Spatial Statistics 66 (2025) 100881

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Clustering of compound events based on multivariate comonotonicity

Fabrizio Durante [a], Sebastian Fuchs [b], Roberta Pappadà [c,*]

[a] Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Lecce, Italy
[b] Department for Artificial Intelligence and Human Interfaces (AIHI), Paris Lodron University of Salzburg, Salzburg, Austria
[c] Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche "B. de Finetti", Università degli Studi di Trieste, Trieste, Italy

A R T I C L E I N F O

Keywords:
Copula
Comonotonicity
Unsupervised learning

A B S T R A C T

Driven by the goal of generating risk maps for flood events—characterized by various physical variables such as peak flow and volume, and measured at specific geographic locations—this work proposes several dissimilarity functions for use in unsupervised learning problems and, specifically, in clustering algorithms. These dissimilarities are rank-based, relying on the dependence occurring among the random variables involved, and assign the smallest values to pairs of subsets that are π-comonotonic. This concept is less restrictive than classical comonotonicity but, in the multivariate case, can offer a more intuitive understanding of compound phenomena.

An application of these measures is presented through the analysis of flood risks using data from the Po river basin, with results compared to similar studies found in the literature.

1. Introduction

Developing stochastic methods to analyze multivariate time series data is currently a demanding challenge, especially when non-linear dependencies and extremes arise, as it is common in various fields such as environmental science. In particular, the study of the linkage among random phenomena is of crucial importance to understand the risks associated with a complex system and to mitigate their potential severe impacts on society (see Saunders et al., 2021).

A recent stream of research on multivariate data has focused on the development of unsupervised learning techniques in order to group different time series according to their behavior; see, for instance, Caiado et al. (2016) and Maharaj et al. (2019). Clustering of time series can be important for two main reasons. First, it allows us to reveal joint effects, which eventually facilitates the implementation of risk mitigation strategies. Secondly, determining similarities can improve the specification of the stochastic behavior in those situations where the information is scarce and the data have been poorly recorded.

Interestingly, recent time-series clustering methods take into account the cross-sectional dependence; that is, they aim at finding groups of time series that are closely correlated to each other and separating them from groups that have null or smaller correlation with these series. See, for instance, Alonso and Peña (2019) and Alonso et al. (2021) for methods based on linear dependence. From a copula perspective, hierarchical clustering methods have been especially considered in the literature (see, e.g., Di Lascio et al. (2017) and Fuchs et al. (2021), and the references therein). In this respect, it should be emphasized that copulas allow addressing various kinds of dependencies of the joint distribution (Durante and Sempi, 2016; Salvadori et al., 2007), such as asymmetries and tail dependence, which could be quite appealing for clustering purposes, especially when extreme behavior is of interest (see,

* Corresponding author.
E-mail address: rpappada@units.it (R. Pappadà).

https://doi.org/10.1016/j.spasta.2025.100881
Received 12 November 2024; Received in revised form 17 January 2025; Accepted 20 January 2025
Available online 27 January 2025

e.g., De Luca and Zuccolotto, 2011, 2017; Disegna et al., 2017; Durante et al., 2015; D'Urso et al., 2023; Di Lascio et al., 2023; Benevento et al., 2024; De Keyser and Gijbels, 2024; Fuchs and Wang, 2024).

As a matter of fact, most of these clustering algorithms assume that each random phenomenon is described by a single random variable. Recent studies suggest, instead, that a multivariate approach should be followed (Viroli, 2011; Puccetti, 2022). In particular, for risk management purposes, the concurrent occurrence of multiple and possibly interdependent hazards can lead to impacts that are larger than any of the individual hazards (see the compound event approach in Zscheischler et al., 2018, 2020 and Bevacqua et al., 2021). For instance, in Pappadà et al. (2018), hazard scenarios (in the sense of Salvadori et al., 2016) are used in order to develop a procedure to cluster flood phenomena (as characterized by peak and volume). Moreover, in Vignotto et al. (2021) a novel nonparametric divergence measure is used to capture similarities in the tail behavior of bivariate distributions, with an application to the identification of compound precipitation and wind extremes.

In this work, we introduce a clustering method to group random phenomena characterized by a vector of features of fixed dimension $p \geq 1$. The methodology is based on the introduction of a suitable dissimilarity measure between random vectors that is hence used to provide input values to a classical unsupervised learning algorithm designed for clustering problems (such as agglomerative hierarchical algorithms, see Murtagh and Contreras, 2012, 2017). Moreover, the procedure is rank-invariant and, hence, based entirely on the copula of the involved continuous random variables.

For $p = 1$, copula-based dissimilarity measures can be constructed by quantifying the discrepancy between the copula C_{XY} associated with the random pair (X, Y) and the comonotonicity copula (see Fuchs et al., 2021). When $p \geq 2$, however, the problem does not admit a unique extension, since various notions of comonotonicity between two random vectors are possible (see Puccetti and Scarsini, 2010 and Puccetti and Wang, 2015). Here, we focus on the definition of π–comonotonicity, which is a weaker notion of multivariate comonotonicity ensuring that, if two random vectors are π–comonotonic, then they have the same copula.

As it will be clarified in the following, random vectors in each cluster are similar (according to π–comonotonicity) in the sense that they tend to have a similar dependence structure (i.e. the copula) and, moreover, the ith variable of one vector is comonotone with the ith variable of the other vector for each component i. Thus, such method is able to detect compound phenomena that are homogeneous in the dependence structure and, at the same time, comonotone with respect to each individual variable. For instance, two flood events (characterized by peak and volume) observed at different sites are clustered together according to π–comonotonicity when: (a) the copula between peak and volume is similar at the two sites; (b) the peak (respectively, volume) variable in one site is comonotonic with the peak (respectively, volume) in the other one. This similarity is different from strong comonotonicity, which would also require that peak and volume are comonotone (which may not be necessary in the present context).

The manuscript is organized as follows. In Section 2, we present the novel proposed methodology of this work. Then, Section 3 presents a case study that is related to the detection of spatially correlated concurrent floods as considered in Pappadà et al. (2018). Section 4 concludes the paper.

2. The methodology

Let $n \geq 2$ and $p \geq 1$ be integers. In the following, we consider a set $\mathscr{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ of p–dimensional continuous random vectors defined on the same probability space that represents typically a random phenomenon (e.g. a flood) observed at n different sites (e.g., gauge stations).

Each random vector $\mathbf{X}_i$ is associated to a random sample $(\mathbf{x}_{ti})_{t=1,\ldots,T}$ observed over time $t = 1, \ldots, T$.

Remark 1. In this setting, the samples are considered as independent realizations. Otherwise, one may first filter the original time series (e.g., for removing trends and seasonality effects), and then investigate the dependence among the resulting residuals. This procedure is described in its full generality in Patton (2012) and Neumeyer et al. (2019), and is applied to cluster analysis in Benevento et al. (2024) among others.

Our aim is to develop a clustering procedure based on a suitably defined similarity criterion, to identify a partition of $\mathscr{X}$ into non-empty and non-overlapping subsets that exploits the dependence between the elements of $\mathscr{X}$. In particular, following Fuchs et al. (2021), we assume that the similarity/dissimilarity between the vectors will only depend on the probability law of the random vectors and is rank-invariant, i.e. it only depends on the copula associated with the probability distribution functions. Thus, without loss of generality, we can assume that $\mathscr{X}$ is a subset of $L^0(\mathbb{I}^p)$, the space of all p–dimensional random vectors with uniform margins on $\mathbb{I} := [0, 1]$, i.e. it corresponds to the space of p–dimensional copulas.

The proposed clustering procedure can be implemented in the following way.

1. A dissimilarity measure is introduced to compare random vectors in $\mathscr{X}$.
2. Next, an $(n \times n)$ matrix is constructed, capturing all pairwise dissimilarities between elements in $\mathscr{X}$.
3. Finally, this dissimilarity matrix serves as the input for a clustering algorithm, such as linkage-based hierarchical clustering.

Concerning the first step, we introduce a notion of dissimilarity measures between random vectors that extends a similar concept introduced in Fuchs et al. (2021, Definition 2.1).

When $p = 1$, according to Fuchs et al. (2021), a measure of dissimilarity usually quantifies some discrepancy between C_{XY}, the unique copula between X and Y, and the Fréchet–Hoeffding upper bound copula $M(u, v) = \min(u, v)$, for all $(u, v) \in \mathbb{I}^2$. As it is well known, if $C_{XY} = M$, then X and Y are comonotonic, i.e. both X and Y are (almost surely) an increasing transformation of a common random variable Z (see, e.g., Durante and Sempi, 2016). Thus, zero dissimilarity corresponds to a monotone (and deterministic) relationship among X and Y.

When $p \geq 2$, instead, the problem does not admit a unique extension, since various notions of comonotonicity between two random vectors are possible (see Puccetti and Scarsini, 2010). A first natural generalization is given by the notion of strong comonotonicity, which is expressed below (see, e.g., Durante and Sempi, 2016 and Puccetti and Scarsini, 2010).

Theorem 1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors in* $L^0(\mathbb{I}^p)$ *with copula* $C_{(\mathbf{X},\mathbf{Y})}$. *Then the following statements are equivalent:*

(a) $(\mathbf{X},\mathbf{Y})$ *is strongly comonotonic.*
(b) The copula $C_{(\mathbf{X},\mathbf{Y})}$ *can be written in the form*

$$C_{(\mathbf{X},\mathbf{Y})}(\mathbf{u},\mathbf{v}) = M(\mathbf{u},\mathbf{v}) = \min(u_1,\dots,u_p,v_1,\dots,v_p) \tag{1}$$

for all $(\mathbf{u},\mathbf{v}) \in \mathbb{I}^{2p}$.
(c) The copula associated with any pair extracted from $(\mathbf{X},\mathbf{Y})$ *is given by* M.

As a matter of fact, the previous notion is somehow excessively strong. In fact, it requires not only that any element of $\mathbf{X}$ is comonotonic with any element of $\mathbf{Y}$, but also that the elements within $\mathbf{X}$ (respectively, $\mathbf{Y}$) are pairwise comonotonic. In some sense, this notion considers comonotonicity both between vectors and within them.

Instead, we focus here on the definition of π–comonotonic random vectors, which is formalized in Puccetti and Scarsini (2010, Theorem 3.7) and is expressed in the following equivalent form.

Theorem 2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors in* $L^0(\mathbb{I}^p)$ *with copula* $C_{(\mathbf{X},\mathbf{Y})}$. *Then the following statements are equivalent:*

(a) $(\mathbf{X},\mathbf{Y})$ *is* π*-comonotonic.*
(b) The copula $C_{(\mathbf{X},\mathbf{Y})}$ *can be written in the form*

$$C_{(\mathbf{X},\mathbf{Y})}(\mathbf{u},\mathbf{v}) = C\big(M(u_1,v_1),\dots,M(u_p,v_p)\big) \tag{2}$$

for all $(\mathbf{u},\mathbf{v}) \in \mathbb{I}^{2p}$, *where* $C_{\mathbf{X}} = C_{\mathbf{Y}} = C$ *for a* p*–dimensional copula* C.
(c) For all $i \in \{1,\dots,p\}$, $C_{(X_i,Y_i)} = M$.

Proof. See Appendix. □

As clarified in Puccetti and Scarsini (2010) and apparent from Eq. (2), π-comonotonicity does not imply that $(\mathbf{X},\mathbf{Y})$ has copula $C_{(\mathbf{X},\mathbf{Y})} = M$, where $M(\mathbf{u}) = \min(u_1,\dots,u_p)$ for every $\mathbf{u} \in \mathbb{I}^p$. In particular, under strong comonotonicity it holds that $\mathbf{X}$ (or $\mathbf{Y}$) has itself the comonotone copula M, while this is not the case for π-comonotonicity.

Remark 2. Theorem 2 provides a general way to sample observations $(\mathbf{x},\mathbf{y})$ from a π-comonotonic copula $(\mathbf{X},\mathbf{Y})$:

1. Sample $\mathbf{x} = (x_1,\dots,x_p)$ from a p–dimensional copula C.
2. Define $\mathbf{y} = \mathbf{x}$.
3. Return $(\mathbf{x},\mathbf{y})$.

Notice that the copula C comes from Eq. (2).

Now, we give the following definition of a dissimilarity function that is coherent with the notion of π–comonotonicity introduced above.

Definition 1. A mapping $d_\pi : L^0(\mathbb{I}^p)\times L^0(\mathbb{I}^p) \to [0,+\infty[$ is called dissimilarity function based on π–comonotonicity if it verifies the following properties:

(A1) $d_\pi(\mathbf{X},\mathbf{Y}) = 0$ holds for all $\mathbf{X}$ and $\mathbf{Y}$ that are π–comonotonic;
(A2) $d_\pi(\mathbf{X},\mathbf{Y}) = d_\pi(\mathbf{Y},\mathbf{X})$.

When $p = 1$, the dissimilarity function d_π from Definition 1 coincides with the $(1,1)$-dissimilarity function $d^{1,1}$ considered in Fuchs et al. (2021). For completeness, we recall that a $(1,1)$-dissimilarity function $d^{1,1}$ is a mapping from $L^0(\mathbb{I}) \times L^0(\mathbb{I})$ to $[0,+\infty[$ that is law-invariant, symmetric in its arguments, and verifies $d^{1,1}(X,Y) = 0$ whenever the random pair (X,Y) is coupled by the copula M.

According to Theorem 2(c), a dissimilarity function d_π among two p–dimensional vectors can be determined by measuring some discrepancy between the copula $C_{(X_i,Y_i)}$ of the random pair (X_i,Y_i) and the 2-copula M for each $i = 1,\dots,p$. Thus, examples of dissimilarity functions from Definition 1 can be obtained in the following way.

Proposition 3. *Let* $d^{1,1}$ *be a* $(1,1)$*-dissimilarity function. Then the following mappings, defined for each pair of continuous* p*–dimensional random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *by*

$$d_\pi^{\max}(\mathbf{X},\mathbf{Y}) = \max_{i=1,2,\dots,p} d^{1,1}(X_i,Y_i), \tag{3}$$

$$d_\pi^{\text{ave}}(\mathbf{X},\mathbf{Y}) = \frac{1}{p}\sum_{i=1}^{p} d^{1,1}(X_i,Y_i), \tag{4}$$

$$d_\pi^{\min}(\mathbf{X},\mathbf{Y}) = \min_{i=1,2,\dots,p} d^{1,1}(X_i,Y_i), \tag{5}$$

are dissimilarity functions in the sense of Definition 1.

Proof. It is an immediate consequence of the assumption on $d^{1,1}$ and Theorem 2(c). □

Various choices of $d^{1,1}$ may produce different dissimilarity functions. Here, we collect some examples already considered in the literature that are based on some measure of association between random variables (see Fuchs et al., 2021). Denoting by C the copula between the random variables X and Y, we consider:

$$d_\beta^{1,1}(X,Y) = M\left(\frac{1}{2},\frac{1}{2}\right) - C\left(\frac{1}{2},\frac{1}{2}\right), \tag{6}$$

$$d_\phi^{1,1}(X,Y) = \int_{\mathbb{I}} (M(u,u) - C(u,u))du, \tag{7}$$

$$d_\tau^{1,1}(X,Y) = \int_{\mathbb{I}^2} M(u,v)dM(u,v) - \int_{\mathbb{I}^2} C(u,v)dC(u,v), \tag{8}$$

$$d_\rho^{1,1}(X,Y) = \int_{\mathbb{I}^2} (M(u,v) - C(u,v))dudv. \tag{9}$$

These dissimilarity functions are related, respectively, to the medial correlation coefficient (also known as Blomqvist's beta), the Spearman's footrule, the Kendall's tau, and Spearman's rho (see, e.g., Nelsen, 2006 and Durante and Sempi, 2016). Moreover, we also define

$$d_{\text{UTD}}^{1,1}(X,Y) = 1 - \lambda_U(C),$$
$$d_{\text{LTD}}^{1,1}(X,Y) = 1 - \lambda_L(C),$$

where λ_U and λ_L denote, respectively, the upper tail dependence (UTD) and lower tail dependence (LTD) coefficients (whenever they exist).

In particular, when choosing the dissimilarity function $d_\pi^{\max}$ in Eq. (3) (respectively, d_π^{ave} in Eq. (4)) with some specific $(1,1)$-dissimilarity function, the following result holds.

Corollary 4. *Given $d_\pi^{\max}$ in Eq. (3) (respectively, d_π^{ave} in Eq. (4)) with $d^{1,1} \in \{d_\phi^{1,1}, d_\tau^{1,1}, d_\rho^{1,1}\}$, $d_\pi^{\max}(\mathbf{X},\mathbf{Y}) = 0$ (respectively, $d_\pi^{\text{ave}}(\mathbf{X},\mathbf{Y}) = 0$) if, and only if, $(\mathbf{X},\mathbf{Y})$ is π–comonotonic.*

Proof. It follows from the fact that, for the above considered dissimilarity functions, $d^{1,1}(X,Y) = 0$ if, and only if, X and Y are coupled by M. □

Notice that the previous result does not apply to $d_\beta^{1,1}$ since $d_\beta^{1,1}(C)$ can be 0 also when C is not the comonotonic copula. Consider, for instance, the case when C is an ordinal sum (see, e.g., Durante and Sempi, 2016) of copulas with respect to the intervals $[0,1/2]$ and $[1/2,1]$.

The following examples show that, as $p \geq 2$, the proposed methodology based on π-comonotonicity is different from that of strong comonotonicity (Example 1) and from previous linkage-based dissimilarity (Example 2) considered in Fuchs et al. (2021).

Example 1. Given a $(1,1)$-dissimilarity function $d^{1,1}$, consider the dissimilarity $d_S^{\max}(\mathbf{X},\mathbf{Y})$ equal to the maximum of all values $d^{1,1}(Z_i,Z_j)$, where Z_i is any random variable in $(\mathbf{X},\mathbf{Y})$. Such a dissimilarity function is related to strong comonotonicity, since $d_S^{\max}(\mathbf{X},\mathbf{Y}) = 0$ if $(\mathbf{X},\mathbf{Y})$ is strongly comonotonic. Clearly, $d_S^{\max}$ need not be equal to $d_\pi^{\max}$. Consider, for instance, $d_\rho^{1,1}$ of Eq. (9). Consider a bivariate vector $\mathbf{X}$ with Frank copula C_θ ($\theta \in \mathbb{R}$), and $\mathbf{Y} = \mathbf{X}$ almost surely. Then $d_\pi^{\max}(X,Y) = 0 < d_S^{\max}(X,Y)$ for any $\theta \in \mathbb{R}$.

Example 2. Let $d^{1,1}$ be a $(1,1)$-dissimilarity function and let $d_\pi^{\max}$ be given in Eq. (3). For random pairs $\mathbf{X}$ and $\mathbf{Y}$ it can be easily checked that

$$d_\pi^{\max}(\mathbf{X},\mathbf{Y}) = \max\{d^{1,1}(X_1,Y_1), d^{1,1}(X_2,Y_2)\}.$$

Instead, if we consider $d_{\max}^{2,2}$, which is the dissimilarity measure based on complete linkage and on $d^{1,1}$ considered in Fuchs et al. (2021), then it holds

$$d_{\max}^{2,2}(\mathbf{X},\mathbf{Y}) = \max\{d^{1,1}(X_1,Y_1), d^{1,1}(X_2,Y_2), d^{1,1}(X_1,Y_2), d^{1,1}(X_2,Y_1)\}.$$

Thus, $d_\pi^{\max}$ and $d_{\max}^{2,2}$ are, in general, different. Consider, for instance, $d^{1,1} = d_\beta^{1,1}$ and the case when the copula of $(\mathbf{X},\mathbf{Y})$ is given by

$$C_{(\mathbf{X},\mathbf{Y})}(u_1,u_2,v_1,v_2) = M(u_1,v_1)M(u_2,v_2).$$

Then, $d_\pi^{\max} = 0$ and $d_{\max}^{2,2} = 1/4$.

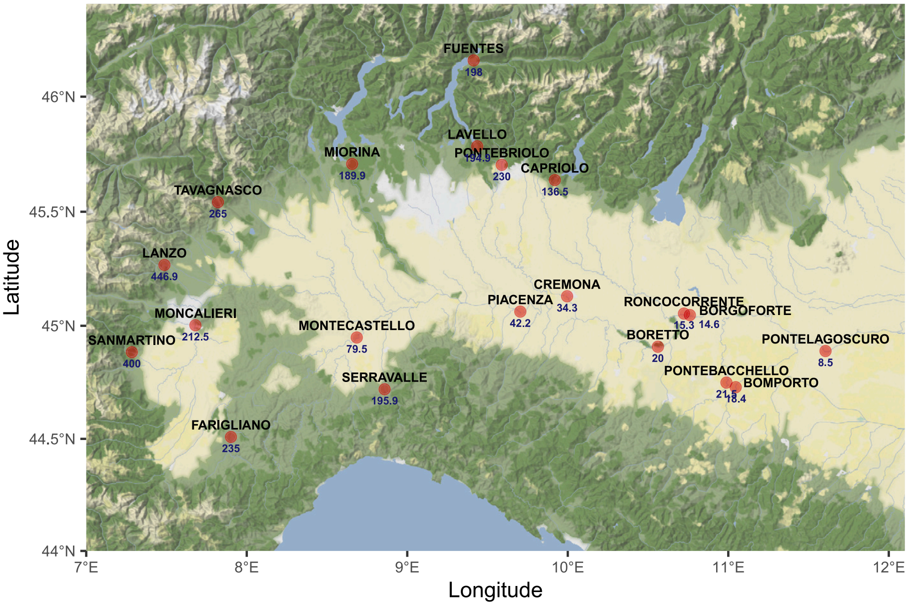

Fig. 1. Map of the stations of the Po river basin where the analyzed data were collected; elevation (m asl) for each station is reported on the map.

Remark 3. For each pair of continuous p–dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$, we can also consider the mapping $d_\pi(\mathbf{X},\mathbf{Y})$ given by the distance (for instance in the L^2 or L^∞ norm) between the copula $C_{(\mathbf{X},\mathbf{Y})}$ and the copula given by Eq. (2), where C is the average of the copulas $C_\mathbf{X}$ and $C_\mathbf{Y}$, i.e. $C = (C_\mathbf{X} + C_\mathbf{Y})/2$. It can be easily seen that such a mapping d_π is a dissimilarity measure in the sense of Definition 1. Clearly, the adoption of such a dissimilarity measure may not be desirable in practice, since it requires the estimation of a $(2p)$–dimensional copula rather than the estimation of p bivariate copulas as for the dissimilarities in Proposition 3.

Now, having chosen a way to calculate the dissimilarity among all the elements of $\mathscr{X}$, the matrix of pairwise dissimilarities can be used as an input for various clustering algorithms. Here, we prefer to use hierarchical agglomerative clustering, which provides an intuitive representation of the dependence among the involved vectors and does not require the number of clusters as input.

3. An illustration

We analyze the data used in Pappadà et al. (2018) consisting of discharge measurements on flood events recorded at $d = 20$ geographical sites spread across the Po river basin in Northern Italy in the period 1923–2008 (see Fig. 1). Data are recorded on a daily basis, and the realizations concern the occurrence of a flood episode. A flood episode is characterized by a discharge that exceeds the "Q300" threshold (alarm threshold in standard guidelines) and involves several dependent variables. Here, we consider the flood peak (in m^3/s), P, and the flood volume (in 10^6 m^3), V. Moreover, as is typical in hydrology, we assume that successive flood episodes are separated by at least 3 days (the discharge falls short of the considered threshold). For each station S_j, $j \in \{1, \dots, 20\}$, the observed events are realizations of the random vector $\mathbf{X}_j = (P_j, V_j)$, where P_j and V_j are time series of length t_j.

To build the $d \times d$ dissimilarity matrix, we proceed as follows:

(i) For each pair of stations, we first extract the observations of temporally concurrent episodes that can be assumed as generated by the same meteorological event over the Po river basin, acting at (short) different times at (close) spatially separated sites. We empirically set the temporal interval for potential overlap equal to 5 days.
(ii) We compute pairwise dissimilarities for each pair $\mathbf{X}_j = (P_j, V_j)$ and $\mathbf{X}_{j'} = (P_{j'}, V_{j'})$, $j, j' \in \{1, \dots, 20\}$, by using Eq. (3)–(5), where $d^{1,1}$ is the dissimilarity function for the pairs $(P_j, P_{j'})$ and $(V_j, V_{j'})$ defined as in Eq. (6)–(9), where the copula C is estimated nonparametrically via the empirical copula (see, e.g. Fuchs and Tschimpke, 2024).

The dissimilarity matrices obtained in step (ii) can be used as input for any hierarchical agglomerative algorithm with a specified linkage method. Here, for instance, we run the `agnes` algorithm (see Kaufman and Rousseeuw (1990, chapter 5) and the R package cluster by Maechler et al., 2023) with complete and average linkages (the single linkage is discarded because of its poor performance).

In what follows, we assume that the clustering is performed by adopting the dissimilarity $d_\pi^{\max}$ (Eq. (3)) combined with $d_\tau^{1,1}$ (Eq. (8)), where the algorithm uses complete and average linkage, respectively. The number of clusters, K, may be chosen via classical validation statistics. In particular, among various indices, we consider a Dunn-like index computed as the ratio of minimum average dissimilarity between two clusters and maximum average within cluster dissimilarity (see Hennig, 2023).

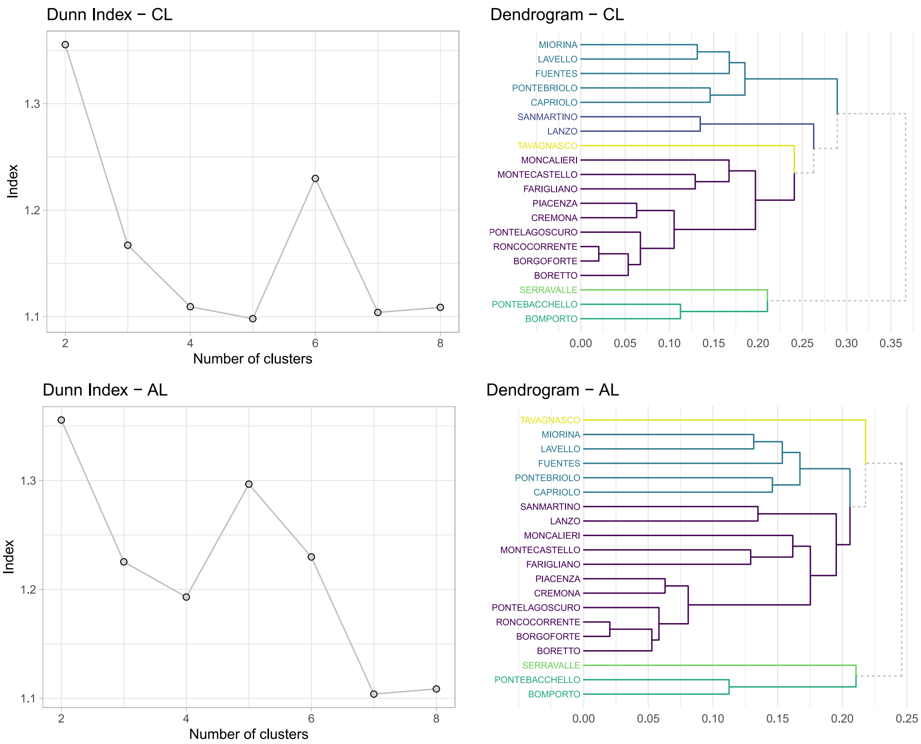

Fig. 2. Dunn Index and dendrogram of the hierarchical clustering of flood data based on $d_\pi^{\max}$ and $d_\tau^{1,1}$, with complete linkage (CL, top) and average linkage (AL, bottom). The branches and labels of the two trees are colored according to the partition into $K = 6$ and $K = 5$ clusters, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Cluster members in the partitions based on $d_\pi^{\max}$ and $d_\tau^{1,1}$, with complete linkage ($K = 6$ groups) and average linkage ($K = 5$ groups), respectively (see Fig. 2).

Cluster	$d_\pi^{\max}$, $d_\tau^{1,1}$, CL
1	BOMPORTO, PONTEBACCHELLO
2	BORETTO, BORGOFORTE, CREMONA, FARIGLIANO, MONCALIERI, MONTECASTELLO, PIACENZA, PONTELAGOSCURO, RONCOCORRENTE
3	CAPRIOLO, FUENTES, LAVELLO, MIORINA, PONTEBRIOLO
4	LANZO, SANMARTINO
5	SERRAVALLE
6	TAVAGNASCO
Cluster	$d_\pi^{\max}$, $d_\tau^{1,1}$, AL
1	BOMPORTO, PONTEBACCHELLO
2	BORETTO, BORGOFORTE, CREMONA, FARIGLIANO, LANZO, MONCALIERI, MONTECASTELLO, PIACENZA, PONTELAGOSCURO, RONCOCORRENTE, SANMARTINO
3	CAPRIOLO, FUENTES, LAVELLO, MIORINA, PONTEBRIOLO
4	SERRAVALLE
5	TAVAGNASCO

Remark 4. A fast implementation of $d_\tau^{1,1}$ in Eq. (8) is provided by the R function `cor.fk` from the R package pcaPP (Filzmoser et al., 2024), enabling a computation time of $O(n\,log(n))$ instead of the typical $O(n^2)$.

Fig. 2 reports the value of such index for the partitions into K clusters, $K \in \{2, \dots, 8\}$, where complete and average linkages were used, respectively. Note that a higher index indicates a better clustering in terms of compactness and separation of the clusters. As can be seen from Fig. 2, $K = 2$ is the value that maximizes the index in both cases; however, the solutions with $K = 6$ and $K = 5$, respectively, may be adopted to avoid a too coarse separation into two groups. Fig. 2 also displays the associated dendrograms, were the branches and labels of the trees are colored according to the resulting groups.

The 6- and 5-cluster solutions based on the complete linkage and average linkage, respectively, are shown on the map in Fig. 3 and in Table 1.

As can be seen, complete and average linkage methods show a large agreement and, on this dataset, both allow us to identify clusters of stations that are also spatially coherent. When adopting complete linkage, the largest central cluster is split into two

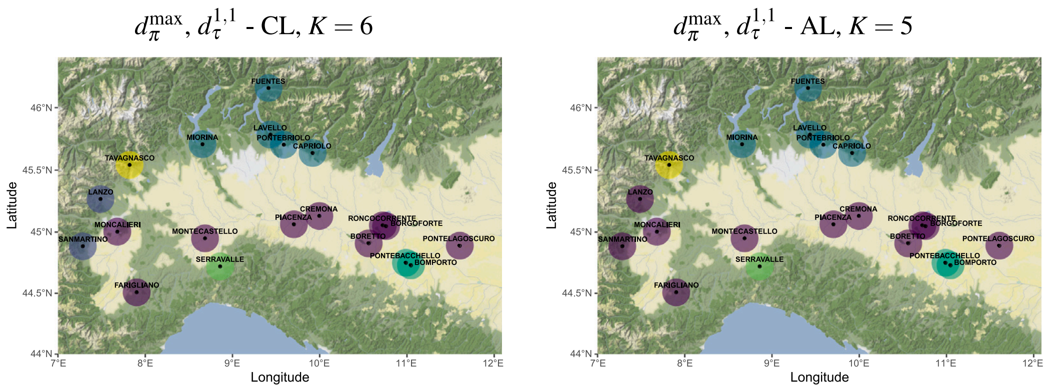

Fig. 3. Map of clusters obtained via agglomerative hierarchical clustering based on $d_\pi^{\max}$ and $d_\tau^{1,1}$, with complete linkage (CL, left) and average linkage (AL, right).

groups, resulting in the formation of a cluster consisting only of stations Lanzo and San Martino, characterized by the two largest elevation values (see Fig. 1). Notice that the same clusters are obtained when the dissimilarity measure is based on Spearman's ρ.

For comparison purposes, we also consider two other methods that are grounded on the strong notion of multivariate comonotonicity, introduced in Pappadà et al. (2018) and denoted by d_S^{Kendall} and d_S^{AND}, respectively.

1. First, we consider the dissimilarity function $d_S^{\text{Kendall}}(\mathbf{X},\mathbf{Y})$ that is based on the L^2 distance of the Kendall distribution of the comonotonicity copula M and the Kendall distribution of the copula $C_{(\mathbf{X},\mathbf{Y})}$.
2. Second, we consider $d_S^{\text{AND}}(\mathbf{X},\mathbf{Y})$ that is the L^2 distance between the diagonal section of the comonotonicity copula M and the diagonal section of the copula $C_{(\mathbf{X},\mathbf{Y})}$.

These dissimilarity functions are related to the concept of Hazard Scenarios (HS), namely the Kendall and the AND ones, which are commonly used to study multivariate events. Events of interest under the Kendall HS are those exceeding a level set $\mathscr{L}_t$, which can be regarded as a critical multivariate threshold; for an event of the AND HS, it is necessary that all the variables exceed a thresholds given by a high quantile (see Salvadori et al., 2016 for further details).

For both dissimilarity functions $d_S(\mathbf{X},\mathbf{Y}) = 0$ if, and only if, $(\mathbf{X},\mathbf{Y})$ is strongly comonotonic. In fact, the diagonal section (respectively, the Kendall function) completely characterizes the copula in the comonotonic case (see, for instance, Durante and Fernández-Sánchez, 2011; Nelsen et al., 2003 and Nelsen et al., 2009).

Fig. 4 shows the dendrogram resulting from agglomerative hierarchical clustering with complete and average linkage methods with dissimilarity function $d_\pi^{\max}$ combined with $d_\tau^{1,1}$, d_S^{Kendall} and d_S^{AND}, respectively.

As expected, the kind of dissimilarity measure appears to have an impact on the final partition, although there are subgroups of stations that are allocated in the same cluster irrespective of the dissimilarity measure, denoting a stronger similarity that is preserved in the dendrogram structure for all the compared methods. We also note that $d_\pi^{\max}$ produces a dendrogram that demonstrates greater agreement with trees based on d_S^{AND} compared to d_S^{Kendall}.

This tendency is confirmed by Table 2, reporting Goodman and Kruskal's Gamma (γ) coefficient (Kruskal and Goodman, 1954) that measures the similarity between two dendrograms for the considered combinations of dissimilarity functions and linkage methods (such coefficient ranges from -1 to $+1$, where the latter value indicates perfect agreement between the two classifications, see Baker, 1974). Within the pairs sharing the same linkage, the most dissimilar trees are those based on π-comonotonicity and the Kendall's strong comonotonicity measure; moreover, such dissimilarity is stronger in the average linkage case ($\gamma = 0.798$) than under the complete one ($\gamma = 0.86$). As a result, the partitions based on the proposed $d_\pi^{\max}$ measure and on the Kendall's HS derived by cutting the associated dendrograms at a suitable height, are expected to present relevant differences, which can be explained by the fact that the notion of π-comonotonicity is less restrictive than classical comonotonicity.

4. Conclusions

We have introduced a set of dissimilarity measures for random vectors to be used in clustering procedures that extend the classical approach for pairs of random variables. Such measures exploit a suitable characterization of the comonotonicity between two random vectors in terms of their copula and, in particular, we have focused on the notion of π-comonotonicity. A major advantage of such a measure is that it already assigns the minimum value zero to a π-comonotonic vectors and no longer only when the entire vector behaves strictly comonotonic.

By reviewing their properties, we have considered and compared the dissimilarity functions grounded on the notion of π-comonotonicity and strong-comonotonicity in the framework of hierarchical agglomerative clustering. Some insights were provided for the use of such measures and an application to the spatial clustering of flooding discharge measurements has been developed to illustrate how the proposed methodology can be used to investigate compound phenomena. Given the ability of the proposed approach to identify spatial clusters based on the comonotonic behavior of the involved variables, the resulting clusters may represent a key information for determining areas where compound events may occur simultaneously.

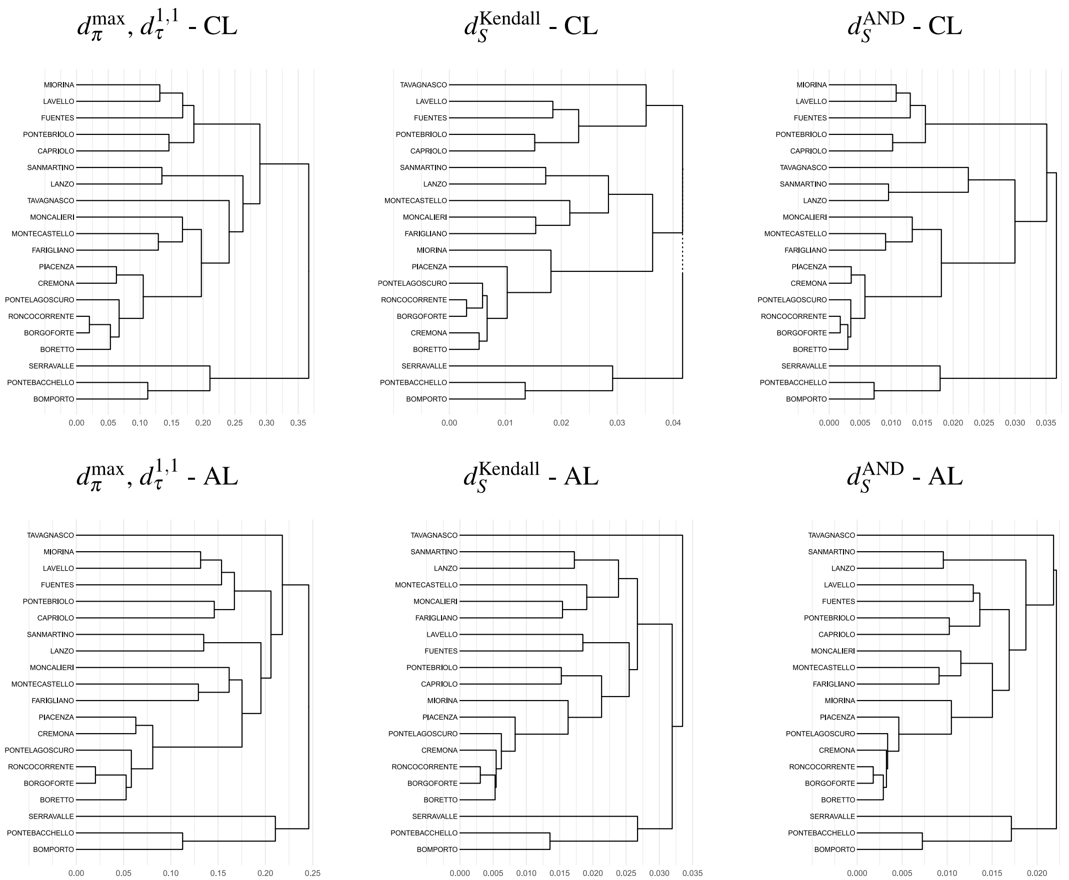

Fig. 4. Dendrograms obtained via agglomerative hierarchical clustering with complete linkage (CL) and average linkage (AL), based on π-comonotonicity ($d_\pi^{\max}$, $d_\tau^{1,1}$) and strong comonotonicity (d_S^{Kendall}, d_S^{AND}), respectively.

Table 2
Goodman and Kruskal's Gamma assessing the similarity between hierarchical clustering based on π-comonotonicity ($d_\pi^{\max}$, $d_\tau^{1,1}$) and strong comonotonicity (d_S^{Kendall}, d_S^{AND}), with average linkage (AL) and complete linkage (CL).

	$d_\pi^{\max}$, $d_\tau^{1,1}$ AL	d_S^{Kendall} AL	d_S^{AND} AL	$d_\pi^{\max}$, $d_\tau^{1,1}$ CL	d_S^{Kendall} CL	d_S^{AND} CL
$d_\pi^{\max}$, $d_\tau^{1,1}$ - AL	1.0000	0.7981	0.8936	0.9178	0.8710	0.9217
d_S^{Kendall} - AL		1.0000	0.8429	0.5947	0.6997	0.6053
d_S^{AND} - AL			1.0000	0.8068	0.8463	0.8066
$d_\pi^{\max}$, $d_\tau^{1,1}$ - CL				1.0000	0.8600	0.9975
d_S^{Kendall} - CL					1.0000	0.8689
d_S^{AND} - CL						1.0000

Acknowledgments

The authors thank the Reviewers for their thorough and constructive feedback, which greatly improved the quality of this manuscript. FD has been partially supported by MUR-PRIN 2022 PNRR, Project "Stochastic Modeling of Compound Events" (No. P2022KZJTZ) funded by European Union – Next Generation EU. SF gratefully acknowledges the support of the Austrian Science Fund (FWF) project P 36155-N "ReDim: Quantifying Dependence via Dimension Reduction" and the support of the WISS 2025 project "IDA-lab Salzburg" (20204-WISS/225/197-2019 and 20102-F1901166-KZP). RP has been supported by MUR-PRIN 2022, Project "Modelling Non-standard data and Extremes in Multivariate Environmental Time series" (No. 20223CEZSR) funded by European Union – Next Generation EU.

Appendix

We denote by $\mathbf{0}$ the vector with all entries equal to 0, by $\mathbf{1}$ the vector with all entries equal to 1 and by Q^C the probability measure associated with the copula C.

The following technical result will be needed in the following and it is reported here.

Lemma 5.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two random vectors in $L^0(\mathbb{I}^p)$ with copula $C_{(\mathbf{X},\mathbf{Y})}$. For all $i \in \{1, \dots, p\}$, suppose $C_{(X_i,Y_i)} = M$. Then:

(i) $Q^{C_{(\mathbf{X},\mathbf{Y})}}\left(\left\{(\mathbf{u},\mathbf{v}) \in \mathbb{I}^p \times \mathbb{I}^p \,\middle|\, \mathbf{u} = \mathbf{v}\right\}\right) = 1.$

(ii) The identity

$$\int_{\mathbb{I}^p\times\mathbb{I}^p} f(\mathbf{u},\mathbf{v})\, \mathrm{d}Q^{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{u},\mathbf{v}) = \int_{\mathbb{I}^p} f(\mathbf{u},\mathbf{u})\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{u}) = \int_{\mathbb{I}^p} f(\mathbf{v},\mathbf{v})\, \mathrm{d}Q^{C_\mathbf{Y}}(\mathbf{v})$$

holds for every measurable function $f : \mathbb{I}^p \times \mathbb{I}^p \to \mathbb{R}$.

(iii) $C_\mathbf{X} = C_\mathbf{Y}$.

Proof. We first obtain

$$\begin{aligned}
0 &\le Q^{C_{(\mathbf{X},\mathbf{Y})}}\left(\bigcup_{i=1}^{p}\left\{(\mathbf{u},\mathbf{v}) \in \mathbb{I}^p \times \mathbb{I}^p \,\middle|\, u_i \neq v_i\right\}\right)\\
&\le \sum_{i=1}^{p} Q^{C_{(\mathbf{X},\mathbf{Y})}}\left(\left\{(\mathbf{u},\mathbf{v}) \in \mathbb{I}^p \times \mathbb{I}^p \,\middle|\, u_i \neq v_i\right\}\right)\\
&= \sum_{i=1}^{p} Q^{C_{(X_i,Y_i)}}\left(\left\{(u_i,v_i) \in \mathbb{I} \times \mathbb{I} \,\middle|\, u_i \neq v_i\right\}\right)\\
&= \sum_{i=1}^{p} Q^{M}\left(\left\{(u_i,v_i) \in \mathbb{I} \times \mathbb{I} \,\middle|\, u_i \neq v_i\right\}\right) = 0
\end{aligned}$$

which proves (i). Applying (i) and disintegration and denoting by $K_{C_{(\mathbf{X},\mathbf{Y})}}(.,.)$ the regular conditional distribution corresponding to the copula $C_{(\mathbf{X},\mathbf{Y})}$ then gives

$$\begin{aligned}
\int_{\mathbb{I}^p\times\mathbb{I}^p} f(\mathbf{u},\mathbf{v})\, \mathrm{d}Q^{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{u},\mathbf{v}) &= \int_{\mathbb{I}^p\times\mathbb{I}^p} f(\mathbf{u},\mathbf{u})\, \mathrm{d}Q^{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{u},\mathbf{v})\\
&= \int_{\mathbb{I}^p}\int_{\mathbb{I}^p} f(\mathbf{u},\mathbf{u})\, K_{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{u},\mathrm{d}\mathbf{v})\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{u})\\
&= \int_{\mathbb{I}^p} f(\mathbf{u},\mathbf{u})\, K_{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{u},\mathbb{I}^p)\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{u})\\
&= \int_{\mathbb{I}^p} f(\mathbf{u},\mathbf{u})\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{u})
\end{aligned}$$

We refer to Kasper et al. (2021) and Mroz et al. (2021) for more background on a copula's regular conditional distribution. Finally, (ii) yields

$$\begin{aligned}
C_\mathbf{X}(\mathbf{u}) = C_{(\mathbf{X},\mathbf{Y})}(\mathbf{u},\mathbf{1}) &= \int_{\mathbb{I}^p\times\mathbb{I}^p} \mathbb{1}_{[\mathbf{0},\mathbf{u}]\times[\mathbf{0},\mathbf{1}]}(\mathbf{s},\mathbf{t})\, \mathrm{d}Q^{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{s},\mathbf{t})\\
&= \int_{\mathbb{I}^p} \mathbb{1}_{[\mathbf{0},\mathbf{u}]\times[\mathbf{0},\mathbf{1}]}(\mathbf{t},\mathbf{t})\, Q^{C_\mathbf{Y}}(\mathbf{t}) = C_\mathbf{Y}(\mathbf{u})
\end{aligned}$$

for all $\mathbf{u} \in \mathbb{I}^p$. □

Proof of Theorem 2. It follows from Puccetti and Scarsini (2010, Theorem 3.7), that (a) is equivalent to (b) which implies (c). Now, assume that (c) holds. Then, by Lemma 5, $C_\mathbf{X} = C_\mathbf{Y}$ and

$$\begin{aligned}
C_{(\mathbf{X},\mathbf{Y})}(\mathbf{u},\mathbf{v}) &= \int_{\mathbb{I}^p\times\mathbb{I}^p} \mathbb{1}_{[\mathbf{0},\mathbf{u}]\times[\mathbf{0},\mathbf{v}]}(\mathbf{s},\mathbf{t})\, \mathrm{d}Q^{C_{(\mathbf{X},\mathbf{Y})}}(\mathbf{s},\mathbf{t})\\
&= \int_{\mathbb{I}^p} \mathbb{1}_{[\mathbf{0},\mathbf{u}]\times[\mathbf{0},\mathbf{v}]}(\mathbf{s},\mathbf{s})\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{s})\\
&= \int_{\mathbb{I}^p} \prod_{i=1}^{p} \mathbb{1}_{[0,M(u_i,v_i)]}(s_i)\, \mathrm{d}Q^{C_\mathbf{X}}(\mathbf{s})\\
&= C_\mathbf{X}\big(M(u_1,v_1),\dots,M(u_p,v_p)\big)
\end{aligned}$$

for all $(\mathbf{u},\mathbf{v}) \in \mathbb{I}^p \times \mathbb{I}^p$. This proves the result. □

References

Alonso, A.M., D'Urso, P., Gamboa, C., Guerrero, V., 2021. Cophenetic-based fuzzy clustering of time series by linear dependency. Int. J. Approx. Reason. 137, 114–136.

Alonso, A.M., Peña, D., 2019. Clustering time series by linear dependency. Stat. Comp. 29, 655–676.

Baker, F.B., 1974. Stability of two hierarchical grouping techniques case I: sensitivity to data errors. J. Amer. Statist. Assoc. 69 (346), 440–445.

Benevento, A., Durante, F., Pappadà, R., 2024. Tail-dependence clustering of time series with spatial constraints. Environ. Ecol. Stat. 31, 801–817.

Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A.F.S., Ramos, A.M., Vignotto, E., Bastos, A., Blesić, S., Durante, F., Hillier, J., Oliveira, S.C., Pinto, J.G., Ragno, E., Rivoire, P., Saunders, K., van der Wiel, K., Wu, W., Zhang, T., Zscheischler, J., 2021. Guidelines for studying diverse types of compound weather and climate events. Earth's Futur. 9 (11), e2021EF002340.

Caiado, J., Maharaj, E., D'Urso, P., 2016. Time-series clustering. In: Handbook of Cluster Analysis. CRC Press, Boca Raton, FL, pp. 241–263.

De Keyser, S., Gijbels, I., 2024. Hierarchical variable clustering via copula-based divergence measures between random vectors. Int. J. Approx. Reason. 165, 22, Id/No 109090.

De Luca, G., Zuccolotto, P., 2011. A tail dependence-based dissimilarity measure for financial time series clustering. Adv. Data Anal. Classif. 5 (4), 323–340.

De Luca, G., Zuccolotto, P., 2017. Dynamic tail dependence clustering of financial time series. Stat. Pap. 58, 641–657.

Di Lascio, F., Durante, F., Pappadà, R., 2017. Copula–based clustering methods. In: Úbeda Flores, M., de Amo, E., Durante, F., Fernández Sánchez, J. (Eds.), Copulas and Dependence Models with Applications. Springer International Publishing, Cham, pp. 49–67.

Di Lascio, F., Menapace, A., Pappadà, R., 2023. A spatially-weighted AMH copula-based dissimilarity measure for clustering variables: An application to urban thermal efficiency. Env. e2828.

Disegna, M., D'Urso, P., Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. Spat. Stat. 21 (part A), 209–225.

Durante, F., Fernández-Sánchez, J., 2011. On the classes of copulas and quasi-copulas with a given diagonal section. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 19 (1), 1–10.

Durante, F., Pappadà, R., Torelli, N., 2015. Clustering of time series via non–parametric tail dependence estimation. Stat. Pap. 56 (3), 701–721.

Durante, F., Sempi, C., 2016. Principles of Copula Theory. CRC Press, Boca Raton, FL.

D'Urso, P., De Luca, G., Vitale, V., Zuccolotto, P., 2023. Tail dependence-based fuzzy clustering of financial time series. Ann. Oper. Res. http://dx.doi.org/10.1007/s10479-023-05744-1.

Filzmoser, P., Fritz, H., Kalcher, K., 2024. pcaPP: Robust PCA by projection pursuit. URL https://CRAN.R-project.org/package=pcaPP. R package version 2.0-5.

Fuchs, S., Di Lascio, F.M.L., Durante, F., 2021. Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Comput. Statist. Data Anal. 107201.

Fuchs, S., Tschimpke, M., 2024. A novel positive dependence property and its impact on a popular class of concordance measures. J. Multivariate Anal. 200, 21, Id/No 105259.

Fuchs, S., Wang, Y., 2024. Hierarchical variable clustering based on the predictive strength between random vectors. Int. J. Approx. Reason. 170, 25, Id/No 109185.

Hennig, C., 2023. Fpc: Flexible procedures for clustering. URL https://CRAN.R-project.org/package=fpc. R package version 2.2-10.

Kasper, T., Fuchs, S., Trutschnig, W., 2021. On weak conditional convergence of bivariate archimedean and extreme value copulas, and consequences to nonparametric estimation. Bernoulli 27, 2217–2240.

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

Kruskal, W.H., Goodman, L., 1954. Measures of association for cross classifications. J. Amer. Statist. Assoc. 49 (268), 732–764.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2023. Cluster: Cluster analysis basics and extensions. URL https://CRAN.R-project.org/package=cluster. R package version 2.1.6.

Maharaj, E.A., D'Urso, P., Caiado, J., 2019. Time Series Clustering and Classification. CRC Press, Boca Raton, FL.

Mroz, T., Fuchs, S., Trutschnig, W., 2021. How simplifying and flexible is the simplifying assumption in pair-copula constructions – analytic answers in dimension three and a glimpse beyond. Electron. J. Stat. 15 (1), 1951–1992.

Murtagh, F., Contreras, P., 2012. Algorithms for hierarchical clustering: an overview. WIRE S Data Min. Knowl. Discov. 2 (1), 86–97.

Murtagh, F., Contreras, P., 2017. Algorithms for hierarchical clustering: an overview, II. WIRE S Data Min. Knowl. Discov. 7 (6), e1219.

Nelsen, R.B., 2006. An Introduction to Copulas, second ed. Springer Ser. Stat., Springer, New York, NY.

Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M., 2003. Kendall distribution functions. Statist. Probab. Lett. 65 (3), 263–268.

Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M., 2009. Kendall distribution functions and associative copulas. Fuzzy Sets and Systems 160 (1), 52–57.

Neumeyer, N., Omelka, M., Hudecová, S., 2019. A copula approach for dependence modeling in multivariate nonparametric time series. J. Multivariate Anal. 171, 139–162.

Pappadà, R., Durante, F., Salvadori, G., De Michele, C., 2018. Clustering of concurrent flood risks via hazard scenarios. Spat. Stat. 23, 124–142.

Patton, A., 2012. A review of copula models for economic time series. J. Multivariate Anal. 110, 4–18.

Puccetti, G., 2022. Measuring linear correlation between random vectors. Inform. Sci. 607, 1328–1347.

Puccetti, G., Scarsini, M., 2010. Multivariate comonotonicity. J. Multivariate Anal. 101 (1), 291–304.

Puccetti, G., Wang, R., 2015. Extremal dependence concepts. Stat. Sci. 30 (4), 485–517.

Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R., 2007. Extremes in Nature. An Approach Using Copulas. Water Science and Technology Library, vol. 56, Springer, Dordrecht (NL).

Salvadori, G., Durante, F., De Michele, C., Bernardi, M., Petrella, L., 2016. A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. Water Resour. Res. 52 (5), 3701–3721.

Saunders, K.R., Stephenson, A.G., Karoly, D.J., 2021. A regionalisation approach for rainfall based on extremal dependence. Extremes 24 (2), 215–240.

Vignotto, E., Engelke, S., Zscheischler, J., 2021. Clustering bivariate dependencies of compound precipitation and wind extremes over great britain and Ireland. Weather. Clim. Extrem. 32, 100318.

Viroli, C., 2011. Model based clustering for three-way data structures. Bayesian Anal. 6 (4), 573–602.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R.M., van den Hurk, B., AghaKouchak, A., Jźq́uel, A., Mahecha, M.D., Maraun, D., Ramos, A.M., Ridder, N.N., Thiery, E., 2020. A typology of compound weather and climate events. Nat. Rev. Earth Env. 1 (7), 333–347.

Zscheischler, J., Westra, S., van den Hurk, B.J.J.M., Seneviratne, S.I., Ward, P.J., Pitman, A., AghaKouchak, A., Bresch, D.N., Leonard, M., Wahl, T., Zhang, X., 2018. Future climate risk from compound events. Nature Clim. Change 8 (6), 469–477.