

Article

Wasserstein Dissimilarity for Copula-Based Clustering of Time Series with Spatial Information

Alessia Benevento  and Fabrizio Durante * 

Dipartimento di Scienze dell'Economia, Università del Salento, 73100 Lecce, Italy; alessia.benevento@unisalento.it

* Correspondence: fabrizio.durante@unisalento.it

Abstract: The clustering of time series with geo-referenced data requires a suitable dissimilarity matrix interpreting the comovements of the time series and taking into account the spatial constraints. In this paper, we propose a new way to compute the dissimilarity matrix, merging both types of information, which leverages on the Wasserstein distance. We then make a quasi-Gaussian assumption that yields more convenient formulas in terms of the joint correlation matrix. The method is illustrated in a case study involving climatological data.

Keywords: clustering; copula; dissimilarity matrix; optimal transport; time series; Wasserstein distance

MSC: 62H30; 60E15; 49Q22

1. Introduction

Time-series cluster analysis collects various unsupervised learning techniques for organizing data points collected over time into groups based on their similarity. The general objective is to maximize data similarity within clusters and minimize it across clusters (see, for instance, [1,2]).

Usually, such methods depend on the different ways of defining the similarity among time series, which can be based on (a) the observed time series values or suitable transformations thereof; (b) specific features extracted from the time series, such as autocorrelation, periodograms, etc.; or (c) the data-generating stochastic process (see [3]).

Copula-based clustering methods for time series tend to group time series according to the degree of cross-sectional dependence among different patterns (see [4] and references therein). Typical choices include rank-correlation measures [5], tail dependence coefficients [6–8], and copula distances and/or divergences [9–13]. As their specific feature, the obtained cluster composition is invariant under a monotone transformation of the original time series, hence it is usually more robust against the presence of outliers. Copula-based algorithms, hence, focus on the joint comovements among time series and, as such, are particularly used in risk analysis due to their ability to detect extreme scenarios that occur when both time series tend to fall/rise at the same time) [6,7,14–16].

When time series are collected at different geographic locations, i.e., geo-referenced data are considered where a variable is observed over time at a static spatial location, copula-based algorithms may, hence, detect the impacts of extreme events across space [17]. In such cases, it may be of interest to obtain clusters that show geographical proximity in order to enhance their interpretability. See, for instance, [18,19]. These situations occur not only in environmental science, but also in the financial context. In this latter case, the attributes may describe the economic distance between the markets and can provide deeper insights into the dependence structure of stock returns (see, e.g., [20–22]).

The goal of this paper is to consider clustering of time series for geo-referenced data. In particular, we focus on the construction of a novel pairwise dissimilarity among time series that takes into account the spatial information.



Citation: Benevento, A.; Durante, F. Wasserstein Dissimilarity for Copula-Based Clustering of Time Series with Spatial Information. *Mathematics* **2024**, *12*, 67. <https://doi.org/10.3390/math12010067>

Academic Editors: Antonio Di Crescenzo and Virginia Giorno

Received: 13 November 2023

Revised: 12 December 2023

Accepted: 21 December 2023

Published: 24 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

To fix ideas, consider $n \geq 2$ real-valued time series $\mathbf{x}_i = (x_{i1}, \dots, x_{iT})$ of length T , where x_{it} represents the observed value of the i -th time series ($i = 1, \dots, n$) at the t -th period ($t = 1, \dots, T$). Each time series is equipped with additional information on the phenomenon under consideration, represented by a p -dimensional vector of attributes, $\mathbf{s}_i = (s_{i1}, \dots, s_{ip}) \in \mathbb{R}^p$, where s_{ik} represents the value of the i -th unit ($i = 1, \dots, n$) at the k -th attribute ($k = 1, \dots, p$). The additional attributes are typically geographic coordinates of the location at which the i -th time series is observed (thus, $p = 2$ and latitude and longitude are collected). For this reason, we will refer to these vectors as the spatial information.

The classical strategy in spatio-temporal clustering (see, e.g., [23]) is to start with a $(n \times n)$ matrix of pairwise dissimilarities among $\mathbf{x}_1, \dots, \mathbf{x}_n$. This latter matrix is, hence, modified by a nonlinear function of the separating distances among $\mathbf{s}_1, \dots, \mathbf{s}_n$. Finally, the modified dissimilarity matrix is used as the input to obtain the clustering partition via algorithms such as (agglomerative) hierarchical clustering, PAM, and fuzzy C-means (see, e.g., [24–26]). In the copula-based framework, such an approach has been adopted, for instance, in [27,28], which considers the temporal matrix induced by the (pairwise) copula parameters. In [29], instead, the correlation-based dissimilarity matrix is transformed by means of the spatial information, taking into account the intrinsic geometry of the space of correlation matrices. According to [30], all these methods have the tendency to produce a smooth dissimilarity matrix without, however, reinforcing the spatial contiguity between the resulting clusters.

In [11] (see also [31]), a different strategy has been proposed. It consists of associating with the spatial information (usually contained in the Euclidean space) a suitable copula (the spatial copula). Such a latter object is, hence, combined with the temporal copula to obtain an element of the copula space that merges temporal and spatial information, and is used for the calculation of the associated dissimilarity.

The present work exploits tools from optimal transport and the Wasserstein metric (see, e.g., [32,33]) to modify the framework of [11] by introducing the following main novelties: (a) the dissimilarity measure is constructed by equipping the space of bivariate copulas \mathcal{C} with the Wasserstein metric $d_{\mathcal{W}_2}$; (b) the temporal and spatial information is merged (at the level of pairwise association) via geodesic curves [34] (also known as weighted barycenters [35]) in $(\mathcal{C}, d_{\mathcal{W}_2})$. Such an approach has the following advantages:

- The Wasserstein distance allows for a meaningful comparison between distributions also without density. This property is not shared by the most common distances and divergences, such as the total variation distance, the Hellinger distance, or the Kullback–Leibler divergence (see, e.g., [36]).
- The Wasserstein metric seems to be more appropriate to measure distance between copulas since it does not lead to counter-intuitive clusters (see, e.g., [37]).

As possible limitations, however, the calculations cannot always be explicitly made. For this reason, we rely on a quasi-Gaussian approach of the presented framework, as in [38].

The paper is organized as follows. In Section 2, we consider some tools from optimal transport theory that will be used in the manuscript. In Section 3, we present the methodology underlying the computation of a dissimilarity matrix. In Section 4, we present a quasi-Gaussian approximation of the dissimilarity matrix extraction. Section 5 illustrates the effect of the spatial dependence to the whole procedure. Finally, Section 6 presents an empirical application. Section 7 concludes the paper.

2. Background on Optimal Transport, Wasserstein Distance, and Copulas

In order to extract the spatial information and to merge the temporal and spatial dependencies, we will leverage on the 2-Wasserstein distance between probability measures [39,40] and on optimal transport theory [32,33]. To this end, we recall some basic facts that will be used in this paper (see, e.g., [32,33,39]).

Given a domain $\Omega \subseteq \mathbb{R}^d$ (that we take to be compact and convex for simplicity), we define the space $\mathcal{W}_2(\Omega)$ as the space of probability measures on Ω with finite moments of order 2 endowed with the distance $d_{\mathcal{W}_2}$, defined through

$$d_{\mathcal{W}_2}(\mu_X, \mu_Y) = \left(\inf_{\gamma \in \Gamma(\mu_X, \mu_Y)} \int_{\Omega \times \Omega} |x - y|^2 d\gamma(x, y) \right)^{1/2}, \tag{1}$$

where $\Gamma(\mu_X, \mu_Y)$ denotes the collection of all joint measures on $\Omega \times \Omega$ whose marginals are μ_X and μ_Y , respectively.

It is possible to prove that the above minimization problem has a solution, which is unique (and it is concentrated on the graph of a map $f^{\text{OT}}: \Omega \rightarrow \Omega$, called the optimal transport map) if μ_X is absolutely continuous with respect to the Lebesgue measure (actually, it is enough that μ_X vanishes on small sets [32]). Moreover, $d_{\mathcal{W}_2}$ is a distance in $\mathcal{W}_2(\Omega)$.

The space $\mathcal{W}_2(\Omega)$ is a geodesic space where, for two measures μ_X and μ_Y , the unique geodesic curve connecting them is obtained through

$$\mu^{(\alpha)} = ((1 - \alpha)id + \alpha f^{\text{OT}})_{\#} \mu_X, \tag{2}$$

where f^{OT} is the optimal transport map from μ_X to μ_Y , id is the identity mapping on \mathbb{R}^d , and $\alpha \in [0, 1]$. Here, $T_{\#}\mu$ denotes the push-forward of the measure μ under the mapping T . Such a curve is also known as the *displacement interpolation* [34]. This provides a useful interpolation between μ_X and μ_Y which is, in general, different from the convex combination of the two measures. As a matter of fact, $\mu^{(\alpha)}$ coincides with the weighted *barycenter* [35] of $(\mu_X, 1 - \alpha)$ and (μ_Y, α) , i.e., it solves the minimization problem

$$\mu^{(\alpha)} = \arg \min_{\mu} \left((1 - \alpha) d_{\mathcal{W}_2}^2(\mu_X, \mu) + \alpha d_{\mathcal{W}_2}^2(\mu_Y, \mu) \right) \tag{3}$$

over all possible probability measures μ (see also [41,42]). Wasserstein barycenters have found various applications in machine learning, especially in dealing with image recognition (see, for instance, [39]).

The actual computation related to the Wasserstein distances is largely simplified when we deal with Gaussian measures. In fact, the Wasserstein distance between centered Gaussian distributions is well known and it is given by the so-called Bures–Wasserstein distance between their covariance matrices. Specifically, let $\mathbf{X} \sim N(\mathbf{0}, \Sigma_1)$ and $\mathbf{Y} \sim N(\mathbf{0}, \Sigma_2)$ be two Gaussian random vectors with positive semi-definite covariance matrices Σ_1 and Σ_2 , respectively. Then, the following holds (see, e.g., Theorem 2.2 in [43]):

$$d_{\mathcal{W}_2}^2(\mathbf{X}, \mathbf{Y}) = d_{\mathcal{W}_2}^2(\Sigma_1, \Sigma_2) = \text{tr}(\Sigma_1) + \text{tr}(\Sigma_2) - 2\text{tr} \left(\left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right). \tag{4}$$

Such a formula also extends to the case when Σ_1 and Σ_2 are singular (see [44], p. 239). In the particular case that \mathbf{X} and \mathbf{Y} are also two-dimensional, the Wasserstein distance has the following expression (see [44], p. 239):

$$d_{\mathcal{W}_2}^2(\mathbf{X}, \mathbf{Y}) = \text{tr}(\Sigma_1) + \text{tr}(\Sigma_2) - 2 \left(\text{tr}(\Sigma_1 \Sigma_2) + 2 \sqrt{\det(\Sigma_1 \Sigma_2)} \right)^{1/2}. \tag{5}$$

The (weighted) barycenter among the Gaussian vectors \mathbf{X} and \mathbf{Y} is a Gaussian vector whose covariance matrix is of the form (see, e.g., Lemma 2.3 in [43])

$$\Sigma^{(\alpha)} = \Sigma_1 \#_{\alpha} \Sigma_2 = \Sigma_1^{-1/2} \left((1 - \alpha) \Sigma_1 + \alpha \left(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \right)^{1/2} \right) \Sigma_1^{-1/2}. \tag{6}$$

Standard references for the study of the Wasserstein distance for Gaussian measures include [44–47].

Beyond the Gaussian cases, explicit expressions for the Wasserstein distance and/or the optimal couplings for multivariate measures are rare. However, the involved probability measures can be approximated by taking consistent empirical versions of the input measures. See, for instance, [39,40].

In the following, we will mainly consider doubly stochastic measures [48], i.e., probability measures on $[0, 1]^2$ whose margins coincide with the Lebesgue measure on $[0, 1]$. These measures are better known as *copulas*. The space \mathcal{C} of a bivariate copula can be ordered with respect to pointwise order among functions. It contains the Fréchet–Hoeffding upper-bound copula $M(u, v) = \min(u, v)$, which interprets the maximal degree of similarity (comonotonicity) between two random variables. Moreover, the Fréchet–Hoeffding lower bound is given by the copula $W(u, v) = \max(u + v - 1, 0)$, which interprets the counter-monotonic behavior of two variables. Another notable copula is $\Pi(u, v) = uv$, which represents the independence of random variables.

The set \mathcal{C} , considered as a compact subset of $\mathcal{W}_2([0, 1]^2)$ with respect to weak convergence, will be equipped with the metric $d_{\mathcal{W}_2}$. In general, the use of the Wasserstein metric for capturing dependence aspects has been recently employed in [38,49,50], among others. Typically, copula-based dependence measurements are expressed as a discrepancy of the estimated copula from the independent case. Here, instead, we will consider the Wasserstein distance between a copula and the comonotonicity copula M . This perspective has been also recently considered in [36] in a Bayesian setting.

Copulas have been recently adopted in the construction of various dissimilarity measures that are used as input to clustering algorithms (see, for instance, [51] and references therein). The use of a dissimilarity based on the Wasserstein distance between copulas appeared, for instance, in [37] in order to measure the distances among different Gaussian copulas for clustering purposes (see also [52]). Here, we extend the range of applications to the case of time series with geo-referenced information.

3. The Methodology

Our aim is to cluster $n \geq 2$ time series, each of them equipped with a (static) p -dimensional vector of attributes associated with each unit, which are typically geographic coordinates of the location at which the time series are observed.

Given the input time series and the (static) attributes, and a fixed weighting parameter $\alpha \in [0, 1]$, we aim at finding an $(n \times n)$ dissimilarity matrix $\Delta^{(\alpha)}$ that merges all the previous information. The (i, j) -entry $\Delta_{ij}^{(\alpha)}$ of $\Delta^{(\alpha)}$ interprets the temporal dependence between units i and j by taking into account the similarity between \mathbf{s}_i and \mathbf{s}_j . The relative importance of each of the two criteria is summarized by the parameter α , i.e., $\alpha = 0$ corresponds to no influence of the spatial component.

For every pair (i, j) with $i \neq j$, $\Delta_{ij}^{(\alpha)}$ is constructed in the following way:

1. Determine the copula C_{ij}^{ts} that describes the temporal dependence between the i -th and j -th time series.
2. Determine the copula C_{ij}^{sp} that interprets the spatial proximity between the attribute vectors \mathbf{s}_i and \mathbf{s}_j associated with the i -th and j -th time series.
3. Merge C_{ij}^{ts} and C_{ij}^{sp} into one single copula $C_{ij}^{(\alpha)}$ that represents their weighted barycenter. This copula depends on the tuning parameter α . Then, define

$$\Delta_{ij}^{(\alpha)} = d_{\mathcal{W}_2}(M, C_{ij}^{(\alpha)}),$$

i.e., the distance of $C_{ij}^{(\alpha)}$ from the comonotonicity copula M , which represents maximal concordance.

Once the dissimilarity matrix is obtained, it can be used as an input of various algorithms like hierarchical agglomerative methods, medoids-based procedures, etc.

The previously described procedure is illustrated in detail below.

3.1. Extract the Temporal Dependence

In order to capture the dependence among the time series, it is usual in the context of copula-based clustering (see, for instance, [4]) to proceed in two steps. First, we filter serial dependence using univariate time-series models, like the ARMA-GARCH family, and hence, we model the cross-sectional dependence using a copula for the residuals (see, for instance, [53–56]).

To fix ideas, we assume that the multivariate time series follows the model $\mathbf{X}_t = (X_{1t}, \dots, X_{nt})$, where

$$\mathbf{X}_t = \boldsymbol{\mu}_t(\theta) + \boldsymbol{\sigma}_t(\theta)\boldsymbol{\varepsilon}_t, \quad (t = 1, \dots, T) \tag{7}$$

where the innovations $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})$ are independent and identically distributed, with $\mathbb{E}(\varepsilon_{it}) = 0$ and $\mathbb{V}(\varepsilon_{it}) = 1$ for $i = 1, \dots, n$, with the continuous joint distribution function H . Moreover, $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are the (time-varying) conditional mean and standard deviation, respectively, and they are both \mathcal{F}_{t-1} -measurable and independent of $\boldsymbol{\varepsilon}_t$. Here, \mathcal{F}_{t-1} contains information from the past and possibly information from exogenous variables as well (see, e.g., [55]). Since the distribution function H is continuous, there exists a unique copula C so that for all $\mathbf{x} \in \mathbb{R}^n$,

$$H(\mathbf{x}) = C(F_1(x_1), \dots, F_n(x_n)),$$

where F_i is the distribution function of ε_{it} for every t . Defining $\mathbf{U}_t = (F_1(\varepsilon_{1t}), \dots, F_n(\varepsilon_{nt}))$, one obtains that $\mathbf{U}_1, \dots, \mathbf{U}_T$ are independent and identically distributed with distribution function C . Since the marginal distributions are unknown, \mathbf{U}_t is not observable. However, given an estimator $\hat{\theta}$ of θ , we can compute the residuals

$$\mathbf{e}_t = \frac{\mathbf{x}_t - \boldsymbol{\mu}_t(\hat{\theta})}{\boldsymbol{\sigma}(\hat{\theta})} \tag{8}$$

for $t = 1, \dots, n$. The ranks associated with the residuals contain the information about the copula among the time series for any fixed t and, as such, they can be used to capture the cross-sectional dependence (see, e.g., [56]). Specifically, for any $t = 1, \dots, T$, let r_{it} be the rank of e_{it} among the residuals e_{i1}, \dots, e_{iT} of the i -th time series. The multivariate scaled ranks

$$\mathbf{u}_t = \left(\frac{r_{1t}}{T+1}, \dots, \frac{r_{nt}}{T+1} \right) \tag{9}$$

are the so-called *pseudo-observations*.

Now, for every pair (i, j) , $i \neq j$, the copula C_{ij}^{ts} that describes the cross-sectional dependence between the i -th and j -th time series can be obtained from the associated pseudo-observations $(u_{it}, u_{jt})_{t=1, \dots, T}$. Here, one can adopt:

- a parametric approach, i.e., one assumes that C_{ij}^{ts} belongs to the same specific family of copulas, whose parameter can be fitted via, e.g., maximum likelihood techniques. See, e.g., [57].
- a non-parametric approach, which assumes that C_{ij}^{ts} coincides with one smoothed version of the empirical copula associated with the pseudo-observations, such as the empirical checkerboard copula [58,59] or the empirical beta copula [60].

In the following, if not otherwise stated, we assume that every C_{ij}^{ts} is absolutely continuous (with respect to the Lebesgue measure).

3.2. Extract the Spatial Dependence

In order to describe the geographical proximity, we preliminarily select a specific family of copulas that can be parameterized as $(C_\theta)_\Theta$, with $\Theta = [\theta_{\min}, \theta_{\max}] \subset \mathbb{R}$. In particular, we assume that $M \in \{C_{\theta_{\min}}, C_{\theta_{\max}}\}$, i.e., the family includes the Fréchet–Hoeffding upper bound as a limiting case. Moreover, the following technical condition is assumed

$$\theta \mapsto d_{\mathcal{W}_2}(M, C_\theta) \quad \text{is continuous and strictly monotone in } \theta. \tag{10}$$

The continuity should be interpreted in the sense of topology of uniform convergence, which is equivalent to pointwise convergence (see, e.g., [48]). Roughly speaking, condition (10) implies a one-to-one correspondence between the copula parameter and the Wasserstein distance between M and any member of the given parametric family.

Now, for every pair (i, j) , $i \neq j$, we assume that the copula C_{ij}^{SP} that describes the proximity among the attributes \mathbf{s}_i and \mathbf{s}_j belongs to $(C_\theta)_\Theta$, i.e., $C_{ij}^{\text{SP}} = C_{\theta_{ij}}$. Moreover, we assume that θ_{ij} only depend on the (normalized) distance d_{ij} between \mathbf{s}_i and \mathbf{s}_j .

Furthermore, we want to ensure that $C_{\theta_{ij}}$ approaches M , i.e., it is close to the comonotonic case, when the normalized geographic distance is close to zero. To this end, we propose to select, for every pair (i, j) , θ_{ij} as the unique value satisfying the equality:

$$\frac{d_{\mathcal{W}_2}(M, C_\theta)}{\max_{\theta' \in \Theta} d_{\mathcal{W}_2}(M, C_{\theta'})} = \frac{\text{dist}(\mathbf{s}_i, \mathbf{s}_j)}{\max_{i', j'} \text{dist}(\mathbf{s}_{i'}, \mathbf{s}_{j'})} (= d_{ij}). \tag{11}$$

The existence and uniqueness of such a parameter is guaranteed by (10).

Various copula families satisfy the conditions stated above.

Example 1. Consider the parametric family of copulas of type

$$C_\theta^{\text{Fre}}(u, v) = (1 - \theta)M(u, v) + \theta W(u, v) \tag{12}$$

where $\theta \in [0, 1]$, M and W are the Fréchet–Hoeffding upper- and lower-bound copulas, respectively. Copulas of type (12) belong to the so-called Fréchet class, which was suggested as a possible model for spatial dependence in [11]. For such copulas, we have (see Appendix A)

$$d_{\mathcal{W}_2}(M, C_\theta^{\text{Fre}}) = \sqrt{\frac{\theta}{3}}.$$

Moreover, the family is continuous in θ with respect to uniform convergence.

Now, according to Equation (11), if such a family is used to model spatial dependence, then the parameter is chosen so that

$$\frac{\sqrt{\theta_{ij}/3}}{\sqrt{1/3}} = \frac{\text{dist}(\mathbf{s}_i, \mathbf{s}_j)}{\max_{i', j'} \text{dist}(\mathbf{s}_{i'}, \mathbf{s}_{j'})} = d_{ij},$$

which gives $\theta_{ij} = d_{ij}^2$.

3.3. Create the Dissimilarity Measure

In order to merge the temporal and spatial information, for a fixed $\alpha \in [0, 1]$, we associate to each pair of units (i, j) , $i \neq j$, the copula $C_{ij}^{(\alpha)}$ associated with the displacement interpolation of (2). Namely, $C_{ij}^{(\alpha)}$ is the copula associated with the probability measures $\mu^{(\alpha)}$ that solves the minimization problem

$$\mu^{(\alpha)} = \arg \min_{\mu \in \mathcal{W}_2([0,1]^2)} \left((1 - \alpha)d_{\mathcal{W}_2}^2(\mu_{C_{ij}^{\text{ts}}}, \mu) + \alpha d_{\mathcal{W}_2}^2(\mu_{C_{ij}^{\text{sp}}}, \mu) \right), \tag{13}$$

where $\mu_{C_{ij}^{\text{ts}}}$ and $\mu_{C_{ij}^{\text{sp}}}$ are the measures associated with the copulas C_{ij}^{ts} and C_{ij}^{sp} , respectively. The existence of such a copula is guaranteed under the assumption that C_{ij}^{ts} is absolutely continuous. Moreover, in such a case, $\mu^{(\alpha)}$ is also absolutely continuous and it admits a unique copula in view of Sklar’s theorem [48].

Remark 1. Generally, given two copulas C_1 and C_2 , the copula of their weighted barycenter, given by (2), does not coincide with the convex combination of the two copulas. This fact follows from, e.g., [34], and it is illustrated in Figure 1.

Finally, according to a general way of calculating dissimilarity as a distance from the comonotonic case [51], we define the dissimilarity between the time series i and j as

$$\Delta_{ij}^{(\alpha)} = d_{\mathcal{W}_2}(M, C_{ij}^{(\alpha)}). \tag{14}$$

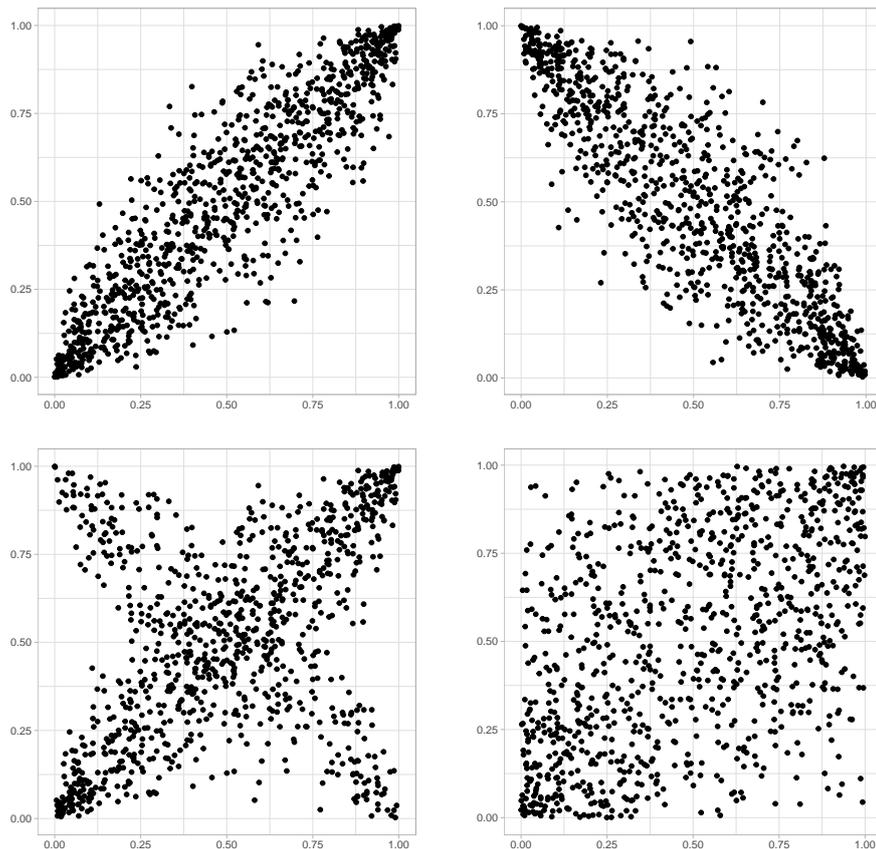


Figure 1. Random sample from a Gaussian copula C_1 with parameter 0.9 (upper left), a Gaussian copula C_2 with parameter -0.9 (upper right), $(C_1 + C_2)/2$ (lower left), and the (equally weighted) copula associated with the barycenter of C_1 and C_2 (lower right).

4. The Quasi-Gaussian Approach

Although theoretically appealing, the actual computation related to the dissimilarity in (14) is involved. In fact, generally, $\mu^{(\alpha)}$ (and, hence, $C^{(\alpha)}$) cannot be calculated in a closed form. Therefore, following seminal ideas in [38], we use a quasi-Gaussian approach based on correlation matrices in order to define a modified dissimilarity. The main idea is to replace the copula space with the space of all bivariate Gaussian distributions with standard marginals. Such distributions are called G -copulas in [38] and their space will be denoted by \mathcal{G} . In fact, in this latter case, the calculations can be obtained only in terms of the corresponding correlation matrices, as recalled in Section 2.

We recall that each G -copula is characterized by a Gaussian vector with mean $(0, 0)$ and correlation matrix

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}.$$

Then, for every pair (i, j) , $i \neq j$, we proceed as follows.

1. The copula C_{ij}^{ts} from Section 3.1 is replaced with the G-copula with correlation matrix $\Sigma(\hat{\theta}_{ij}^{ts})$, where $\hat{\theta}_{ij}^{ts}$ equals the estimation of the normal score correlation among the involved observations (as suggested in [38]).
2. The copula C_{ij}^{sp} from Section 3.2 is replaced with the G-copula with correlation matrix $\Sigma(\hat{\theta}_{ij}^{sp})$, where $\hat{\theta}_{ij}^{sp}$ is the unique value θ that solves

$$\frac{d_{W_2}(\Sigma(1), \Sigma(\theta))}{\max_{\theta' \in [0,1]} d_{W_2}(\Sigma(1), \Sigma(\theta'))} = d_{ij},$$

where d_{ij} is the normalized distance between s_i and s_j . Notice that $\Sigma(1)$ is the (singular) correlation matrix showing maximal dependence (comonotonic case). For the existence and uniqueness of $\hat{\theta}_{ij}^{sp}$ see Remark 2. In particular, we notice that since $\theta' \in [0, 1]$, the maximal spatial distance is interpreted by the vector of independent components.

3. For a fixed $\alpha \in]0, 1[$, the copula $C_{ij}^{(\alpha)}$ from Section 3.3 is replaced with the G-copula having the correlation matrix

$$\Sigma^{(\alpha)} = D^{-1/2} \Sigma' D^{-1/2}, \tag{15}$$

where $\Sigma' = \Sigma(\hat{\theta}_{ij}^{ts}) \#_{\alpha} \Sigma(\hat{\theta}_{ij}^{sp})$ is the weighted barycenter of $\Sigma(\hat{\theta}_{ij}^{ts})$ and $\Sigma(\hat{\theta}_{ij}^{sp})$ given by (6) and D is the diagonal matrix associated with Σ' . In other words, we transform the covariance matrix of the weighted barycenter into a correlation matrix with the natural projection.

Summarizing, in the quasi-Gaussian approach, we define the modified dissimilarity between the time series i and j as

$$\tilde{\Delta}_{ij}^{(\alpha)} = d_{W_2}(\Sigma(1), \Sigma^{(\alpha)}), \tag{16}$$

i.e., in terms of the Wasserstein distance between two correlation matrices associated with elements of \mathcal{G} . This is analogous to Equation (14) since both matrices $\Sigma(1)$ and M represent the comonotonic case, while $C_{ij}^{(\alpha)}$ and $\Sigma^{(\alpha)}$ correspond to the weighted barycenters in their respective spaces.

Remark 2. In the bivariate case, the Wasserstein distance among two Gaussian distributions is obtained via Formula (5). In particular, if we consider the singular correlation matrix $\Sigma(1)$, we obtain that

$$d_{W_2}^2(\Sigma(1), \Sigma(\theta)) = 4 - 2\sqrt{2 + 2\theta}.$$

The plot of $\theta \mapsto d_{W_2}(\Sigma(1), \Sigma(\theta))$ is visualized in Figure 2.

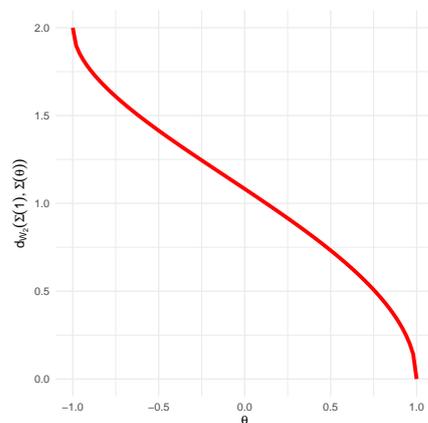


Figure 2. Graph of $\theta \mapsto d_{W_2}(\Sigma(1), \Sigma(\theta))$ where $\theta \in [-1, 1]$.

5. An Illustration with a Fuzzy-PAM Algorithm

Given an $(n \times n)$ dissimilarity matrix among time series, various algorithms can be exploited to provide a suitable clustering. Here, we will apply a (fuzzy) partitioning-around-medoid (PAM) clustering method for time series whose output expresses the membership degree of each time series to a cluster (see [11]). As known, the main advantage of PAM is that the prototypes (i.e., medoids) of each cluster are time series actually observed and not average time series, which is often very appealing for the interpretation of the selected clusters (see, e.g., [61]). For a fixed number K of clusters, the fuzzy PAM algorithm can be formalized as follows:

$$\begin{aligned} \min_{u_{ij}} \quad & \sum_{i=1}^n \sum_{k=1}^K u_{ik}^p \Delta_{ik} \\ \text{s.t.} \quad & \sum_{k=1}^K u_{ik} = 1, \quad u_{ik} \geq 0, \end{aligned}$$

where u_{ik} indicates the membership degree of the i -th unit to the k -th cluster ($k = 1, \dots, K$); and $p > 1$ is a weighting exponent that controls the fuzziness of the obtained partition (hereinafter, $p = 1.5$). Here, Δ_{ik} is a suitable dissimilarity between the time series of the i -th unit and the time series of the k -th medoid.

To illustrate the proposed algorithm, we consider the fuzzy PAM algorithm applied to the dissimilarity given in (16). We consider a scenario similar to the one in Section 3.1 of [11]. Specifically, we consider $n = 48$ time series of innovations of length $T = 100$. We assume that for each $i = 1, \dots, 48$, the i -th time series has been observed with a vector of attributes \mathbf{s}_i so that $\text{dist}(\mathbf{s}_i, \mathbf{s}_j) = \text{dist}(\mathbf{s}_j, \mathbf{s}_i) = 1$ if $1 \leq i \leq 12$ and $13 \leq j \leq 48$; otherwise, $\text{dist}(\mathbf{s}_i, \mathbf{s}_j) = 0$. Roughly speaking, the time series can be either contiguous (distance equal to 0) or not (distance equal to 1).

Moreover, the time series have a temporal dependence given by the following copula model:

$$C(u_1, \dots, u_{48}) = C_1(u_1, \dots, u_{24}) \cdot C_2(u_{25}, \dots, u_{48}), \quad (17)$$

where C_1 and C_2 are copulas belonging to the Frank family (see, e.g., [48]), with a pairwise Kendall's $\tau \in \{0.25, 0.50, 0.75\}$.

Figure 3 reports the membership degree of unit 1 to the same cluster of unit 13 at different levels of α (which is the weight assigned to the spatial component). Clearly, when $\alpha = 0$ (i.e., no spatial information) units 1 and 13 tend to belong to the same cluster, since they are positively associated via the copula C_1 . However, when α increases, the spatial component plays a major role and, roughly speaking, it moves unit 13 far from unit 1, i.e., into a different cluster.

It is important to highlight the effects of different choices of the spatial component. When we assume that the maximal spatial distance between two units corresponds to a zero correlation in the Gaussian model (Figure 3, on the left), the membership degree of unit 1 to its natural temporal cluster tends to decline sharply only for α close to 0.5. Instead, the decline comes earlier when the maximal spatial distance between two units corresponds to a correlation equal to -1 in the Gaussian model (Figure 3, on the right). We notice that this latter choice was also adopted in [11] when the Fréchet family of copulas was used.

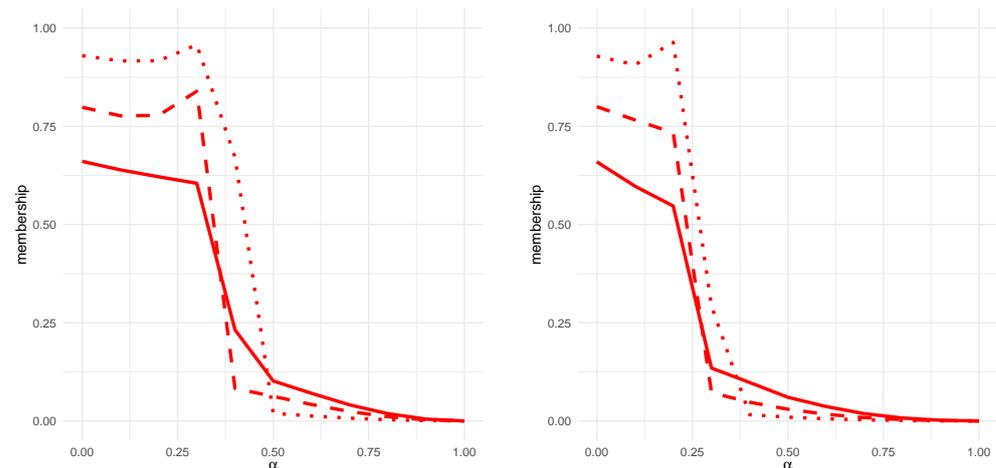


Figure 3. Membership degree of unit 1 to the same cluster of unit 13 for different values of the dissimilarity matrix of (16). Solid line: $\tau = 0.25$; dashed line: $\tau = 0.50$; dotted line: $\tau = 0.75$. (Left) Maximal spatial correlation equal to 0. (Right) Maximal spatial correlation equal to -1 . The results are mean values over $R = 25$ replications from model (17).

6. An Empirical Application

In this section, we apply the fuzzy-PAM clustering algorithm on time series representing the summer temperature maxima of Italy from 1971 to 2023. The data have been downloaded from the Climate Data Store (<https://cds.climate.copernicus.eu/>, accessed on 1 September 2023) which collects global climate and weather data of the past eight decades. Specifically, we focus on JJA (June–July–August) maxima of daily maximum temperatures over a grid of the Italian land. The considered data are $n = 527$ grid points. Similarly to [62], we detrend the observed temperatures from the long-term warming trend, following a two-step procedure. First, we remove the multi-year climatological average from the daily temperature maxima within the dataset. Then, from these temperature residuals, we remove the 92-day running average. After this detrending process, we consider the maximum value of each season for each grid point, leading to a collection of time series of length $T = 53$. Given the set of detrended time series and the geographical locations, we compute the temporal and the spatial dependencies, as described in Section 4. Specifically, for every pair of $(i, j), i \neq j$, we use the estimation of the normal score correlation to compute the temporal correlation matrix $\Sigma(\hat{\theta}_{ij}^t)$ as in step 1, and we use the normalized distances d_{ij} to compute the spatial correlation matrix $\Sigma(\hat{\theta}_{ij}^{sp})$ according to step 2. Initially, we consider the fuzzy-PAM algorithm applied to the dissimilarity given in (16) for $\alpha = 0$ representing the pure temporal case. We select the number of clusters via the fuzzy silhouette (FS) index [63], a suitable measure for the fuzzy clustering algorithm that is computed as the weighted average of the individual silhouettes. The better the units are assigned to the clusters, at the same time as minimizing intra-cluster distance and maximizing inter-cluster distance, the higher the value of FS. We represent the obtained clusters in the pure temporal case ($\alpha = 0$) in Figure 4. From the maximization of the FS we have that the optimal number of clusters is $K = 3$.

In the pure temporal case, time series belonging to the same cluster may be not contiguous. This aspect can be reduced by combining the two dissimilarities via Equation (16) for some $\alpha \in [0, 1]$. As an illustration, we show the cluster composition when $\alpha = 0.30$ (clearly, other values are possible, although a range of 0.05–0.30 seems generally reasonable, as suggested in [30]).

In order to choose the optimal number of clusters for $\alpha = 0.30$, we proceed with the fuzzy silhouette index maximization. Table 1 shows the evolution of the fuzzy silhouette index for different values of $K \in \{2, \dots, 10\}$. Thus, $K = 3$ is selected.

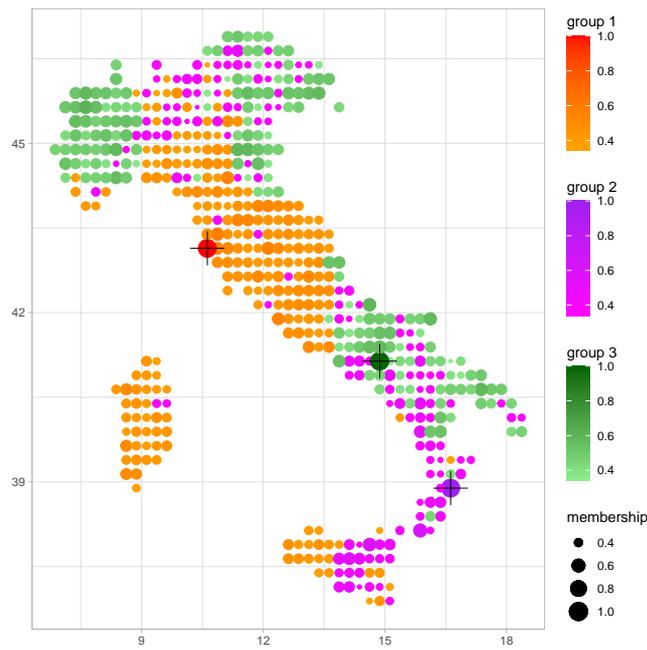


Figure 4. Membership representation of the pure temporal clustering. Darker colors and bigger points represent a higher membership degree. The crossed points are the medoids of each group.

Table 1. Fuzzy silhouette index for $\alpha = 0.30$ and $K \in \{2, \dots, 10\}$.

$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
0.176	0.208	0.194	0.163	0.120	0.154	0.118	0.105	0.102

The corresponding clustering configuration is shown in Figure 5.

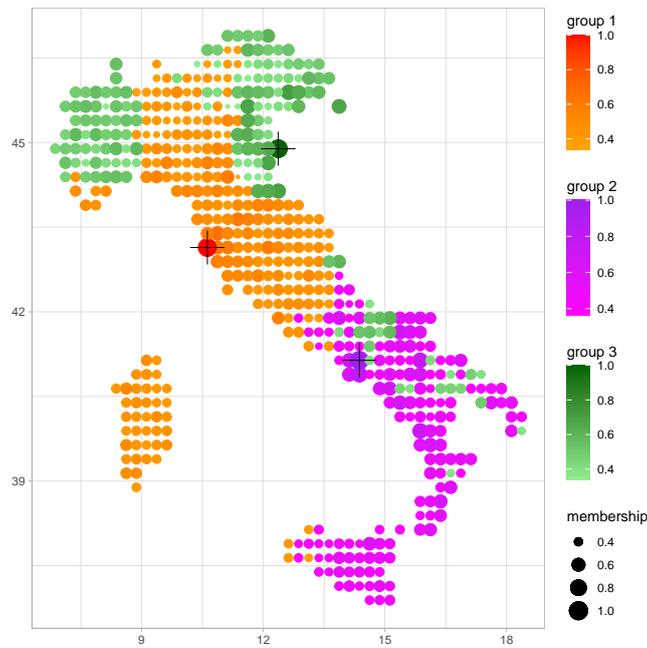


Figure 5. Membership representation of the clustering for $\alpha = 0.30$. Darker colors and bigger points represent a higher membership degree. The crossed points are the medoids of each group.

We note that when α increases, the cluster composition increases the spatial contiguity among its elements. Clearly, however, some time series still exhibit a behavior that is not driven by geographic proximity. Such time series are particularly of interest in order to detect spatial anomalies in some geographic area, and their behavior should be the object of in-depth investigations.

7. Conclusions

In this paper, we proposed a way to define a suitable dissimilarity matrix interpreting the comovements of time series subjected to spatial constraints. The ultimate goal was the clustering of time series with geo-referenced information. The proposed extraction of the dissimilarity matrix leverages on the 2-Wasserstein distance between probability measures and on optimal transport theory. Moreover, as the computational aspects of such an extraction are involved in practical applications, we proposed a quasi-Gaussian approximation of the correlation matrices in order to define a modified dissimilarity and reduce the complexity and the computational burden of the procedure. In the second part of the paper, we presented an illustration with a fuzzy-PAM algorithm to show the effect of the spatial dependence on the whole procedure. Lastly, we proposed a case study showing the effects of the proposed method on real data involving climatic time series.

Author Contributions: All authors contributed equally to this work; A.B. mainly focused on validation and investigation; F.D. mainly focused on investigation and visualization. All authors have read and agreed to the published version of the manuscript.

Funding: A.B. acknowledges the support of Regione Puglia via the Programma Regionale “RIPARTI (assegni di Ricerca per riPARTire con le Imprese)”-research project “FIRST: a Framework for Innovation in Risk management to support Territories” (code: c19a5daa). F.D. has been supported by MIUR-PRIN 2017, Project “Stochastic Models for Complex Systems” (No. 2017JFFHSH) and by MIUR-PRIN 2022, Project “Statistical Mechanics of Learning Machines: from algorithmic and information theoretical limits to new biologically inspired paradigms” (No. 20229T9EAT). The work of F. D. has been carried out with the partial financial support from ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union NextGenerationEU (CUP F83C22000740001). F.D. is also member of the group GNAMPA of INdAM (Istituto Nazionale di Alta Matematica).

Data Availability Statement: The datasets that were used in this study are available online on the following link, <https://cds.climate.copernicus.eu/>, accessed on 1 September 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Here we compute the Wasserstein distance $d_{W_2}(M, (1 - \theta)M + \theta W)$. Due to the probabilistic interpretation of the copulas M , W and of their convex combination (see, e.g., [48]), this corresponds to computing

$$\inf \left\{ \mathbb{E} \left((U, U) - (V, 2ZV - V - Z + 1) \right)^2 \right\}^{1/2}$$

where the infimum is taken over all possible joint probability distribution of the involved random variables. Here, U and V are random variables uniformly distributed on $[0, 1]$, while Z is a Bernoulli variable, so that $Z = 1$ with probability θ , and $Z = 0$ with probability $1 - \theta$. Thus, $(U, U) \sim M$ and $(V, 2ZV - V - Z + 1) \sim C_\theta^{\text{Fre}}$. We start by computing the expression to be minimized.

$$\begin{aligned}
& \mathbb{E}|(U, U) - (V, 2ZV - V - Z + 1)|^2 \\
&= \mathbb{E}((U - V)^2 + (U - (2ZV - V - Z + 1))^2) \\
&= \mathbb{E}(U^2 + V^2 - 2UV + U^2 + 4Z^2V^2 + V^2 + Z^2 + 1 \\
&\quad - 4ZUV + 2UV + 2UZ - 2U - 4ZV^2 - 4Z^2V + 4ZV + 2VZ - 2V - 2Z) \\
&= \mathbb{E}(2U^2 + 2V^2 + Z^2 + 1 + 6VZ + 4Z^2V^2 - 4ZUV + 2UZ \\
&\quad - 2U - 2V - 2Z - 4ZV^2 - 4Z^2V) \\
&= \mathbb{E}(4U^2 + Z^2 + 1 + 6\theta V + 4Z^2V^2 - 4\theta UV + 2\theta U - 1 - 1 - 2\theta - 4\theta V^2 - 4Z^2V) \\
&= 4\mathbb{E}(U^2) + \mathbb{E}(Z^2) - 1 + 2\theta + 4\mathbb{E}(Z^2)\mathbb{E}(V^2) - 4\theta\mathbb{E}(UV) - 4\theta\mathbb{E}(V^2) - 2\mathbb{E}(Z^2) \\
&= \frac{4}{3} + \theta - 1 + 2\theta + \frac{4}{3}\theta - 4\theta\mathbb{E}(UV) - \frac{4}{3}\theta - 2\theta \\
&= \frac{1}{3} + \theta - 4\theta\mathbb{E}(UV) \\
&= \frac{1}{3} + \theta - 4\theta \cdot \frac{\rho + 3}{12} \\
&= \frac{1}{3} + \theta - \frac{\theta}{3}(\rho + 3)
\end{aligned}$$

The minimum value of $\frac{1}{3} + \theta - \frac{\theta}{3}(\rho + 3)$ is reached for $\rho = 1$ and corresponds to $\frac{1}{3} - \frac{1}{3}\theta$. For the previous chain of equalities we used the following facts:

- $\mathbb{E}(Z) = \theta, \mathbb{E}(Z^2) = \theta$;
- $\mathbb{V}(U) = \frac{1}{12}, \mathbb{E}(U^2) = \mathbb{E}(V^2) = \frac{1}{3}$;
- $\mathbb{E}(UV) = \frac{\rho+3}{12}$, where ρ is the Pearson's correlation between U and V .

References

1. Hennig, C.; Meila, M.; Murtagh, F.; Rocci, R. *Handbook of Cluster Analysis*; Chapman and Hall/CRC Handbook of Modern Statistical Methods; CRC Press: Boca Raton, FL, USA, 2016.
2. Maharaj, E.A.; D'Urso, P.; Caiado, J. *Time Series Clustering and Classification*; Chapman Hall/CRC Computing and Data Analytics Series; CRC Press: Boca Raton, FL, USA, 2019.
3. Caiado, J.; Maharaj, E.A.; D'Urso, P. Time-series clustering. In *Handbook of Cluster Analysis*; Hennig, C., Meila, M., Murtagh, F., Rocci, R., Eds.; CRC Press: Boca Raton, FL, USA, 2016; pp. 241–263.
4. Di Lascio, F.; Durante, F.; Pappadà, R. Copula-based clustering methods. In *Copulas and Dependence Models with Applications*; Úbeda Flores, M., de Amo, E., Durante, F., Fernández Sánchez, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 49–67.
5. Marti, G.; Nielsen, F.; Bińkowski, M.; Donnat, P. A Review of Two Decades of Correlations, Hierarchies, Networks and Clustering in Financial Markets. In *Progress in Information Geometry: Theory and Applications*; Nielsen, F., Ed.; Springer: Cham, Switzerland, 2021; pp. 245–274.
6. De Luca, G.; Zuccolotto, P. A tail dependence-based dissimilarity measure for financial time series clustering. *Adv. Data Anal. Classif.* **2011**, *5*, 323–340. [\[CrossRef\]](#)
7. De Luca, G.; Zuccolotto, P. Hierarchical time series clustering on tail dependence with linkage based on a multivariate copula approach. *Internat. J. Approx. Reason.* **2021**, *139*, 88–103. [\[CrossRef\]](#)
8. Durante, F.; Pappadà, R.; Torelli, N. Clustering of time series via non-parametric tail dependence estimation. *Statist. Pap.* **2015**, *56*, 701–721. [\[CrossRef\]](#)
9. Bonanomi, A.; Nai Ruscone, M.; Osmetti, S.A. Dissimilarity measure for ranking data via mixture of copulae. *Stat. Anal. Data Min. Asa Data Sci. J.* **2019**, *12*, 412–425. [\[CrossRef\]](#)
10. De Keyser, S.; Gijbels, I. Hierarchical variable clustering via copula-based divergence measures between random vectors. *Int. J. Approx. Reason.* **2023**, *21*, 109090. [\[CrossRef\]](#)
11. Disegna, M.; D'Urso, P.; Durante, F. Copula-based fuzzy clustering of spatial time series. *Spat. Stat.* **2017**, *21*, 209–225. [\[CrossRef\]](#)
12. Kojadinovic, I. Agglomerative hierarchical clustering of continuous variables based on mutual information. *Comput. Stat. Data Anal.* **2004**, *46*, 269–294. [\[CrossRef\]](#)
13. Zhang, B.; An, B. Clustering time series based on dependence structure. *PLoS ONE* **2018**, *13*, e0206753. [\[CrossRef\]](#)
14. De Luca, G.; Zuccolotto, P. A double clustering algorithm for financial time series based on extreme events. *Stat. Risk Model.* **2017**, *34*, 1–12. [\[CrossRef\]](#)

15. Durante, F.; Pappadà, R.; Torelli, N. Clustering of financial time series in risky scenarios. *Adv. Data Anal. Classif.* **2014**, *8*, 359–376. [[CrossRef](#)]
16. Pappadà, R.; Durante, F.; Salvadori, G.; De Michele, C. Clustering of concurrent flood risks via Hazard Scenarios. *Spat. Stat.* **2018**, *23*, 124–142. [[CrossRef](#)]
17. Saunders, K.R.; Stephenson, A.G.; Karoly, D.J. A regionalisation approach for rainfall based on extremal dependence. *Extremes* **2021**, *24*, 1386–1999. [[CrossRef](#)]
18. Fouedjio, F. Clustering of multivariate geostatistical data. *WIREs Comput Stat.* **2020**, *12*, e1510. [[CrossRef](#)]
19. Kopczewska, K. Spatial machine learning: New opportunities for regional science. *Ann. Reg. Sci.* **2022**, *68*, 713–755. [[CrossRef](#)]
20. Asgharian, H.; Hess, W.; Liu, L. A spatial analysis of international stock market linkages. *J. Bank. Financ.* **2013**, *37*, 4738–4754. [[CrossRef](#)]
21. Fernández-Avilés, G.; Montero, J.M.; Orlov, A.G. Spatial modeling of stock market comovements. *Fin. Res. Lett.* **2012**, *9*, 202–212. [[CrossRef](#)]
22. Hüttner, A.; Scherer, M.; Gräler, B. Geostatistical modeling of dependent credit spreads: Estimation of large covariance matrices and imputation of missing data. *J. Bank. Financ.* **2020**, *118*, 105897. [[CrossRef](#)]
23. Oliver, M.A.; Webster, R. A geostatistical basis for spatial weighting in multivariate classification. *Math. Geol.* **1989**, *21*, 15–35. [[CrossRef](#)]
24. Coppi, R.; D’Urso, P.; Giordani, P. A fuzzy clustering model for multivariate spatial time series. *J. Class.* **2010**, *27*, 54–88. [[CrossRef](#)]
25. Fouedjio, F. A hierarchical clustering method for multivariate geostatistical data. *Spat. Stat.* **2016**, *18*, 333–351. [[CrossRef](#)]
26. D’Urso, P.; Vitale, V. A robust hierarchical clustering for georeferenced data. *Spat. Stat.* **2020**, *35*, 100407. [[CrossRef](#)]
27. Di Lascio, F.; Menapace, A.; Pappadà, R. A spatially-weighted AMH copula-based dissimilarity measure for clustering variables: An application to urban thermal efficiency. *Environmetrics* **2023**, e2828. [[CrossRef](#)]
28. Zuccolotto, P.; De Luca, G.; Metulini, R.; Carpita, M. Modeling and clustering of traffic flows time series in a flood prone area. In *Proceedings of the Statistics and Data Science Conference*; Cerchiello, P., Agosto, A., Osmetti, S., Spelta, A., Eds.; EGEA: Pavia, Italy, 2023; pp. 113–118.
29. Benevento, A.; Durante, F. Correlation-based hierarchical clustering of time series with spatial constraints. *Spat. Stat.* **2024**, *59*, 100797. [[CrossRef](#)]
30. Romary, T.; Ors, F.; Rivoirard, J.; Deraisme, J. Unsupervised classification of multivariate geostatistical data: Two algorithms. *Comput. Geosci.* **2015**, *85*, 96–103. [[CrossRef](#)]
31. Benevento, A.; Durante, F.; Pappadà, R. An approach to cluster time series extremes with spatial constraints. In *Proceedings of the Book of the Short Papers SIS 2023*; Chelli, F., Ciommi, M., Ingrassia, S., Mariani, F., Recchioni, M., Eds.; Pearson: London, UK, 2023; pp. 679–684.
32. Villani, C. *Optimal Transport Old and New*; Grundlehren der mathematischen Wissenschaften; Springer: Berlin, Germany, 2009; Volume 338.
33. Santambrogio, F. Optimal transport for applied mathematicians. *Birkhäuser* **2015**, *55*, 94.
34. McCann, R.J. A convexity principle for interacting gases. *Adv. Math.* **1997**, *128*, 153–179. [[CrossRef](#)]
35. Agueh, M.; Carlier, G. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **2011**, *43*, 904–924. [[CrossRef](#)]
36. Catalano, M.; Lijoi, A.; Prünster, I. Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *Ann. Stat.* **2021**, *49*, 2916–2947. [[CrossRef](#)]
37. Marti, G.; Andler, S.; Nielsen, F.; Donnat, P. Optimal transport vs. Fisher-Rao distance between copulas for clustering multivariate time series. In *Proceedings of the 2016 IEEE statistical signal processing workshop (SSP)*, Palma de Mallorca, Spain, 26–29 June 2016; pp. 1–5.
38. Mordant, G.; Segers, J. Measuring dependence between random vectors via optimal transport. *J. Multivar. Anal.* **2022**, *189*, 104912. [[CrossRef](#)]
39. Peyré, G.; Marti, M. Computational Optimal Transport: With Applications to Data Science. *Found. Trends® Mach. Learn.* **2019**, *11*, 355–607. [[CrossRef](#)]
40. Panaretos, V.M.; Zemel, Y. Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **2019**, *6*, 405–431. [[CrossRef](#)]
41. Cuturi, M.; Doucet, A. Fast computation of Wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning*; Xing, E., Jebara, T., Eds.; Proceedings of Machine Learning Research: Beijing, China, 2014; Volume 32, pp. 685–693.
42. Puccetti, G.; Rüschendorf, L.; Vanduffel, S. On the computation of Wasserstein barycenters. *J. Multivar. Anal.* **2020**, *176*, 16. [[CrossRef](#)]
43. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
44. Givens, C.; Shortt, R.M. A class of Wasserstein metrics for probability distributions. *Mich. Math. J.* **1984**, *31*, 231–240. [[CrossRef](#)]
45. Dowson, D.C.; Landau, B.V. The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.* **1982**, *12*, 450–455. [[CrossRef](#)]
46. Knott, M.; Smith, C.S. On the optimal mapping of distributions. *J. Optim. Theory Appl.* **1984**, *43*, 39–49. [[CrossRef](#)]
47. Olkin, I.; Pukelsheim, F. The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.* **1982**, *48*, 257–263. [[CrossRef](#)]
48. Durante, F.; Sempi, C. *Principles of Copula Theory*; CRC Press: Boca Raton, FL, USA, 2016.

49. Nielsen, F.; Marti, G.; Ray, S.; Pyne, S. Clustering patterns connecting COVID-19 dynamics and human mobility using optimal transport. *Sankhyā Ser. B* **2021**, *83*, 167–184. [[CrossRef](#)]
50. Wiesel, J. Measuring association with Wasserstein distances. *Bernoulli* **2022**, *28*, 2816–2832. [[CrossRef](#)]
51. Fuchs, S.; Di Lascio, F.M.L.; Durante, F. Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. *Comput. Statist. Data Anal.* **2021**, *159*, 107201. [[CrossRef](#)]
52. Marti, G.; Andler, S.; Nielsen, F.; Donnat, P. Exploring and measuring non-linear correlations: Copulas, Lightspeed Transportation and Clustering. In *Proceedings of the NIPS 2016 Time Series Workshop*; PMLR: Beijing, China, 2017; pp. 59–69.
53. Chen, X.; Fan, Y. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *J. Econom.* **2006**, *135*, 125–154. [[CrossRef](#)]
54. Patton, A. A review of copula models for economic time series. *J. Multivar. Anal.* **2012**, *110*, 4–18. [[CrossRef](#)]
55. Rémillard, B. Goodness-of-Fit Tests for Copulas of Multivariate Time Series. *Econometrics* **2017**, *5*, 13. [[CrossRef](#)]
56. Nasri, B.R.; Rémillard, B.N. Copula-based dynamic models for multivariate time series. *J. Multivar. Anal.* **2019**, *172*, 107–121. [[CrossRef](#)]
57. Hofert, M.; Kojadinovic, I.; Mächler, M.; Yan, J. *Elements of Copula Modeling with R*; Springer: Cham, Switzerland, 2018.
58. Genest, C.; Nešlehová, J.G.; Rémillard, B. Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivar. Anal.* **2017**, *159*, 82–110. [[CrossRef](#)]
59. Pfeifer, D.; Mändle, A.; Ragulina, O.; Girschig, C. New copulas based on general partitions-of-unity. III: The continuous case. *Depend. Model.* **2019**, *7*, 181–201. [[CrossRef](#)]
60. Segers, J.; Sibuya, M.; Tsukahara, H. The empirical beta copula. *J. Multivar. Anal.* **2017**, *155*, 35–51. [[CrossRef](#)]
61. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
62. Bador, M.; Naveau, P.; Gilleland, E.; Castellà, M.; Arivelo, T. Spatial clustering of summer temperature maxima from the CNRM-CM5 climate model ensembles & E-OBS over Europe. *Weather Clim. Extrem.* **2015**, *9*, 17–24.
63. Campello, R.J.; Hruschka, E.R. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets Syst.* **2006**, *157*, 2858–2875. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.