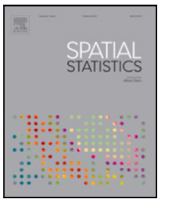
Spatial Statistics 59 (2024) 100797

Contents lists available at ScienceDirect

Spatial Statistics

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



Correlation-based hierarchical clustering of time series with spatial constraints

Alessia Benevento *, Fabrizio Durante

Dipartimento di Scienze dell'Economia, Università del Salento, Lecce, Italy

ARTICLE INFO

Keywords:
Copula
Correlation matrix
Clustering
Riemannian manifold
Time series

ABSTRACT

Correlation-based hierarchical clustering methods for time series typically are based on a suitable dissimilarity matrix derived from pairwise measures of association. Here, this dissimilarity is modified in order to take into account the presence of spatial constraints. This modification exploits the geometric structure of the space of correlation matrices, i.e. their Riemannian manifold. Specifically, the temporal correlation matrix (based on van der Waerden coefficient) is aggregated to the spatial correlation matrix (obtained from a suitable Matérn correlation function) via a geodesic in the Riemannian manifold. Our approach is presented and discussed using simulated and real data, highlighting its main advantages and computational aspects.

1. Introduction

The study of stochastic dependence among random variables is one of the first fundamental steps when dealing with a data analysis problem, especially in high dimensions. As is well known, starting with the Pearson's linear correlation coefficient, various alternatives have been provided that enable the detection of non-linear dependencies among variables. In particular, when these dependence measures are assumed to be invariant under monotone transformation of the involved variables, the concept of copula naturally comes into play (Joe, 2015; Nelsen, 2006).

When many variables are considered, a preliminary way to summarize the information about the (pairwise) association is provided by clustering algorithms (see, e.g., Maharaj et al. (2019)), that are able to identify clusters of variables such that within-cluster dependence is larger than between-cluster dependence (see Di Lascio et al. (2017), Fuchs et al. (2021) and references therein). By highlighting possibly nonlinear interdependencies among variables, such algorithms can be exploited as a first explanatory analysis, for example, in the construction of multivariate stochastic models (see, e.g., Czado et al. (2012), Dißmann et al. (2013), Côté and Genest (2015), Górecki et al. (2017, 2021), Palacios-Rodriguez et al. (2023)). Moreover, clustering procedures have been used to detect comovements of time series. In fact, starting with seminal works on correlation-based hierarchies of financial returns (Mantegna, 1999; Marti et al., 2021b), such procedures have been provided helpful information for managing financial risks, especially for portfolio diversification, when the tail association is of interest (see, e.g., De Luca and Zuccolotto (2011, 2021), Durante et al. (2015)).

According to Fouedjio (2020), when general-purpose clustering methods are used in the case of geo-referenced data, the resulting clusters can be spatially scattered over the geographic domain, even if spatial coordinates are considered as attributes. Such a situation may be undesirable for many applications.

In the literature, the spatial information is mainly considered in two ways. One approach groups individuals that are both similar and contiguous, often by applying a contiguity constraint to some other grouping strategy (see. e.g., Murtagh (1985)

* Corresponding author.
E-mail addresses: alessia.benevento@unisalento.it (A. Benevento), fabrizio.durante@unisalento.it (F. Durante).

https://doi.org/10.1016/j.spasta.2023.100797
Received 3 April 2023; Received in revised form 22 September 2023; Accepted 22 November 2023
Available online 30 November 2023

and also Guénard and Legendre (2022) for hierarchical algorithms). An alternative viewpoint, also known as soft contiguity or soft constraints, applies some form of spatial weighting function to the separating distances/dissimilarities. In this case, instead of enforcing the constraint strictly, a penalty term is often added to control the strength of the constraint. This allows for more flexibility in the clustering process and can result in more reasonable solutions when the constraints are difficult to satisfy exactly (see, e.g., Chavent et al. (2018), Distefano et al. (2020), D'Urso and Vitale (2020), D'Urso et al. (2023), Fouedjio (2016), Oliver and Webster (1989), Romary et al. (2015)). In a copula-based framework, this latter approach has been adopted for time series in Disegna et al. (2017) (see also Palacios-Rodriguez et al. (2023)), for extremal observations in Benevento et al. (2023), Zuccolotto et al. (2023), and for panel data in Di Lascio et al. (2021).

Following these lines of investigations, the goal of this paper is to propose a hierarchical clustering method for time series that accounts for capturing the temporal dependence with a (soft) spatial constraints. To this end, we glue together two objects: a correlation matrix whose entries correspond to the pairwise association between time series (measured with the van der Waerden coefficient); a correlation matrix that reflects the spatial information. This aggregation is tuned by means of a suitable transformation that preserves the unique mathematical structure of correlation matrices, known as the Riemannian manifold, as recently considered in David and Gu (2019, 2022), Thanwerdas and Pennec (2021). In fact, mathematical operations on the correlation matrices may not necessarily be coherent (i.e., the structure of the correlation matrix may not be preserved). Thus, a (spatially-driven) modification of a correlation matrix may destroy its structure that contains more information as a whole than the sum of independent pairwise correlation coefficients (You and Park, 2022).

This contribution is organized as follows. Section 2 presents the proposed algorithm for correlation-based hierarchical clustering with spatial constraints (Spatial-CHC). As a matter of fact, the algorithm can be also used when no spatial constraint is present. Thus, Section 3.1 presents a simulation study to compare its performance with existing linkage-based clustering methods already presented in Fuchs et al. (2021). An illustration of the Spatial-CHC is instead presented in Section 3.2. In Section 4 the methodology is applied to an empirical analysis of the extreme temperature trends in Apulia region (Italy), combined with geographical information, i.e., the actual locations of the weather stations. Section 5 concludes.

2. The methodology

We are interested in clustering a set of time series, represented by a $(T \times n)$-data matrix, where n is the number of units and T is the time period length. The data matrix $\mathbf{X}$ is hence represented in the form

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{T1} & \cdots & x_{Tn} \end{bmatrix}$$

where x_{ti} is a generic element that represents the value of the ith time series ($i = 1, \dots, n$) at the tth period ($t = 1, \dots, T$). We set $\mathbf{x}_i = (x_{1i}, \dots, x_{Ti})^\top$.

For $i = 1, \dots, n$ the ith time series is associated with a p-dimensional vector $\mathbf{s}_i$ of features that contains additional information on the phenomenon under consideration. In our setting, we will refer to $\mathbf{s}_i$ as the spatial information, since it will typically contain the information about the geographic coordinates of the location at which the ith time series is observed. However, this vector may contain also other (deterministic) variables. For instance, for some financial stock i, $\mathbf{x}_i$ may represent the log-returns observed at times $1, \dots, T$, while $\mathbf{s}_i$ may collect information about the financial sector, balance sheet, etc.

Given the input matrices $\mathbf{X}$ (temporal matrix) and $(\mathbf{s}_1^\top, \dots, \mathbf{s}_n^\top)$ (spatial matrix), as well as a weighting parameter $\alpha \in [0, 1]$, the Spatial-CHC algorithm will provide as output a hierarchical representation (Contreras and Murtagh, 2015) of the original time series.

The Spatial-CHC method is grounded on two main ideas. First, temporal and spatial information should be represented in terms of correlation matrices. In fact, according to Fuchs et al. (2021), the dissimilarity between two variables can be obtained by some transformation of association measures that assign the maximal value to the comonotonic case. Second, merging temporal and spatial information should reflect the intrinsic geometry of the space of such correlation matrices. In this respect, the geometric structure of the set of correlation matrices will be considered (David and Gu, 2019, 2022). The whole procedure will mainly require five steps, as illustrated in detail below.

2.1. Data preprocessing

Since we are interested in the rank-invariant dependence among the time series, we should take into account that, in order to provide fully valid results, $\mathbf{x}_1, \dots, \mathbf{x}_n$ should be regarded as i.i.d. sample from a joint continuous distribution function H. However, the iid assumption can fails in various cases.

In such a case, as common in this framework (see, e.g., Durante et al. (2015)), we can consider that each matrix $\mathbf{X}$ is a realization of the stochastic volatility model (Rémillard, 2017)

$$\mathbf{X}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\sigma}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t, \tag{1}$$

where the innovations $\boldsymbol{\varepsilon}_t$ are iid with marginal mean equal to 0 and marginal unit variance, with 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 the information from the past and possibly information from exogenous variables as well. Since the distribution function H is continuous, there exists a unique copula C so that for all

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_d(x_d)),$$

where $F_1, \dots, F_n$ are the distribution functions of the innovations. As is well-known, the copula C contains full information about the monotone association between the involved variables (see, e.g., Durante and Sempi (2016)).

Now, given an estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, the parameters' vector of the stochastic model (1), we can compute the residuals

$$\mathbf{e}_t = (\mathbf{x}_t - \boldsymbol{\mu}_t(\widehat{\boldsymbol{\theta}}))/\boldsymbol{\sigma}_t(\widehat{\boldsymbol{\theta}}).$$

The normalized ranks of the residuals are called pseudo-observations and can be used to calculate the copula C and its related functionals (like measures of association). In fact, even if the stochastic volatility model is appropriate, the residuals are not i.i.d. However, the empirical copula process based on the residuals has the same asymptotic behaviour as the empirical copula process for i.i.d. observations as shown in Rémillard (2017). Thus, one can apply the rank-based copula inference procedures on the estimated residuals as if they were the innovations.

In applications, frequent choices for the marginal models are ARMA models for the conditional means, and GARCH models for the conditional variances. The name copula-GARCH model is often used to refer to such types of models (Chen and Fan, 2006; Jondeau and Rockinger, 2006).

2.2. Extract the temporal dependence

Once the time series have been filtered via suitable marginal models, we have to extract the information about their pairwise dependence. Among various possible alternatives, here we focus on the van der Waerden correlation coefficient ζ (see, e.g., Genest and Verret (2005)), also known as normal score correlation or Gaussian rank correlation. 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, 2020). Interestingly, van der Waerden coefficient is also a concordance measure, that can be expressed as an integral of a function with respect to a copula measure (Genest and Verret, 2005).

Our main interest in this coefficient stands in the following fact. Given a random vector of dimension $n \geq 2$, the $(n \times n)$ matrix that collects the (pairwise) normal score correlations is semidefinite positive, since ζ is a Pearson's correlation. Moreover, the converse implication is also true in the following sense: for any symmetric, positive semi-definite matrix $\mathbf{M}$ in $[-1, 1]^{n \times n}$ with diagonal elements one, there exists a n-dimensional random vector whose normal score correlation matrix coincides with $\mathbf{M}$. In other words, the set of van der Waerden's matrices coincides with the set of all correlation matrices (Hofert and Koike, 2019, Proposition 5).

In the following, we denote by Corr(n) the class of all correlation matrices, i.e. symmetric, positive semi-definite matrices in $[-1, 1]^{n \times n}$ with diagonal elements one. Clearly, Corr(n) $\subset \mathrm{Sym}^+(n)$, the class of symmetric, positive semi-definite matrices.

Remark. Notice that Corr(n) does not coincides with the set of all correlation matrices derived from other popular measures of concordance like Kendall's τ and Blomqvist's β (Hofert and Koike, 2019). Moreover, Corr(n) coincides with the set of all Spearman's rank correlation matrices only when $n \leq 9$, but not when $n \geq 12$, while the case where $n \in \{10, 11\}$ remains to be settled (Devroye and Letac, 2015; McNeil et al., 2022; Wang et al., 2019).

Thus, given the pseudo-observations extracted from the original time series, we denote by $\mathbf{M}_T$ the $(n \times n)$ matrix that collects all the estimated pairwise values of van der Waerden coefficient among the time series, i.e.

$$\mathbf{M}_T = \begin{bmatrix} \widehat{\zeta}_{11} & \cdots & \widehat{\zeta}_{1n} \\ \cdots & \cdots & \cdots \\ \widehat{\zeta}_{n1} & \cdots & \widehat{\zeta}_{nn} \end{bmatrix} \tag{2}$$

where $\widehat{\zeta}_{ij}$ denotes the estimated van der Waerden's coefficient between the residuals $\mathbf{e}_i$ and $\mathbf{e}_j$ extracted from the ith and jth time series. Specifically,

$$\widehat{\zeta}_{ij} = \frac{1}{n} \sum_{k=1}^{T} \Phi^{-1}\left(\frac{R_{ki}}{n+1}\right) \Phi^{-1}\left(\frac{R_{kj}}{n+1}\right), \tag{3}$$

where Φ is the standard Gaussian distribution function, and R_{ki} is the rank of e_{ki} among $(e_{1i}, \dots, e_{Ti})$. For more details, see Koike and Hofert (2020).

As written above, $\mathbf{M}_T \in$ Corr(n) and represents the dependence similarity. In particular, $(\mathbf{M}_T)_{ij} = 1$ whenever the ith time series is comonotonic with the jth time series.

2.3. Extract the spatial dependence

Analogously to the temporal case, we would like to construct a matrix $\mathbf{M}_S \in \text{Corr(n)}$ that interprets the spatial information given by $\mathbf{s}_1, \dots, \mathbf{s}_n$. With a little abuse of language, we will refer to such a matrix as the spatial dependence matrix.

To this end, we fix a suitable distance d between the features associated to each time series. In most cases, the Euclidean distance will be used when the features are geographical coordinates.

A correlation matrix $\mathbf{M}_S$ can be hence obtained by transforming the distance among points to correlations so that the correlation between $\mathbf{s}_i$ and $\mathbf{s}_j$ decreases as their distance $d(\mathbf{s}_i, \mathbf{s}_j)$ increases. In this context, a popular choice is to use the *Matérn family of correlation* (Diggle and Ribeiro, 2007; Stein, 1999; Abramowitz and Stegun, 1965). In such a case, the correlation function depends only on the distance u between the points and it is given by

$$\rho_{\text{Matern}}(u) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{u}{\phi}\right)^{\nu} K_{\nu}\left(\frac{u}{\phi}\right), \tag{4}$$

where ϕ is a positive length-scale parameter, $\Gamma(\cdot)$ is the standard Gamma function and $K_{\nu}(\cdot)$ is a modified Bessel function (Abramowitz and Stegun, 1965). The positive parameter ν effectively controls the shape of the function.

The correlation matrix interpreting the spatial information is hence the $n \times n$ matrix

$$\mathbf{M}_S = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \cdots & \cdots & \cdots \\ \rho_{n1} & \cdots & \rho_{nn} \end{bmatrix} \tag{5}$$

where $\rho_{ij} = \rho_{\text{Matern}}(d(\mathbf{s}_i, \mathbf{s}_j))$ with $i, j \in \{1, \dots, n\}$. Since the matrix $\mathbf{M}_S$ is by definition positive semi-definite and it has all the elements in the diagonal equal to one, we have $\mathbf{M}_S \in \text{Corr(n)}$.

Clearly, the Matérn correlation function can generate very different correlation structure according to the chosen parameters. In the present context, we suggest to calibrate its parameters so that a given practical range is fixed, which we define as the distance u_0 at which the correlation is 0.05, i.e. $\rho_{\text{Matern}}(u_0) = 0.05$ (Diggle and Ribeiro, 2007).

In the following, we will mainly consider two main cases: (a) the exponential case $\rho_{\text{Matern}}(u) = \exp(-u/\phi)$; (b) the Gaussian case $\rho_{\text{Matern}}(u) = \exp(-(u/\phi)^2)$. In the simulation and in the case study, the effects of these different choices are discussed.

2.4. Merging temporal and spatial dependence

Now, suppose that both the temporal matrix $\mathbf{M}_T$ and the spatial matrix $\mathbf{M}_S$ are obtained. As known, both matrices belong to Corr(n). Moreover, hereinafter, we assume that they have full rank. Our aim is to construct a matrix $\mathbf{M}_\alpha$ as a function that depends on $\mathbf{M}_T$, $\mathbf{M}_S$ and a given $\alpha \in [0, 1]$, that, roughly speaking, can interpolate the two input matrices. In fact, notice that the linear combination $(1-\alpha)\mathbf{M}_T + \alpha\mathbf{M}_S$ may not be a convenient choice, as showed for instance in Marti et al. (2021b). To this end, two genuine questions should be answered:

(a) how far is $\mathbf{M}_T$ from $\mathbf{M}_S$?
(b) is there any optimal trajectory that pushes $\mathbf{M}_T$ forward to $\mathbf{M}_S$?

To answer both questions we rely on the geometric structure of the space of positive definite matrices (see, e.g., Bhatia (2009)) and of related subspace of correlation matrices (David and Gu, 2019; Thanwerdas and Pennec, 2021, 2022). As is known (see, e.g., Bhatia (2009, section 6)), the space $\text{Sym}^+(n)$ is a manifold with a natural Riemannian structure. It can be equipped with the distance

$$d_{\text{Sym}^+}(\mathbf{A}, \mathbf{B}) = \| \operatorname{Log}\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right) \|_F, \tag{6}$$

where $\| \cdot \|_F$ represents the Frobenius norm. Moreover, there exists a unique geodesic joining two matrices $\mathbf{A}$ and $\mathbf{B}$ (Bhatia, 2009, Theorem 6.1.6) and it is given by

$$\gamma_{\mathbf{A},\mathbf{B}}(t) = \mathbf{A}^{1/2} \operatorname{Exp}\left(t \operatorname{Log}\left(\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\right)\right) \mathbf{A}^{1/2}, \quad t \in [0, 1]. \tag{7}$$

Furthermore,

$$d_{\text{Sym}^+}(\mathbf{A}, \gamma_{\mathbf{A},\mathbf{B}}(t)) = t \cdot d_{\text{Sym}^+}(\mathbf{A}, \mathbf{B}) \tag{8}$$

for every $t \in [0, 1]$ (Bhatia, 2009, eq. (6.12)).

The middle point of the geodesic $\gamma_{\mathbf{A},\mathbf{B}}(0.5) \in \text{Sym}^+(n)$ corresponds to the (Riemannian) mean of the two matrices (see Moakher (2005, Proposition 3.5) and also (Kubo and Ando, 1980)). In general, it does not coincide with $(\mathbf{A}+\mathbf{B})/2$. Moreover, it also satisfies

$$\arg \min_{\mathbf{C} \in \text{Sym}^+(n)} \left(d^2_{\text{Sym}^+}(\mathbf{C}, \mathbf{A}) + d^2_{\text{Sym}^+}(\mathbf{C}, \mathbf{B}) \right). \tag{9}$$

Now, mathematical operations of the correlation matrices under the $\text{Sym}^+(n)$ geometry may not necessarily be coherent (i.e., the structure of the correlation matrix may not be preserved), necessitating a post-hoc normalization (see, e.g., Marti et al. (2021a)). To this end, we consider the space $\text{Diag}_+(n)$ of $(n \times n)$ diagonal matrices with positive entries. We denote by $\pi : \text{Sym}^+(n) \to \text{Corr(n)}$

the function given by $\pi(\mathbf{A}) = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the diagonal matrix whose entries are the diagonal entries of $\mathbf{A}$. According to David and Gu (2022) (see also (Huckemann et al., 2010)), we can define the following distance on Corr(n):

$$d_{\text{Corr}}(\mathbf{A},\mathbf{B}) = \inf_{\mathbf{D}\in \text{Diag}_+(n)} d_{\text{Sym}^+}(\mathbf{A},\mathbf{DBD}). \tag{10}$$

Notice that d_{Corr} is a distance on Corr(n), thus $d_{\text{Corr}}(\mathbf{A},\mathbf{B}) = d_{\text{Corr}}(\mathbf{B},\mathbf{A})$ even if different $\mathbf{D}$ satisfy Eq. (10). The geodesic joining $\mathbf{A}$ and $\mathbf{B}$ can be taken as the projection of the geodesic connecting $\mathbf{A}$ and $\widetilde{\mathbf{B}}$, which is the unique element in $\pi^{-1}(\mathbf{B})$ that has minimal d_{Sym^+}-distance from $\mathbf{A}$. This can be written as

$$\gamma^{\text{Corr}}_{\mathbf{A},\mathbf{B}}(t) = \pi\left(\mathbf{A}^{1/2}\,\text{Exp}(t\,\text{Log}(\mathbf{A}^{-1/2}\mathbf{D}^*\mathbf{B}\mathbf{D}^*\mathbf{A}^{-1/2}))\mathbf{A}^{1/2}\right) \tag{11}$$

where

$$\mathbf{D}^* = \arg\inf_{\mathbf{D}\in \text{Diag}_+(n)} d_{\text{Sym}^+}(\mathbf{A},\mathbf{DBD}). \tag{12}$$

A closed-form expression of the matrix $\mathbf{D}^*$ is difficult to obtain Thanwerdas (2022). However, some numerical procedures based on gradient descent method have been suggested in David and Gu (2019) (see also (You and Park, 2022)). See Algorithm 1.

Thus, for any fixed $\alpha \in [0,1]$ a way to merge temporal and spatial dependence is obtained by the matrix

$$\mathbf{M}_\alpha = \gamma^{\text{Corr}}_{\mathbf{M}_T,\mathbf{M}_S}(\alpha), \tag{13}$$

which belongs to the geodesic joining the two matrices. Clearly, $\mathbf{M}_0 = \mathbf{M}_T$ and $\mathbf{M}_1 = \mathbf{M}_S$.

Remark. Numerical computations in Marti et al. (2021b) seems to indicate that $\mathbf{M}_\alpha$ coincides with the solution of the minimization problem

$$\mathbf{M}_\alpha = \arg\min_{\mathbf{C}\in \text{Corr(n)}} \left((1-\alpha)\cdot d^2_{\text{Corr}}(\mathbf{C},\mathbf{A}) + \alpha\cdot d^2_{\text{Corr}}(\mathbf{C},\mathbf{B})\right), \tag{14}$$

i.e. it is the weighted (Riemannian) mean of $\mathbf{A}$ and $\mathbf{B}$. This conjecture seems to be confirmed by the procedure described in Riquelme and Ortiz (2023, Algorithm 3) (see also (Riquelme, 2002)). However, we have been not able to provide a formal proof of this statement.

The main steps for calculating the matrix $\mathbf{M}_\alpha$ are summarized below in Algorithm 1 (see also the supplementary material in You and Park (2022)). The convergence of the computation of $\widehat{\mathbf{D}^*}$ follows from the convergence of the gradient descent algorithm as shown in David (2019, chapter 3).

Algorithm 1 Calculation of $\mathbf{M}_\alpha$.

input: the temporal matrix $\mathbf{M}_T$, the spatial matrix $\mathbf{M}_S$, the mixing parameter $\alpha \in [0,1]$

calculate $\widehat{\mathbf{D}^*} \approx \mathbf{D}^*$ **in** (12)
  initialize $\mathbf{D}^{(0)} = I_n$, a stepsize $\tau > 0$, stop_tol $= \epsilon$
  while $\|\mathbf{D}^{(t+1)} - \mathbf{D}^{(t)}\|_F \le \epsilon$ **do**:
    $\mathbf{B}^{(t)} \leftarrow I_n \bullet Symm[\mathbf{D}^{(t)} Log(\mathbf{M}_T\mathbf{D}^{(t)}\mathbf{M}_S^{-1}\mathbf{D}^{(t)})]$
    $\mathbf{D}^{(t+1)} \leftarrow \mathbf{D}^{(t)} Exp(-2\tau(\mathbf{D}^{(t)})^{-1}\mathbf{B}^{(t)})$
    where $\bullet$ is the Hadamard product, and $Symm$ extracts symmetric component of a matrix
  end while

calculate $\mathbf{M}_\alpha$ **in** (13)
  $\Gamma_\alpha \leftarrow \mathbf{M}_T^{1/2}\,\text{Exp}(\alpha\,\text{Log}(\mathbf{M}_T^{-1/2}\widehat{\mathbf{D}^*}\mathbf{M}_S\widehat{\mathbf{D}^*}\mathbf{M}_T^{-1/2}))\mathbf{M}_T^{1/2}$
  initialize $\mathbf{D}_\alpha = \mathbf{0}_{n\times n}$
  $diag(\mathbf{D}_\alpha) \leftarrow diag(\Gamma_\alpha)$
  $\mathbf{M}_\alpha \leftarrow \mathbf{D}_\alpha^{-1/2}\Gamma_\alpha\mathbf{D}_\alpha^{-1/2}$

return: $\mathbf{M}_\alpha$

2.5. Apply the hierarchical clustering algorithm

Given $\alpha \in [0,1]$ and given the correlation matrix $\mathbf{M}_\alpha$, we have to transform $\mathbf{M}_\alpha$ into a dissimilarity matrix $\boldsymbol{\Delta}^\alpha = (\delta^\alpha_{ij})$, which can represent the input of various (agglomerative) hierarchical clustering algorithms. As is known, $\boldsymbol{\Delta}^\alpha$ must satisfy:

(a) $\delta^\alpha_{ij} \ge 0$ for every i, j;
(b) $\delta^\alpha_{ii} = 0$ for every i;
(c) $\delta^\alpha_{ij} = \delta^\alpha_{ji}$ for every i, j.

Thus, a natural choice is to consider $\boldsymbol{\Delta}^{\alpha} = 1 - \mathbf{M}_{\alpha}$.

Remark. In the literature, various alternatives have been provided to transform a correlation matrix $\mathbf{M}$ by applying to its entries a strictly decreasing $f : [-1, 1] \to [0, +\infty[$ with $f(1) = 0$. For instance, $f(x) = \sqrt{1 - x^2}$ and $f(x) = \sqrt{0.5(1 - x)}$ (Van Dongen and Enright, 2012). However, such modifications are mainly used when the triangle inequality must be satisfied, i.e. a distance rather than a dissimilarity is required.

The matrix $\boldsymbol{\Delta}^{\alpha}$ can be hence used to perform any hierarchical clustering algorithm that requires as input a dissimilarity matrix.

The dissimilarity $\boldsymbol{\Delta}^{\alpha}$ depends on the chosen hyper-parameter α. Thus, if α approaches 0, then the algorithm output will result in a partition that is closer to the one that could be obtained with the temporal correlation only. Conversely, choosing an α closer to 1 will lead to a partition more similar to the one related to the spatial information.

As a main feature of the Spatial-CHC algorithm, the parameter α describes the relative influence between the spatial and the temporal dependence in a convenient Riemannian metric. So, it may provide a way to directly assign the weight of the spatial effects on the procedure (provided it is available to the decision maker). Thus, the value $\alpha = 0.5$ can be adopted as the way to summarize spatial and temporal dependence when no preference among the two criteria is expressed.

In general, the problem of selecting a suitable value for $\alpha \in [0, 1]$ has been crucial in analogous studies. In Romary et al. (2015), it is heuristically recommended to put 5-30% on the geographical coordinates and the remainder to the other attributes. However, in many cases, the algorithm is run for different values of α (and a fixed number K of clusters) and a choice is made according to some external criteria. In Chavent et al. (2018), for instance, the value α is selected so that it increases the geographical homogeneity of a partition in K clusters without adversely affecting the (non-spatial) feature homogeneity. To this end, a concept of pseudo within-cluster inertias is introduced. In Le (2021), α is determined so that it maximizes the spatial auto-correlation among post-cluster units. This approach allows the cluster formation to be largely based on the temporal dependence, while the spatial information only fine-tunes the final cluster partition. Similar criteria are adopted in the present context.

Specifically, inspired by Zuccolotto et al. (2023), in order to select a suitable $\alpha \in [0, 1]$ we proceed as follows.

(1) We define a discrete set of α's values in $[0, 1]$ and a maximum number of clusters $K_{\max}$;
(2) For each α
    (a) we perform a hierarchical cluster algorithm from the dissimilarity $\boldsymbol{\Delta}^{\alpha} = 1 - \mathbf{M}_{\alpha}$;
    (b) we choose the number of clusters K_{α} that maximizes a given internal validation index (i.e. Dunn Index) over different values of $K \in \{2, \dots, K_{\max}\}$;
    (c) given the optimal K_{α}, we evaluate the cluster composition by the same internal validation index but calculated with reference to the dissimilarity matrices $1 - \mathbf{M}_T$ and/or $1 - \mathbf{M}_S$;
(3) For the given α values, we plot the internal validation indices calculated with respect to $\mathbf{M}_T$ and/or $\mathbf{M}_S$. Finally, we choose the optimal value α that approximately preserves the internal validity index with respect to the temporal dissimilarity by giving, at the same time, an improvement on the performances with respect to the spatial component.

An illustration of the method is given in Sections 4 and 3.2.

3. A simulation study

Here, we present a simulation study to show the performance of the proposed dissimilarity matrix in two cases: (a) when $\alpha = 0$, i.e. the spatial information is not relevant; (b) when α belongs in $[0, 1]$.

3.1. The purely temporal case

In Fuchs et al. (2021) the performance of hierarchical clustering methods has been studied, when pairwise dissimilarity matrices are based on various measures of association (like Kendall's tau, Spearman's rho and Blomqvist's beta), and the classical linkage functions are considered. Here, we would like to compare these methods with the hierarchical clustering method based on van der Waerden's coefficient (without any spatial constraint), which has not been considered in the previous studies.

Thus, we consider the following simulation setup. A random vector $\mathbf{X}$ of dimension $n \in \{60, 120\}$ is constructed in the following way:

- it is formed by $k \in \{3, 6, 12\}$ independent subvectors $\mathbf{X}_k$ of an equal and fixed size;
- each $\mathbf{X}_k$ is distributed according to a copula generated from two different copula models, namely Clayton and Gumbel (for the definition of these families and their simulation, see, e.g., Mai and Scherer (2017)), with pairwise Kendall's tau equal to τ.

For $B = 500$ replications, $N \in \{50, 100\}$ independent realizations from $\mathbf{X}$ are simulated with $\tau \in \{0.1, 0.2, 0.3\}$. Hence, for each simulated scenario:

- we apply a hierarchical clustering algorithm with linkage function equal to average or maximum (complete linkage) and pairwise dissimilarity derived from Blomqvist's β, Spearman's footrule ϕ, Kendall's τ, Spearman's ρ and van der Waerden ζ;

Table 1
A comparison of the mean and standard deviation of the ARI obtained after 500 repetitions of the method. The data are simulated from an $n = 60$ dimensional random vector with K independent groups having a Clayton copula with Kendall's τ equal to 0.1, 0.2, and 0.3 within each group and number of clusters equal to $K = 3$, $K = 6$, and $K = 12$, respectively. Each column represents the ARI with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ and (ii) linkage method among the average, and complete (maximum) one. Sample size equal to 100 points.

	$n = 60,\ N = 100$, Clayton model					
	$\tau = 0.1;\ K = 3$		$\tau = 0.2;\ K = 6$		$\tau = 0.3;\ K = 12$	
	Mean	σ	Mean	σ	Mean	σ
β average	0.51	0.18	0.85	0.10	0.96	0.06
β complete	0.30	0.17	0.68	0.14	0.90	0.09
ϕ average	0.79	0.15	0.98	0.03	1.00	0.00
ϕ complete	0.62	0.21	0.96	0.06	1.00	0.01
τ average	0.80	0.14	0.99	0.03	1.00	0.00
τ complete	0.64	0.21	0.96	0.06	1.00	0.00
ρ average	0.80	0.14	0.98	0.03	1.00	0.00
ρ complete	0.64	0.20	0.97	0.05	1.00	0.01
ζ **average**	0.83	0.13	0.99	0.02	1.00	0.00
ζ **complete**	0.70	0.20	0.98	0.05	1.00	0.00

- we derive the corresponding partition in K clusters by cutting the dendrogram at a suitable height;
- we calculate the Adjusted Rand Index (Hubert and Arabie, 1985) (ARI, hereafter) to measure the agreement between the obtained partition and the true one. We remind that a larger Adjusted Rand Index means a higher agreement between two partitions and the maximum value of the index is 1.

Notice that, since the choice of the copula family seems to be irrelevant in the previous simulation studies (Fuchs et al., 2021), we do not extend the study to other families. Analogously, we do not consider the minimum linkage (single method), since it generally shows poor performance.

The distribution of ARI for each simulated scenario is shown in the figures in Appendix. In Table 1 we summarize the values of the mean and the standard deviation for an excerpt of the simulated scenarios.

From the simulation study, the following interpretation can be done:

- As one could have expected, the lower is the degree of dependence among the variables of a group, the harder is for the hierarchical clustering algorithm to identify the true partition. Moreover, the larger is the sample size, the better are the results for a given dependence degree. Thus, the use of van der Waerden coefficient is in agreement with similar methods.
- However, the method based on the ζ coefficient seems to perform similarly/slightly better than the methods based on Kendall's τ and Spearman's ρ.
- Similarly to other studies (Fuchs et al., 2021), the average linkage method appears to be more satisfactory than the complete one.
- At the increase of number of variables to cluster, we do not observe any substantial changes for $\tau = 0.1$, while the interquartile range seems to be lower for higher number of variables for $\tau \in \{0.2, 0.3\}$, which represents a lower spread of data.

Overall, the finite-sample performance of van der Waerden coefficients for correlation-based clustering is in agreement with/slightly better than other popular rank-correlation coefficients. Moreover, as we have explained above, its application in presence of spatial constraints has the advantage to be grounded on a consistent mathematical framework (related to the geometric structure of the space of correlation matrices).

Remark. The simulation study has been developed by using the R packages (Hofert et al., 2022) and Asquith (2022) for the copula simulations and estimation of the measures of association. For the general hierarchical clustering algorithms, we also use (Maechler et al., 2022).

3.2. An illustration with spatial constraints with simulated data

The spatial clustering techniques illustrated in Section 2 could not be directly validated through simulation procedures, since the final clustering solution, which represents the main task of the method, is determined on the ground of the method itself. However, we could assess the impact on the temporal clustering of the clusters arising from distinct spatial information. This could be another way to highlight how relevant the spatial information is in characterizing the clusters' composition. To this end, we illustrate the Spatial-CHC algorithm in the following simulated scenario.

Similarly to Section 3.1, we consider $n = 54$ time series of length $T = 100$. For simplicity, the time series are grouped in $K = 3$ (equally sized) clusters and, specifically, they are obtained from three independent 18-dimensional Archimedean copulas of the Clayton family with pairwise Kendall's $\tau = 0.2$ as one of the cases of Section 3.1 (the Gumbel case can be done analogously and

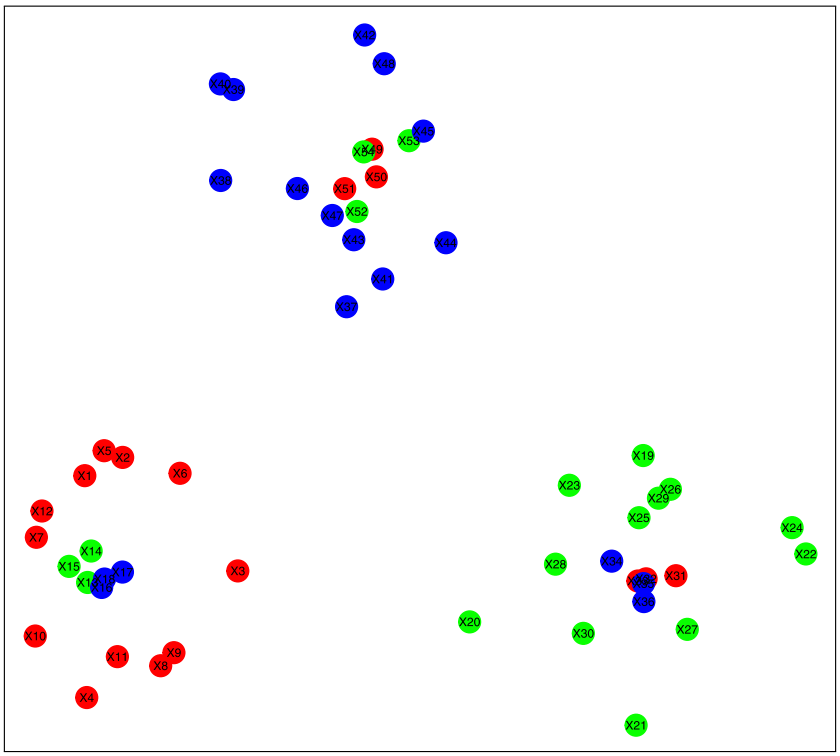


Fig. 1. The plot shows the geographic sites where the time series have been collected. Each colour represents a different temporal cluster to which the related time series belongs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

produces similar results). Moreover, we assume that each time series is associated with some geographic coordinates that can be interpreted as the location of the sites where the observations have been collected.

Fig. 1 shows the geographic proximity of the observation sites, which can be separated into three regions. The colours represent the true 3-group cluster composition induced by the temporal data generating model. As can be seen, the spatial proximity only partially overlap with the temporal structure. For instance, 12 time series of the red temporal cluster are located in the south-west region, while 6 time series belong to other geographic areas.

Temporal dependence From the observed time series, we calculate the temporal matrix $\mathbf{M}_T$ as in Section 2.2. Fig. 2 shows the dendrogram derived from the hierarchical clustering (average linkage) applied to the dissimilarity matrix $1 - \mathbf{M}_T$. The corresponding 3-group cluster composition is in agreement with the configuration in Fig. 1.

Spatial dependence By using the spatial information given by geographic coordinates, as in Section 2.3, we compute two different Matérn correlation functions: (a) the exponential case with parameters $\phi = -\frac{u_0}{\ln(\rho_{\text{Matern}}(u_0))}$ and $\nu = 0.5$; (b) the Gaussian case with parameters $\phi = \frac{u_0}{\sqrt{-\ln(\rho_{\text{Matern}}(u_0))}}$ and $\nu = +\infty$. In both cases, we empirically set $u_0 = 4$ to have a common practical range.

Now, we want to observe how the cluster partition is modified when we gradually increase the influence of the spatial information by replacing the dissimilarity matrix $1 - \mathbf{M}_T$ (which coincides with $\boldsymbol{\Delta}^0$) with the matrix $\boldsymbol{\Delta}^\alpha = 1 - \mathbf{M}_\alpha$, letting α go from 0 to 1.

We start with the exponential case. For the sake of illustration, we consider $\alpha = 0.5$. First, in Fig. 3 we show how the matrix $\boldsymbol{\Delta}^{0.5}$ smoothly mixes the characteristics of $\boldsymbol{\Delta}^0$ and $\boldsymbol{\Delta}^1$. The dendrogram deriving from the hierarchical clustering (average linkage) with dissimilarity matrix $\boldsymbol{\Delta}^{0.5}$ is shown in Fig. 4. As can be seen, it partially overlaps with the dendrogram in Fig. 2, while rearranging the other variables also according to spatial constraints.

Now, we follow the procedure described in Section 2.5 to select a suitable $\alpha \in [0, 1]$ based on an internal validity index. Here, we use the Dunn Index, which is a scalar that formalizes the idea of a ratio between between-cluster separation and within-cluster compactness for general dissimilarity input data and a fixed number of clusters (Dunn, 1974; Hennig et al., 2015). Specifically, for a discrete set of α's values in $[0, 1]$ we choose the number of clusters K by maximizing the Dunn Index over different values of $K \in \{3, \dots, 10\}$ with respect to the matrix $\mathbf{M}_\alpha$. Then, we use the selected K_α to compute the Dunn Index with respect to $\mathbf{M}_T$ and $\mathbf{M}_S$. We show the results of the procedure in Fig. 5. The graph in Fig. 5 suggests that all the values of α between 0 and 0.3 are suitable choices because the temporal cluster compactness does not have a clear loss. Alternatively, for $\alpha = 0.35$ the temporal cluster compactness is not at its minimum value yet, but it shows an improvement in the spatial compactness. Based on the observation of the plot, we perform the hierarchical clustering with the dissimilarity matrices $\Delta^{0.15}$ and $\Delta^{0.35}$, in addition to the one with $\Delta^{0.5}$.

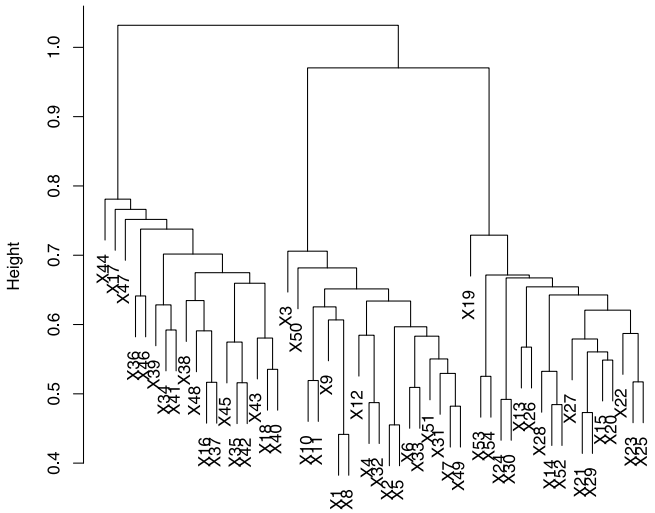


Fig. 2. Structure of the dendrogram representing the hierarchical clustering of the temporal dependencies.

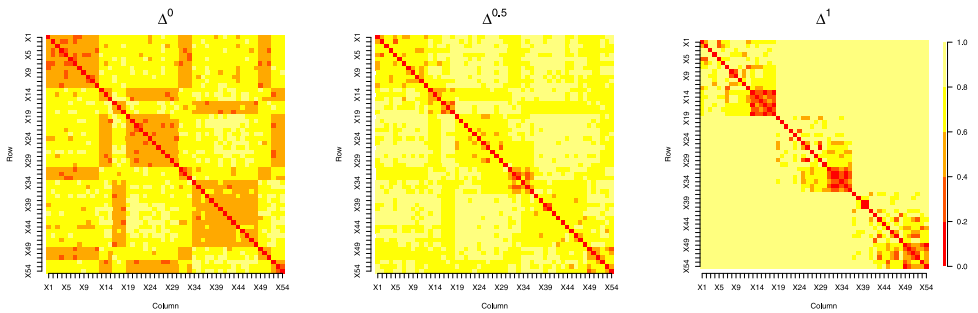


Fig. 3. The left panel represents the temporal dissimilarities in the data set. The central panel represents the dissimilarities $\boldsymbol{\Delta}^{0.5}$. The right panel represents the spatial dissimilarities. Darker colours represent stronger similarities among the data, which means that the darker squares are groups of data with similar characteristics. This reflects the behaviour of the clustering.

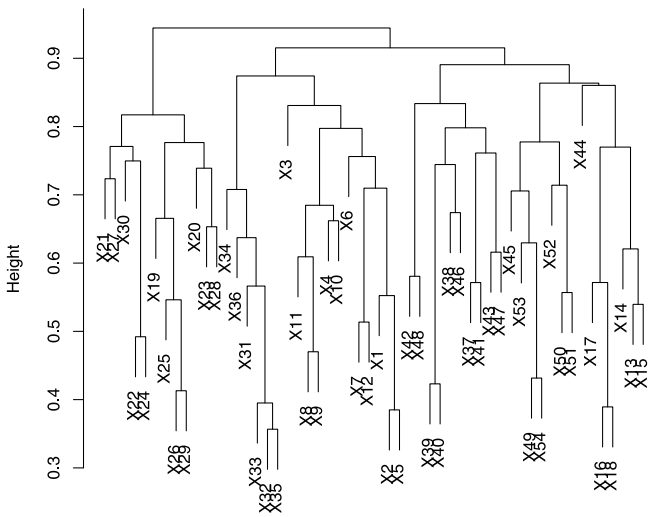


Fig. 4. Structure of the dendrogram representing the hierarchical clustering of $\boldsymbol{\Delta}^{0.5}$.

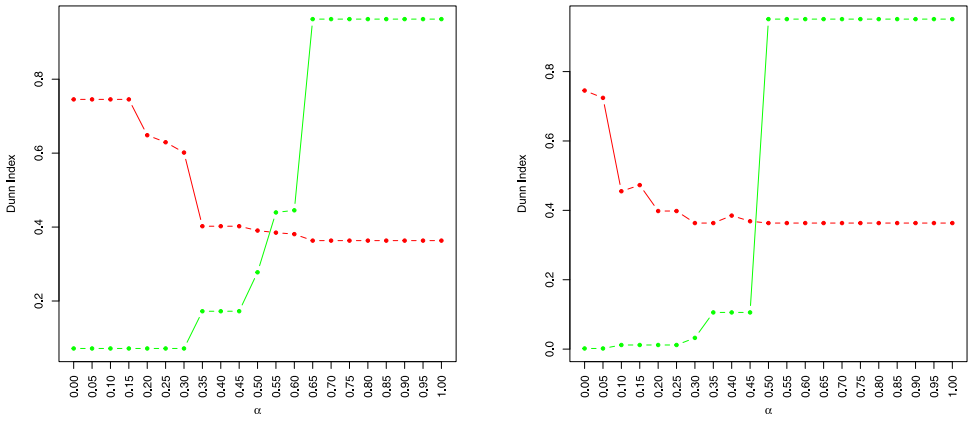


Fig. 5. Evolution of the Dunn Index for $\alpha \in [0,1]$, in the exponential case (left) and in the Gaussian case (right). The red points represent the Dunn Index computed with respect to the temporal matrix; the green points represent the Dunn Index computed with respect to the spatial matrix. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison of the partitions obtained with $\boldsymbol{\Delta}^{0.15}$ or $\boldsymbol{\Delta}^{0.5}$ and $\boldsymbol{\Delta}^{0}$ (temporal dependence) and $\boldsymbol{\Delta}^{1}$ (spatial dependence) respectively. In the first two columns we compare the cophenetic index; in the last two, we compare the ARI index.

		$\boldsymbol{\Delta}^{0.15}$	$\boldsymbol{\Delta}^{0.5}$	$\boldsymbol{\Delta}^{0.15}$	$\boldsymbol{\Delta}^{0.5}$
Exp	$\boldsymbol{\Delta}^{0}$	Coph = 0.978	Coph = 0.412	ARI = 1.000	ARI = 0.308
	$\boldsymbol{\Delta}^{1}$	Coph = 0.245	Coph = 0.726	ARI = 0.221	ARI = 0.435
Gauss	$\boldsymbol{\Delta}^{0}$	Coph = 0.796	Coph = 0.161	ARI = 0.673	ARI = 0.220
	$\boldsymbol{\Delta}^{1}$	Coph = 0.325	Coph = 0.850	ARI = 0.309	ARI = 1.000

From the maximization of the Dunn index we obtain that the optimal number of clusters performed with $\boldsymbol{\Delta}^{0.15}$ is $K^{\star}_{0.15} = 3$ while for $\boldsymbol{\Delta}^{0.35}$ is $K^{\star}_{0.35} = 6$ and for $\boldsymbol{\Delta}^{0.5}$ is $K^{\star}_{0.5} = 10$. The composition of the identified clusters is shown in Fig. 6. Analogously, the optimal number of clusters computed for $\boldsymbol{\Delta}^{0}$ and computed for $\boldsymbol{\Delta}^{1}$ is 3 in both cases, and, hence, in agreement with the data generating process.

Summarizing, at the increase of α in $[0, 0.5]$, the procedure tends to split a larger (and spatially dispersed) cluster into two or more clusters that are smaller but more spatially concentrated.

We repeat the simulation in the case the spatial matrix is obtained from the Gaussian kernel function. In Fig. 5 we show the evolution of the Dunn Index computed according with Section 2.5. The figure suggests that the compactness of temporal clusters is lost for very small values of α. A suitable choice in this example could be $\alpha = 0.15$ as the temporal compactness has not reached its minimum value yet, while the spatial one has started growing.

In Fig. 7 we show the distribution of the elements in the clusters with respect to the dissimilarities $\Delta^{0.15}$ and $\Delta^{0.5}$. When $\alpha = 0.15$ each group is preserving the temporal behaviour even if some elements are influenced by the space (i.e., the element 44 or 19). Moreover, as expected, when $\alpha = 0.5$ the cluster distribution corresponds to the purely spatial one.

To complete the illustration, in Table 2 we compare the values of the Adjusted Rand Index (ARI) between the (optimal) cluster composition obtained from the different dissimilarity matrices, both in the exponential and in the Gaussian case. Moreover, the table also shows the value of the Cophenetic index between the related dendrograms. We recall that the cophenetic index is a measure for comparing two dendrograms, i.e., it measures the cophenetic correlation between two trees (Galili, 2015). The value can range between -1 and 1.

4. A case study with climatological data

Here we exploit the Spatial-CHC algorithm for finding a common behaviour of temperatures trends. Specifically, given a geographic region having various weather stations that collect temperature data, we aim at identifying agglomerations of cities characterized by similar temperature behaviour over time and, eventually, by a geographic proximity.

In this analysis, we consider daily temperature levels (in Celsius) in 23 municipalities located in Apulia region (Southern Italy) collected by ARPA Puglia[1] from January 2019 to December 2022. The Weather Service has performed the process of collecting and validating the weather data of the Telemetry Network, made up of 5 automatic stations located at its provincial offices in ARPA Puglia and 18 weather stations belonging to the Regional Air Quality Network (RRQA). The temperature parameters are

[1] http://www.webgis.arpa.puglia.it/meteo/index.php

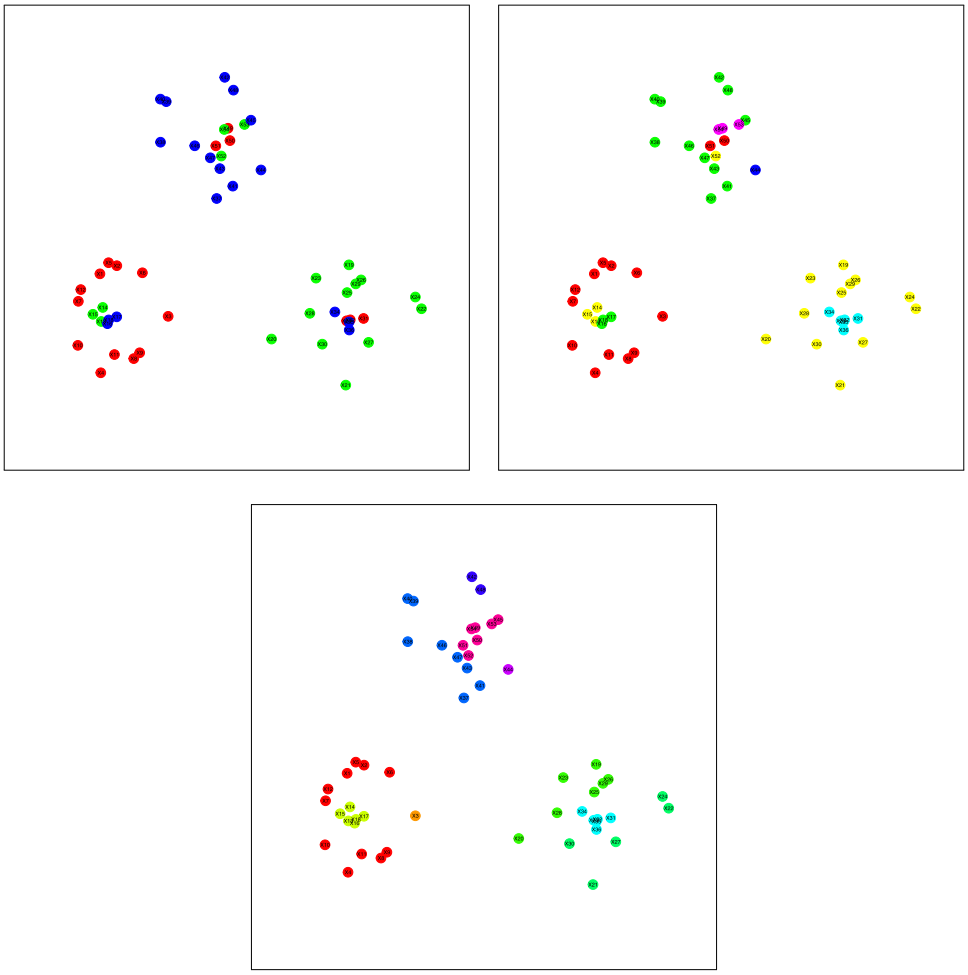


Fig. 6. [Exponential case] Distribution of the elements in the clusters with respect to the dissimilarities $\boldsymbol{\Delta}^{0.15}$ with $K^{\star}_{0.15} = 3$ (on the top-left), with respect to $\boldsymbol{\Delta}^{0.35}$ with $K^{\star}_{0.35} = 6$ (on the top-right) and with respect to $\boldsymbol{\Delta}^{0.5}$ with $K^{\star}_{0.5} = 10$ (on the bottom). Each colour represents a different cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

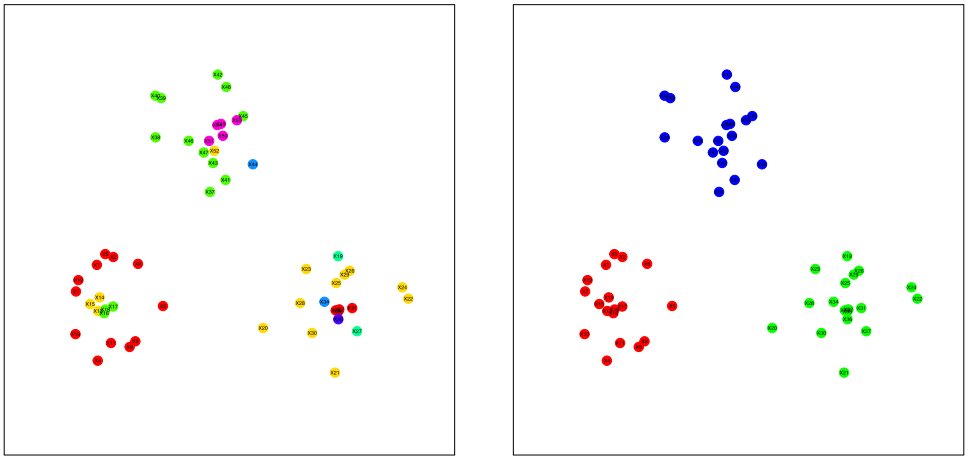


Fig. 7. [Gaussian case] Distribution of the elements in the clusters with respect to the dissimilarities $\boldsymbol{\Delta}^{0.15}$ with $K^{\star}_{0.15} = 7$ (on the left) and with respect to $\boldsymbol{\Delta}^{0.5}$ with $K^{\star}_{0.5} = 3$ (on the right). Each colour represents a different cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

	Station	Lat	Lon
1	Andria	41.23	16.30
2	Bari	41.12	16.89
3	Barletta	41.32	16.23
4	Brindisi	40.62	17.93
5	Candela	41.17	15.52
6	Carbonara	41.08	16.87
7	Casamassima	40.95	16.92
8	Cisternino	40.74	17.42
9	Foggia	41.46	15.55
10	Francavilla F.	40.53	17.59
11	Galatina	40.17	18.17
12	Grottaglie	40.54	17.42
13	Lecce	40.34	18.18
14	Maglie	40.12	18.29
15	Manfredonia	41.63	15.91
16	Massafra	40.59	17.12
17	Molfetta	41.20	16.60
18	Monopoli	40.95	17.29
19	San Pancrazio Sal.	40.42	17.85
20	San Severo	41.69	15.38
21	Taranto	40.48	17.20
22	Taranto San Vito	40.42	17.25
23	Torchiarolo	40.49	18.05

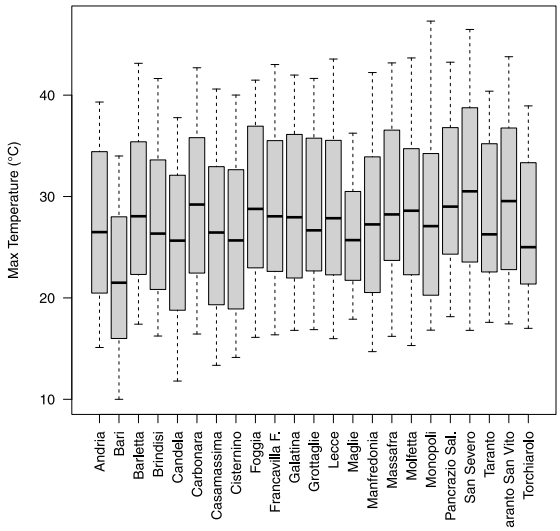


Fig. 8. (Left) Geographical coordinates of the locations in the data set. See Fig. 10 for the corresponding locations on the map. (Right) Distribution of maximum monthly temperature for the 23 stations in the period 2019–2022.

validated according to "Guidelines for checking the validity of hydro-meteorological data" (Barbero et al., 2017), developed within the National System for the Protection of the Environment. The website also provides a map with the locations of all the weather stations (longitude and latitude) that have been used to compute the geographical distances and the spatial matrix as described in Section 2.3. Furthermore, we accessed another weather website[2] containing weather data to acquire the missing data of some days. From these time series, temperatures monthly maxima were computed, resulting in a data set with $n = 23$ time series and $T = 48$ observations. The geographical coordinates of the locations as well as the distribution of the maximum monthly temperatures for the 23 stations are shown in Fig. 8.

According to Section 2.1, individual time series have been preprocessed in order to remove seasonal effects. The resulting residuals from the time series are hence used to compute the pseudo-observations. The filtered data are then used to determine the pairwise dependencies as described in Section 2.2. The dissimilarity matrix related to temporal dependence is hence visualized in Fig. 9. The corresponding clustering composition in $K = 8$ groups is illustrated in Fig. 10. Here, the average linkage is used and the number of clusters is selected via Dunn index.

In order to compute the spatial dependence, according to Section 2.3, we use two different Matérn correlation functions: (a) the exponential case; (b) the Gaussian case. In both cases, we empirically set $u_0 = 45$ to have a common practical range.

We hence perform the Spatial-CHC algorithm with these two different pairs of Matérn correlation functions according to Section 2.4. The corresponding merging matrix $\mathbf{M}_\alpha$ is obtained via (13). As the hierarchical clustering algorithm, we use the average linkage function, which has proved to perform well in the simulation study.

In order to select a suitable $\alpha \in [0, 1]$, we proceed as in Section 2.5. Specifically, for a discrete set of α's values in $[0, 1]$ we choose the number of clusters K by maximizing the Dunn Index over different values of $K \in \{3, \dots, 8\}$ with respect to the matrix $\mathbf{M}_\alpha$. Then, we use the selected K_α to compute the Dunn Index with respect to $\mathbf{M}_T$ and $\mathbf{M}_S$. The obtained results are shown in Fig. 11.

In the exponential case a suitable α to pick is $\alpha = 0.5$, which maximizes the Dunn index with respect to the temporal matrix by giving an improvement on the performances with respect to the spatial matrix. Analogously, in the Gaussian case, the most suitable α value is $\alpha = 0.3$.

Focusing on the final cluster solutions in the exponential case (see Fig. 12), we can notice that the Spatial-CHC algorithm mostly preserves the temporal cluster configuration, but also incorporates some stations that are very similar with respect to their geographical locations. For example, the northern stations in cluster 1 ($\alpha = 0.5$) were separated in the purely temporal case.

In the Gaussian case, first we notice that the cluster composition obtained with the purely spatial information coincides with the exponential case, represented in Fig. 12(a). Then, the obtained output is given in Fig. 13 for $\alpha = 0.3$ (optimal choice). Moreover, for the sake of illustration, we also include the case with $\alpha = 0.5$, which can be considered as an intermediate case between temporal and spatial information. As can be seen, the clusters obtained with the Spatial-CHC for $\alpha = 0.5$ are geographically connected, although, in general, the method does not force the spatial contiguity between elements. For instance, the cluster 1 with $\alpha = 0.3$ is not geographically connected (see 13(a)).

[2] https://www.ilmeteo.it/portale/archivio-meteo

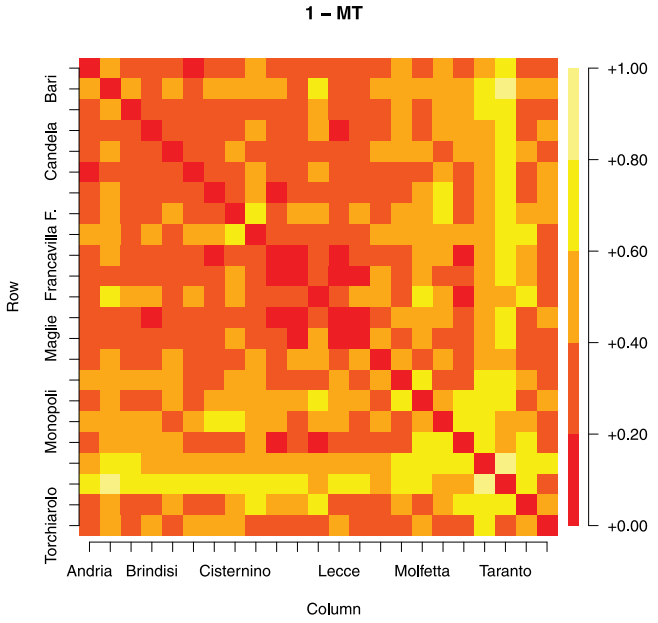


Fig. 9. Temporal dissimilarities in the data set. The 23 stations are alphabetically ordered. Only some of those are labelled in the picture.

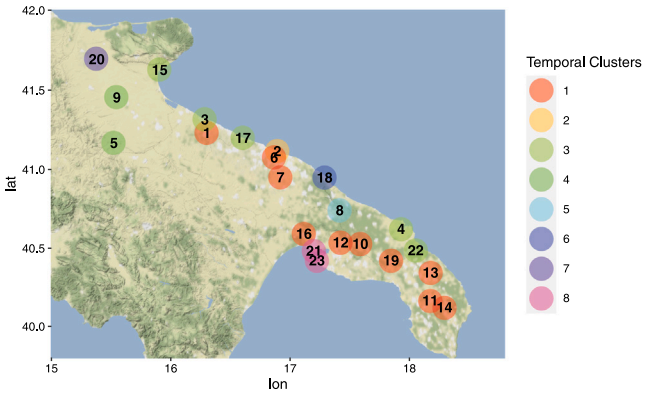


Fig. 10. Optimal cluster composition of time series derived from hierarchical clustering (average linkage) and purely temporal dissimilarity matrix $\boldsymbol{\Delta}^0$.

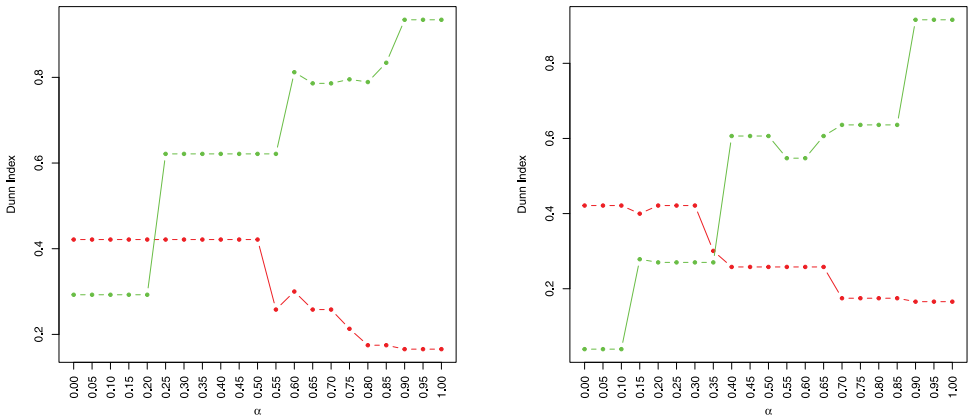


Fig. 11. Evolution of the Dunn Index for $\alpha \in [0, 1]$, in the exponential case (left) and in the Gaussian case (right). The red points represent the Dunn Index computed with respect to the temporal matrix; the green points represent the Dunn Index computed with respect to the spatial matrix. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

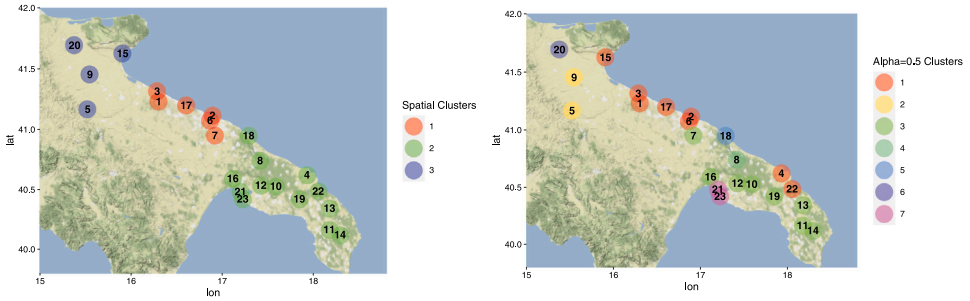


Fig. 12. [Exponential case] (a) Optimal cluster composition of time series derived from hierarchical clustering (average linkage) and pure spatial similarities (geographic locations); (b) optimal cluster composition of time series derived from hierarchical clustering (average linkage) and similarities $\mathbf{M}_{0.5}$.

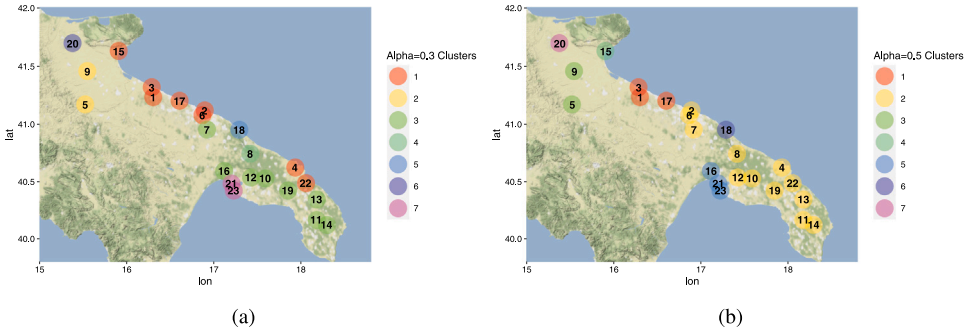


Fig. 13. [Gaussian case] (a) cluster composition of time series derived from hierarchical clustering (average linkage) and similarities $\mathbf{M}_{0.3}$; (b) optimal cluster composition of time series derived from hierarchical clustering (average linkage) and similarities $\mathbf{M}_{0.5}$.

Interestingly, both the choices of the Matérn correlation function preserve the evidence of the different behaviour of some particular stations by keeping them in separate clusters, which is not visible in the purely spatial configuration (see e.g. stations 18 or 20).

5. Conclusions

This paper introduces the Spatial-CHC algorithm that incorporates spatial/ geographical information in the clustering of time series. The algorithm requires the construction of two matrices which lie in the space of correlation matrices. In particular, the information about the time series's pairwise dependence is extracted via van der Waerden's correlation coefficients, while the spatial information are interpreted by a correlation matrix exploiting the Matérn correlation function. The two matrices are then glued through a geodesic function depending on a mixing parameter $\alpha \in [0, 1]$ that preserves the geometric properties of correlation matrices. In particular, to build the geodesic, we rely on the geometric structure of the space of positive definite matrices and of related subspace of correlation matrices. The geodesic is such that it is a correlation matrix itself for each $\alpha \in [0, 1]$. A hierarchical clustering method is performed with a dissimilarity matrix constructed from the correlation matrix depending on α.

From a theoretical point of view, to the best of our knowledge, the computational aspects about the properties of the distance in the space of all correlation matrices Corr(n) have not been fully explained (see, e.g., Thanwerdas (2022, Theorem 6.6)), although the numerical simulations about their use are quite promising (see also (Marti et al., 2021a; You and Park, 2022)).

Different simulation studies and a real case study have been presented to illustrate the usefulness and effectiveness of the suggested clustering method for time series. In particular, the findings of the simulation studies suggest that the use of van der Waerden coefficient is in agreement with similar methods. As regards the interpretation of the spatial information, in each example we provide two different examples of Matérn correlation functions, namely the exponential and the Gaussian. Even if there are some differences, based also on the choice of the parameters, the optimal cluster composition of the time series derived from the hierarchical clustering with pure spatial similarities in the two cases have shown a similar behaviour in all the simulations. A reason for this may be that the spatial correlation matrix only provides a soft constraint to the whole procedure and, hence, different choices for the spatial model does not provide substantial changes in the whole procedure.

Moreover, the simulations clearly show the effect that the gradual introduction of the spatial constraint has to the clusters configuration. The application of the Spatial-CHC method to the real case study demonstrates how it may be used to pick groups that are both dependent and spatially close, improving the applicability of the cluster analysis results. As regards the choice of α,

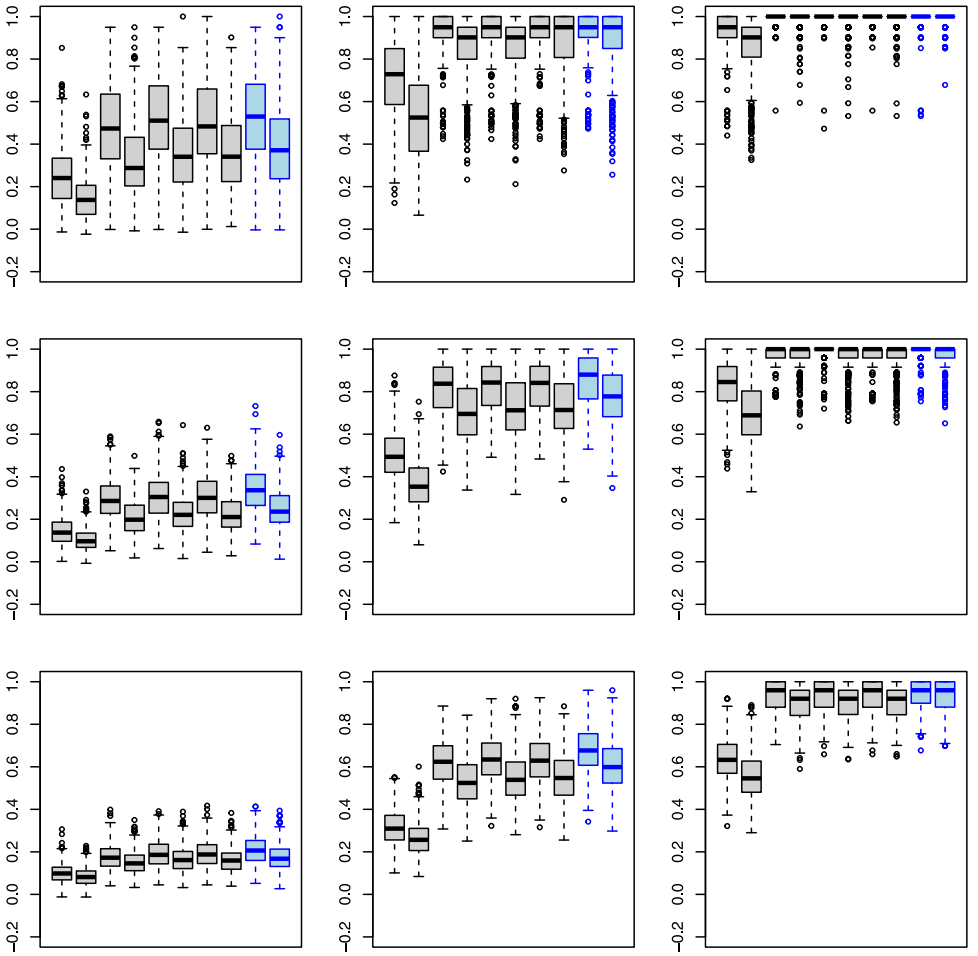


Fig. 14. Data simulated from an $n = 60$ dimensional random vector with K independent groups having a Clayton copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 50 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we provide both a theoretical insight based on the definition of the distance on Corr(n), and a practical observable criterion, based on the Dunn Index.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

A.B. acknowledges the support of Regione Puglia (Italy) 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). 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).

Both authors would like to thank Roberta Pappadà for useful discussions and Y. Thanwerdas for helpful clarifications.

Appendix

See Figs. 14–21.

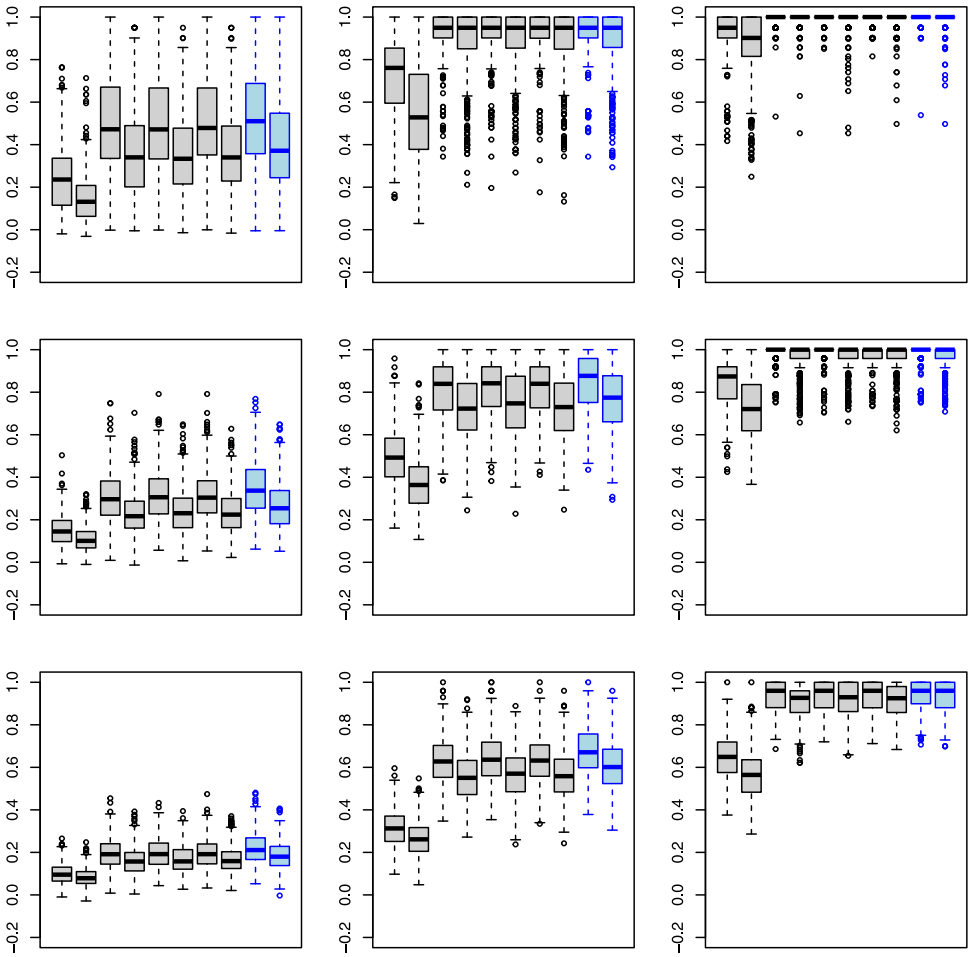


Fig. 15. Data simulated from an $n = 60$ dimensional random vector with K independent groups having a Gumbel copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 50 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

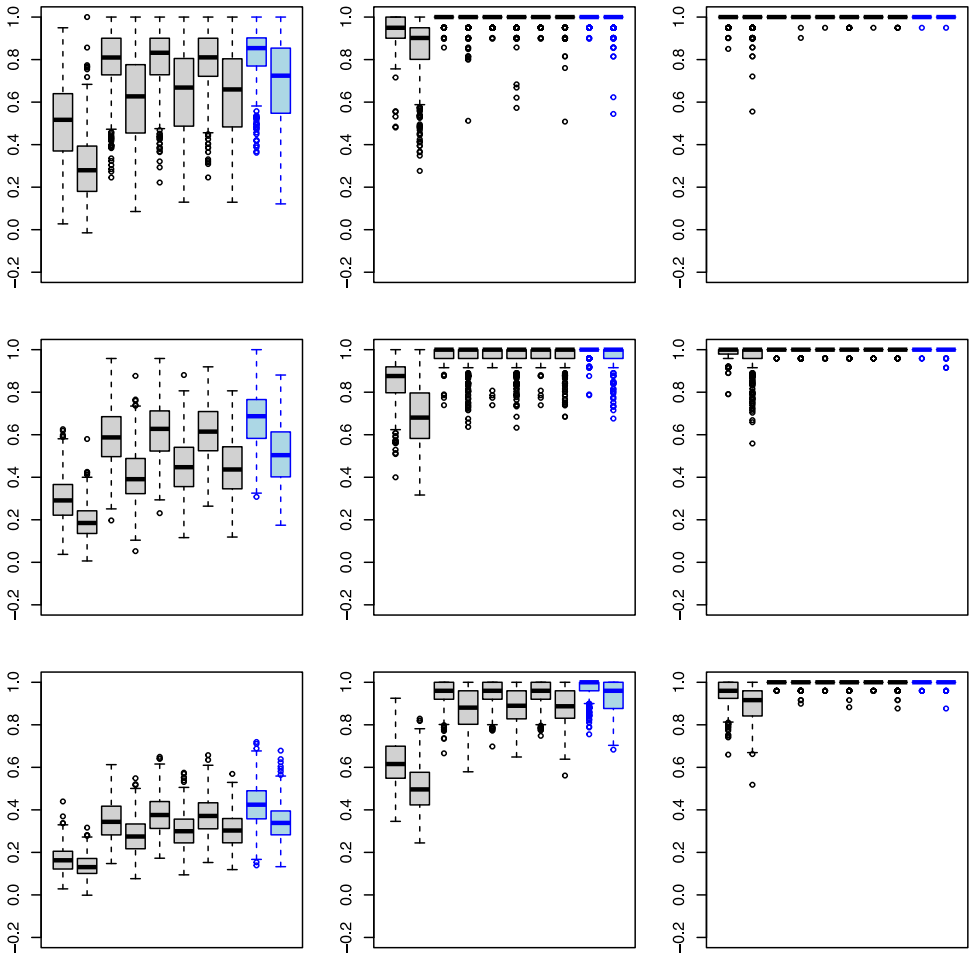


Fig. 16. Data simulated from an $n = 60$ dimensional random vector with K independent groups having a Clayton copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 100 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

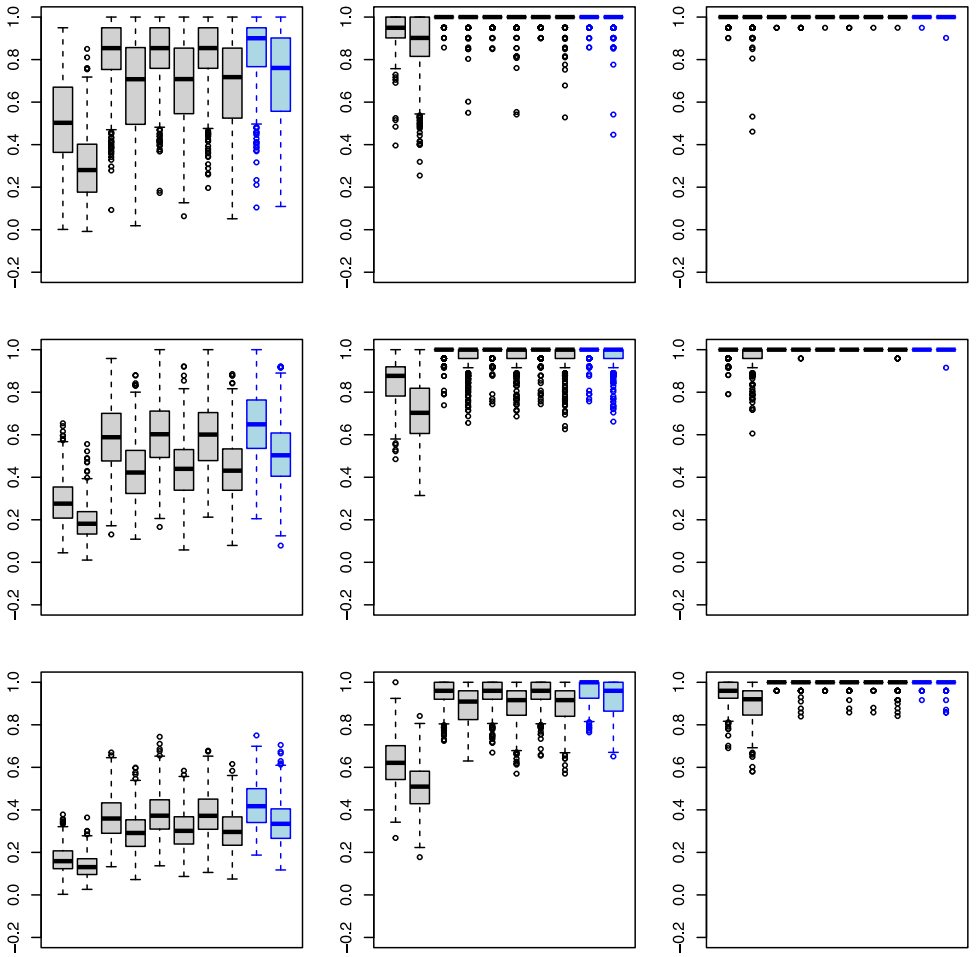


Fig. 17. Data simulated from an $n = 60$ dimensional random vector with K independent groups having a Gumbel copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 100 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

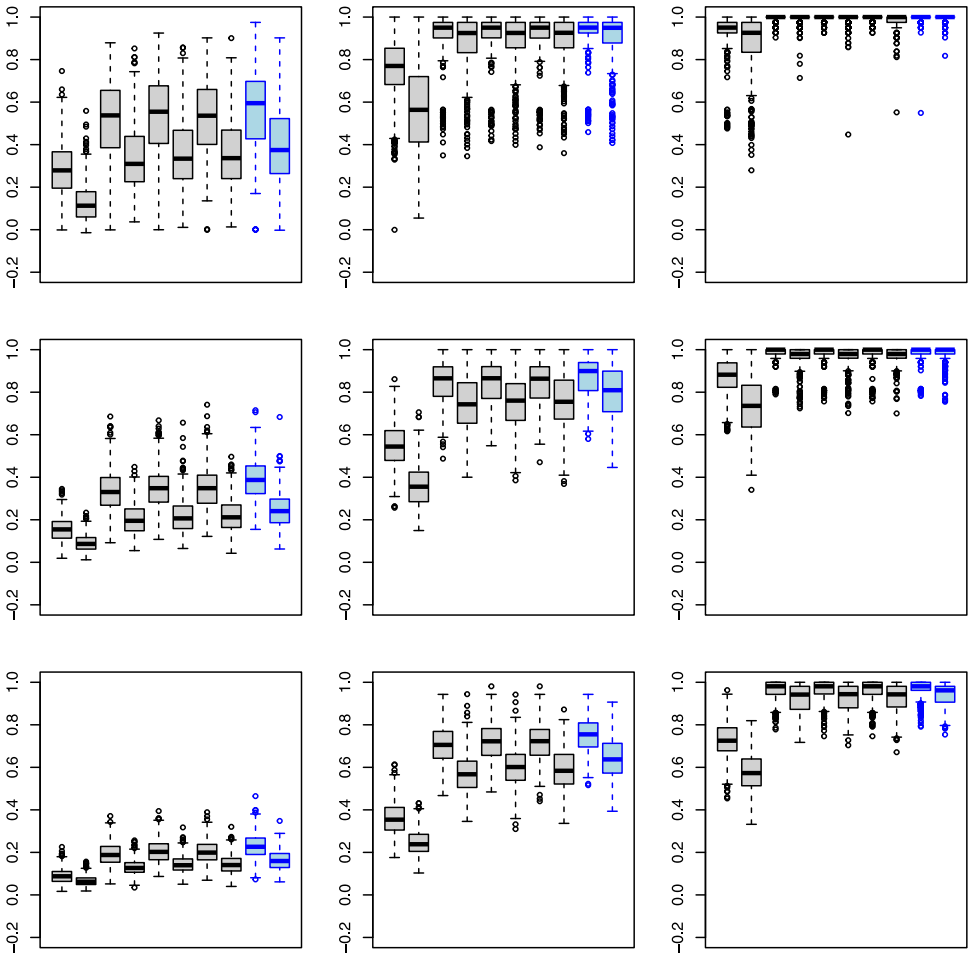


Fig. 18. Data simulated from an $n = 120$ dimensional random vector with K independent groups having a Clayton copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 50 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

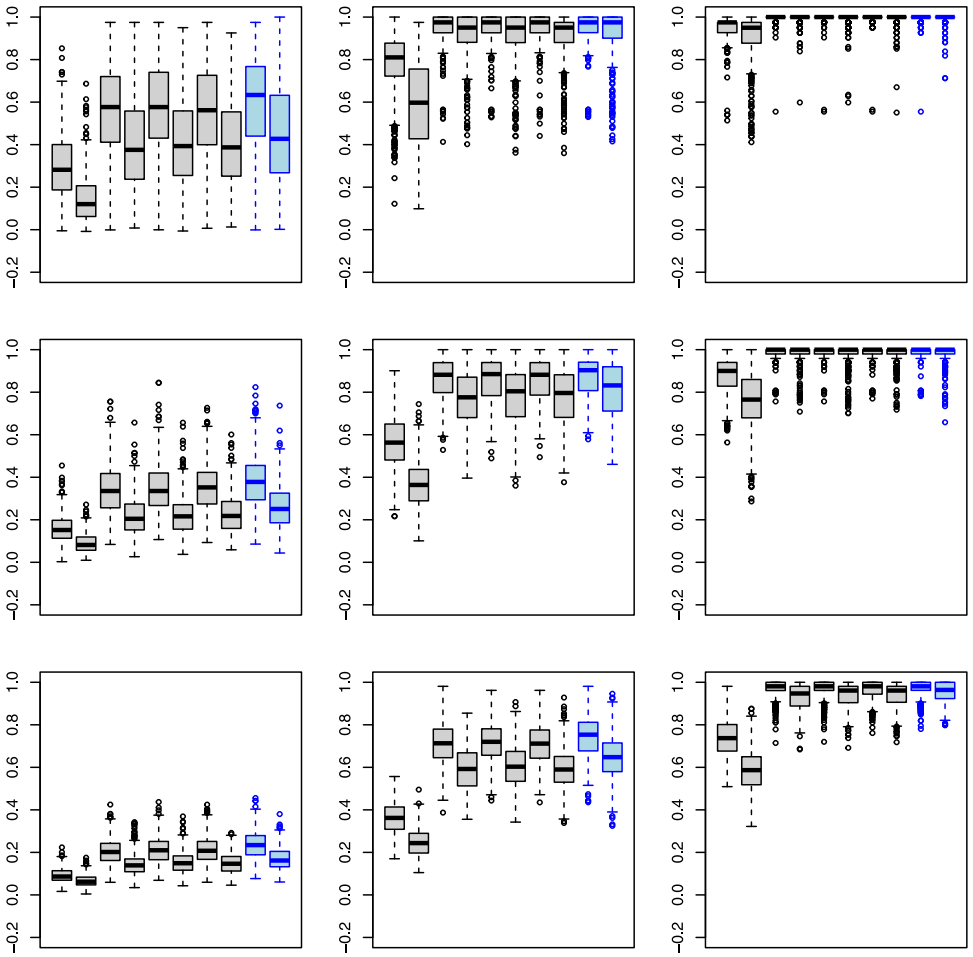


Fig. 19. Data simulated from an $n = 120$ dimensional random vector with K independent groups having a Gumbel copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (*i*) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (*ii*) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 50 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

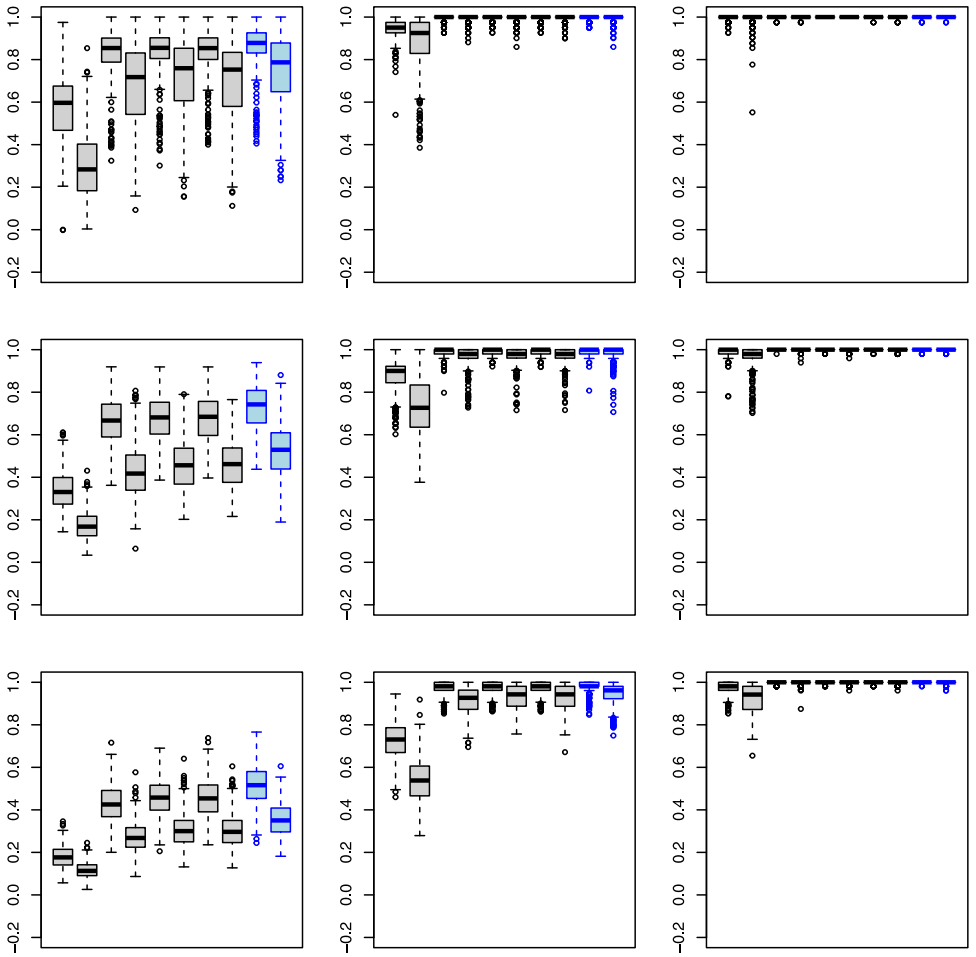


Fig. 20. Data simulated from an $n = 120$ dimensional random vector with K independent groups having a Clayton copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 100 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

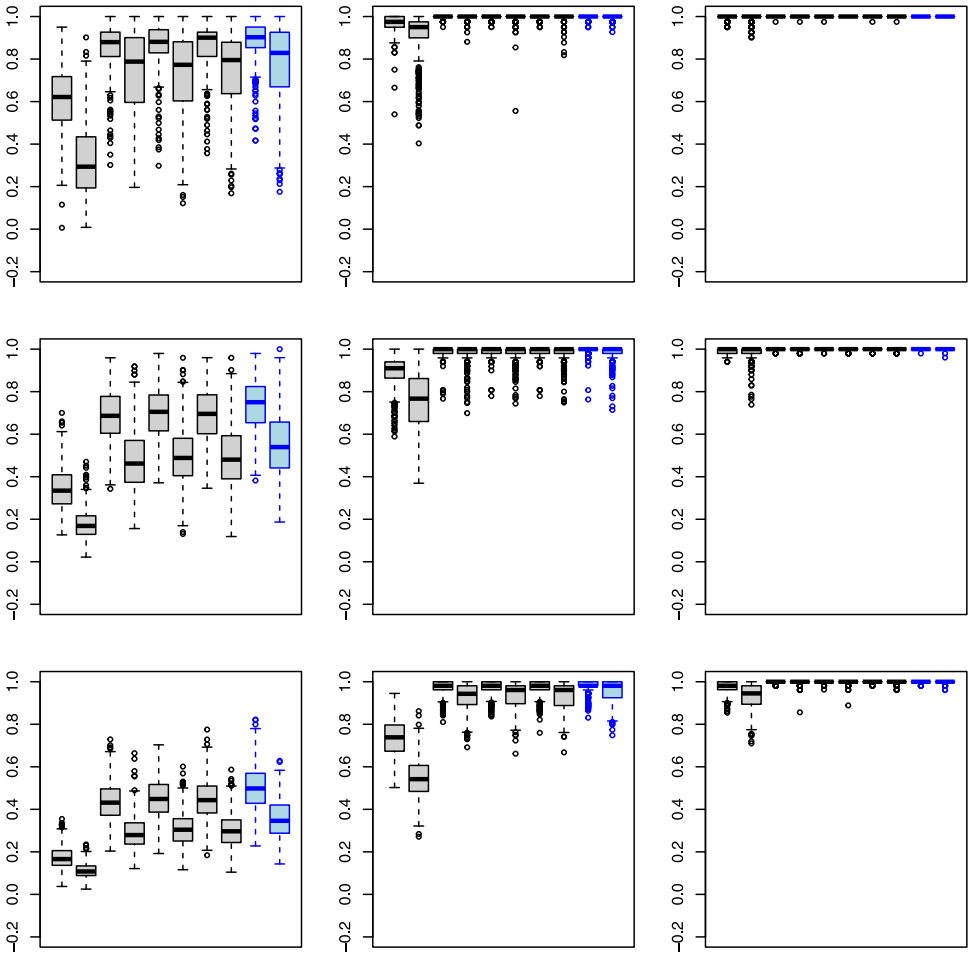


Fig. 21. Data simulated from an $n = 120$ dimensional random vector with K independent groups having a Gumbel copula with Kendall's τ equal to 0.1 (left), 0.2 (middle), and 0.3 (right) within each group. Each row corresponds to a number of clusters equal to $K = 3$ (up), $K = 6$ (middle) and $K = 12$ (down). Each boxplot represents the ARI (y-axis) with respect to the true partition by varying (i) the pairwise dissimilarity measure among Blomqvist's β, Spearmans' ϕ, Kendall's τ, Spearman's ρ, and van der Waerden's ζ (in blue) and (ii) linkage method among the average, and complete (maximum) one. In each blox plot the x-axis starting with the average linkage and β, continues with the complete linkage and β, and ends with the complete linkage and ζ. Sample size equal to 100 points (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. Dover Publications, New York.

Asquith, W.H., 2022. copBasic—General bivariate copula theory and many utility functions. R package version 2.1.9.

Barbero, S., Zaccagnino, S.M., Mariani, S., Lastoria, B., Braca, G., Bussettini, M., Casaioli, M., Marsico, L., Rotundo, R., Pavan, V., et al., 2017. Linee guida per il controllo di validità dei dati idro-meteorologici. (Guidelines for the quality check of hydrometeorological data). Available at Website https://www.certifico.com/component/attachments/download/5512.

Benevento, A., Durante, F., Pappadà, R., 2023. An approach to cluster time series extremes with spatial constraints. In: Chelli, F.M., Ciommi, M., Ingrassia, S., Mariani, F., Recchioni, M.C. (Eds.), Book of Short Papers SEAS IN 2023. Pearson, pp. 679–684.

Bhatia, R., 2009. Positive Definite Matrices. Princeton University Press.

Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J., 2018. ClustGeo: an R package for hierarchical clustering with spatial constraints. Comput. Statist. 33 (4), 1799–1822.

Chen, X., Fan, Y., 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econometrics 135 (1–2), 125–154.

Contreras, P., Murtagh, F., 2015. Hierarchical clustering. In: Hennig, C., Meila, M., Murtagh, F., Rocci, R. (Eds.), Handbook of Cluster Analysis. Chapman & Hall.

Côté, M.P., Genest, C., 2015. A copula-based risk aggregation model. Can. J. Stat. 43 (1), 60–81.

Czado, C., Schepsmeier, U., Min, A., 2012. Maximum likelihood estimation of mixed C-vines with application to exchange rates. Stat. Model. 12 (3), 229–255.

David, P., 2019. A Riemannian Quotient Structure for Correlation Matrices with Applications to Data Science (Ph.D. thesis). The Claremont Graduate University.

David, P., Gu, W., 2019. A Riemannian structure for correlation matrices. Oper. Matrices 13 (3), 607–627.

David, P., Gu, W., 2022. Anomaly detection of time series correlations via a novel Lie group structure. Stat 11 (1), e494.

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

De Luca, G., Zuccolotto, P., 2021. Hierarchical time series clustering on tail dependence with linkage based on a multivariate copula approach. Internat. J. Approx. Reason. 139, 88–103.

Devroye, L., Letac, G., 2015. Copulas with prescribed correlation matrix. In: Memoriam Marc Yor – Séminaire de Probabilités XLVII. Springer, pp. 585–601.
Di Lascio, F.M.L., Durante, F., Pappadà, R., 2017. Copula–based clustering methods. In: Úbeda Flores, M., de Amo, E., Durante, F., Fernández Sánchez, J. (Eds.), Copulas and Dependence Models with Applications. Springer International Publishing, pp. 49–67.
Di Lascio, F.M.L., Menapace, A., Pappadà, R., 2021. A spatial AMH copula-based dissimilarity measure to cluster variables in panel data. BEMPS - Bozen Economics & Management Paper Series BEMPS89, Faculty of Economics and Management at the Free University of Bozen, URL https://EconPapers.repec.org/RePEc:bzn:wpaper:bemps89.
Diggle, P.J., Ribeiro, P.J., 2007. Model-Based Geostatistics. Springer, New York.
Disegna, M., D'Urso, P., Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. Spat. Stat. 21 (part A), 209–225.
Dißmann, J., Brechmann, E.C., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Comput. Statist. Data Anal. 59, 52–69.
Distefano, V., Mameli, V., Poli, I., 2020. Identifying spatial patterns with the Bootstrap ClustGeo technique. Spat. Stat. 38, 100441.
Dunn, J.C., 1974. Well-separated clusters and optimal fuzzy partitions. J. Cybern. 4 (1), 95–104.
Durante, F., Pappadà, R., Torelli, N., 2015. Clustering of time series via non–parametric tail dependence estimation. Statist. Papers 56 (3), 701–721.
Durante, F., Sempi, C., 2016. Principles of Copula Theory. CRC Press, Boca Raton, FL.
D'Urso, P., De Giovanni, L., Federico, L., Vitale, V., 2023. Fuzzy clustering of spatial interval-valued data. Spat. Stat. 57, 100764.
D'Urso, P., Vitale, V., 2020. A robust hierarchical clustering for georeferenced data. Spat. Stat. 35, 100407.
Fouedjio, F., 2016. A hierarchical clustering method for multivariate geostatistical data. Spat. Stat. 18, 333–351.
Fouedjio, F., 2020. Clustering of multivariate geostatistical data. Wiley Interdiscip. Rev. Comput. Stat. 12 (5), e1510.
Fuchs, S., Di Lascio, F.M.L., Durante, F., 2021. Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Comput. Statist. Data Anal. 107201.
Galili, T., 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31 (22), 3718–3720.
Genest, C., Verret, F., 2005. Locally most powerful rank tests of independence for copula models. J. Nonparametr. Stat. 17 (5), 521–539.
Górecki, J., Hofert, M., Holená, M., 2017. Kendall's tau and agglomerative clustering for structure determination of hierarchical Archimedean copulas. Depend. Model. 5 (1), 75–87.
Górecki, J., Hofert, M., Okhrin, O., 2021. Outer power transformations of hierarchical Archimedean copulas: Construction, sampling and estimation. Comput. Statist. Data Anal. 155, 107109.
Guénard, G., Legendre, P., 2022. Hierarchical clustering with contiguity constraint in R. J. Stat. Softw. 103 (1), 1–26.
Hennig, C., Meila, M., Murtagh, F., Rocci, R., 2015. Handbook of Cluster Analysis. CRC Press.
Hofert, M., Koike, T., 2019. Compatibility and attainability of matrices of correlation-based measures of concordance. ASTIN Bull. 49 (3), 885–918.
Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2022. copula: Multivariate dependence with copulas. R package version 1.1-1.
Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classification 2, 193–218.
Huckemann, S., Hotz, T., Munk, A., 2010. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. Statist. Sinica 20, 1–58.
Joe, H., 2015. Dependence Modeling with Copulas. In: Monogr. stat. appl. probab., vol. 134, CRC Press, Boca Raton, FL.
Jondeau, E., Rockinger, M., 2006. The Copula–GARCH model of conditional dependencies: An international stock market application. J. Int. Money Fin. 25 (5), 827–853.
Koike, T., Hofert, M., 2020. Estimation and comparison of correlation-based measures of concordance. arXiv preprint:2006.13975.
Kubo, F., Ando, T., 1980. Means of positive linear operators. Math. Ann. 246, 205–224.
Le, T.H., 2021. An Analysis of Changing Dietary Trends and the Implications for Global Health (Ph.D. thesis). Bournemouth University (UK).
Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2022. Cluster: Cluster analysis basics and extensions.
Maharaj, E.A., D'Urso, P., Caiado, J., 2019. Time Series Clustering and Classification. CRC Press, Boca Raton, FL.
Mai, J.F., Scherer, M., 2017. Simulating Copulas. Stochastic Models, Sampling Algorithms and Applications, second ed. In: Ser. Quant. Finance, vol. 6, World Scientific, Hackensack, NJ.
Mantegna, R.N., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B 11 (1), 193–197.
Marti, G., Goubet, V., Nielsen, F., 2021a. cCorrGAN: Conditional correlation GAN for learning empirical conditional distributions in the elliptope. In: Geometric Science of Information: 5th International Conference, GSI 2021, Paris, France, July 21–23, 2021, Proceedings 5. Springer, pp. 613–620.
Marti, G., Nielsen, F., Bińkowski, M., Donnat, P., 2021b. A review of two decades of correlations, hierarchies, networks and clustering in financial markets. In: Nielsen, F. (Ed.), Progress in Information Geometry: Theory and Applications. Springer International Publishing, Cham, pp. 245–274.
McNeil, A.J., Nešlehová, J.G., Smith, A.D., 2022. On attainability of Kendall's tau matrices and concordance signatures. J. Multivariate Anal. 191, 105033.
Moakher, M., 2005. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. 26 (3), 735–747.
Murtagh, F., 1985. A Survey of Algorithms for Contiguity-constrained Clustering and Related Problems. Comput. J. 28 (1), 82–88.
Nelsen, R.B., 2006. An Introduction to Copulas, second ed. In: Springer Series in Statistics, Springer, New York, p. xiv+269.
Oliver, M.A., Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. Math. Geol. 21 (1), 15–35.
Palacios-Rodriguez, F., Bernardino, E. Di, Mailhot, M., 2023. Smooth copula-based generalized extreme value model and spatial interpolation for extreme rainfall in Central Eastern Canada. Environmetrics 34 (3), e2795.
Rémillard, B., 2017. Goodness-of-fit tests for copulas of multivariate time series. Econometrics 5 (1).
Riquelme, Á.I., 2002. Multivariate Simulation Using A Locally Varying Coregionalization Model (Ph.D. thesis). Queen's University at Kingston, URL https://qspace.library.queensu.ca/handle/1974/30443.
Riquelme, Á.I., Ortiz, J.M., 2023. A Riemannian tool for clustering of geo-spatial multivariate data. Math. Geosci. 1–21.
Romary, T., Ors, F., Rivoirard, J., Deraisme, J., 2015. Unsupervised classification of multivariate geostatistical data: Two algorithms. Comput. Geosci. 85, 96–103.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer Science & Business Media.
Thanwerdas, Y., 2022. Riemannian and Stratified Geometries on Covariance and Correlation Matrices (Ph.D. thesis). Université Côte d'Azur, URL https://hal.science/tel-03698752.
Thanwerdas, Y., Pennec, X., 2021. Geodesics and curvature of the quotient-affine metrics on full-rank correlation matrices. In: Nielsen, Frank, Barbaresco, Frédéric (Eds.), Geometric Science of Information. Springer International Publishing, Cham, pp. 93–102.
Thanwerdas, Y., Pennec, X., 2022. Theoretically and computationally convenient geometries on full-rank correlation matrices. SIAM J. Matrix Anal. Appl. 43 (4), 1851–1872.
Van Dongen, S., Enright, A.J., 2012. Metric distances derived from cosine similarity and Pearson and Spearman correlations. arXiv preprint arXiv:1208.3145.
Wang, B., Wang, R., Wang, Y., 2019. Compatible matrices of Spearman's rank correlation. Statist. Probab. Lett. 151, 67–72.
You, K., Park, H.J., 2022. Geometric learning of functional brain network on the correlation manifold. Sci. Rep. 12 (1), 17752.
Zuccolotto, P., De Luca, G., Metulini, R., Carpita, M., 2023. Modeling and clustering of traffic flows time series in a flood prone area. In: Cerchiello, P., Agosto, A., Osmetti, S., Spelta, A. (Eds.), Proceedings of the Statistics and Data Science Conference. Pavia University Press, Pavia, pp. 113–118.