, G., Cuturi, M.: Computational optimal transport with applications to data sciences. Foundations and Trends in Machine Learning **11**(5–6), 355–607 (2019) https://doi.org/10.1561/2200000073

Resnick, S.I.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer Series in Operations Research and Financial Engineering. Springer, New York (2007). https://doi.org/10.1007/978-0-387-45024-7

Schuhmacher, D., Bähre, B., Bonneel, N., Gottschlich, C., Hartmann, V., Heinemann, F., Schmitzer, B., Schrieber, J.: transport: Computation of Optimal Transport Plans and Wasserstein Distances. (2024). https://doi.org/10.32614/CRAN.package.transport . R package version 0.15-4. https://CRAN.R-project.org/package=transport

Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris **8**(3), 229–231 (1959)

Sabourin, A., Naveau, P.: Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. Computational Statistics & Data Analysis **71**, 542–567 (2014) https://doi.org/10.1016/j.csda.2013.04.021

Sabourin, A., Renard, B.: Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data. Water Resources Research **51**(12), 9646–9664 (2015) https://doi.org/10.1002/2015WR017320

Steinley, D.: Properties of the Hubert–Arabie adjusted Rand index. Psychological Methods **9**(3), 386–396 (2004) https://doi.org/10.1037/1082-989X.9.3.386

Erven, T., Harremoës, P.: Rényi divergence and Kullback–Leibler divergence. IEEE Transactions on Information Theory **60**(7), 3797–3820 (2014) https://doi.org/10.1109/TIT.2014.2320500

Vignotto, E., Engelke, S., Zscheischler, J.: Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland. Weather and Climate Extremes **32**, 100318 (2021) https://doi.org/10.1016/j.wace.2021.100318

Villani, C.: Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften, vol. 338. Springer, Berlin and Heidelberg (2009). https://doi.org/10.1007/978-3-540-71050-9

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R.M., Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M.D., Maraun, D., Ramos, A.M., Ridder, N.N., Thiery, W., Vignotto, E.: A typology of compound weather and climate events. Nature Reviews Earth & Environment **1**(7), 333–347 (2020) https://doi.org/10.1038/s43017-020-0060-z