Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Comonotonic-Based Time Series Clustering With Constraints: A Review and a Conceptual Framework

Alessia Benevento[1] | Fabrizio Durante[1] | Roberta Pappadà[2]

[1]Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Lecce, Italy | [2]Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche "B. de Finetti", Università degli Studi di Trieste, Trieste, Italy

Correspondence: Roberta Pappadà (rpappada@units.it)

Received: 7 April 2025 | **Revised:** 31 July 2025 | **Accepted:** 9 October 2025

Funding: This work was supported by MUR-PRIN 2022 PNRR, Project "Stochastic Modeling of Compound Events" (No. P2022KZJTZ), and by MUR-PRIN 2022, Project "Modelling Non-standard data and Extremes in Multivariate Environmental Time series" (No. 20223CEZSR), both funded by European Union – Next Generation EU. Partial financial support from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by EU – Next Generation EU (CUP F83C22000740001) is also acknowledged.

ABSTRACT

Time series clustering is a widely used unsupervised learning approach that identifies groups of similar time series to uncover hidden patterns in complex datasets. In recent years, this technique has gained traction in the analysis of geo-referenced time series, where spatial information must be incorporated into the dissimilarity measure to achieve meaningful results. This paper offers a thorough review of dissimilarity-based clustering methods with soft spatial constraints, i.e., approaches that integrate spatial context into the clustering process without enforcing strict spatial proximity within clusters. Our focus is on copula-based clustering techniques, which effectively capture comovements among time series without requiring explicit modeling of their marginal distributions. We first introduce a general framework for copula-based time series clustering and then explore how spatial constraints can be embedded into the clustering process. Finally, we propose a general framework, called TRIPLE-C, which provides two comprehensive model architectures that address this challenge through either a dissimilarity fusion step or a copula aggregation approach.

1 | Introduction

Advancements in data storage and acquisition technologies have led to a significant surge in the collection of large-scale datasets. Time series data, which consist of one or more variables (such as satellite images) that change over time, are widely recorded and analyzed across diverse fields, including environmental science, engineering, medicine, economics, and finance. In the absence of sufficient prior knowledge about group structures, clustering serves as a powerful data mining technique to classify these temporal datasets into meaningful related groups. For a review, see Hennig et al. (2015) and Aghabozorgi et al. (2015).

In environmental sciences, various statistical methods have been developed to simplify the complex spatio-temporal variability of variables by summarizing them into a few key statistical quantities that capture the current state and future trends of the climate system (see, e.g., Straus 2018). In particular, clustering methods are commonly employed to identify regions where time series behave similarly (according to a given criterion).

The clustering process typically relies on a well-defined notion of dissimilarity between time series (Liao 2005; Maharaj et al. 2019), which is commonly established through features such as autocorrelation, partial autocorrelation, periodograms, quantiles, and

so forth (Caiado et al. 2006; Pértega Díaz and Vilar 2010; Maharaj and D'Urso 2011; Musau et al. 2022; Gaetan et al. 2025). Alternatively, in a model-based setting, the time series are assumed to be generated by a stochastic process, and the clustering is based on the idea that time series produced by the same model are likely to exhibit similar patterns. Thus, time series are typically grouped according to parameter estimates or residuals from fitted models. Specific methods include ARMA/ARIMA models, GARCH models, and so forth. Examples are discussed by Piccolo (1990), Kalpakis et al. (2001), Caiado and Crato (2010), D'Urso et al. (2013). Novel approaches that treat time series as functional data are also of particular interest, especially in climatological science (see, for instance, Giraldo et al. 2012; Haggarty et al. 2015).

Here, as stressed by Davis et al. (2023), we focus on dissimilarity measures that capture the dependence structure between individual components, rather than just their marginal distributions. In environmental sciences, for instance, such methods have been introduced to detect comovements of time series, especially related to joint extremes, for example, maxima of precipitations (Bernard et al. 2013), temperature (Bador et al. 2015), or to model flood risks (Pappadà et al. 2018). Multivariate extensions have also been recently presented by Boulin et al. (2025), Durante et al. (2025), and Vignotto et al. (2021). These methods utilize a rank-invariant dissimilarity measure that depends on the proximity of the copula to the comonotonicity case (i.e., perfect positive dependence).

Clustering methods for time series are commonly employed in unsupervised learning scenarios (Hennig et al. 2015), where the aim is to uncover hidden structures in the data. However, the problem of identifying regions with common temporal patterns is usually complicated by the spatial dimension. In fact, when general-purpose clustering methods are used for partitioning geo-referenced time series, the resulting clusters are scattered over the spatial domain of the study, and this even if spatial coordinates are considered as attributes (Fouedjio 2020). Thus, numerous studies have emphasized the importance of including spatial information for geographically referenced data since, in these scenarios, forming clusters that also reflect geographical proximity can significantly improve the interpretability of the results (Fouedjio 2020; Kopczewska 2022).

Copula-based clustering methods with constraints have been introduced by Disegna et al. (2017) to integrate in the clustering process the information related to (non-temporal) proximity among time series, which may be derived from spatial information or other covariates. Notably, these methods do not require that the resulting clusters strictly adhere to proximity constraints. Rather, the clustering process can, but does not have to, group variables that are geographically distant but otherwise similar (Romary et al. 2015). This flexibility is due to the so-called soft proximity constraints (Chavent et al. 2018), which differ from hard constraints that have been considered, for example, by Pawitan and Huang (2003) and Guénard and Legendre (2022).

The focus of this paper is on copula-based clustering methods with soft constraints. In particular, it aims to consolidate the recent knowledge on such methods. The various methodologies on the subject are collected and systematically organized, considering both past and recent literature on spatial clustering (see Oliver and Webster 1989; Fouedjio 2020). This leads to the definition of a general framework for clustering time series based on three paradigms: (C1) the use of copulas to focus on the dependence among time series; (C2) the detection of comovements that are interpreted as proximity of the pairwise copula between two time series to the comonotonicity case; (C3) the presence of some constraints that may guide the clustering process from a semi-supervised viewpoint. Thus, this framework is named TRIPLE-C, since it focuses on three main aspects: Copula, Comonotonicity, and Constraints. The main architecture of the TRIPLE-C algorithm is introduced, and several new examples are produced to guide the interested reader in the choice of their favorite variants according to specific domain knowledge needs.

The manuscript is organized as follows. Section 2 contains a review of copula-based time series clustering without any constraints. The TRIPLE-C algorithm is hence illustrated in Section 3 together with its two main variants. Section 4 provides an illustration of the algorithm with synthetic data as well as with climatological data. Section 5 concludes with some final remarks and comments on future developments.

2 | The Conceptual Framework for Copula-Based Time Series Clustering

We start by setting the notation and formally reviewing the main aspects of copula-based time series clustering in the unsupervised setting.

For every natural $n \geq 2$, let $\mathcal{X}$ be a set of n objects. A clustering algorithm is any procedure that divides these objects into K ($2 \leq K \leq n$) non-empty subsets $G_1, \ldots, G_K$, called *clusters*, so that

$$\bigcup_{k=1}^{K} G_k = \mathcal{X} \quad \text{and} \quad G_k \cap G_{k'} = \emptyset, \text{ for } k \neq k'.$$

Each partition in K clusters is represented by a membership matrix $\mathbf{U}$ of order $(K \times n)$ so that each entry U_{ki} belongs to $\{0, 1\}$ and the sum of the entries in each column is 1. Each U_{ki} indicates whether the i-th object belongs to the cluster G_k (value equals 1) or not (value equals 0). The set of such matrices is denoted by $\text{Memb}(K, n)$. Notice that, in the case of fuzzy clustering (Ruspini et al. 2019), any $U_{ki} \in [0, 1]$ indicates the membership degree of the i-th object to the k-th cluster.

A number of clustering procedures (Hennig et al. 2015) determine the cluster partition from a matrix that collects the (pairwise) degree of dissimilarity among the n objects of interest. As is well known, such a matrix evaluates how distinct or different the objects are from one another. We denote by $\text{Diss}(n)$ the set of all dissimilarity matrices, that is, $(n \times n)$ symmetric matrices with non-negative entries and diagonal entries equal to 0. Moreover, we call *dissimilarity-based algorithm* (DBA, for short) any mapping Alg that associates to each $\boldsymbol{\Delta} \in \text{Diss}(n)$ a suitable partition of the objects into K clusters, that is, $\text{Alg}(\boldsymbol{\Delta}) = \mathbf{U} \in \text{Memb}(K, n)$.

When the objects to be clustered are components of a time series, copula-based methods can be adopted to group those series components that tend to comove (see Di Lascio et al. (2024) and references therein). These methods typically start with the set

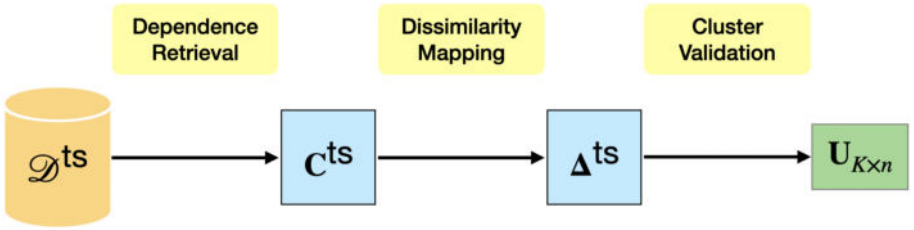

FIGURE 1 | Copula-based time series clustering: Model architecture.

of observations $\mathcal{D}^{ts} = \{(x_{t1}, x_{t2}, \ldots, x_{tn}), t = 1, \ldots, T\}$ that can be considered as a sample (of size T) generated from a stochastic model $(\mathbf{X}_t)_{t\in\mathbb{N}}$. It represents the time series observations about the variables to be clustered, the so-called *temporal data*. Typically, the set $\mathcal{D}^{ts}$ is transformed into a dissimilarity matrix. Such a matrix is, hence, used as input of the DBA to obtain a partition of the time series represented by the membership matrix $\mathbf{U}$. Before being converted into a dissimilarity matrix, the temporal data can be conveniently treated to extract relevant information about the dependence; this latter aspect is usually captured by a matrix of pairwise (empirical) copulas that is associated with the time series. These steps are represented in Figure 1 and are detailed in the next subsections.

Before analyzing each step of the algorithm, it is important to emphasize that the dissimilarity-based algorithm (DBA) must be chosen a priori and serves as an input to the process. The selection of a suitable clustering method is, in fact, inherently tied to the specific objectives of the clustering task. Below, we outline several clustering methods discussed in the literature and examine how they align with the desirable properties typically expected in copula-based approaches. The commonly adopted methods include:

Hierarchical algorithms. They are typically employed when a complete hierarchy of clusters is desired—often with the intent of extracting a single partition by cutting the hierarchy at a certain level. As is well known, different hierarchical methods can lead to significantly different clustering outcomes. For example, single linkage emphasizes separation by ensuring that the closest points from different clusters remain distant, whereas complete linkage prioritizes compactness by minimizing the largest dissimilarity within a cluster. Most other hierarchical approaches, such as average linkage, strike a balance between these two extremes. The performance of these methods when combined with various copula-based dissimilarity measures has been thoroughly reviewed by Fuchs et al. (2021).

Partitioning Around Medoids (PAM) algorithms. They enhance interpretability by selecting actual observed time series as cluster centers, rather than relying on synthetic averages. This feature is especially valuable when preserving the original data semantics is crucial. In the context of extreme-value clustering, PAM-based methods have been adopted by Bernard et al. (2013) and Bador et al. (2015). Additionally, Bien and Tibshirani (2011) introduce the hierarchical clustering algorithm with prototypes that integrates the hierarchical structure with the medoid concept, and has been explored further by Benevento et al. (2024).

Fuzzy clustering algorithms. These allow time series to belong partially to multiple clusters, making them well-suited for modeling uncertainty and smooth transitions between cluster boundaries, i.e., an important feature when dealing with real-world time series that often exhibit overlapping patterns. These methods also tend to be more computationally efficient, as they are less prone to abrupt changes in cluster membership during estimation and are generally more robust to local optima and convergence issues. In the context of copula-based approaches, fuzzy clustering has been explored in works by Disegna et al. (2017), De Luca and Zuccolotto (2021b), and D'Urso et al. (2023).

2.1 | Dependence Retrieval

As clarified above, the clustering procedure adopts a copula framework and, as such, it is invariant under increasing transformations of the involved time series components and does not depend on the marginals. However, the copula of the multivariate time series cannot be directly estimated from the original time series, since it is necessary to disentangle the dependence from the marginal effects. Thus, it is necessary to proceed with some preliminary steps that are usually grounded on some assumptions on the multivariate time series (see Patton 2012; Nasri and Rémillard 2019; Neumeyer et al. 2019).

Specifically, we assume that each time series, $(x_{ti})_{t=1,\ldots,T}$, $i \in \{1, \ldots, n\}$, has potentially time-varying conditional mean and variance and assume that the standardized residual series has a continuous, time-invariant, marginal distribution (with null mean and unit variance, for identification). Then, due to Sklar's theorem, the joint conditional distribution of the time series, given some information set $\mathcal{F}_{t-1}$, is completely specified by the individual conditional mean and variance functions, the marginal distributions of the residuals, and their copula C.

More specifically, for all $i = 1, \ldots, n$, we will assume that the i-th time series follows a marginal model of the form

$$\begin{aligned} X_{ti} &= f_i(\mathbf{Z}_{t-1}; \boldsymbol{\beta}_f) + a_{ti}, \quad \mathbf{Z}_{t-1} \in \mathcal{F}_{t-1}, \\ a_{ti} &= g_{ti}\epsilon_{ti}, \\ g_{ti} &= g_i(\mathbf{Z}_{t-1}; \boldsymbol{\beta}_g), \end{aligned} \tag{1}$$

where $f_i(\cdot)$ and $g_i(\cdot)$ are two known functions with finite-dimensional parameter vectors $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_g$, respectively, and $(\epsilon_{ti})_t$ is a sequence of independent and identically distributed random variables with mean zero and variance one (the so-called

innovations). Notice that the vector of innovations $(\varepsilon_{t1}, \ldots, \varepsilon_{tn})$ is independent of past and present information provided by $\mathcal{F}_s$, $s \leq t$. Model (1) encompasses many commonly used models in the literature. For instance, if $g(\cdot) = \beta_1$ is a positive constant, and $f(\cdot) = \sum_{l=1}^{p} \phi_l X_{t-l} + \sum_{m=1}^{q} \psi_m a_{t-m}$, then one gets the well-known ARMA model (see Box et al. 2015). The functions $f_i(\cdot)$ and $g_i(\cdot)$ in model (1) govern the conditional mean and variance of the i-th time series, that is, $\mathrm{E}(X_{ti}|\mathbf{Z}_{t-1}, \boldsymbol{\theta}_i) = f_i(\mathbf{Z}_{t-1})$ and $\mathrm{Var}(X_{ti}|\mathbf{Z}_{t-1}, \boldsymbol{\theta}_i) = g_{ti}^2$, where $\boldsymbol{\theta}_i$ is the set of all parameters of model (1).

Now, if $F_{i\varepsilon}$ denotes the distribution of the standardized residuals for the i-th time series, the joint conditional distribution of the observations, given $\mathcal{F}_{t-1}$, is

$$P(X_{t1} \leq x_1, \ldots, X_{tn} \leq x_n|\mathcal{F}_{t-1}) = P(\varepsilon_{t1} \leq v_1, \ldots, \varepsilon_{tn} \leq v_n) = C(F_{1\varepsilon}(v_1), \ldots, F_{n\varepsilon}(v_n)),$$

where $v_i = (x_i - f_i(\mathbf{z}_{t-1}; \boldsymbol{\beta}_f))/g_i(\mathbf{z}_{t-1}; \boldsymbol{\beta}_g)$, for $i \in \{1, \ldots, n\}$. Thus, the joint conditional distribution of the observations, given the past information, is completely specified by the individual conditional mean and variance functions, the marginal distributions of the innovations, and their copula (Neumeyer et al. 2019).

To retrieve the copula information, we proceed as follows.

i. First, we fit the univariate time series model of type (1) to filter out serial dependence, trends, and/or seasonal cycles from each time series, by obtaining the vector of estimated parameters for the models of the conditional mean and conditional variance, $\hat{\boldsymbol{\theta}}_i$.

ii. The inference about the copula C is thus based on the estimated residuals extracted from the previous step, given by

$$\hat{\varepsilon}_{ti} = \frac{x_{ti} - f_i(\mathbf{z}_{t-1}; \hat{\boldsymbol{\theta}}_i)}{g_i(\mathbf{z}_{t-1}; \hat{\boldsymbol{\theta}}_i)}, \tag{2}$$

which are transformed into the *pseudo-observations*, $r_{ti} = F_{i\varepsilon}(\hat{\varepsilon}_{ti})$, where $F_{i\varepsilon}$ may be estimated from a parametric model (Gaussian, Student-t, etc.) or by using the empirical distribution function.

As a result, $(r_{t1}, \ldots, r_{tn})_{t=1,\ldots,T}$ contains the information about the link (i.e., the copula) among the time series under consideration (Rémillard 2017) and can be used to describe the dependence structure of the n time series. The validity of this procedure has also been outlined by Neumeyer and Omelka (2025) under very general assumptions on the marginal process.

Remark 2.1. As for point (i), it should be stressed that the univariate time series model is of a general form and can be different for each marginal time series. Its selection is grounded on classical methods from univariate time series. For instance, in the ARMA-GARCH setting, the model selection can be carried out via classical criteria, such as AIC and BIC. To validate the model, homoscedasticity and uncorrelatedness tests can be conducted to verify that the residuals approximate an i.i.d. sample. In the case the marginal time series are affected by trends and seasonal components, suitable transformations (like local mean) should be done depending on the variables under consideration (see Erhardt and Czado 2018).

Once the time series observations have been transformed into pseudo-observations, we have collected the information needed to learn the n-dimensional copula C among all the n individual time series. However, especially for a large n, it is difficult to provide a full parametric copula model, and clustering procedures can be helpful to guide the process of model building (see Dißmann et al. 2013; Czado et al. 2012; Górecki et al. 2017, 2021; Palacios-Rodriguez et al. 2023).

A convenient way to extract meaningful information is to focus on the pairwise information and, hence, to consider the copula associated with the i-th and j-th pseudo-observations $(r_{ti}, r_{tj})_{t=1,\ldots,T}$, say C_{ij}. Notice that, in most algorithms presented in the literature, the copula C_{ij} is not explicitly indicated since it is replaced by its empirical counterpart obtained from the pseudo-observations. For instance, one can consider the empirical copula (Rüschendorf 2009; Segers 2012) or some of its variants, such as checkerboard copulas and empirical beta copulas (Segers et al. 2017).

Now, the collection of all the bivariate copulas associated with the n time series will be denoted by $\mathbf{C}^{\mathrm{ts}} = (C_{ij}^{\mathrm{ts}})$. By convention, $C_{ij}^{\mathrm{ts}} = M$ for every $i = j$, $(i, j \in \{1, \ldots, n\})$, where $M(u, v) = \min\{u, v\}$ for all $(u, v) \in [0, 1]^2$ denotes the comonotonicity copula. In general, the class of all matrices of pairwise copulas associated with an n-dimensional random vector will be denoted by $\mathrm{Cop}(n)$, and its generic element will be called *copula matrix*.

Below, we illustrate two examples from the literature where some parametric assumption is made on the copula matrix.

Example 2.1. As discussed by De Luca and Zuccolotto (2011) (see also De Luca and Zuccolotto 2021b), the extraction of the temporal copula matrix is done in the following way. First, a univariate Student-t AR-GARCH model is fitted to each time series to remove autocorrelation and heteroskedasticity effects. Then, the standardized residuals are used to estimate the copula matrix $\mathbf{C}^{\mathrm{ts}}$. In particular, it is assumed that, for all (i, j), $i \neq j$, the copula C_{ij}^{ts} belongs to the Joe-Clayton two-parameter copula class $C_{\kappa,\theta}$ (see Joe 1997) that is able to model differences in the lower tail and upper tail of the distribution. D'Urso et al. (2023) modify the procedure described above to allow the selection, for each $i \neq j$, of the best parametric copula (from a given set of families) according to Akaike Information Criterion.

Example 2.2. Durante, Pappadà, and Torelli (2015) assume that each element of the copula matrix belongs to the extreme-value class (Gudendorf and Segers 2010). Thus, the identification of the copula is equivalent to the identification of the associated Pickands dependence function. In particular, Benevento et al. (2024) extract the pairwise Pickands function from the estimation of the n-dimensional Pickands function associated with the multivariate vector of residuals via madogram estimators (Marcon et al. 2017; Gijbels et al. 2020).

Remark 2.2 (Missing data). When dealing with data from environmental sciences, one can make the observation that missing values occur rather often. To handling them, specific

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix} \longrightarrow D^{1,1}(\mathbf{C}) = \begin{pmatrix} d^{1,1}(C_{11}) & d^{1,1}(C_{12}) & \cdots & d^{1,1}(C_{1n}) \\ \vdots & \vdots & \ddots & \vdots \\ d^{1,1}(C_{n1}) & d^{1,1}(C_{n2}) & \cdots & d^{1,1}(C_{nn}) \end{pmatrix}$$

FIGURE 2 | Dissimilarity mapping.

copula-based estimators have been proposed; see, for instance, Segers (2015), Liebscher (2024), and Boulin et al. (2022) in an extreme-value framework. Moreover, when appropriate, suitable techniques for data imputation can be applied as well (see Di Lascio et al. 2025).

Remark 2.3 (Caveat emptor!). In most cases, the copula matrix $\mathbf{C}^{ts}$ is only a proxy of the true n-dimensional dependence structure C among the time series, that is, obviously, unknown. In particular, it is neither required nor guaranteed that, for a copula matrix $\mathbf{C}^{ts}$, it is possible to construct an n-dimensional copula model whose all bivariate marginals belong to $\mathbf{C}^{ts}$. Such a problem is usually very difficult to handle and goes under the name "compatibility problem" in the copula literature (Joe 1997; Nelsen 2006).

Finally, it should be stressed that, in general, a copula-based algorithm may focus not only on the pairwise dependence information, but also on higher-dimensional marginals. Examples are provided by De Luca and Zuccolotto (2021b), Fuchs et al. (2021), and Fuchs and Wang (2024).

2.2 | The Dissimilarity Mapping

Given the copula matrix $\mathbf{C}^{ts} = (C_{ij}^{ts}) \in \mathrm{Cop}(n)$, it is necessary to define a convenient dissimilarity matrix $\mathbf{\Delta}^{ts} = (\Delta_{ij}^{ts}) \in \mathrm{Diss}(n)$ that embeds each C_{ij}^{ts} into a numerical value Δ_{ij}^{ts}. According to the comonotone-based clustering procedure illustrated by Fuchs et al. (2021), the dissimilarity value should reflect a kind of distance between C_{ij}^{ts} and the comonotonicity copula M.

In general, the transformation from a copula matrix to a dissimilarity matrix can be done by considering a $(1,1)$-dissimilarity function $d^{1,1}$ (in the sense of Fuchs et al. 2021) that is a mapping from the space of bivariate copulas to $[0, +\infty]$ satisfying the following properties:

D1. $d^{1,1}(C) = 0$ if C coincides with the comonotonicity copula M;

D2. $d^{1,1}(C) = d^{1,1}(C^T)$ for any bivariate copula C, where $C^T(u,v) = C(v,u)$ for every $(u,v) \in [0,1]^2$.

Notice that the notation $d^{1,1}$ refers to the fact that we are comparing two one-dimensional random vectors (for dissimilarity functions among random vectors in higher dimensions, see Fuchs et al. 2021; Durante et al. 2025). Moreover, property (D1) distinguishes the present methodology from other clustering algorithms that focus on detecting functional dependence and/or independent groups (Kojadinovic 2004, 2010; Fuchs and Wang 2024; De Keyser and Gijbels 2024).

Now, we denote by $D^{1,1}$ the mapping that associates to each copula matrix $\mathbf{C}$ the dissimilarity matrix obtained by applying the function $d^{1,1}$ to each element of $\mathbf{C}$. Thus, for every (i,j), we have

$$(D^{1,1}(\mathbf{C}))_{ij} = d^{1,1}(C_{ij}).$$

See Figure 2.

In the literature, there are different possible choices for $d^{1,1}$ according to the specific criterion the analyst would like to adopt.

A first natural proposal for $d^{1,1}$ is to consider pairwise dissimilarities that are related to popular measures of association (Schmid et al. 2010). For instance, some examples already considered in the literature (see Fuchs et al. 2021) are:

$$d^{1,1}(C_{ij}) = M\left(\frac{1}{2},\frac{1}{2}\right) - C_{ij}\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{4}(1-\beta(C_{ij})), \quad (3)$$

$$d^{1,1}(C_{ij}) = \int_0^1 (M(u,u) - C_{ij}(u,u))du = \frac{1}{6}(1-\phi(C_{ij})), \quad (4)$$

$$d^{1,1}(C_{ij}) = \int_0^1\int_0^1 M(u,v)dM(u,v) - \int_0^1\int_0^1 C_{ij}(u,v)dC_{ij} = \frac{1}{4}(1-\tau(C_{ij})), \quad (5)$$

$$d^{1,1}(C_{ij}) = \int_0^1\int_0^1 (M(u,v) - C_{ij}(u,v))dudv = \frac{1}{12}(1-\rho(C_{ij})). \quad (6)$$

These dissimilarity functions are related, respectively, to the medial correlation coefficient β (also known as Blomqvist's beta), Spearman's footrule ϕ, Kendall's correlation τ, and Spearman's correlation ρ. They can be estimated non-parametrically by relying on classical estimation of measures of association (Schmid et al. 2010), or by replacing C_{ij} with its corresponding empirical copula (Rüschendorf 2009; Segers 2012). Fuchs et al. (2021) demonstrate that, in the context of linkage-based hierarchical clustering, the proposed dissimilarity functions perform well even under moderate dependence. In contrast, the dissimilarity defined in Equation (3) shows poor performance in cases of weak dependence. Furthermore, among the considered linkage methods, average linkage yields the best results.

Another example is derived from the van der Waerden correlation coefficient ζ (see Genest and Verret 2005), also known as normal score correlation or Gaussian rank correlation. In fact, if (U,V) is distributed as the copula C, then we can consider the $(1,1)$-dissimilarity function given by

$$d^{1,1}(C_{ij}) = 1 - \rho_P(\Phi^{-1}(U), \Phi^{-1}(V)), \quad (7)$$

where ρ_P denotes the linear Pearson's correlation, while Φ denotes the cumulative distribution function of the standard

Gaussian distribution (Genest and Verret 2005; Koike and Hofert 2024). The advantage of this coefficient is that the associated dissimilarity matrix can be embedded into a peculiar space of correlation matrices (Benevento and Durante 2024).

As a consequence of the Fréchet-Hoeffding bounds for copulas (Durante and Sempi 2016), all the coefficients from Equations (3) to (7) assume non-negative values and, in particular, take the value 0 when $C_{ij} = M$. However, for different reasons, the value of the dissimilarity can be slightly transformed with a decreasing real function f taking non-negative values. Popular dissimilarities of this type are obtained as a function of a measure of association. For instance, $f(t) = \sqrt{1-t^2}$, where t is Kendall's tau or Spearman's rho, defines a $(1,1)$-dissimilarity function that is also a pseudo-metric (see Côté and Genest 2015; Van Dongen and Enright 2012).

Remark 2.4. Notice that, for most of the one-parameter copula families $(C_\theta)_{\theta\in\Theta}$, it holds that there is a bijection $f : \Theta \to A \subset [-1,1]$, given by $f(\theta) = \kappa(C_\theta)$, where κ is a measure of association. In such a case, a transformation of the copula parameter (as estimated from observations assuming that the copula model holds) can be used to express the dissimilarity. See, for instance, Di Lascio et al. (2024).

Example 2.3. Since most of the previous dissimilarity functions from Equations (3) to (7) are expressed in terms of discrepancy between the copula M and C_{ij}, another possible approach (Disegna et al. 2017) is to calculate directly a dissimilarity of type

$$d^{1,1}(C_{ij}) = \mathrm{dist}(M, C_{ij})$$

where dist is the L^2 or L^∞ distance in the space of copulas. Another possibility is to consider in the space of bivariate copulas equipped with the Wasserstein metric $d_{\mathcal{W}_2}$ (see Benevento and Durante 2023; Marti et al. 2016; Nielsen et al. 2021).

When the joint behavior of the time series in extreme scenarios is of special interest, the tail dependence coefficients (Durante, Fernández-Sánchez, and Pappadà 2015; Joe 1997) are considered as measures of association. Here, a classical choice (see De Luca and Zuccolotto 2011) is to consider

$$\begin{aligned} d^{1,1}(C_{ij}) &= -\ln(\lambda_U(C_{ij})), \\ d^{1,1}(C_{ij}) &= -\ln(\lambda_L(C_{ij})), \end{aligned} \tag{8}$$

where λ_U and λ_L denote the upper and lower tail dependence coefficients, respectively (whenever they exist). In such a case, the estimation of the dissimilarity may suffer from a large uncertainty when dealing with the tail of the distribution, so that some assumptions about the copulas are usually made. Related to this approach, conditional measures of association have also been considered (Durante et al. 2014) by focusing on the conditional copula (Durante and Jaworski 2010) associated with the tails of the distribution. It should be stressed that the approach based on tail dependence coefficients has also been adopted in the context of extreme value distributions (see, for instance, Bador et al. 2015; Bernard et al. 2013; Maume-Deschamps et al. 2025; Saunders et al. 2021).

Finally, a dissimilarity can be derived by projecting C_{ij} to a real-valued one-dimensional function, and hence the dissimilarity is computed from this lower-dimensional object. In the following, we present some examples of such a methodology.

Example 2.4. Sometimes, it is important to examine the comovements of time series not only in the asymptotic case, but also under moderate scenarios. In such situations, assessing tail-dependence dissimilarity through appropriate joint quantiles can be particularly useful. For instance, Durante, Fernández-Sánchez, and Pappadà (2015) propose a dissimilarity grounded on the so-called tail concentration function, defined in terms of the diagonal section of the copula C, as the function $q_C : (0,1) \to [0,1]$ given by $q_C(t) = (C(t,t)/t)\mathbf{1}_{(0,0.5]} + ((1-2t+C(t,t))/(1-t))\mathbf{1}_{(0.5,1)}$, with $\mathbf{1}_S$ denoting the indicator function of the set S. The dissimilarity between the i-th and j-th time series is thus defined as the distance between the function $q_{C_{ij}}$ associated with the copula C_{ij} and the tail concentration function of the comonotone copula M (for which $q_M(t) = 1$ for every $t \in (0,1)$):

$$d^{1,1}(C_{ij}) = \int_0^1 (q_{C_{ij}}(t) - q_M(t))^2 dt.$$

When the interest is in the lower or upper tail of the joint distribution, the above integral can be computed on a restricted domain. For instance, in the work by Durante, Fernández-Sánchez, and Pappadà (2015) the aim is to cluster financial time series and, hence, the pairwise dissimilarity is computed using the empirical tail concentration function estimated on the interval $(0, 0.5]$ to capture joint losses on the considered assets.

Example 2.5. Durante and Pappadà (2015) (see also Castrovilli et al. 2025) use the Kendall distribution function associated with a copula to derive a dissimilarity measure. This choice is mainly motivated by the use of Kendall hazard scenarios in hydrology and environmental sciences (see Salvadori et al. 2013, 2016). Let K_C denote the Kendall's function associated with the bivariate copula C, that is, $K_C(t) = P(C(U,V) \le t)$, with $t \in [0,1]$. Here (U,V) is a random pair distributed according to C. The dissimilarity between the i-th and j-th time series is defined as a suitable distance between the Kendall distribution function associated with C_{ij} and the one related to the comonotone copula M:

$$d^{1,1}(C_{ij}) = \int_0^1 (K_{C_{ij}}(t) - K_M(t))^2 dt,$$

where $K_M(t) = t$ for every $t \in (0,1)$. The dissimilarities are computed via the non-parametric estimation of the function $K_{C_{ij}}$ (see Durante and Pappadà 2015).

Remark 2.5. It is worth noticing that a slightly modified approach has been proposed by De Luca and Zuccolotto (2011). Here, the dissimilarity matrix has been converted to a distance matrix via non-metric multidimensional scaling techniques. The main advantage is that, in such a way, algorithms requiring that the objects are represented as points in a high-dimensional Euclidean space (like K-means) can also be applied.

In summary, the choice of the dissimilarity function $d^{1,1}$ plays a crucial role in the clustering process. However, it largely

depends on the analyst's perspective and the specific characteristics of the phenomenon being studied. Each dissimilarity measure proposed here highlights a particular aspect of dependence, and ultimately, the nature of the problem should guide the selection.

2.3 | Cluster Validation

Once the dissimilarity matrix has been provided, the determination of the cluster partition depends on the DBA that has been chosen as input. In many cases, the DBA requires setting certain hyper-parameters, most notably the number of clusters. The main challenge addressed here is determining the appropriate number of clusters, denoted as K. This task is commonly referred to as cluster validation, which encompasses a range of techniques designed to evaluate and identify the most suitable number of clusters by relying on indices that capture key aspects of clustering quality, such as cluster separation (Hennig et al. 2015). While prior knowledge about the data can sometimes guide the choice of K (e.g., Coppi et al. 2010), in the absence of such information, K is typically determined by solving an optimization problem. In the literature, copula-based methods for time series have mainly considered the following cluster validation indices:

Hierarchical algorithms. Cluster validation has been based either on the silhouette index (Durante et al. 2014, Durante, Pappadà, and Torelli 2015) or on the Dunn index (Di Lascio et al. 2024, Benevento and Durante 2024). Zuccolotto et al. (2023) consider three indices for cluster validation, namely the average silhouette width, the Dunn index, and the Calinski and Harabasz index. Furthermore, the optimal number of clusters is obtained by an automated procedure called cutMOB (De Luca and Zuccolotto 2023; Carpita et al. 2024). Finally, De Luca and Zuccolotto (2021a) use the entropy of the distribution of cluster memberships as an internal validation index, since it penalizes the presence of clusters composed of one stock only and, consequently, methods that tend to generate chaining phenomena.

Partitioning Around Medoids (PAM) algorithms. The number of clusters is usually selected via the classical silhouette index (see Bernard et al. 2013; Bador et al. 2015) or Dunn-like index (Benevento et al. 2024).

Fuzzy algorithms. The COFUST algorithm proposed by Disegna et al. (2017) selects the number of clusters via the fuzzy silhouette index (Campello and Hruschka 2006); the latter is also considered by Benevento and Durante (2023) to select a partition obtained through the fuzzy-PAM algorithm, and by De Luca and Zuccolotto (2021b), where a fuzzy K-means clustering algorithm is proposed. D'Urso et al. (2023) focus on fuzzy C-medoids methods, and the fuzzy silhouette index is adopted to find the optimal number of clusters.

It should be noticed that many of the previously cited validation indices are based on a combination of measures of within-cluster variation and between-cluster separation. For this reason, De Keyser and Gijbels (2024) and Fuchs and Wang (2024) address the selection of the optimal number of clusters by graphical comparisons of two-coordinate points representing both the intra-cluster similarity and the inter-cluster similarity.

3 | Embedding the Spatial Constraints Into the Framework

Now, we assume that the observations $D^{\text{ts}} = \{(x_{t1}, x_{t2}, \ldots, x_{tn}), t = 1, \ldots, T\}$ are collected with a set of feature vectors $D^{\text{sp}} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ (e.g., geographic locations) so that $\mathbf{s}_i$ refers to the time series $(x_{ti})_{t=1,\ldots,T}$ for $i = 1, \ldots, n$. Notice that these feature vectors are deterministic and fixed across time. Moreover, since they are usually related to the geographic information about the sites where the time series have been collected, with a little abuse of language, this set will always be referred to as the *spatial data*. Typically, we are hence interested in geo-referenced time series that record time-changing values of one attribute at fixed locations and consistent time intervals (Kisilevich et al. 2010) (e.g., hourly total precipitations observed at a network of ground monitoring stations).

One way to perform clustering of time series with spatial information consists of using general-purpose clustering methods based on a spatio-temporal dissimilarity measure (Oliver and Webster 1989; Coppi et al. 2010; Fouedjio 2016; D'Urso and Vitale 2020) to explicitly take into account the spatial dependency between data locations. This procedure has also been adopted in the copula-based clustering community and can be formalized in the TRIPLE-C algorithm that we formalize here. This algorithm is a procedure that, starting with temporal as well as spatial data, provides a dissimilarity matrix that is used to obtain the clustering partition via DBAs.

The idea of TRIPLE-C is to merge both the temporal and the spatial information into a dissimilarity matrix $\mathbf{\Delta}^{\alpha}$ by means of a suitable hyper-parameter $\alpha \geq 0$. Such an α describes how much the spatial information influences the clustering process, with $\alpha = 0$ meaning no spatial influence. To this end, two main strategies can be adopted to modify the framework of Figure 1:

- The dissimilarity matrix $\mathbf{\Delta}^{\text{ts}}$ can be joined to a suitable dissimilarity matrix $\mathbf{\Delta}^{\text{sp}}$ that interprets the spatial information. This aggregation is done with a suitable operation in $\text{Diss}(n)$ that returns as output $\mathbf{\Delta}^{\alpha} \in \text{Diss}(n)$. See Figure 3.
- The copula matrix $\mathbf{C}^{\text{ts}}$ is combined with a copula matrix $\mathbf{C}^{\text{sp}}$ that interprets the spatial information. The resulting output copula matrix $\mathbf{C}^{\alpha}$ is hence transformed via a dissimilarity mapping $D^{1,1}$ into $\mathbf{\Delta}^{\alpha} \in \text{Diss}(n)$. See Figure 4.

In both cases, the obtained dissimilarity matrix $\mathbf{\Delta}^{\alpha}$ is used as the input of the DBA in the cluster validation step.

The algorithms illustrated in Figures 3 and 4 present, hence, some novel tasks compared to the existing literature (as schematized in Figure 1). These will be illustrated in detail in the following sections.

3.1 | Spatial Proximity Retrieval

A key message of the clustering procedure considered in this work is that the additional information contained in the set D^{sp}

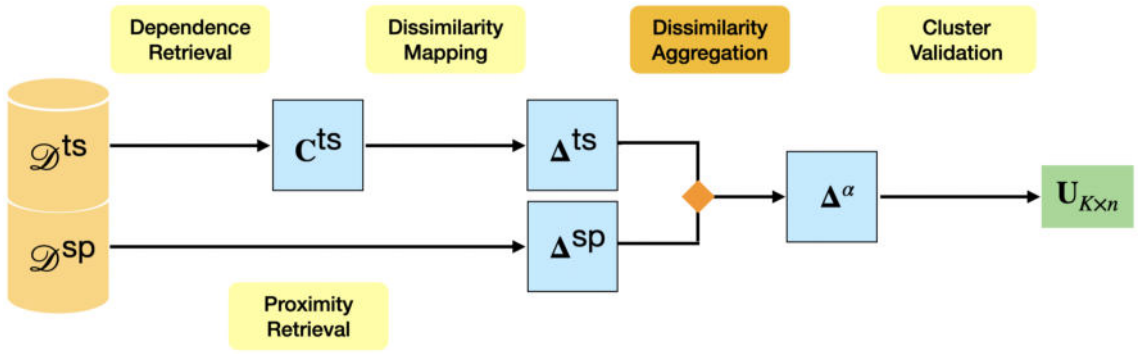

FIGURE 3 | TRIPLE-C algorithm based on dissimilarity aggregation: Model architecture.

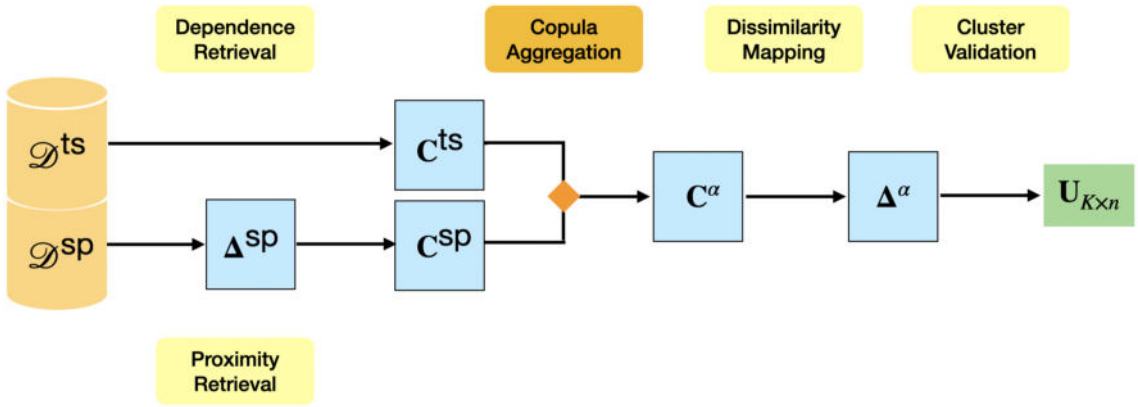

FIGURE 4 | TRIPLE-C algorithm based on copula aggregation: Model architecture.

is associated with the time series data. Here, we assume that D^{sp} is included in $\mathbb{R}^p$; otherwise, suitable transformations may be needed to project the spatial information onto the Euclidean space. For instance, each $\mathbf{s}_i$ represents a geographic location expressed with longitude and latitude coordinates.

The spatial information D^{sp} is converted into an element in Diss(n). In particular, two main types of matrices can be obtained:

- a binary matrix $\boldsymbol{\Delta}^{sp} = (\Delta_{ij}^{sp}) \in \{0,1\}^n$, where each entry Δ_{ij}^{sp} indicates whether the i-th and j-th time series components are related (value equal to 0) or not (value equal to 1). In particular, $\Delta_{ij}^{sp} = 0$ whenever $i = j$.
- a distance matrix $\boldsymbol{\Delta}^{sp} = (\Delta_{ij}^{sp}) \in [0,+\infty]^n$, where each entry Δ_{ij}^{sp} only depends on the (Euclidean) distance between $\mathbf{s}_i$ and $\mathbf{s}_j$.

The distance matrix is often normalized to $[0,1]$-values. In such a case, one can use the min-max renormalization. Otherwise, each entry Δ_{ij}^{sp} can be expressed as $1 - \exp(-d_{ij})$, where d_{ij} is the Euclidean distance between the i-th and j-th time series.

Both these types of dissimilarity matrices can be naturally interpreted through graph theory, where they represent the structure and properties of a graph $G = (V, E)$, where V is the set of nodes corresponding to the set of time series, and E is the set of edges. Specifically, when $\boldsymbol{\Delta}^{sp}$ is the binary matrix, $1 - \boldsymbol{\Delta}^{sp}$ corresponds to the adjacency matrix of an unweighted graph, where an edge $(v_i, v_j) \in E$ exists if and only if $\Delta_{ij}^{sp} = 0$. In the case of $\boldsymbol{\Delta}^{sp}$ is the distance matrix, each edge $(v_i, v_j) \in E$ corresponds to the distance between the corresponding components. Such a graph is called, e.g., *the sampling graph* (Romary et al. 2015).

When dealing with geo-referenced time series, the spatial data can mainly be either raster-based or sparse. Raster data refers to spatial information represented as a grid of cells, where each cell holds a value corresponding to a specific attribute. This structure is particularly suited for continuous time series (i.e., time series that can virtually be recorded at any point in space) and is widely used in research in geography. Sparse spatial data, instead, refers to the case when the time series are collected at discrete and irregularly spaced locations (or gauge stations).

Example 3.1. The conversion of raster data to a binary matrix is a well-documented process in spatial analysis and is utilized in fields such as Geographic Information Systems and remote sensing. In such a case, the grid cells are treated as the nodes of the graph, and their spatial relationships are encoded as edges. This transformation involves identifying spatial relationships (adjacencies) between raster grid cells, which are often defined by *rook* (4-neighbor) or *queen* (8-neighbor) connectivity (see, e.g., Lloyd 2010). In the context of clustering of geo-referenced data, raster data have been considered, for example, by Wang et al. (2024), Bador et al. (2015), and Benevento et al. (2024).

In the case of sparse spatial data, constructing a binary matrix requires a different approach compared to raster data. To avoid

creating a fully connected graph, where every node is connected to every other node, a threshold distance d_{max} must be set. This threshold determines which points are considered neighbors in the sampling graph.

In the algorithm described in Figure 4, the spatial dissimilarity matrix $\mathbf{\Delta}^{\text{sp}}$ must be furthermore converted into a copula matrix $\mathbf{C}^{\text{sp}} = (C_{ij}^{\text{sp}})$. To this end, any element C_{ij}^{sp} is assumed to belong to a parametric class of bivariate copulas $\{C_\theta : \theta \in \Theta = [\theta_{\min}, \theta_{\max}] \subseteq [-\infty, +\infty]\}$. To ensure the identifiability of the copula, the mapping $\theta \mapsto C_\theta$ is supposed to be injective. Typically, the parameter space Θ is transformed through a monotone function into the set of realized values of $\mathbf{\Delta}^{\text{sp}}$, or into $[0,1]$ if this set is normalized. Moreover, $M \in \{C_{\theta_{\min}}, C_{\theta_{\max}}\}$, that is, the family of copulas includes the comonotonicity copula, and it is associated with the case $\Delta_{ij}^{\text{sp}} = 0$. Some examples of spatial copula matrices are considered at the end of Section 3.3.

3.2 | Aggregation of the Dissimilarity Matrices

Now, we describe the crucial aggregation step in the algorithm illustrated in Figure 3 about the merging of temporal and spatial information expressed in terms of dissimilarity matrices $\mathbf{\Delta}^{\text{ts}}$ and $\mathbf{\Delta}^{\text{sp}}$ in $\text{Diss}(n)$. Such a general strategy is inspired by early works in spatial clustering (e.g., Oliver and Webster 1989) that modify the matrix of pairwise (temporal) dissimilarities by a nonlinear function of the distances among $\mathbf{s}_1, \ldots, \mathbf{s}_n$. However, it could equivalently be represented in terms of an operation in $\text{Diss}(n)$.

Specifically, spatial and temporal matrices are joined by considering an operation ψ_α on the class of dissimilarity matrices $\text{Diss}(n)$, that is,

$$\psi_\alpha : \text{Diss}(n) \times \text{Diss}(n) \to \text{Diss}(n), \qquad \psi_\alpha(\mathbf{\Delta}^{\text{ts}}, \mathbf{\Delta}^{\text{sp}}) = \mathbf{\Delta}^{\alpha} \tag{9}$$

for $\alpha \geq 0$. The operation ψ_α depends on a suitable (hyper-)parameter α, tuning the relative influence of the spatial component on the whole procedure. As a matter of fact, $\psi_0(\mathbf{\Delta}^{\text{ts}}, \mathbf{\Delta}^{\text{sp}}) = \mathbf{\Delta}^{\text{ts}}$.

The following examples illustrate possible choices for the operation ψ_α appeared in the literature.

Example 3.2. A simple way to join the two dissimilarity matrices is to consider a convex combination of them, so that

$$\mathbf{\Delta}^{\alpha} = (1-\alpha)\mathbf{\Delta}^{\text{ts}} + \alpha\mathbf{\Delta}^{\text{sp}}, \qquad \alpha \in [0,1], \tag{10}$$

as discussed for instance in De Carvalho et al. (2023), Example 1. Notice that $\Delta_{ij}^{\alpha} = \Delta_{ij}^{\text{ts}}$ if $\Delta_{ij}^{\text{ts}} = \Delta_{ij}^{\text{sp}}$, for some (i,j). In some cases, temporal and spatial matrices should be rescaled before performing their combination to sum quantities that are comparable in magnitude. For instance, Deb and Karmakar (2023) consider the following operation

$$\mathbf{\Delta}^{\alpha} = (1-\alpha)\frac{\mathbf{\Delta}^{\text{ts}}}{\|\mathbf{\Delta}^{\text{ts}}\|_F} + \alpha\frac{\mathbf{\Delta}^{\text{sp}}}{\|\mathbf{\Delta}^{\text{sp}}\|_F}, \qquad \alpha \in [0,1], \tag{11}$$

where $\|\cdot\|_F$ indicates the Frobenius norm of a matrix.

Example 3.3. Consider the clustering procedure introduced by Zuccolotto et al. (2023) (see also Carpita et al. 2024), where the temporal dissimilarity matrix takes into account the co-occurrence of extreme events. In such a case, the spatio-temporal dissimilarity is defined (see Equation (7) in Carpita et al. 2024) for every $\alpha \geq 0$, by

$$\mathbf{\Delta}^{\alpha} = \mathbf{\Delta}^{\text{ts}} + \alpha\mathbf{\Delta}^{\text{sp}}, \tag{12}$$

where $\mathbf{\Delta}^{\text{ts}}$ is defined in Equation (8) and $\mathbf{\Delta}^{\text{sp}}$ is a binary matrix. Notice that, for every (i,j), $\Delta_{ij}^{\alpha} \geq \Delta_{ij}^{\text{ts}}$. Thus, the spatial information may only add a penalty to the temporal dissimilarity for time series that are spatially far away. A similar procedure is adopted in a fuzzy scenario by Gelb and Apparicio (2021).

Example 3.4. Another way to construct the dissimilarity matrix $\mathbf{\Delta}^{\alpha}$ is to consider the Hadamard product (i.e., element-wise product) between $\mathbf{\Delta}^{\text{ts}}$ and $\mathbf{\Delta}^{\text{sp}}$, that is,

$$\mathbf{\Delta}^{\alpha} = \mathbf{\Delta}^{\text{ts}} \circ (\alpha\mathbf{\Delta}^{\text{sp}}).$$

In particular, to ease the interpretability, it could be convenient that $\mathbf{\Delta}^{\text{sp}} \in [0,1]^{n\times n}$, that is, the spatial information is normalized so that it can be interpreted as a weight to the temporal matrix. Such an operation is considered, for instance, by Legendre and Legendre (1998) and Legendre and Gauthier (2014). Notice that, if $\Delta_{ij}^{\text{sp}} = 0$ for some $i \neq j$, then $\Delta_{ij}^{\alpha} = 0$ regardless of α. This would correspond in the clustering procedure to a must link constraint (see, e.g., Cai et al. 2023), since it would imply that the i-th and j-th time series must be in the same group.

An element-wise operation is also proposed by Webster and Burrough (1972), see Equation (3) (see also Equation (1) in Oliver and Webster 1989). Translated in the present notation, it corresponds to the mapping that gives

$$\Delta_{ij}^{\alpha} = \Delta_{ij}^{\text{ts}}\left(1 - \exp\left(-\frac{\Delta_{ij}^{\text{sp}}}{\alpha}\right)\right). \tag{13}$$

Here, Δ_{ij}^{sp} coincides with the distance between the feature vectors $\mathbf{s}_i$ and $\mathbf{s}_j$. Note that, for every (i,j), $\Delta_{ij}^{\alpha} \leq \Delta_{ij}^{\text{ts}}$. Thus, the spatial information may allow a reduction of the temporal dissimilarity for time series that are spatially close.

The operation ψ_α may be restricted to operate on a specific subset of $\text{Diss}(n)$. For instance, if $1 - \mathbf{\Delta}^{\text{ts}}$ is a correlation matrix, as when the dissimilarity mapping is derived from various measures of association (see Hofert and Koike 2019; McNeil et al. 2022), then it could be convenient to require that also $1 - \mathbf{\Delta}^{\alpha}$ is a correlation matrix. The following example illustrates this aspect.

Example 3.5. In the approach proposed by Benevento and Durante (2024), both $1 - \mathbf{\Delta}^{\text{ts}}$ and $1 - \mathbf{\Delta}^{\text{sp}}$ belong to the space of correlation matrices $\text{Corr}(n)$, that is, symmetric matrices in $[-1,1]^{n\times n}$ that are positive semi-definite matrices with diagonal entries equal to 1. Specifically, $\mathbf{\Delta}^{\text{ts}}$ has been obtained as in Equation (3) in Benevento and Durante (2024), via van der Werden coefficient. Instead, $\mathbf{\Delta}^{\text{sp}}$ has been obtained as in Equation (5) in Benevento and Durante (2024). In such a case, both matrices have been combined by exploiting the geometric structure of the

space of correlation matrices (David and Gu 2019; Thanwerdas and Pennec 2022). Specifically,

$$\mathbf{\Delta}^{\alpha} = 1 - \gamma_{\alpha}(1 - \mathbf{\Delta}^{\text{ts}}, 1 - \mathbf{\Delta}^{\text{sp}}), \tag{14}$$

where γ_{α} is a geodesic in the Riemannian manifold Corr(n).

3.3 | Aggregation of Copula Matrices

Now, we describe how to join temporal and spatial information expressed in terms of copula matrices $\mathbf{C}^{\text{ts}}$ and $\mathbf{C}^{\text{sp}}$ in Cop(n), as illustrated in Figure 4. Such a general strategy has been introduced by Disegna et al. (2017). The main idea is to consider an operation Ψ_{α} on the class of copula matrices Cop(n). The most intuitive operation is the convex combination (see Disegna et al. 2017). However, when the copulas are from some subclasses, some alternatives are possible (see, e.g., Benevento et al. 2024; Benevento et al. 2025). Here, we collect some examples where the procedure schematized in Figure 4 is adopted.

Example 3.6. Disegna et al. (2017) adopt the following specific choices:

- Dependence retrieval: Each entry of the copula matrix $\mathbf{C}^{\text{ts}}$ is obtained from the empirical copula associated with each pair of time series.
- Spatial proximity retrieval: A copula matrix $\mathbf{C}^{\text{sp}}$ is constructed so that, for every (i, j), $C^{\text{sp}}_{ij} = \Delta^{\text{sp}}_{ij} W + (1 - \Delta^{\text{sp}}_{ij})M$, where M and W are the comonotonicity and countermonotonicity copula, respectively.
- Copula aggregation: The copula matrix C^{α}_{ij} is the convex combination of $\mathbf{C}^{\text{ts}}$ and $\mathbf{C}^{\text{sp}}$ previously defined. Thus, in this setting Ψ_{α} is the convex combination of its inputs.
- Dissimilarity mapping: Finally, the dissimilarity function $d^{1,1}$ is defined as a distance between the copula M and the input copula C via an increasing and continuous real-valued function f and a suitable norm $\|\cdot\|$ in the space of bivariate copulas.

Summarizing, the (i,j)-entry of $\mathbf{\Delta}^{\alpha}$ can be represented as

$$\Delta^{\alpha}_{ij} = f(\|M - (1 - \alpha)C^{\text{ts}}_{ij} - \alpha(\Delta^{\text{sp}}_{ij} W + (1 - \Delta^{\text{sp}}_{ij})M)\|) \tag{15}$$

where $\alpha \in [0, 1]$. The function $f(\cdot)$ is such that $f(0) = 0$, hence the dissimilarity is zero for spatially adjacent pairs of time series ($\Delta^{\text{sp}}_{ij} = 0$) that are comonotone ($C^{\text{ts}}_{ij} = M$). Moreover, for a fixed α and a fixed copula C^{ts}_{ij}, the dissimilarity increases as the spatial separation increases.

Example 3.7. Benevento et al. (2024) suggest the following specific choices:

- Dependence retrieval: Each entry of the copula matrix $\mathbf{C}^{\text{ts}}$ is a bivariate margin of an n-dimensional extreme-value copula expressed in terms of the Pickands dependence function.
- Spatial proximity retrieval: Each entry in $\mathbf{C}^{\text{sp}}$ is a bivariate margin of a subclass of Hüsler–Reiss n-dimensional extreme-value copula so that the parameter values only depend on the pairwise distances among spatial features.
- Copula aggregation: The copula matrix C^{α}_{ij} is obtained via the element-wise operation $C^{\alpha}_{ij}(u, v) = C^{\text{ts}}_{ij}(u^{1-\alpha}, v^{1-\alpha})C^{\text{sp}}_{ij}(u^{\alpha}, v^{\alpha})$ for $\alpha \in [0, 1]$ (see, e.g., Khoudraji 1995; Durante 2009; Liebscher 2008).
- Dissimilarity mapping: The dissimilarity function $d^{1,1}$ is defined as in Equation (8).

By virtue of specific properties of the extreme-value class of copulas in combination with the considered operation (Genest et al. 1998), the (i, j)-entry of $\mathbf{\Delta}^{\alpha}$ can be represented as

$$\Delta^{\alpha}_{ij} = -\ln((1 - \alpha)\lambda_U(C^{\text{ts}}_{ij}) + \alpha\lambda_U(C^{\text{sp}}_{ij})) \tag{16}$$

where $\alpha \in [0, 1]$.

3.4 | Cluster Validation and Hyper-Parameter Selection

Another thorny aspect of the TRIPLE-C algorithm is the selection of the hyper-parameters K and α, namely the number of clusters and the weight used to merge the temporal and the spatial information, respectively. These hyper-parameters are used to configure various aspects of the learning algorithms and may largely affect the resulting model and its performance (Claesen and De Moor 2015).

In practice, finding the optimal values for K and α involves evaluating the clustering results for multiple combinations of these parameters. For each pair (α, K), the quality of clustering can be assessed using internal and external validation criteria such as the silhouette index, Dunn index, and so forth.

In most cases, the optimization process is carried out in two steps: First, the optimal number of clusters K and the cluster partition are chosen while keeping α fixed in a given interval (here $[0, 1]$), and then α is selected by comparing the obtained partitions. This process is illustrated below in Algorithm 1 (see also Carpita et al. 2024, Algorithm 1).

ALGORITHM 1 | Hyper-parameter selection.

Require: The dissimilarity matrices $\mathbf{\Delta}^{\text{ts}}, \mathbf{\Delta}^{\text{sp}}, \mathbf{\Delta}^{\alpha}$, a DBA Alg, a validity index Val.

1: Define a sequence A of possible values of $\alpha \geq 0$.
2: **for** each α in A **do**
3: Apply Alg to the dissimilarity matrix $\mathbf{\Delta}^{\alpha}$.
4: Identify the optimal number of clusters K^{α} and the optimal partition $\mathbf{U}^{\alpha} \in \text{Memb}(K^{\alpha}, n)$.
5: Compute Val($\mathbf{\Delta}^{\alpha}, \mathbf{U}^{\alpha}$), Val($\mathbf{\Delta}^{\text{ts}}, \mathbf{U}^{\alpha}$) and Val($\mathbf{\Delta}^{\text{sp}}, \mathbf{U}^{\alpha}$).
6: **end for**
7: Compare the values of the validation index (with different dissimilarity matrices $\mathbf{\Delta}^{\alpha}, \mathbf{\Delta}^{\text{ts}}, \mathbf{\Delta}^{\text{sp}}$) versus α, and decide its optimal value α^*.
8: Return (α^*, K^{α^*}).

In general, the choice of the optimal value of α of Algorithm 1 is driven by a graphical representation. In the following, some examples from the literature are collected. It is worth noticing that the choice of α is not always provided and, sometimes, the user has to decide by comparing the outputs according to some a priori knowledge of the specific application domain (Deb and Karmakar 2023).

Example 3.8. Coppi et al. (2010) (see also López-Oriona et al. 2021), set the selection of the number of clusters based on a prior knowledge about the data (in their example $K = 3$ since the Italian socio-demographic structure is usually distinguished with respect to three main areas: Northern Italy, Central Italy, and Southern Italy). The parameter α is selected according to a validation index related to the so-called spatial autocorrelation of the obtained cluster partition $\mathbf{U}^\alpha$ for any α.

Example 3.9. Chavent et al. (2018) (see also Aguiar et al. 2020; Mattera and Franses 2024) adopt the following specifications. First, $K^\alpha = K^0$ for each α, that is, the selection of the optimal number of clusters only depends on the dissimilarity $\mathbf{\Delta}^{\text{ts}}$. Then, the optimal value α is chosen graphically as the value that is a trade-off between the loss of temporal homogeneity (as indicated by $\text{Val}(\mathbf{\Delta}^{\text{ts}}, \mathbf{U}^\alpha)$) and a larger geographical cohesion (as indicated by $\text{Val}(\mathbf{\Delta}^{\text{sp}}, \mathbf{U}^\alpha)$). Chavent et al. (2018) consider Val related to the so-called inertia of the algorithm.

Example 3.10. In Morelli et al. (2025), a parameter α is selected for each K as the one that jointly maximizes the amount of pseudo inertia explained from both information, weighted by the cumulated pseudo inertia embedded in the data (thus, the criterion is indicated by $\text{Val}(\mathbf{\Delta}^\alpha, \mathbf{U}^\alpha)$) and is related to the inertia defined by Chavent et al. (2018). Then, the optimal K, depending on the corresponding α, is selected according to one or more hierarchical clustering criteria such as the silhouette index.

Example 3.11. Benevento et al. (2024) define Val as the so-called connectedness index, which is related to the representation of a cluster in a partition as an undirected graph, where edges are placed between stations whose distance is below a suitable threshold, which determines the neighbors based on geographical coordinates. For a partition $\mathbf{U}$ into K clusters, $G_1, \ldots, G_K$, the connectedness index is calculated by taking the total number of connected components associated with the clusters G_k ($k = 1, \ldots, K$) and dividing by K. This index can be regarded as a measure of spatial cohesion, which is larger when the index is smaller. The optimal number of clusters K is first selected for each value of α as the one that maximizes an index of the Dunn family (Dunn 1974). Then, the connectedness index is computed for different values of α. To account for both the spatial and temporal information, the optimal α is selected by choosing the value immediately before a sharp decrease of the index with respect to α, that is, the last value for which temporal dependence has a non-negligible impact on the final clustering.

Remark 3.1. Pappadà et al. (2018) use another approach to cluster validation. In fact, the DBA is applied uniquely to the temporal dissimilarity $\mathbf{\Delta}^{\text{ts}}$, while the choice of the number of clusters is made by using the ratio of the within-cluster to the between-cluster dispersion, defined in terms of the dissimilarity matrix $\mathbf{\Delta}^{\text{sp}}$. In this case, hence, the hyper-parameter α is not of direct use.

4 | Illustration

In this section, we illustrate the different steps of the TRIPLE-C algorithm with both synthetic and real data.

4.1 | Dissimilarity-Based Aggregation

Firstly, we provide a step-by-step application of the TRIPLE-C algorithm (dissimilarity aggregation, see Figure 3) to synthetic data. As explained in Section 2, the clustering algorithm (DBA) is required by the process and must be chosen in advance by the user. Here, we consider the classical hierarchical agglomerative clustering and adopt the average linkage (Average-HAC), which represents a compromise between single and complete linkages. The main steps of the algorithm in Figure 3 are detailed below.

Data. We simulate $n = 20$ time series (random sample) of length $T = 250$, in such a way that they form two equally sized groups of 10 time series. The two groups are independent, while the within-group dependence is modeled using a 10-dimensional Clayton copula with a fixed pairwise Kendall's tau value. Here, we consider $\tau = 0.5$ for each group. The complete dataset is then obtained by concatenating these two groups and assigning to each time series a pair of spatial coordinates. The two groups of time series simulated from the two Clayton copulas C and C' are represented in Figure 5. As can be seen, the data points are spatially distant so as to resemble a set of locations belonging to two different regions (e.g., two islands).

Dependence retrieval and dissimilarity mapping. The temporal similarity between each pair of time series is obtained using the van der Waerden correlation coefficient, also known as normal score correlation (Genest and Verret 2005). We recall that such a coefficient is defined, for any continuous random pair (X, Y) with copula C and marginals F_X and F_Y by

$$\zeta(X, Y) = \rho_P(\Phi^{-1}(F_X(X)), \Phi^{-1}(F_Y(Y))),$$

where ρ_P denotes the linear Pearson's correlation, while Φ denotes the standard Gaussian distribution (Koike and Hofert 2024). Interestingly, the van der Waerden coefficient is also a concordance measure; thus, it assumes the value 1 when both variables are comonotonic. The dissimilarity matrix is then obtained by subtracting these similarity values from one. Thus, comonotonic time series have dissimilarity value equal to 0 (as required by the TRIPLE-C algorithm). The dissimilarity mapping is estimated non-parametrically, starting with the empirical copula associated with data (Genest and Verret 2005).

Spatial proximity retrieval. The spatial information is collected into a dissimilarity matrix, $\mathbf{\Delta}^{\text{sp}}$, in two ways:

A. by using the Euclidean distance between the points
B. by assigning value 0 to time series that are collected from sites within the same green area (see Figure 5)

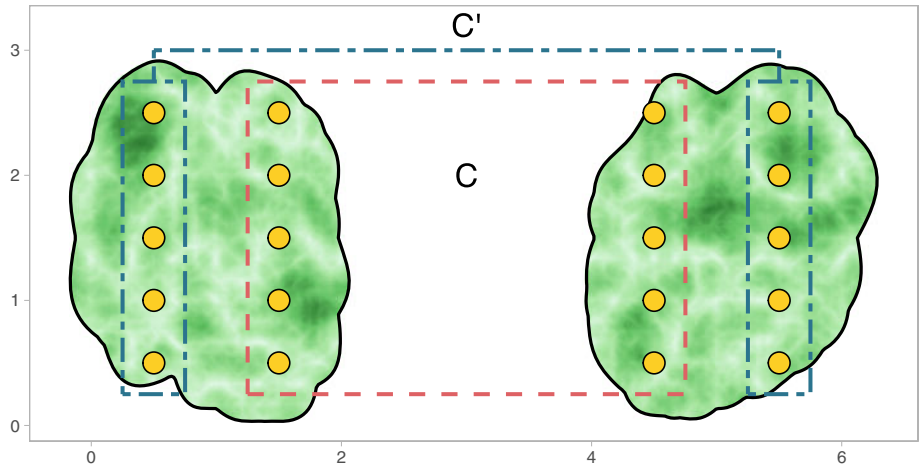

FIGURE 5 | Synthetic data. The 10 data points inside the red dashed box are simulated from the 10-dimensional Clayton copula C. The other 10 points in the teal dot-dashed boxed are simulated from the 10-dimensional Clayton copula C'.

and value 1 to time series from different geographical regions (the two green islands). Then, the dissimilarity matrix is a symmetric (0, 1)-matrix with zeros on its diagonal.

Dissimilarity aggregation. As discussed in Section 3.2, the aggregation of the dissimilarity matrices from the temporal and spatial information is a crucial step for which several alternative approaches have been explored. Here, we consider the convex combination in Equation (10), yielding the matrix $\mathbf{\Delta}^{\alpha}$ with entries

$$\Delta_{ij}^{\alpha} = (1-\alpha)\Delta_{ij}^{\text{ts}} + \alpha\Delta_{ij}^{\text{sp}}, \quad \alpha \in [0,1]$$

where Δ_{ij}^{ts} is the dissimilarity obtained via the van der Waerden coefficient, that is, $\Delta_{ij}^{\text{ts}} = 1 - \zeta_{ij}$, and Δ_{ij}^{sp} is the spatial dissimilarity obtained via method (A) or (B) from the previous step. In the former case, the Euclidean distances, d_{ij}, are transformed using the Matérn correlation function of the form $\rho_{\text{Matern}}(u) = \exp(-(u/\phi)^2)$ (Diggle and Ribeiro 2007), where $\phi = 2.31$. Then, the spatial dissimilarity is given, for each pair (i, j), by $\Delta_{ij}^{\text{sp}} = 1 - \rho_{\text{Matern}}(d_{ij})$ (see Example 3.5).

Cluster Validation and Hyper-parameter selection. The matrix $\mathbf{\Delta}^{\alpha}$ obtained in the previous step depends on the hyper-parameter $\alpha \in [0, 1]$, in such a way that a pure temporal dissimilarity is obtained for $\alpha = 0$, while $\alpha = 1$ yields the pure spatial case. We note that, in such extreme cases, the dendrogram resulting from the Average-HAC algorithm clearly indicates the partition of the data points into two groups, each formed of 10 points. As expected, these two clusters correspond to the points generated from the copulas C and C', respectively, in the pure temporal case; conversely, in the pure spatial clustering, the partition coincides with the groups of points belonging to two of the green areas in Figure 5.

Given a discrete set of values of $\alpha \in (0, 1)$ and the associated matrix $\mathbf{\Delta}^{\alpha}$, the optimal number of clusters, K^{α}, is selected as the value K that maximizes the silhouette index over different values of $K \in \{2, \ldots, 10\}$. The selected K^{α} is then used to compute the silhouette index with respect to the full temporal (respectively, spatial) dissimilarity matrix $\mathbf{\Delta}^{\text{ts}}$ (respectively, $\mathbf{\Delta}^{\text{sp}}$). The results of this procedure are shown in Figure 6, where the left and right panels refer to the spatial proximity phase based on (A) the Euclidean distances, and (B) the binary matrix, respectively. The evolution of the silhouette index in Figure 6 suggests that $\alpha = 0.45$ and $\alpha = 0.3$ may represent suitable choices for the scenarios (A) and (B), respectively, as they allow the spatial component to gain compactness without overly penalizing the temporal aspect. Such optimal values of α correspond to the optimal number of clusters $K^{\alpha} = 3$ in (A) and $K^{\alpha} = 4$ in (B).

Output. The final clusters obtained through the proposed procedure are illustrated in Figures 7 and 8. In Figure 7, the influence of the Euclidean distance between the two islands is not negligible; in this case, the optimal clustering preserves the internal structure of Cluster 2 while splitting the more distant data points into two separate clusters (1 and 3). These clusters exhibit strong temporal internal dependencies but are spatially distant. In contrast, when using the binary matrix, as in Figure 8, the two islands appear clearly spatially separated. As a result, the optimal configuration splits even the central columns of points, producing four distinct groups, each characterized by strong internal correlations.

4.2 | Copula Aggregation

This section illustrates the step-by-step application of the TRIPLE-C algorithm (copula aggregation, see Figure 4) using climatological data.

In general, statistical climatology primarily aims to simplify the complex spatio-temporal variability of atmospheric and oceanic variables by providing geographical regions that have a similar behavior. In particular, from a risk assessment perspective, extreme weather events that are spatially and temporally correlated are of particular concern due to their high potential for widespread destruction. To address this, clustering methods are commonly employed to reduce the dimensionality of the problem

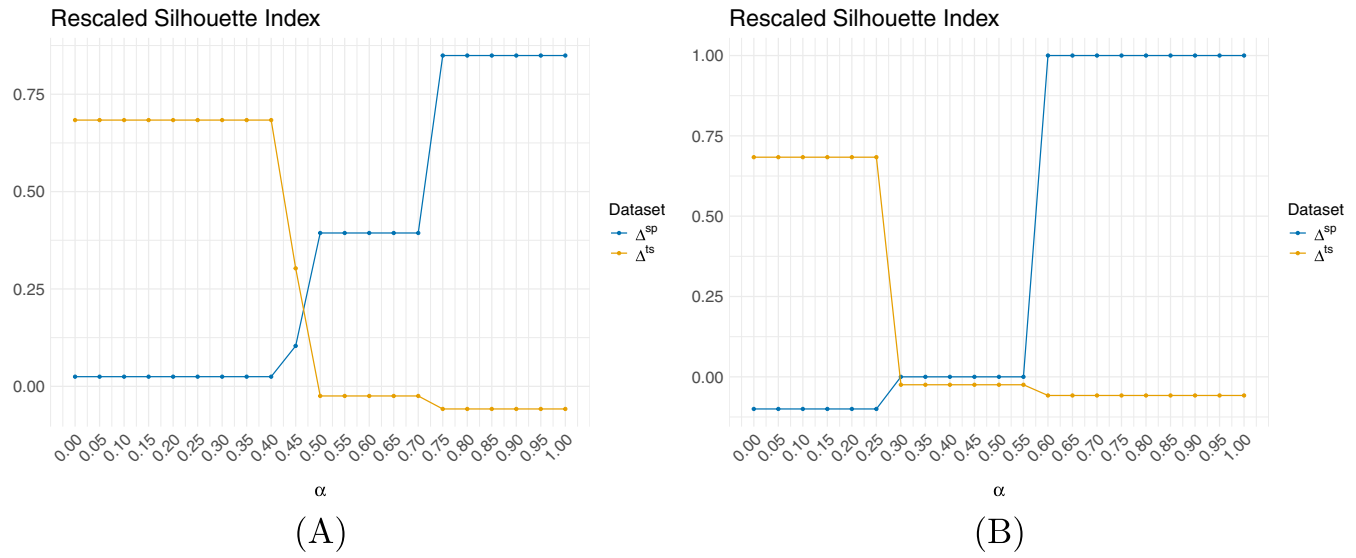

FIGURE 6 | Dissimilarity aggregation: Evolution of the average silhouette index for $\alpha \in [0, 1]$. The indices are computed with respect to the temporal matrix (orange) and the spatial matrix (blue). The left panel corresponds to spatial proximity based on Euclidean distances, while the right panel uses the adjacency matrix to define spatial proximity.

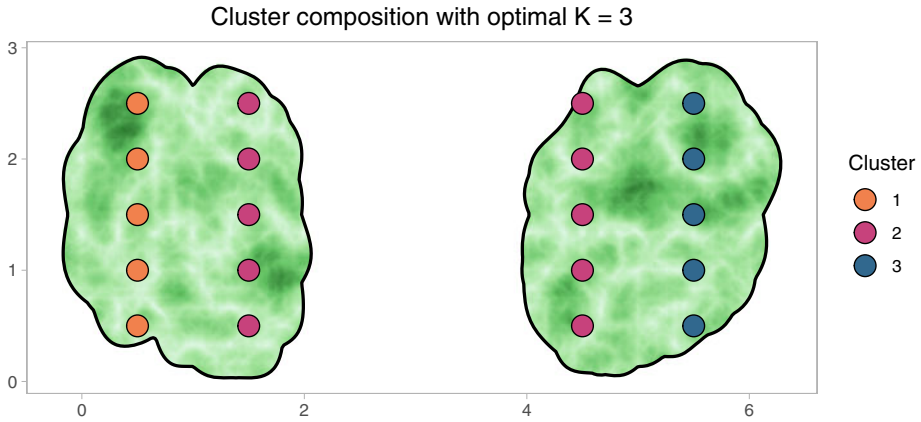

FIGURE 7 | Cluster composition for the optimal pair of hyper-parameters $\alpha = 0.45$, $K^{(0.45)} = 3$. The spatial proximity is based on the Euclidean distances.

and identify regions where extreme events are likely to have a joint impact (see, for instance, Saunders et al. 2021; Straus 2018).

In this framework, inspired by Bador et al. (2015), we consider the problem of clustering summer temperature maxima over Europe. In fact, as is well known, high temperatures, especially during heat waves, can significantly affect society, leading to increased mortality, reduced productivity, and strain on infrastructure (see, e.g., Ballester et al. 2023; Ebi et al. 2021).

Before applying the proposed algorithm, it is necessary to define the analyst's preferences, which must be selected a priori and serve as input to the algorithm. Based on the motivations provided by Bernard et al. (2013) and Bador et al. (2015), the following choices are made:

- The copulas considered should belong to the extreme-value class, as this family is naturally suited for modeling phenomena involving joint maxima.
- The dissimilarity measure should be capable of capturing the extreme behavior of the joint distribution. A common and effective approach is to focus on tail dependence coefficients, which provide relevant information in the tails of the distribution.
- The clustering algorithm (DBA) should be flexible, meaning that it should not require a priori knowledge of the number of clusters and it should be able to identify representative elements of each group (e.g., medoids). In particular, algorithms based on averaging are not suitable, as averaging violates the property of max-stability: The mean of two maxima is no longer a maximum, rendering the interpretation based on tail dependence coefficients invalid.

This application leverages two clustering algorithms: The minimax linkage hierarchical clustering (Minimax-HAC) by Bien and Tibshirani (2011), and the non-hierarchical method Partitioning Around Medoids (PAM) (Kaufman and Rousseeuw 1990). Both

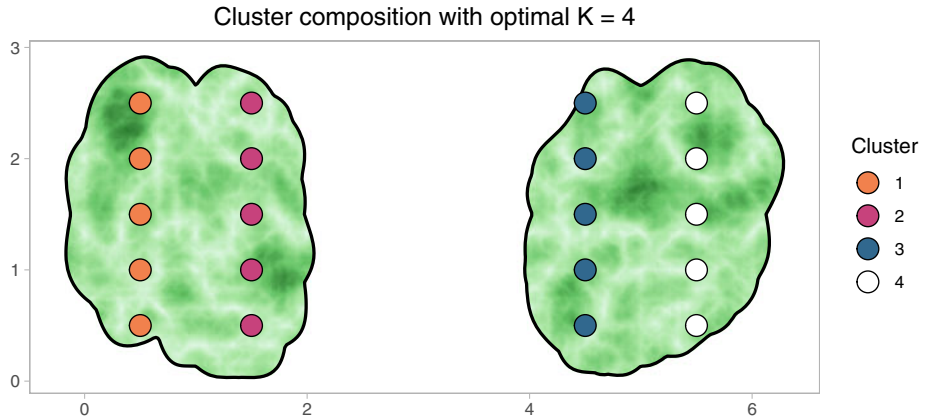

FIGURE 8 | Cluster compositions for the optimal pair of hyper-parameters $\alpha = 0.3$, $K^{(0.3)} = 4$. The spatial proximity is based on the adjacency matrix.

methods associate a prototype (or medoid) chosen from the original dataset with every cluster in the final partition. These prototypes can be used to enhance the interpretability of the clustering.

Now, with reference to the algorithm described in Figure 4, the main steps can be considered as follows:

Data. We focus on summer temperature maxima across Europe. Specifically, we analyze time series of hourly maximum temperatures for the months of June, July, and August (JJA) from 1960 to 2024 downloaded from the Climate Data Store,[1] which collects global climate and weather data of the past 8 decades (the temporal data). The time series data are collected over a regular latitude–longitude grid of 0.25 degrees (the spatial data).

Dependence retrieval. From the time series data set, we have extracted the seasonal maxima and the data recorded over the land, leading to $n = 1418$ time series of length $T = 65$. The obtained time series have no seasonal component, but may present a trend due to climate change. Thus, we estimate and remove a linear trend fitted by regression to obtain the detrended time series of seasonal maxima. From the residuals, we extract the pseudo-observations and derive the $(n \times n)$ copula matrix that expresses the pairwise dependence. Since we are mainly interested in joint maxima, a natural framework is to consider that each entry of the copula matrix is an extreme-value copula, whose Pickands' dependence function is estimated via madogram estimator (Gijbels et al. 2020).

Spatial Proximity retrieval. Our spatial information is represented as raster data, where time series are collected over a grid of cells defined by latitude and longitude coordinates. From this structure, we derive a spatial dissimilarity matrix based on the Euclidean distance between cell locations. Specifically, we employ the bivariate copulas derived from the Hüsler–Reiss n-dimensional extreme-value copula, as outlined by Benevento et al. (2024). In this setting, the copula parameters depend solely on the pairwise distances between the stations.

Copula aggregation. The copula aggregation should reflect the fact that both the temporal as well as the spatial copulas belong to the extreme-value class. To ensure this, we perform the aggregation within the same class. Specifically, the copula matrix $\mathbf{C}^{\alpha}$ is obtained via the element-wise operation $C^{\alpha}_{ij}(u,v) = C^{\text{ts}}_{ij}(u^{1-\alpha}, v^{1-\alpha})C^{\text{sp}}_{ij}(u^{\alpha}, v^{\alpha})$ for every $\alpha \in [0,1]$, as illustrated in Example 3.7.

Dissimilarity mapping. The dissimilarity function $d^{1,1}$ should reflect the extreme behavior of the copula in the upper tail. Therefore, a natural choice is to consider the dissimilarity function defined as in Equation (8), by using λ_U. As an alternative choice, one can simply consider transforming the coefficients $\lambda_U(C^{\alpha}_{ij}(u,v))$ via the mapping $\Delta^{\alpha}_{ij} = 1 - \lambda_U(C^{\alpha}_{ij}(u,v))$. In such a case, due to the particular properties of the extreme-value class of copulas combined with the applied operation (Benevento et al. 2024), the (i,j)-entry of $\mathbf{\Delta}^{\alpha}$ can be expressed as

$$\Delta^{\alpha}_{ij} = 1 - ((1-\alpha)\lambda_U(C^{\text{ts}}_{ij}) + \alpha\lambda_U(C^{\text{sp}}_{ij}))$$

with $\alpha \in [0,1]$.

Cluster Validation and Hyper-parameter selection. The previous step of dissimilarity mapping produces a collection of matrices $\mathbf{\Delta}^{\alpha}$ that depend on the value of α. Here, we let this parameter vary in the set $\{0, 0.05, 0.1, \ldots, 0.95, 1\}$ and, for each value in this set, we perform the two clustering approaches mentioned above, namely the Minimax-HAC and the PAM algorithm. For the latter, the number of clusters is required as input. Thus, we obtain the partitions into K clusters, where we let the number of clusters vary from $K = 2$ to $K = 30$.

To tackle the issue of selecting α in the considered set, we adopt as a criterion the average number of connected components, as illustrated in Example 3.11. Namely, for a fixed $\alpha \in \{0, 0.05, 0.1, \ldots, 0.95, 1\}$ and a given K ($K = 2, \ldots, 30$), Figure 9 shows the boxplots of the distribution of the average number of connected components for the partitions with K clusters ($K = 2, \ldots, 30$), for all possible values of α and the two selected DBAs. When increasing α from the minimum value of zero, the relevance of temporal dependence decreases, which may produce a loss of information. On the other hand, while α becomes larger, the average number of connected components decreases,

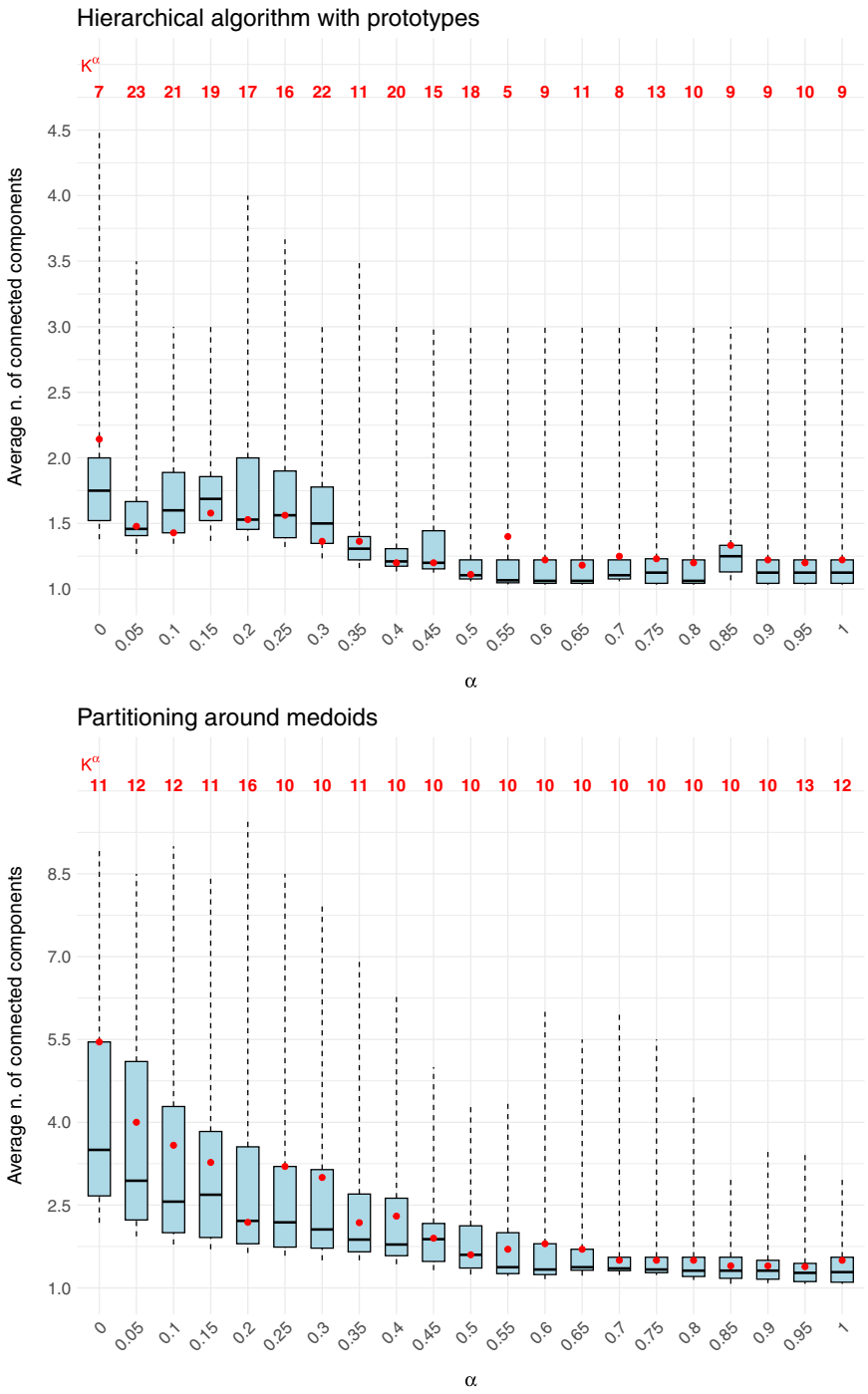

FIGURE 9 | Boxplots of average number of connected components (y-axis) of the partition into K clusters ($K \in \{2, \ldots, 30\}$), for $\alpha \in \{0, 0.05, \ldots, 1\}$ (x-axis). For each α, the optimal K^α selected via the silhouette index is reported in red in the upper part of the plot, and the associated index values are the points highlighted in red. The DBA used is the hierarchical clustering with prototypes (upper panel) and the partitioning around medoids (lower panel).

producing clustering results that are characterized by a stronger spatial compactness. Hence, an optimal α comes at the best trade-off of both aspects. Observing Figure 9, we could consider $\alpha = 0.35$ and $\alpha = 0.40$ as appropriate choices for the hierarchical algorithm with prototypes and the PAM algorithm, respectively. Finally, for the chosen α, we select the optimal number of clusters K^α as the one that maximizes the silhouette index, yielding the values $K^\alpha = 11$ for Minimax-HAC and $K^\alpha = 10$ for PAM.

Output. The partitions resulting from the selected α and K are shown in Figure 10 (see the Supporting Information for all the maps obtained from the clustering of $\mathbf{\Delta}^\alpha$, where α varies in the considered set). Although we have a finite set of representative stations, the map is displayed using continuous colors to provide a full spatial representation across Europe. Each station is associated with a rectangular area (pixel), allowing the entire map to be fully covered. As can be seen, the

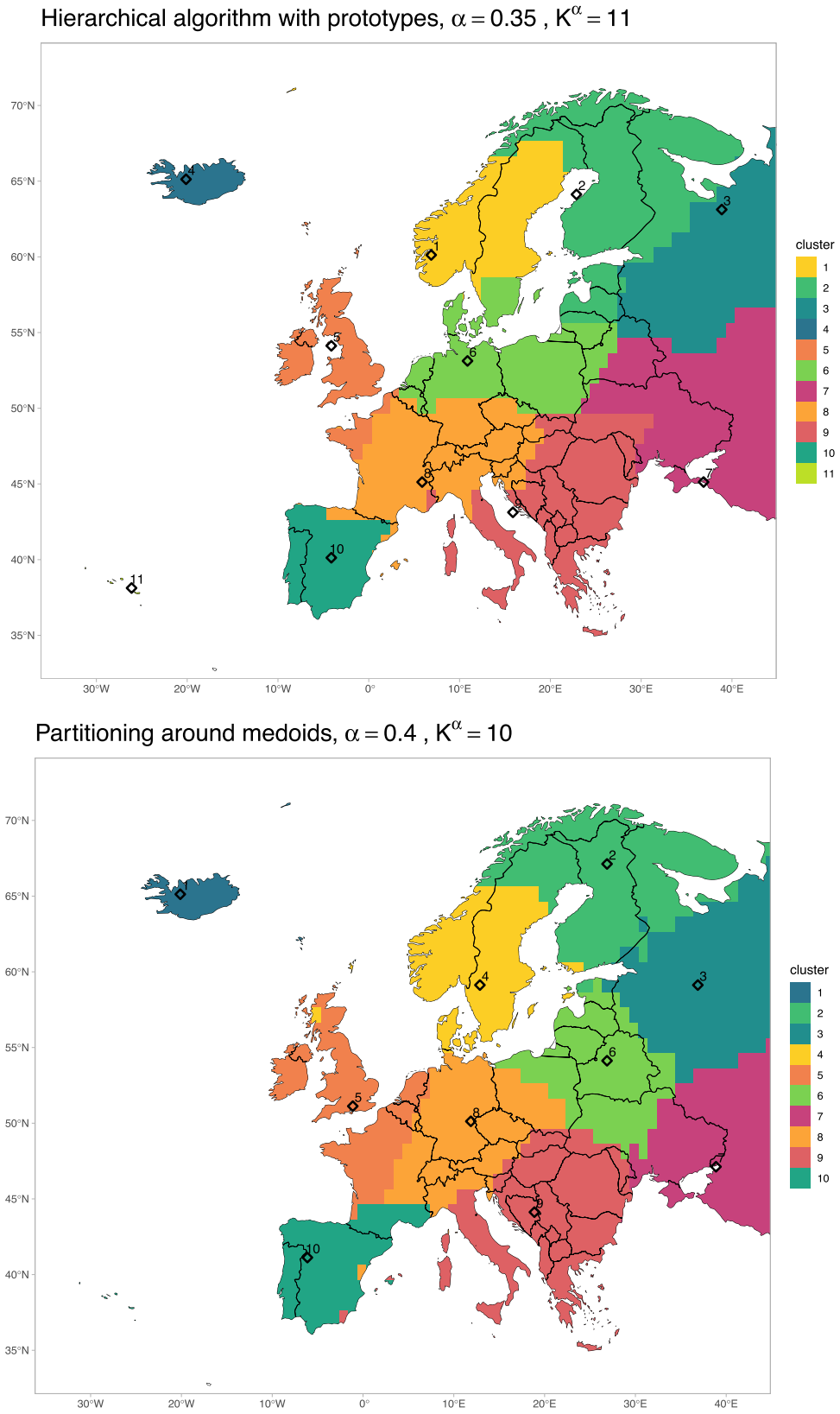

FIGURE 10 | Clustering of JJA maximum temperatures (de-trended) obtained via Minimax-HAC (upper panel) and PAM (lower panel). Colors refer to the $K^{\alpha} = 11$ and $K^{\alpha} = 10$ clusters, respectively. Diamond-shaped locations represent the Minimax-HAC prototypes and PAM medoids, respectively.

two clusters exhibit some overlap and are both characterized by a high degree of spatial cohesion. In particular, the Minimax-HAC algorithm isolates the Azores Islands, which form the eleventh cluster, while the clusters obtained using PAM appear similar in size and roughly spherical.

5 | Conclusions

In this paper, we have investigated the existing literature on clustering time series using dissimilarity-based algorithms, with a specific focus on methods that account for detecting comovements of time series, taking into account the spatial structure of the data. Interestingly, copula functions offer a natural way to describe joint comovements among time series and, as such, a growing number of contributions in the literature have proposed clustering algorithms based on copulas with different fields of application. In particular, we illustrated the application of the proposed methodology on climatological data, highlighting how it can support the understanding of complex environmental patterns.

Motivated by the exploration of recent and potential future developments along this direction, we derived a conceptual framework to deal with clustering of spatial data, which can support the user in the clustering process and provide some insights on the methodology to adopt. This is especially relevant in environmental sciences, where the abundance of spatio-temporal data presents unprecedented opportunities for analyzing and understanding local and regional dynamics.

Our proposed architecture—the TRIPLE-C algorithm—is grounded on key aspects that are crucial in the field of clustering with proximity constraints in many recent works in the literature. The proposed Algorithm is able to distinguish the main steps, starting from the data and including cluster validation, for many proposed approaches in spatio-temporal clustering, depending on how the aggregation of the temporal and spatial information is performed.

It is worth noting that the choice of an appropriate clustering strategy is often strongly dependent on the application itself. The framework provided by the TRIPLE-C algorithm offers great flexibility in dealing with both the temporal and the spatial information, guaranteeing that such decisions are tailored to the user's needs.

In our opinion, at least two main aspects deserve particular attention in future works.

In current practice, clustering outputs are predominantly evaluated using validation indices designed for specific contexts, such as Euclidean spaces, where clustering properties are closely tied to shape characteristics such as compactness, separation, and roundness (see also De Luca and Zuccolotto 2021a). However, the applicability of these indices in time series clustering is unclear, particularly when the dissimilarity measure is not derived from traditional distance metrics. This also pertains to the selection of the hyper-parameter α that is usually inspired by graphical tools.

Moreover, looking at the non-constrained case in Figure 1, clustering algorithms with spatial constraints have mainly modified the first (dependence retrieval) or the second phase of the process (dissimilarity mapping). However, a promising avenue could be to intervene directly on the cluster validation phase, for instance, by including a spatial penalty term in the objective function of the clustering method. This approach is quite typical in fuzzy clustering (Pham 2001; Coppi et al. 2010; D'Urso et al. 2019) and could be further developed in the copula-based framework.

Acknowledgments

A.B. and F.D. have been supported by MUR-PRIN 2022 PNRR, Project "Stochastic Modeling of Compound Events" (No. P2022KZJTZ) funded by European Union – Next Generation EU. The work of F.D. has been carried out with partial financial support from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by EU – Next Generation EU (CUP F83C22000740001). R.P. has been supported by MUR-PRIN 2022, Project "Modeling Non-standard data and Extremes in Multivariate Environmental Time series" (No. 20223CEZSR) funded by European Union – Next Generation EU. Open access publishing facilitated by Università degli Studi di Trieste, as part of the Wiley - CRUI-CARE agreement.

Data Availability Statement

The data that support the findings of this study are openly available in ERA5 hourly data on single levels from 1940 to the present at https://cds.climate.copernicus.eu/.

Endnotes

[1] https://cds.climate.copernicus.eu/.

References

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah. 2015. "Time-Series Clustering–a Decade Review." *Information Systems* 53: 16–38.

Aguiar, C. D., R. Gutiérrez Sánchez, and E. L. Silva Camêlo. 2020. "Hierarchical Clustering With Spatial Constraints and Standardized Incidence Ratio in Tuberculosis Data." *Mathematics* 8, no. 9: 1478.

Bador, M., P. Naveau, E. Gilleland, M. Castellà, and T. Arivelo. 2015. "Spatial Clustering of Summer Temperature Maxima From the CNRM-CM5 Climate Model Ensembles & E-OBS Over Europe." *Weather and Climate Extremes* 9: 17–24.

Ballester, J., M. Quijal-Zamorano, R. F. Méndez Turrubiates, et al. 2023. "Heat-Related Mortality in Europe During the Summer of 2022." *Nature Medicine* 29, no. 7: 1857–1866.

Benevento, A., and F. Durante. 2023. "Wasserstein Dissimilarity for Copula-Based Clustering of Time Series With Spatial Information." *Mathematics* 12, no. 1: 67.

Benevento, A., and F. Durante. 2024. "Correlation-Based Hierarchical Clustering of Time Series With Spatial Constraints." *Spatial Statistics* 59: 100797.

Benevento, A., F. Durante, and R. Pappadà. 2024. "Tail-Dependence Clustering of Time Series With Spatial Constraints." *Environmental and Ecological Statistics* 31: 801–817.

Benevento, A., F. Durante, and R. Pappadà. 2025. "Semi-Supervised Time-Series Clustering Using Gaussian Copulas." In *Statistical Methods for Data Analysis and Decision Sciences (SDS 2025). Italian Statistical Society Series on Advances in Statistics*, 1–6. Springer Cham.

Bernard, E., P. Naveau, M. Vrac, and O. Mestre. 2013. "Clustering of Maxima: Spatial Dependencies Among Heavy Rainfall in France." *Journal of Climate* 26, no. 20: 7929–7937.

Bien, J., and R. Tibshirani. 2011. "Hierarchical Clustering With Prototypes via Minimax Linkage." *Journal of the American Statistical Association* 106, no. 495: 1075–1084.

Boulin, A., E. Di Bernardino, T. Laloë, and G. Toulemonde. 2022. "Non-Parametric Estimator of a Multivariate Madogram for missing-Data

and Extreme Value Framework." *Journal of Multivariate Analysis* 192: 105059.

Boulin, A., E. Di Bernardino, T. Laloë, and G. Toulemonde. 2025. "Identifying Regions of Concomitant Compound Precipitation and Wind Speed Extremes Over Europe." *Journal of the Royal Statistical Society Series C: Applied Statistics* 74: 1057–1076.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Cai, J., J. Hao, H. Yang, X. Zhao, and Y. Yang. 2023. "A Review on Semi-Supervised Clustering." *Information Sciences* 632: 164–200.

Caiado, J., and N. Crato. 2010. "Identifying Common Dynamic Features in Stock Returns." *Quantitative Finance* 10, no. 7: 797–807.

Caiado, J., N. Crato, and D. Peña. 2006. "A Periodogram-Based Metric for Time Series Classification." *Computational Statistics & Data Analysis* 50, no. 10: 2668–2684.

Campello, R. J., and E. R. Hruschka. 2006. "A Fuzzy Extension of the Silhouette Width Criterion for Cluster Analysis." *Fuzzy Sets and Systems* 157, no. 21: 2858–2875.

Carpita, M., G. De Luca, R. Metulini, and P. Zuccolotto. 2024. "Traffic Flows Time Series in a Flood-Prone Area: Modeling and Clustering on Extreme Values With a Spatial Constraint." *Stochastic Environmental Research and Risk Assessment* 38: 3109–3125.

Castrovilli, R., F. Durante, D. Gallo, and G. Salvadori. 2025. "Regionalization Methods for Compound Extremes Based on the Wasserstein Distance." In *Methodological and Applied Statistics and Demography III SIS 2024. Italian Statistical Society Series on Advances in Statistics*, edited by A. Pollice and P. Mariani, 360–365. Springer Cham.

Chavent, M., V. Kuentz-Simonet, A. Labenne, and J. Saracco. 2018. "ClustGeo: an R Package for Hierarchical Clustering With Spatial Constraints." *Computational Statistics* 33, no. 4: 1799–1822.

Claesen, M., and B. De Moor. 2015. "Hyperparameter Search in Machine Learning." arXiv preprint arXiv:1502.02127.

Coppi, R., P. D'Urso, and P. Giordani. 2010. "A Fuzzy Clustering Model for Multivariate Spatial Time Series." *Journal of Classification* 27, no. 1: 54–88.

Côté, M., and C. Genest. 2015. "A Copula-Based Risk Aggregation Model." *Canadian Journal of Statistics* 43, no. 1: 60–81.

Czado, C., U. Schepsmeier, and A. Min. 2012. "Maximum Likelihood Estimation of Mixed C-Vines With Application to Exchange Rates." *Statistical Modelling* 12, no. 3: 229–255.

D'Urso, P., L. De Giovanni, M. Disegna, and R. Massari. 2019. "Fuzzy Clustering With Spatial–Temporal Information." *Spatial Statistics* 30: 71–102.

D'Urso, P., G. De Luca, V. Vitale, and P. Zuccolotto. 2023. "Tail Dependence-Based Fuzzy Clustering of Financial Time Series." In *Annals of Operations Research*, 1–27.

D'Urso, P., D. Di Lallo, and E. A. Maharaj. 2013. "Autoregressive Model-Based Fuzzy Clustering and Its Application for Detecting Information Redundancy in Air Pollution Monitoring Networks." *Soft Computing* 17, no. 1: 83–131.

D'Urso, P., and V. Vitale. 2020. "A Robust Hierarchical Clustering for Georeferenced Data." *Spatial Statistics* 35: 100407.

David, P., and W. Gu. 2019. "A Riemannian Structure for Correlation Matrices." *Operators and Matrices* 13, no. 3: 607–627.

Davis, R. A., L. Fernandes, and K. Fokianos. 2023. "Clustering Multivariate Time Series Using Energy Distance." *Journal of Time Series Analysis* 44, no. 5-6: 487–504.

De Carvalho, M., R. Huser, and R. Rubio. 2023. "Similarity-Based Clustering for Patterns of Extreme Values." *Stat* 12, no. 1: e560.

De Keyser, S., and I. Gijbels. 2024. "Hierarchical Variable Clustering via Copula-Based Divergence Measures Between Random Vectors." *International Journal of Approximate Reasoning* 165: 22 Id/No 109090.

De Luca, G., and P. Zuccolotto. 2011. "A Tail Dependence-Based Dissimilarity Measure for Financial Time Series Clustering." *Advances in Data Analysis and Classification* 5: 323–340.

De Luca, G., and P. Zuccolotto. 2021a. "Hierarchical Time Series Clustering on Tail Dependence With Linkage Based on a Multivariate Copula Approach." *International Journal of Approximate Reasoning* 139: 88–103.

De Luca, G., and P. Zuccolotto. 2021b. "Regime Dependent Interconnectedness Among Fuzzy Clusters of Financial Time Series." *Advances in Data Analysis and Classification* 15, no. 2: 315–336.

De Luca, G., and P. Zuccolotto. 2023. "Dynamic Time Series Clustering With Multivariate Linkage and Automatic Dendrogram Cutting Using a Recursive Partitioning Algorithm." *Information Sciences* 649: 119605.

Deb, S., and S. Karmakar. 2023. "A Novel Spatio-Temporal Clustering Algorithm With Applications on COVID-19 Data From the United States." *Computational Statistics & Data Analysis* 188: 107810.

Di Lascio, F. M. L., A. Gatto, and S. Giannerini. 2025. "CoImp: Parametric and Nonparametric Copula-Based Imputation Methods." Free University of Bozen-Bolzano, Italy. R Package Version: 2.1.0. https://CRAN.R-project.org/package=CoImp.

Di Lascio, F. M. L., A. Menapace, and R. Pappadà. 2024. "A Spatially-Weighted AMH Copula-Based Dissimilarity Measure for Clustering Variables: An Application to Urban Thermal Efficiency." *Environmetrics* 35, no. 1: e2828.

Diggle, P. J., and P. J. Ribeiro. 2007. *Model-Based Geostatistics*. Springer.

Disegna, M., P. D'Urso, and F. Durante. 2017. "Copula-Based Fuzzy Clustering of Spatial Time Series." *Spatial Statistics* 21: 209–225.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka. 2013. "Selecting and Estimating Regular Vine Copulae and Application to Financial Returns." *Computational Statistics & Data Analysis* 59: 52–69.

Dunn, J. C. 1974. "Well-Separated Clusters and Optimal Fuzzy Partitions." *Journal of Cybernetics* 4, no. 1: 95–104.

Durante, F. 2009. "Construction of Non-Exchangeable Bivariate Distribution Functions." *Statistical Papers* 50, no. 2: 383–391.

Durante, F., J. Fernández-Sánchez, and R. Pappadà. 2015. "Copulas, Diagonals, and Tail Dependence." *Fuzzy Sets and Systems* 264: 22–41.

Durante, F., S. Fuchs, and R. Pappadà. 2025. "Clustering of Compound Events Based on Multivariate Comonotonicity." *Spatial Statistics* 66: 100881.

Durante, F., and P. Jaworski. 2010. "Spatial Contagion Between Financial Markets: a Copula-Based Approach." *Applied Stochastic Models in Business and Industry* 26, no. 5: 551–564.

Durante, F., and R. Pappadà. 2015. "Cluster analysis of time series via Kendall distribution." In *Strengthening Links Between Data Analysis and Soft computing. Collected Papers Based on the Presentations at the 7th International Conference on Soft Methods in Probability and Statistics, SMPS 2014*, 209–216. Springer.

Durante, F., R. Pappadà, and N. Torelli. 2014. "Clustering of Financial Time Series in Risky Scenarios." *Advances in Data Analysis and Classification* 8: 359–376.

Durante, F., R. Pappadà, and N. Torelli. 2015. "Clustering of Time Series via Non–Parametric Tail Dependence Estimation." *Statistical Papers* 56, no. 3: 701–721.

Durante, F., and C. Sempi. 2016. *Principles of Copula Theory*. Vol. 474. CRC Press Boca Raton.

Ebi, K. L., A. Capon, P. Berry, et al. 2021. "Hot Weather and Heat Extremes: Health Risks." *Lancet* 398, no. 10301: 698–708.

Erhardt, T. M., and C. Czado. 2018. "Standardized Drought Indices: a Novel Univariate and Multivariate Approach." *Journal of the Royal Statistical Society Series C: Applied Statistics* 67, no. 3: 643–664.

Fouedjio, F. 2016. "A Hierarchical Clustering Method for Multivariate Geostatistical Data." *Spatial Statistics* 18: 333–351.

Fouedjio, F. 2020. "Clustering of Multivariate Geostatistical Data." *Wiley Interdisciplinary Reviews: Computational Statistics* 12, no. 5: e1510.

Fuchs, S., F. M. L. Di Lascio, and F. Durante. 2021. "Dissimilarity Functions for Rank-Invariant Hierarchical Clustering of Continuous Variables." *Computational Statistics & Data Analysis* 159: 107201.

Fuchs, S., and Y. Wang. 2024. "Hierarchical Variable Clustering Based on the Predictive Strength Between Random Vectors." *International Journal of Approximate Reasoning* 170: 25 Id/No 109185.

Gaetan, C., P. Girardi, and V. M. Musau. 2025. "Spatial Quantile Clustering of Climate Data." *Advances in Data Analysis and Classification* 19, no. 1: 147–175.

Gelb, J., and P. Apparicio. 2021. "Apport de la Classification Floue c-Means Spatiale en géographie: Essai de Taxinomie Socio-résidentielle et Environnementale à Lyon." *Cybergeo: European Journal of Geography*, no. 972.

Genest, C., K. Ghoudi, and L.-P. Rivest. 1998. "Understanding Relationships Using Copulas by Edward Frees and Emiliano Valdez, January 1998." *North American Actuarial Journal* 2, no. 3: 143–149.

Genest, C., and F. Verret. 2005. "Locally Most Powerful Rank Tests of Independence for Copula Models." *Journal of Nonparametric Statistics* 17, no. 5: 521–539.

Gijbels, I., V. Kika, and M. Omelka. 2020. "Multivariate Tail Coefficients: Properties and Estimation." *Entropy* 22, no. 7: 728.

Giraldo, R., P. Delicado, and J. Mateu. 2012. "Hierarchical Clustering of Spatially Correlated Functional Data." *Statistica Neerlandica* 66, no. 4: 403–421.

Górecki, J., M. Hofert, and M. Holenă. 2017. "Kendall's Tau and Agglomerative Clustering for Structure Determination of Hierarchical Archimedean Copulas." *Dependence Modeling* 5, no. 1: 75–87.

Górecki, J., M. Hofert, and O. Okhrin. 2021. "Outer Power Transformations of Hierarchical Archimedean Copulas: Construction, Sampling and Estimation." *Computational Statistics & Data Analysis* 155: 107109.

Gudendorf, G., and J. Segers. 2010. "Extreme-Value Copulas." In *Copula Theory and Its Applications*, edited by P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, vol. 198, 127–145. Springer.

Guénard, G., and P. Legendre. 2022. "Hierarchical Clustering With Contiguity Constraint in R." *Journal of Statistical Software* 103: 1–26.

Haggarty, R. A., C. A. Miller, and E. M. Scott. 2015. "Spatially Weighted Functional Clustering of River Network Data." *Journal of the Royal Statistical Society Series C: Applied Statistics* 64, no. 3: 491–506.

Hennig, C., M. Meila, F. Murtagh, and R. Rocci. 2015. *Handbook of Cluster Analysis*. CRC Press.

Hofert, M., and T. Koike. 2019. "Compatibility and Attainability of Matrices of Correlation-Based Measures of Concordance." *Astin Bulletin* 49, no. 3: 885–918.

Joe, H. 1997. "Multivariate Models and Dependence Concepts." In *Monographs on Statistics and Applied Probability*, vol. 73. Chapman and Hall.

Kalpakis, K., D. Gada, and V. Puttagunta. 2001. "Distance Measures for Effective Clustering of ARIMA Time-Series." In *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM)*, 273–280. San Jose IEEE.

Kaufman, L., and P. J. Rousseeuw. 1990. "Partitioning Around Medoids (Program PAM)." In *Finding Groups in Data*, edited by L. Kaufman and P. J. Rousseeuw, 68–125. John Wiley & Sons, Ltd.

Khoudraji, A. 1995. *Contributions à l'étude des copules et à la modélisation des valeurs extrêmes bivariées*. PhD thesis. Université Laval, Québec (Canada).

Kisilevich, S., F. Mansmann, M. Nanni, and S. Rinzivillo. 2010. "Spatio-Temporal Clustering." In *Data Mining and Knowledge Discovery Handbook*, edited by L. Rokach and O. Maimon, 855–874. Springer.

Koike, T., and M. Hofert. 2024. "Comparison of Correlation-Based Measures of Concordance in Terms of Asymptotic Variance." *Journal of Multivariate Analysis* 201: Id/No 105265.

Kojadinovic, I. 2004. "Agglomerative Hierarchical Clustering of Continuous Variables Based on Mutual Information." *Computational Statistics & Data Analysis* 46, no. 2: 269–294.

Kojadinovic, I. 2010. "Hierarchical Clustering of Continuous Variables Based on the Empirical Copula Process and Permutation Linkages." *Computational Statistics & Data Analysis* 54, no. 1: 90–108.

Kopczewska, K. 2022. "Spatial Machine Learning: New Opportunities for Regional Science." *Annals of Regional Science* 68, no. 3: 713–755.

Legendre, P., and O. Gauthier. 2014. "Statistical Methods for Temporal and Space-Time Analysis of Community Composition Data." *Proceedings of the Royal Society B: Biological Sciences* 281, no. 1778: 20132728.

Legendre, P., and L. Legendre. 1998. *Numerical Ecology. Transl. and Revised From the 2nd French (1984) Edition*. 2nd ed. Elsevier Science.

Liao, T. W. 2005. "Clustering of Time Series Data – A Survey." *Pattern Recognition* 38, no. 11: 1857–1874.

Liebscher, E. 2008. "Construction of Asymmetric Multivariate Copulas." *Journal of Multivariate Analysis* 99, no. 10: 2234–2250.

Liebscher, E. 2024. "Fitting Copulas in the Case of Missing Data." *Statistical Papers* 65, no. 6: 3681–3711.

Lloyd, C. 2010. *Spatial Data Analysis: an Introduction for GIS Users*. Oxford University Press.

López-Oriona, Á., P. D'Urso, J. A. Vilar, and B. Lafuente-Rego. 2021. "Spatial Weighted Robust Clustering of Multivariate Time Series Based on Quantile Dependence With an Application to Mobility During COVID-19 Pandemic." *IEEE Transactions on Fuzzy Systems* 30, no. 9: 3990–4004.

Maharaj, E. A., and P. D'Urso. 2011. "Fuzzy Clustering of Time Series in the Frequency Domain." *Information Sciences* 181, no. 7: 1187–1211.

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

Marcon, G., S. A. Padoan, P. Naveau, P. Muliere, and J. Segers. 2017. "Multivariate Nonparametric Estimation of the Pickands Dependence Function Using Bernstein Polynomials." *Journal of Statistical Planning and Inference* 183: 1–17.

Marti, G., S. Andler, F. Nielsen, and P. Donnat. 2016. "Optimal Transport vs. Fisher-Rao Distance Between Copulas for Clustering Multivariate Time Series." In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, 1–5. IEEE.

Mattera, R., and P. H. Franses. 2024. "Spatio-Temporal Hierarchical Clustering of Interval Time Series With Application to Suicide Rates in Europe." *Statistical Modelling*: 1471082X241299250.

Maume-Deschamps, V., P. Ribereau, and M. Zeidan. 2025. "Regionalization of the Extremal Dependence Structure Using Spectral Clustering." *Stochastic Environmental Research and Risk Assessment* 39, no. 2: 725–745.

McNeil, A. J., J. G. Nešlehová, and A. D. Smith. 2022. "On Attainability of Kendall's Tau Matrices and Concordance Signatures." *Journal of Multivariate Analysis* 191: 105033.

Morelli, C., S. Boccaletti, P. Maranzano, and P. Otto. 2025. "Multidimensional Spatiotemporal Clustering–an Application to Environmental Sustainability Scores in Europe." *Environmetrics* 36, no. 2: e2893.

Musau, V. M., C. Gaetan, and P. Girardi. 2022. "Clustering of Bivariate Satellite Time Series: A Quantile Approach." *Environmetrics* 33, no. 7: 16.

Nasri, B. R., and B. N. Rémillard. 2019. "Copula-Based Dynamic Models for Multivariate Time Series." *Journal of Multivariate Analysis* 172: 107–121.

Nelsen, R. B. 2006. *An Introduction to Copulas*. 2nd ed. Springer.

Neumeyer, N., and M. Omelka. 2025. "Generalized Hadamard Differentiability of the Copula Mapping and Its Applications." *Bernoulli* 31, no. 2: 1451–1474.

Neumeyer, N., M. Omelka, and Š. Hudecová. 2019. "A Copula Approach for Dependence Modeling in Multivariate Nonparametric Time Series." *Journal of Multivariate Analysis* 171: 139–162.

Nielsen, F., G. Marti, S. Ray, and S. Pyne. 2021. "Clustering Patterns Connecting COVID-19 Dynamics and Human Mobility Using Optimal Transport." *Sankhyā, Series. B* 83, no. 1: 167–184.

Oliver, M., and R. Webster. 1989. "A Geostatistical Basis for Spatial Weighting in Multivariate Classification." *Mathematical Geology* 21: 15–35.

Palacios-Rodriguez, F., E. D. Bernardino, and M. Mailhot. 2023. "Smooth Copula-Based Generalized Extreme Value Model and Spatial Interpolation for Extreme Rainfall in Central Eastern Canada." *Environmetrics* 34, no. 3: e2795.

Pappadà, R., F. Durante, G. Salvadori, and C. De Michele. 2018. "Clustering of Concurrent Flood Risks via Hazard Scenarios." *Spatial Statistics* 23: 124–142.

Patton, A. J. 2012. "A Review of Copula Models for Economic Time Series." *Journal of Multivariate Analysis* 110: 4–18.

Pawitan, Y., and J. Huang. 2003. "Constrained Clustering of Irregularly Sampled Spatial Data." *Journal of Statistical Computation and Simulation* 73, no. 12: 853–865.

Pértega Díaz, S., and J. A. Vilar. 2010. "Comparing Several Parametric and Nonparametric Approaches to Time Series Clustering: a Simulation Study." *Journal of Classification* 27, no. 3: 333–362.

Pham, D. L. 2001. "Spatial Models for Fuzzy Clustering." *Computer Vision and Image Understanding* 84, no. 2: 285–297.

Piccolo, D. 1990. "A Distance Measure for Classifying ARIMA Models." *Journal of Time Series Analysis* 11, no. 2: 153–164.

Rémillard, B. 2017. "Goodness-Of-Fit Tests for Copulas of Multivariate Time Series." *Econometrics* 5, no. 1: 13.

Romary, T., F. Ors, J. Rivoirard, and J. Deraisme. 2015. "Unsupervised Classification of Multivariate Geostatistical Data: Two Algorithms." *Computers & Geosciences* 85: 96–103.

Rüschendorf, L. 2009. "On the Distributional Transform, Sklar's Theorem, and the Empirical Copula Process." *Journal of Statistical Planning and Inference* 139: 3921–3927.

Ruspini, E. H., J. C. Bezdek, and J. M. Keller. 2019. "Fuzzy Clustering: A Historical Perspective." *IEEE Computational Intelligence Magazine* 14, no. 1: 45–55.

Salvadori, G., F. Durante, C. De Michele, M. Bernardi, and L. Petrella. 2016. "A Multivariate Copula-Based Framework for Dealing With Hazard Scenarios and Failure Probabilities." *Water Resources Research* 52, no. 5: 3701–3721.

Salvadori, G., F. Durante, and E. Perrone. 2013. "Semi-Parametric Approximation of Kendall's Distribution Function and Multivariate Return Periods." *Journal de la Société Française de Statistique* 154, no. 1: 151–173.

Saunders, K. R., A. G. Stephenson, and D. J. Karoly. 2021. "A Regionalisation Approach for Rainfall Based on Extremal Dependence." *Extremes* 24, no. 2: 215–240.

Schmid, F., R. Schmidt, T. Blumentritt, S. Gaißer, and M. Ruppert. 2010. "Copula–Based Measures of Multivariate Association." In *Workshop on Copula Theory and Its Applications*, edited by P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, vol. 198, 209–236. Springer.

Segers, J. 2012. "Asymptotics of Empirical Copula Processes Under Non-Restrictive Smoothness Assumptions." *Bernoulli* 18, no. 3: 764–782.

Segers, J. 2015. "Hybrid Copula Estimators." *Journal of Statistical Planning and Inference* 160: 23–34.

Segers, J., M. Sibuya, and H. Tsukahara. 2017. "The Empirical Beta Copula." *Journal of Multivariate Analysis* 155: 35–51.

Straus, D. M. 2018. "Clustering Techniques in Climate Analysis." *Oxford Research Encyclopedia of Climate Science*.

Thanwerdas, Y., and X. Pennec. 2022. "Theoretically and Computationally Convenient Geometries on Full-Rank Correlation Matrices." *SIAM Journal on Matrix Analysis and Applications* 43, no. 4: 1851–1872.

Van Dongen, S., and A. Enright. 2012. "Metric Distances Derived From Cosine Similarity and Pearson and Spearman Correlations," arXiv Preprint arXiv:1208.3145.

Vignotto, E., S. Engelke, and J. Zscheischler. 2021. "Clustering Bivariate Dependencies of Compound Precipitation and Wind Extremes Over Great Britain and Ireland." *Weather and Climate Extremes* 32: 100318.

Wang, H., C. Song, J. Wang, and P. Gao. 2024. "A Raster-Based Spatial Clustering Method With Robustness to Spatial Outliers." *Scientific Reports* 14, no. 1: 4103.

Webster, R., and P. A. Burrough. 1972. "Computer-Based Soil Mapping of Small Areas for Sample Data." *Journal of Soil Science* 23, no. 2: 222–234.

Zuccolotto, P., G. De Luca, R. Metulini, et al. 2023. "Modeling and Clustering of Traffic Flows Time Series in a Flood Prone Area." In *Statistics for Data Science and Artificial Intelligence. Proceedings of the Statistics and Data Science Conference*, 113–118. EGEA, Pavia University Press.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1**: Supporting Information.