

Spatiotemporal clustering of GHGs emissions in Europe: Exploring the role of spatial component[☆]

Caterina Morelli^{a,b}, Paolo Maranzano^{a,c}, Philipp Otto^d

^a Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Piazza dell'Ateneo Nuovo n.1, Milano, 20126, Italy

^b Department of Economic and Political Sciences, Università della Valle d'Aosta - Université de la Vallée d'Aoste, Strada Cappuccini n.2A, Aosta, 11100, Italy

^c Fondazione Eni Enrico Mattei, Corso Magenta n.63, Milano, 20123, Italy

^d School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow, G12 8QQ, Scotland, United Kingdom

ARTICLE INFO

Dataset link: https://github.com/PaoloMaranzano/CM_PM_PO_STClusterGHG.git

Keywords:

Hierarchical spatiotemporal clustering
Spatial overlapping
Clustering simulation
Greenhouse gases emissions
European regions NUTS-2

ABSTRACT

In this study, we propose a novel application of spatiotemporal clustering in the socio-economic and environmental sciences, with a particular focus on regionalized time series of greenhouse gases (GHGs) emissions from a range of economic sectors. Utilizing a hierarchical spatiotemporal clustering methodology, we analyze yearly time series of emissions by gases and sectors from 1990 to 2022 for European regions at the NUTS-2 level. While the clustering algorithm inherently incorporates spatial information based on geographical distance, the extent to which space contributes to the definition of groups still requires further exploration. To address this gap in the literature, we propose a novel indicator, namely the Joint Inertia, which quantifies the contribution of spatial distances when integrated with other features. Through a simulation experiment, we explore the relationship between the Joint Inertia and the relevance of geography in exploiting the groups structure under several configurations of spatial and features patterns, providing insights into the behavior and potential of the proposed indicator. The empirical findings demonstrate the relevance of the spatial component in identifying emission patterns and dynamics, and the results reveal significant heterogeneity across clusters in trends and dynamics by gases and sectors. This reflects the heterogeneous economic and industrial characteristics of European regions. The study highlights the importance of the spatial and temporal dimensions in understanding GHGs emissions, offering baseline insights for future spatiotemporal modeling and supporting more targeted and regionally informed environmental policies.

1. Introduction

The growing availability of spatiotemporal data presents both opportunities and challenges for researchers. In the environmental field, new technologies such as remote sensing, GPS devices, mobile sensors, and satellite imagery allow us to capture environmental phenomena at unprecedented spatial and temporal resolutions. These data enable detailed investigations into processes that evolve over both space and time. A large body of work has focused on developing advanced spatiotemporal models to predict pollutant

[☆] This article is part of a Special issue entitled: 'SPASTA_Metma: health and environment' published in Spatial Statistics.

* Corresponding author at: Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Piazza dell'Ateneo Nuovo n.1, Milano, 20126, Italy.

E-mail addresses: c.morelli12@campus.unimib.it (C. Morelli), paolo.maranzano@unimib.it (P. Maranzano), philipp.otto@glasgow.ac.uk (P. Otto).

<https://doi.org/10.1016/j.spasta.2026.100960>

Received 30 December 2024; Received in revised form 28 January 2026; Accepted 29 January 2026

Available online 12 February 2026

2211-6753/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

concentrations and to assess their dynamics in different areas (Calculli et al., 2015; Najafabadi et al., 2020; Taghavi-Shahri et al., 2020; Maranzano et al., 2023; Yu et al., 2024; Otto et al., 2024; Rowland, 2024). Other contributions evaluate the effects of specific events (Maranzano and Fassó, 2022; Maranzano and Pelagatti, 2024) or policies (Maranzano et al., 2020; Rodeschini et al., 2024; Pisoni et al., 2025) on air quality.

Despite these advances, a critical gap remains: most studies analyze pollutant concentrations, while very few investigate pollutant emissions (Marongiu et al., 2024; Mutascu, 2022; Marongiu et al., 2025; Banja et al., 2025). Concentrations are influenced not only by emissions but also by morphological and geographical factors such as elevation, land use, and natural barriers, which affect dispersion and accumulation. Focusing exclusively on concentrations may thus obscure the fundamental role of emissions themselves—the primary drivers of pollution and climate change. Studying emissions at their source is essential to better understand the mechanisms of environmental degradation and to design effective policies.

To encourage researchers to look at air pollution from a new and challenging perspective, spatiotemporal cluster analysis offers a powerful approach to start bridging this gap, enabling the identification of spatial and temporal patterns of GHGs emissions.

In this study, spatiotemporal clustering has to be intended as the grouping of spatial units (i.e., the European regions in the empirical application) while accounting for their temporal dynamics. Each unit is described by time series of several emission variables and by its fixed geographical location. We build a single, time-invariant partition of the units by minimizing a criterion that combines (i) dissimilarities between unit-level time series (i.e., the temporal/feature information) and (ii) spatial dissimilarities between geographical coordinates. Spatial proximity increases the chance that two regions are clustered together only insofar as it improves this combined objective; however, contiguity is not enforced.

Early works such as (Oliver and Webster, 1989) and Bourgault et al. (1992) proposed the first techniques for cluster analysis of units combining features with spatial information, emphasizing the need to group units that exhibit both similar time series and geographical proximity. Several methodologies for cluster analysis of spatiotemporal data are reported in the literature review by Ansari et al. (2020), who classify them according to a taxonomy based on data types. Among the most widely used methods are density-based algorithms, such as the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) introduced by Ester et al. (1996), together with its extensions (Tietbohl et al., 2008; Rocha et al., 2010; Georgoulas et al., 2013; Chen et al., 2015; Birant and Kut, 2007). These approaches are able to detect clusters of arbitrary shape and varying sizes without requiring the number of clusters to be fixed in advance. However, they are known to be highly sensitive to input parameters such as radius, minimum number of points, or threshold values. With specific regard to time series clustering using dissimilarity measures, Piccolo (1990) first introduced the concept by proposing a distance metric for autoregressive integrated moving-average (ARIMA) invertible models, defined as the Euclidean distance between their autoregressive expansions. Later, Pértega-Díaz and Vilar (2010) developed several non-parametric statistics to test the equality of the log-spectra of two stochastic processes, considering both stationary and non-stationary time series, ARMA processes, and non-linear models. More recently, Montero and Vilar (2014) described and compared a wide range of time series dissimilarity measures for clustering, including those based on raw data, extracted features, underlying parametric models, complexity levels, and forecast behaviors. They also implemented these measures in the R package TSclust. Other more recent methodologies explicitly incorporate spatial information. For instance, Zhou et al. (2019) extended the well known K-means algorithm to network data (like regionalized lattice data) by designing two different adjacency matrices based on the spectral decomposition of the graph information. Fouedjio (2016) developed an agglomerative hierarchical clustering approach for spatially contiguous clusters, relying on a dissimilarity matrix built from a non-parametric kernel estimator of the multivariate spatial dependence structure. Building on this, D'Urso and Vitale (2020) proposed a robust metric approach to neutralize the effect of possible outliers, exploiting an exponential transformation of a dissimilarity measure. More recently, Benevento and Durante (2024), within a hierarchical framework, suggested merging spatial and temporal information into correlation matrices that take into account the underlying geometric structure of correlation matrix space. In epidemiological applications, Jaya and Folmer (2021) propose a two-stage spatiotemporal clustering approach to identify clusters that vary across space and time, while (Jaya and Folmer, 2020) develop a Bayesian spatiotemporal random-effects model with spatiotemporal interactions and demonstrate heterogeneous temporal patterns across spatial units.

Within the same hierarchical clustering framework, which does not require assumptions on data distribution or cluster shapes, Chavent et al. (2018) proposed a spatial hierarchical clustering algorithm. This approach combines two dissimilarity matrices — one feature-based and one geographical — through a weighted scheme. Owing to its flexibility and interpretability, the method has gained popularity in subsequent applications (Jaya et al., 2019; Bucci et al., 2023; Mattera and Franses, 2023; Deb and Karmakar, 2023; Morelli et al., 2025; Zammarchi and Maranzano, 2024). More recently, Morelli et al. (2025) extended this approach by enabling the use of multiple dissimilarity matrices simultaneously. This extension is particularly valuable when analyzing time series of different variables, since it is possible to obtain one dissimilarity matrix for each variable (Montero and Vilar, 2014). The resulting algorithm produces a cluster partition in which the overall dissimilarity within each cluster is minimized. In other words, it allows researchers to identify groups of units that display similar temporal dynamics of the variables while also being geographically close, though not necessarily adjacent.

In this paper, we build on this framework and propose a novel measure, namely the Joint Inertia index, to quantify the contribution of spatial distances relative to non-spatial dissimilarities in the clustering process. Indeed, while the mixing weights determine how dissimilarity matrices are combined, they should not be read as the importance of space; instead, Joint Inertia measures the incremental explanatory power of spatial information beyond features alone. Our contributions are twofold: on the one side we propose a methodological improvement by introducing Joint Inertia as a tool to assess the role of space when combined with other dissimilarities and illustrating its behavior in theoretical scenarios and simulations; on the other side, we provide an empirical contribution by applying the spatio-temporal clustering and the Joint Inertia to the case of georeferenced time series of

greenhouse gases (GHGs) emissions for several types of gases recorded across the European regions at the sectoral level with the primary goal to reveal geographically-relevant spatio-temporal patterns, and evaluate how much the spatial information helps in identifying them.

The relevance of these findings extends to both policymakers and statisticians. For policymakers, the insights gained from this analysis are instrumental in formulating localized strategies to mitigate GHGs emissions, enabling more targeted and effective interventions. For statisticians in the environmental field, this study provides a baseline analysis that highlights critical factors to consider when modeling GHGs emissions data. Specifically, it underscores the importance of incorporating spatial and temporal dimensions and offers guidance on whether spatiotemporal models are necessary for accurate representation and prediction. Together, these contributions lay the groundwork for more robust and nuanced approaches to addressing pressing environmental challenges.

The rest of the paper is structured as follows: in Section 2, we describe the methodology proposed by Chavent et al. (2018) and Morelli et al. (2025) and we show the algorithm that summarizes the methodological framework of the paper, in Section 3, we provide the methodological contribution of the paper, in Section 4 we present the simulation experiment on the performance of the clustering algorithm and the measure to assess the role of the spatial component and in Section 5 we discuss the empirical findings, thus the clusters obtained from the algorithm and the contribution of the spatial distances.

2. Spatiotemporal hierarchical clustering

In this section, we briefly summarize the baseline methodology. First we discuss the hierarchical clustering method proposed by Chavent et al. (2018), in which a partition of clusters is obtained by considering the dissimilarities between units from two different matrices. Second, we describe the extension proposed by Morelli et al. (2025) to more than two dissimilarity matrices, which is proved particularly useful in spatiotemporal application and adds a novel algorithm for the identification of the weights for the spatial and non-spatial matrices. This builds the methodological background for our newly proposed measure, the Joint Inertia, which will be introduced in Section 3. Joint Inertia allow to directly assess the importance of the spatial component in the resulting clustering partition.

2.1. Baseline methodology: spatial hierarchical clustering

Let $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ and $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$ refer respectively to any distance matrix of variables and the spatial distances matrix considering a sample of n units. Also, let w_i with $i = 1, \dots, n$ be the weight of the i th cross-sectional unit to be clustered. When no prior information is available, the weight is commonly set to $w_i = 1/n$. In order to account for different measurement scales, it is necessary to scale D_0 and D_1 with respect to their maximum values so that the clustering distances across observations take values between 0 and 1.

In the spatial clustering of Chavent et al. (2018), namely the ClustGeo algorithm, feature- and space-based information are combined through a mixing parameter $\alpha \in [0, 1]$ that tunes the contribution of spatial dissimilarities: $\alpha = 0$ yields a purely feature-based solution, whereas $\alpha = 1$ yields a purely spatial one. For a partition $\mathcal{P}_K^\alpha = (C_1^\alpha, \dots, C_K^\alpha)$, the mixed within-cluster pseudo-inertia of C_k^α is

$$I_\alpha(C_k^\alpha) = (1 - \alpha) \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{0,ij}^2 + \alpha \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{1,ij}^2, \quad (1)$$

where $\mu_k = \sum_{i \in C_k^\alpha} w_i$. The total mixed within clusters pseudo-inertia of the partition is

$$W_\alpha(\mathcal{P}_K^\alpha) = \sum_{k=1}^K I_\alpha(C_k^\alpha). \quad (2)$$

The agglomerative step merges the two clusters A and B that minimize the increase

$$\delta_\alpha(A, B) = I_\alpha(A \cup B) - I_\alpha(A) - I_\alpha(B). \quad (3)$$

Following Chavent et al. (2018), this Ward-like procedure is implemented via the Lance–Williams recursion on an aggregation matrix rather than on the raw dissimilarities. Define, for $m \in \{0, 1\}$,

$$\delta_{m,ij} = \frac{w_i w_j}{w_i + w_j} d_{m,ij}^2, \quad \Delta_m = [\delta_{m,ij}]_{i,j}, \quad \Delta_\alpha = (1 - \alpha)\Delta_0 + \alpha\Delta_1, \quad (4)$$

so that for singletons $\{i\}, \{j\}$ we have $\delta_\alpha(\{i\}, \{j\}) = (1 - \alpha)\delta_{0,ij} + \alpha\delta_{1,ij} = \frac{w_i w_j}{w_i + w_j} [(1 - \alpha)d_{0,ij}^2 + \alpha d_{1,ij}^2]$. The Lance–Williams update applied to Δ_α yields $\delta_\alpha(A \cup B, C)$, and the hierarchy proceeds by repeatedly merging the pair with minimal $\delta_\alpha(A, B)$. This differs from running Ward directly on $(1 - \alpha)D_0 + \alpha D_1$ and ensures that the optimization targets W_α .

The partition minimizes a within-cluster criterion built from the weighted combination of inertia in D_0 and D_1 . Two units tend to merge when their combined pseudo-inertia is smaller than competing pairs. As α increases, clusters become more spatially contiguous; as α decreases, feature similarity dominates. Thus D_1 modulates — rather than enforces — spatial contiguity within the objective being minimized.

The main issue in the hierarchical clustering methodology concerns the choice of the mixing parameters α and the number of clusters K . [Chavent et al. \(2018\)](#) suggest setting a prior value for K and then finding α such that it is the best compromises between loss of socio-economic and loss of geographical homogeneity. In order to clarify their approach, we recall the notion of the proportion of the total inertia explained by partition $\mathcal{P}_K^\alpha$ in D_0 and D_1 for K clusters, which is computed as

$$Q_{D_0}(\mathcal{P}_K^\alpha) = 1 - \frac{W_{D_0}(\mathcal{P}_K^\alpha)}{W_{D_0}(\mathcal{P}_1)} \quad Q_{D_1}(\mathcal{P}_K^\alpha) = 1 - \frac{W_{D_1}(\mathcal{P}_K^\alpha)}{W_{D_1}(\mathcal{P}_1)} \quad (5)$$

Specifically, $Q_{D_0}(\mathcal{P}_K^\alpha)$ quantifies the proportion of inertia from features dissimilarities (i.e., $W_{D_0}(\mathcal{P}_1)$) explained by partition $\mathcal{P}_K^\alpha$, while $Q_{D_1}(\mathcal{P}_K^\alpha)$ quantifies the amount of geographical inertia (i.e., $W_{D_1}(\mathcal{P}_1)$) explained by partition $\mathcal{P}_K^\alpha$.

To account for potential scale issues in $Q_{D_0}(\mathcal{P}_K^\alpha)$ and $Q_{D_1}(\mathcal{P}_K^\alpha)$, are then normalized with respect to the baseline case of purely-geographical or purely-features clustering, that is, by computing the following ratios:

$$\tilde{Q}_{D_0}(\mathcal{P}_K^\alpha) = \frac{Q_{D_0}(\mathcal{P}_K^\alpha)}{Q_{D_0}(\mathcal{P}_K^0)} \quad \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha) = \frac{Q_{D_1}(\mathcal{P}_K^\alpha)}{Q_{D_1}(\mathcal{P}_K^1)}. \quad (6)$$

This relative formulation allows for a straightforward interpretation of the values. For a given K and a given mixing parameter α , the formulas are expressing the percentage of the explained proportion of inertia in features (or spatial) dissimilarity matrix obtained by using a mixture of features and geographical information to generate the partition $\mathcal{P}_K^\alpha$ (i.e., $Q_{D_0}(\mathcal{P}_K^\alpha)$ (or $Q_{D_1}(\mathcal{P}_K^\alpha)$)) with respect to the proportion of inertia that would be explained by only using features (or spatial) dissimilarities to generate the partition $\mathcal{P}_K^0$ (or $\mathcal{P}_K^1$) in K clusters (i.e., $Q_{D_0}(\mathcal{P}_K^0)$ (or $Q_{D_1}(\mathcal{P}_K^1)$)). For a deeper understanding and comprehensive discussion about the properties of these quantities, we refer the readers to Section 3 in [Chavent et al. \(2018\)](#).

Thus, they suggest to choose α considering a trade-off between the loss of socio-economic homogeneity, thus the normalized proportion of explained inertia from the feature dissimilarities, and greater geographical cohesion measured as the normalized proportion of explained inertia in the spatial dissimilarity.

2.2. Spatiotemporal hierarchical clustering

In our setting, the temporal dimension is incorporated through trajectory-based dissimilarities between regions. Indeed, for each variable p , we build a dissimilarity matrix $D_p = [d_{p,ij}]$ where $d_{p,ij} = \delta(x_{i,p,1990:2022}, x_{j,p,1990:2022})$ measures dissimilarity between the full time series of regions i and j . These time series dissimilarities are then combined with the spatial dissimilarity D_p within the Ward-like mixed pseudo-inertia criterion of Section 2.1, yielding a single, time-invariant partition of the regions. We stress that we aim at clustering spatial units and not the time points; also, spatial proximity influences merges only insofar as it improves the mixed objective, and contiguity is not enforced.

[Morelli et al. \(2025\)](#) introduced a generalized version of spatial hierarchical clustering that allows the use of more than two dissimilarity matrices, then allowing for a spatio-temporal formulation of the clustering task. Also, they identify the optimal combination of these matrices which maximizes the gain induced by using a mixed approach instead of a purely spatial and non-spatial approach in terms of proportion of explained inertia within the clusters without forcing the algorithm to include all the input matrices in the final combination. Such principle holds for both the spatial and non-spatial information.

The need to incorporate multiple dissimilarity matrices arose from our objective of analyzing a dataset consisting of time series for more than one variable and the geographic coordinates of individual units. Indeed, the approach proposed by [Chavent et al. \(2018\)](#) allows to combine the information from the spatial dissimilarity matrix and one features dissimilarity matrix containing, for example, the Euclidean distances of variables observed at a given moment or a matrix representing the distances between time series for a single variable ([Bucci et al., 2023](#); [Deb and Karmakar, 2023](#); [Mattera and Franses, 2023](#)). When dealing with multiple time series, it becomes necessary to compute a dissimilarity matrix for each variable observed over a given period. Therefore, the challenge lies in combining more than two matrices. Notice that the methodology is not limited to spatiotemporal data but can be applied to any scenario where high-dimensional data requires the combination of more than two dissimilarity matrices. Nevertheless, this approach has been developed specifically to address the need for identifying patterns and dynamics in phenomena where both temporal evolution and geographic location are critical.

In our setting, the temporal dimension enters the clustering exclusively through trajectory-based dissimilarities between regions. For each emission variable, we construct a dissimilarity matrix measuring differences between full regional time series, which is combined with a spatial dissimilarity matrix within a Ward-type mixed pseudo-inertia criterion. This yields a time-invariant partition of spatial units: we cluster regions, not time points, and spatial proximity influences merges only insofar as it improves the mixed objective, without enforcing contiguity. Moreover, we adopt the ClustGeo formulation of [Chavent et al. \(2018\)](#), based on linear combinations of pseudo-inertias, and follow the selection strategy of [Morelli et al. \(2025\)](#) to determine both the weighting vector and the number of clusters.

Let $D_p = [d_{p,ij}]_{p=1,\dots,P-1; i,j=1,\dots,n}$ being the dissimilarity matrices containing the distances across time series for P variables for n units. Multiple alternative approaches can be used to compute time series dissimilarities metrics; among others, we refer the reader to [Montero and Vilar \(2014\)](#) for a comprehensive review and comparison of the most recent methodologies. In our empirical application we adopt the Dynamic Time Warping (DTW) ([Sakoe and Chiba, 1978, 1971](#)), which quantifies the (dis)similarity between two time series by (non-linearly) aligning them in time and minimizing the sum of point-wise distances between matched elements subject to monotonicity and continuity constraints. Furthermore, as shown by [Berndt and Clifford \(1994\)](#), the DTW distance is equivalent to the Euclidean distance computed on aligned time series.

Let us assume that the first $P - 1$ matrices are associated with non-spatial features or time series and let $D_p = [d_{p,ij}]_{i,j=1,\dots,n}$ refer to the spatial (or geographical) dissimilarity matrix across the n units, which is typically computed as geodetic distances between coordinates. Moreover, we recall that dissimilarity matrices are normalized with respect to their maximum value. The main issue here is to find a suitable combination of the two clustering hyperparameters, namely the number of group K , and the weighting vector $\underline{\alpha} = [\alpha_p]_{p=1,2,\dots,P} = (\alpha_1, \alpha_2, \dots, \alpha_p, \dots, \alpha_P)$ on the $(P - 1)$ -simplex Δ^{P-1} , that is $\alpha_p > 0$ for all $p = 1, \dots, P$ and $\sum_{p=1}^P \alpha_p = 1$. In practice, this requires constructing a grid of candidate values for $\underline{\alpha}$. The grid can, in principle, be either regular or irregular, but in most applications a regular grid with constant step size $\Delta\alpha$ is preferable, since it guarantees a uniform coverage of the parameter space. Specifically, a regular grid is obtained by allowing each α_p to take values in the discrete set $\{0, \Delta\alpha, 2\Delta\alpha, \dots, 1\}$ and then retaining only the combinations that satisfy the summation constraint $\sum_{p=1}^P \alpha_p = 1$. In this way, the search for the optimal weighting vector is restricted to a finite number of feasible candidates, ensuring both computational tractability and systematic exploration of the parameter space. Let us define a range of number of clusters from $K = 1$ to $K = K_{max}$ and for each $K = 1, \dots, K_{max}$ consider all the possible combinations for a grid of $\underline{\alpha}$. For each combination of K and $\underline{\alpha}$ we consider the partition $\mathcal{P}_K^{\underline{\alpha}}$ obtained applying the Ward hierarchical algorithm to the combination of pseudo-inertia from each dissimilarity matrices according to the vector of weight $\underline{\alpha}$ thus considering $I(C_K^{\underline{\alpha}}) = \sum_{p=1}^P \alpha_p \sum_{i \in C_K} \sum_{j \in C_K} \frac{w_i w_j}{2 \sum_{i \in C_K} w_i} d_{p,ij}^2$.

Note that Morelli et al. (2025) apply Ward directly to the linear combination of dissimilarity matrices, $D(\underline{\alpha}) = \sum_{p=1}^P \alpha_p D_p$. In contrast, we optimize a linear combination of (pseudo) within-cluster inertias in the spirit of Chavent et al. (2018); this is algebraically more involved but aligned with the mathematical foundation underlying Joint Inertia. We nevertheless adopt the hyperparameter selection strategy of Morelli et al. (2025) for $\underline{\alpha}$ and K^* to retain comparability and clarity in the selection step.

Following the criterion defined in Morelli et al. (2025), for every $K = 1, 2, \dots, K_{max}$ we choose the optimal weighting vector $\underline{\alpha} = \underline{\alpha}^*$ which maximizes the weighted average of the explained inertia¹ induced by the partition, given by

$$\bar{Q}(\mathcal{P}_K^{\underline{\alpha}}) = 1 - \frac{\sum_{p=1}^P W_{D_p}(\mathcal{P}_K^{\underline{\alpha}})}{\sum_{p=1}^P W_{D_p}(\mathcal{P}_1^{\underline{\alpha}})}. \quad (7)$$

Its value ranges between 0 and 1, reaching its maximum when partition $\mathcal{P}_K^{\underline{\alpha}}$ allows us to explain the total dissimilarity in each matrix. Notice that the maximization of Eq. (7) is consistent with the underlying idea of Ward's criterion according to which the clustering partition is obtained by minimizing the total within cluster pseudo inertia (Eq. (1)). Indeed, Eq. (7) is maximal when the numerator of the second component, which is the sum of matrix-specific total mixed pseudo-inertia, is minimum. The optimization step conditioning on K is iterated across a user-defined range of number of clusters $K = 1, 2, \dots, K_{max}$ evaluated at the corresponding optimal $\underline{\alpha}_K^*$. Finally, the optimal number of clusters $K^* | \underline{\alpha}_K^*$ is determined according to a user-defined set of clustering criteria, such as the increments in the weighted average proportion of explained inertia induced by the increase in the number of clusters and the Silhouette index (Morelli et al., 2025), the Dunn's index, the C-index, the Calinski-Harabasz's index, and the McClain-Rao's index (Zammarchi and Maranzano, 2024).

In this paper, we consider only the increments in the weighted average proportion of explained inertia as K increases, which can be interpreted as the marginal gain in explained variability attributable to adding one more cluster. This “elbow”-style criterion is commonly used in hierarchical settings and is directly aligned with the mixed-inertia objective optimized by the algorithm (Chavent et al., 2018). By contrast, homogeneity-based indices, such as the silhouette, may be less informative in our context due to the high dimensionality of the time series dissimilarities (multiple variables over a long horizon), which typically yields uniformly modest values and weak contrast across competing partitions. Nevertheless, to verify the empirical validity of the clustering results under different scenarios (i.e., spatiotemporal, temporal-only, and spatial-only), a set of internal validation indices — including the Average Silhouette, the Calinski–Harabasz index, the Within/Between ratio, and Pearson's Gamma — will be employed.

In Algorithm 1 we summarize the spatiotemporal hierarchical clustering methodology.

3. The role of the spatial component in hierarchical clustering

This section presents the methodological contribution of the paper. We provide a detailed reasoning on how the spatial component influences cluster formation and propose a measure to quantify its role. The formula of the indicator is introduced and its interpretation discussed. For additional technical and mathematical details, the reader is referred to Appendix A. Previously, we mentioned that a key advantage of the multi-matrix algorithm is that the choice of the weighting vector $\underline{\alpha}$ does not necessarily require the inclusion of the spatial component but rather incorporates it only when it contributes meaningfully to explaining the overall dissimilarity in the data. However, it is important to note that the weight assigned to the spatial component cannot be interpreted as a direct measure of its relevance. While it may be intuitive to think that as the spatial weight α_p increases, the explained inertia (and thus the proportion and normalization) also increases, it remains unclear how these quantities grow relative to each other. Indeed, if the increase in inertia were proportional to the weights of the matrices, the choice of parameter α considering the case of two dissimilarity matrices, according to Chavent et al. (2018) would always result in $\alpha = 0.5$, and the optimal outcome for (Morelli et al., 2025) could be any value for α .

¹ We refer the reader to the detailed discussion in Morelli et al. (2025) about the analytics and the interpretation of the weighted average of explained mixed pseudo inertia, in particular to its link with the gain in using a mixed approach compared to the purely non-spatial or purely spatial cases.

Algorithm 1 Hierarchical Spatiotemporal Clustering

Define as $D_p = [d_{p,ij}]_{p=1,\dots,P-1; i,j=1,\dots,n}$ the feature dissimilarity matrices

Define as $D_P = [d_{P,ij}]_{i,j=1,\dots,n}$ the spatial dissimilarity matrices

Define the weight of each unit w_i with $i = 1, \dots, n$, with no prior information $w_i = 1/n$

Set as K_{max} the maximum number of clusters

Set as $\Delta\alpha$ the increment of α_p

Compute $\Delta_p = [\delta_{p,ij}]_{p=1,\dots,P; i,j=1,\dots,n}$ as $\delta_{p,ij} = \frac{w_i w_j}{w_i + w_j} d_{p,ij}^2$

for $K = 1, 2, \dots, K_{max}$ **do**

for $\alpha_p (p = 1, \dots, P) \in [0, 1]$ by $\Delta\alpha$ such that $\sum_{p=1}^P \alpha_p = 1$ **do**

Compute the aggregation matrix $\Delta(\underline{\alpha}) = \sum_{p=1}^P \alpha_p \Delta_p$

Compute the partition $\mathcal{P}_K^{\alpha_p}$ in K clusters according to Ward hierarchical algorithm on the combined matrix $\Delta(\underline{\alpha})$, which correspond to apply the Ward algorithm on the combination of inertia

Compute the weighted average explained inertia $\bar{Q}(\mathcal{P}_K^{\alpha_p})$

end for

Select the best $\underline{\alpha}_K^*$ for each K such that $\underline{\alpha}_K^* = \operatorname{argmax}_{\alpha_p (p=1,\dots,P)} \bar{Q}(\mathcal{P}_K^{\alpha_p})$

end for

Choose K^* (evaluated at the corresponding $\underline{\alpha}^*$) according to the first difference in the weighted average proportion of explained pseudo inertia.

To illustrate this unclear behavior, we provide three simulated numerical examples (or scenarios) in which the spatial component is either (a) completely irrelevant, (b) relevant but not related to the phenomenon, or (3) relevant and useful in describing the phenomenon.² Without losing generalities, let us consider the case with only two dissimilarity matrices, the first containing the spatial distances and the second collecting the Euclidean distances across observations for a single feature. In Fig. 1, we provide a graphical representation of the three above scenarios.

In the first case, the clusters present high dissimilarities related to the features, but they perfectly overlap with respect to the spatial coordinates. As the spatial component is not playing any role in describing the phenomenon and is not useful in maximizing the explained inertia, it will be not included in the final optimal combination of weights. In the second scenario, the partition obtained using only the features dissimilarity matrix is completely different from the one obtained using only the spatial dissimilarities. Here, the spatial component leads to the identification of groups that are well separated in space but are not related to the actual pattern of the phenomenon under study. Therefore, we expect the algorithm to return as optimal mixing parameter either $\alpha = 0$ or $\alpha = 1$, depending on which matrix explains a greater portion of the dissimilarity across clusters. In the third scenario, the cluster partition obtained from observing only the features is identical to the one obtained using only the spatial component. Thus, the geographical distribution is strongly related to the phenomenon of interest and will play a crucial role in building the clusters. In this latter case, the spatial component's role is clearly significant, yet any value of α would suffice to maximize the weighted average explained inertia because if the clustering algorithm on D_0 leads to the same partition obtained using only D_1 , any combination of D_0 and D_1 will lead to the same results.

The above simulated example shows that the weight of the spatial component, represented by α , cannot be interpreted as a measure of the actual importance of geographical information for identifying the groups of observations that characterize the patterns of the examined variables. Indeed, it is only the solution obtained from a maximization task. On the one hand, when the spatial component is included in the combined dissimilarity matrix D , thus $1 - \alpha > 0$, or $\alpha_p > 0$ in multidimensional case, this indicates that spatial information is at least partially relevant in explaining the phenomenon. On the other hand, when the spatial information became more and more related to the dissimilarities across features, that is, D_0 and D_1 lead to similar clustering partitions, it is no longer necessary to include the spatial component because feature dissimilarities are able to capture a high proportion the inertia in the spatial dissimilarities and vice versa.

To properly understand the role of the spatial dissimilarity matrix in relation to the phenomenon under study, it is more appropriate to examine how the explained inertia in the resulting clusters changes when combining the two dissimilarity matrices with respect to the cases in which the features or spatial distances are used independently. The normalized proportion of explained inertia defined by Chavent et al. (2018) (Eq. (6)) is particularly useful for this purpose because it provides the percentage of explained variability in the clustering partition obtained from the combined dissimilarity matrix with respect to the case in which the partition is obtained from one dissimilarity matrix only.

When we choose to incorporate the spatial component, we effectively renounce to a proportion of the inertia explained by the features to gain a certain amount of inertia explained within the spatial component. To clarify the mechanism, we revisit the framework of Chavent et al. (2018) describing how the proportion of explained inertia changes when the inertia from two

² Notice that the statement “the spatial component is relevant with respect to the phenomenon” refers to whether the spatial/geographical distribution of the units to be clustered is related to the phenomenon under study and thus can be useful for the purpose of adequately constructing clusters (“it is relevant for the clustering”) or is unrelated to the underlying phenomenon and thus does not contribute in a useful way to the construction of clusters (“it is not relevant for the clustering”).

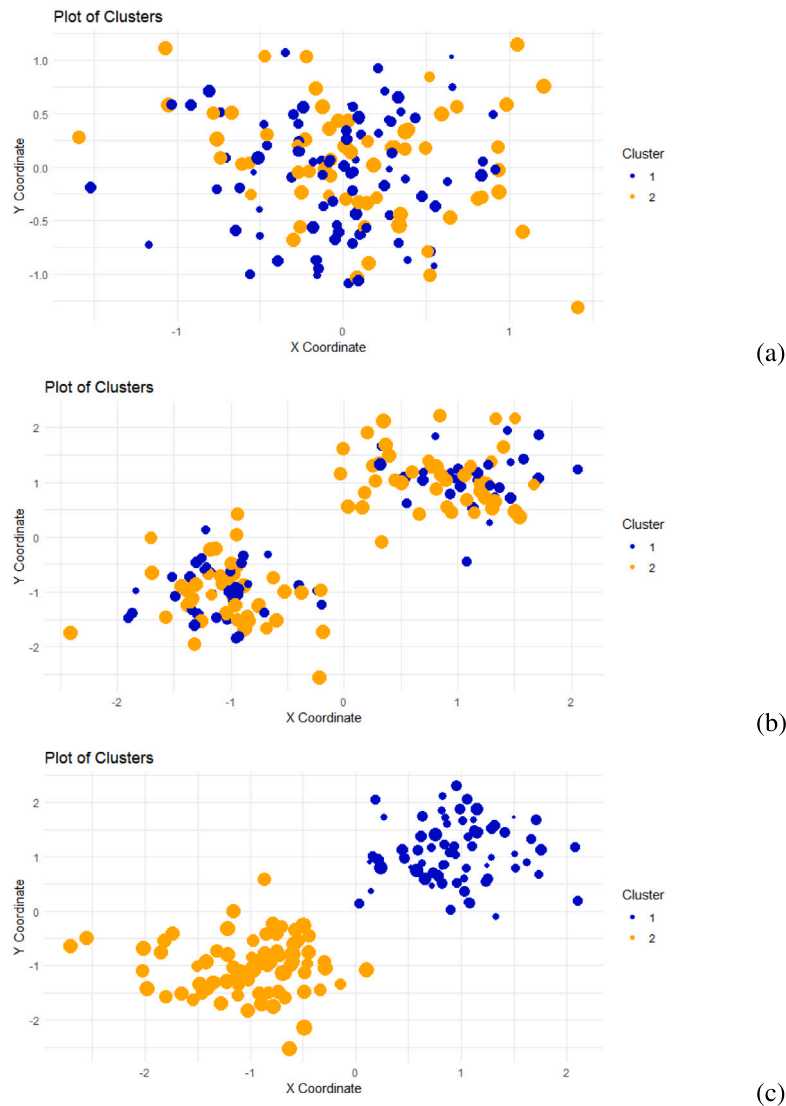


Fig. 1. Example of clusters defined under different spatial conditions. In panel (a), the spatial component is not relevant: observations from different clusters present the same spatial distribution, and clusters are perfectly overlapping in space. In panel (b), the spatial component is relevant but not related to the phenomenon: it is possible to identify groups according to spatial distances, but this is not useful in describing the phenomenon because, in each spatial cluster, we can find observations with different features. In panel (c), the spatial component is relevant and related to the phenomenon: we can identify the same clustering partition either using spatial dissimilarities or feature dissimilarities.

dissimilarity matrices is combined. Let $Q_{D_0}(\mathcal{P}_K^\alpha)$ and $Q_{D_1}(\mathcal{P}_K^\alpha)$ (see (5)) denote the proportion of explained inertia from D_0 and D_1 , respectively, that are captured by the partition $\mathcal{P}_K^\alpha$. These quantities are then normalized relative to the proportions obtained when clustering with D_0 alone and with D_1 alone, thus we obtain the normalized proportion of explained inertia $\bar{Q}_{D_0}(\mathcal{P}_K^\alpha)$ and $\bar{Q}_{D_1}(\mathcal{P}_K^\alpha)$ in (6). If we consider the limit case in which $\alpha = 0$ and thus $\bar{Q}_{D_0}(\mathcal{P}_K^\alpha) = 1$, it is intuitive that the proportion of explained inertia in D_1 will reach its minimum value with respect to all other possible values of α . When combining them using $0 < \alpha < 1$ the value of the proportion of explained inertia in D_0 will decrease (or remain constant) up to its minimum, while the proportion of explained inertia in D_1 will increase (or remain constant), up to one. Both $Q_{D_0}(\mathcal{P}_K^\alpha)$ and $Q_{D_1}(\mathcal{P}_K^\alpha)$ are non strictly monotonic with respect to α , as a consequence $\bar{Q}_{D_0}(\mathcal{P}_K^\alpha)$ and $\bar{Q}_{D_1}(\mathcal{P}_K^\alpha)$ are not strictly monotonic as well. However, the increase or decrease in the explained inertia with respect to α is non-constant and does not depend on α itself. Indeed including D_1 using a certain α may imply a big or small loss in the proportion of explained inertia in D_0 or it can remain unchanged and, at the same time, it may imply a high increase or zero increase in the proportion of explained inertia in D_1 . Moreover, the magnitude of these variations are not related to each other; in other words a big loss (or zero loss) in $Q_{D_0}(\mathcal{P}_K^\alpha)$ is not necessary linked to big or small increase in $Q_{D_1}(\mathcal{P}_K^\alpha)$. Overall, when we take into account an additional dissimilarity matrix in a clustering partition, we are willing to lose some proportion of

explained inertia from the initial matrix, hoping to gain some from the matrix we add. If the loss is smaller than the gain — that is, the decrease in one component is less than the increase in the other — then combining the two matrices is preferable.

From the above intuition, it follows that the actual contribution of the geographical information can be well approximated by the difference between the normalized proportion of inertia explained by the spatial component within the mixed cluster partition and the reduction in the normalized proportion of inertia explained by the features that results from including the spatial component. Specifically, in the case where there are only two matrices D_0 and D_1 , we quantify the importance of spatial features through a novel indicator, namely the Joint Inertia of partition P_K^α , which is computed as

$$JI(P_K^\alpha) = \bar{Q}_{D_1}(P_K^\alpha) - (1 - \bar{Q}_{D_0}(P_K^\alpha)) = \frac{Q_{D_1}(P_K^\alpha)}{Q_{D_1}(P_K^1)} - \left(1 - \frac{Q_{D_0}(P_K^\alpha)}{Q_{D_0}(P_K^0)}\right). \quad (8)$$

This criterion is referred to as Joint Inertia because it captures the overall gain in explained inertia that arises when the spatial and feature dissimilarity matrices are jointly incorporated in the clustering process. Rather than reflecting the contribution of each matrix in isolation, the indicator quantifies how much additional explanatory power is achieved when both sources of information are combined. In this sense, it measures the synergy between spatial and non-spatial components, emphasizing that their joint use can uncover patterns that would remain hidden if either matrix were considered separately.

The generalization to the case of multiple dissimilarity matrices, one of which relates to the spatial component, is straightforward. Recalling the notation previously introduced, let us consider the generic partition induced by a given combination of mixing parameters (i.e., P_K^α) and let $D_{-P} = [d_{p,ij}]_{p=1,\dots,P-1;i,j=1,\dots,n}$ be the dissimilarity matrices containing the distances across the $P-1$ time series for n units and let $D_P = [d_{p,ij}]_{i,j=1,\dots,n}$ be the spatial dissimilarity matrix across the n units, which may be computed as geodetic distances between coordinates. Let $\underline{\alpha} = [\alpha_p]_{p=1,2,\dots,P} \in \Delta^{P-1}$ be the vector of mixing weights used to compute the linear combination of the mixed within pseudo inertia from each dissimilarity matrix and let P_K^α be the resulting clustering partition. Moreover, let us define $\underline{\alpha}' = [\alpha'_p]_{p=1,2,\dots,P-1} \in \Delta^{P-2}$ being the optimal vector of weights resulting from the combination of the pseudo inertia from the non-spatial dissimilarity matrices (i.e. excluding matrix P), leading to partition $P_K^{\alpha'}$. Similarly, define the partition obtained by using only the spatial information (i.e., by fixing $\alpha_P = 1$) as $P_K^{\alpha_P=1}$. If we compute the optimal weighting vector obtained by the combination of the inertia from non-spatial dissimilarity matrices (i.e., $\alpha'_P = 0$), the resulting partition will explain as much as possible the dissimilarities in non-spatial matrices, indeed the vector of weights obtained is the one that maximizes the weighted average explained inertia thus being $\bar{Q}_{D_{-P}}(P_K^{\alpha'|\alpha_P=0})$ greater than any other potential $\bar{Q}_{D_{-P}}(P_K)$ evaluated at $\underline{\alpha}|\alpha_P = 0$.

Considering a generic combination of $\underline{\alpha}$ and K , the multi-matrix Joint Inertia for the spatial dissimilarity matrix D_P can be computed as

$$JI_P(P_K^\alpha) = \bar{Q}_{D_P}(P_K^\alpha) - (1 - \bar{Q}_{D_{-P}}(P_K^\alpha)) = \frac{Q_{D_P}(P_K^\alpha)}{Q_{D_P}(P_K^{\alpha_P=1})} - \left(1 - \frac{\bar{Q}_{D_{-P}}(P_K^\alpha)}{\bar{Q}_{D_{-P}}(P_K^{\alpha'|\alpha_P=0})}\right) \quad (9)$$

where $\bar{Q}_{D_P}(P_K^\alpha)$ is the normalized proportion of explained inertia with respect to the purely-spatial setting and $\bar{Q}_{D_{-P}}(P_K^\alpha)$ is the normalized proportion of explained inertia with respect to the optimal non-spatial setting. Specifically it corresponds to the ratio between the weighted average proportion of inertia in the non-spatial setting, when considering the partition obtained combining the inertia from all the dissimilarity matrices $\bar{Q}_{D_{-P}}(P_K^\alpha)$ and when considering the optimal partition from the non spatial dissimilarities $\bar{Q}_{D_{-P}}(P_K^{\alpha'|\alpha_P=0})$.

As the normalized proportion of explained inertia for every dissimilarity matrix varies between 0 and 1, also the Joint Inertia $JI_P(P_K^\alpha)$ will range from 0 to 1 (i.e., $0 \leq JI_P(P_K^\alpha) \leq 1$), being close to zero when the loss in the normalized proportion of explained inertia from the features dissimilarity due to the inclusion of the spatial component is similar to the normalized proportion of explained inertia in the spatial dissimilarity in the combined partition. Conversely, it will be close to 1 when the clustering partition obtained combining features and spatial dissimilarities allows to explain a high normalized proportion of inertia, both in spatial and non-spatial dissimilarity matrices.

We stress that the Joint Inertia can be utilized to evaluate the contribution of any dissimilarity matrix relative to all other dissimilarity matrices. In this study the Joint Inertia is specifically introduced to assess the role of the spatial component, addressing our initial research objectives. However, in other contexts, the Joint Inertia of any dissimilarity matrix can be computed to gain a comprehensive understanding of which variables exhibit shared explanatory power in terms of their dissimilarities. Further technical and mathematical details about the Joint Inertia, both in the case of two matrices and when considering more than two matrices, are discussed in [Appendix A](#).

Considering the three simulated scenarios described above, in [Table 1](#) we provide the values of the normalized proportion of explained inertia in D_0 and D_1 and the Joint Inertia obtained for every case represented in [Fig. 1](#). When the spatial component is not related to the clustering partition obtained by features (e.g., scenario b), we obtain α being either 0 or 1. Thus, the normalized proportion of explained inertia in the features dissimilarity matrix (or in the spatial dissimilarity matrix) will be equal to 1, while the normalized proportion of explained inertia in the spatial dissimilarity matrix (in the feature dissimilarity matrix) will be close to zero but strictly positive. The Joint Inertia index will be close to zero in cases (a) and (b), which is coherent with the fact that the spatial component plays no role in identifying the actual clustering patterns (conversely, well identified by the features space), while it is equal to 1 in case (c) because the spatial information is strongly related to the underlying phenomenon and exactly

Table 1

Choice of α , normalized proportion of explained inertia and resulting Joint Inertia in different examples of clusters under different spatial conditions. Cases are represented in Fig. 1.

Case (a)	Case (b)		Case (c)
Spatial component is not relevant	Spatial component is relevant but not related to the features		Spatial component is helpful
$\alpha = 0$	if $\alpha = 0$	if $\alpha = 1$	$0 < \alpha < 1$
$\bar{Q}_{D_1}(P_K^\alpha) \approx 0$	$\bar{Q}_{D_1}(P_K^\alpha) \approx 0$	$\bar{Q}_{D_1}(P_K^\alpha) = 1$	$\bar{Q}_{D_1}(P_K^\alpha) = 1$
$\bar{Q}_{D_0}(P_K^\alpha) = 1$	$\bar{Q}_{D_0}(P_K^\alpha) = 1$	$\bar{Q}_{D_0}(P_K^\alpha) \approx 0$	$\bar{Q}_{D_0}(P_K^\alpha) = 1$
$JI \approx 0$	$JI \approx 0$	$JI \approx 0$	$JI = 1$

describes the groups structure.³ When $0 < \alpha < 1$, the dissimilarity matrices are combined because their combination allows us to capture an overall higher proportion of the explained inertia, thus we expect to obtain a higher value of the presented indicator. In the following Section 4 we discuss the connection between the Join Inertia and the spatial information embedded in the data. In particular, we focus on how the index can be used to assess the actual relevance of the geographical dissimilarity in computing the clustering partitions.

4. Simulation experiment

In this section, we present a simulation experiment that compares two rules for selecting the mixing parameter α — the trade-off criterion of Chavent et al. (2018) and the weighted-average explained inertia maximization of Morelli et al. (2025) — while using the same Ward-like hierarchical procedure throughout. We assess their performance under different scenarios for the true contribution of geographical information. In particular, Chavent et al. (2018) propose a practical (though not globally optimal) solution: for a given K , explore $\alpha \in [0, 1]$ and choose the value that best balances the loss of socio-economic and geographical homogeneity, using a plot of the proportions of explained inertia in D_0 and D_1 over the α -grid. To automate this procedure, one can select the α for which these two normalized proportions are as close as possible, i.e., the difference between them is minimized. Morelli et al. (2025) suggest to choose α such that it maximizes the overall explained inertia quantified through the weighted average explained inertia in Eq. (7). We argue that the latter criterion is preferable because selecting α to maximize the total (weighted) explained inertia the procedure adapts to the actual signal present in the dataset, rather than enforcing an a priori balance between spatial and attribute contributions.

Moreover, we provide insights into the role of the spatial component by using an overlapping parameter that controls the distance across the means of the spatial distribution of the true clusters. Under the simulation design, we compare the algorithms in terms of accuracy, precision, sensitivity (Naidu et al., 2023), and the adjusted rand index (Gordon, 1999) for multi-class data.

Using a Monte Carlo simulation scheme with $N = 500$ repetitions, we simulate samples of size $n = 100$ observations from a mixture of three-dimensional Gaussian distribution in which the first dimension refers to a unique non-spatial feature and the other two dimensions refer to the spatial coordinates. We simulate data clustered into $K = 4$ different groups. We indicate with $\pi_1, \dots, \pi_K$ the probability to sample an observation belonging to each cluster $k = 1, \dots, 4$ and, according to the cluster, we simulate the variable Z and the spatial coordinates (X, Y) . Specifically, we define the distribution of the variable as $Z_k \sim N(\mu_k, \sigma)$, where $\mu = (\mu_1, \dots, \mu_4)$ and $\sigma = v$, and the spatial component $(X, Y)_k \sim N(\mu_{sp,k}, \Sigma_{sp})$. The means of the group-specific spatial coordinates are defined according to an overlapping parameter, namely d , which can vary between 0 and 1, and it determines the proximity across the spatial means of the clusters. In particular, $\mu_{sp} = (\mu_{sp,1}, \mu_{sp,2}, \mu_{sp,3}, \mu_{sp,4}) = ((d, d), (-d, d), (d, -d), (-d, -d))$ while $\Sigma_{sp} = v_{sp} \mathbb{I}_2$. Notice that, given this simulation setting, as the overlapping parameter grows from 0 to 1, the centroids of the groups tend to diverge, thus generating well-distanced clusters and enhancing the relevance of the geographical information in clustering the units.

To exemplify the contribution of the overlapping parameter d , in Fig. 2, we depict some examples of the clustering partition that can be obtained for different values of the overlapping parameter and distinguishing among (Chavent et al., 2018) (left column) and Morelli et al. (2025) (right column) criteria. The plots are simulated by setting $\pi_k = 1/K$, $\mu_k = (2, 4, 6, 8)$, $v = 0.4$, $v_{sp} = 0.4$, and considering the following set of values for the overlapping parameter $d \in \{0, 1/3, 2/3, 1\}$ (i.e., sorted top-down).

From the figure, it is possible to observe that the maximum average criterion is able to detect a clustering partition similar to the real one, even when the spatial component is not relevant or less relevant, that is, in panel a and b. When the distance in the spatial location of the cluster is higher (e.g., in panel c and d), clustering performance improves using both methodologies. Results reported in Table 2 confirm the previous statements. Indeed, while (Chavent et al., 2018) methodology allows identifying a reasonable clustering partition, the (Morelli et al., 2025) methodology demonstrates improved clustering performance in terms of accuracy and adjusted rand index. This enhancement is particularly notable in scenarios where the spatial component plays a limited role, such as when clusters fully or partially overlap in space.

³ It should be noted that the case $JI = 1$ holds when the geographical information and the feature space describe the same group structure while being strongly related to the phenomenon under study. This happens when space and features embed the same amount of information associated with the phenomenon and resulting in the same clustering partition. So, in a sense, the geography jointly determines the features and the clusters. However, geography compresses the relevant information using only two dimensions, whereas it is reasonable to expect the same degree of information to be achieved by a much larger number of features.

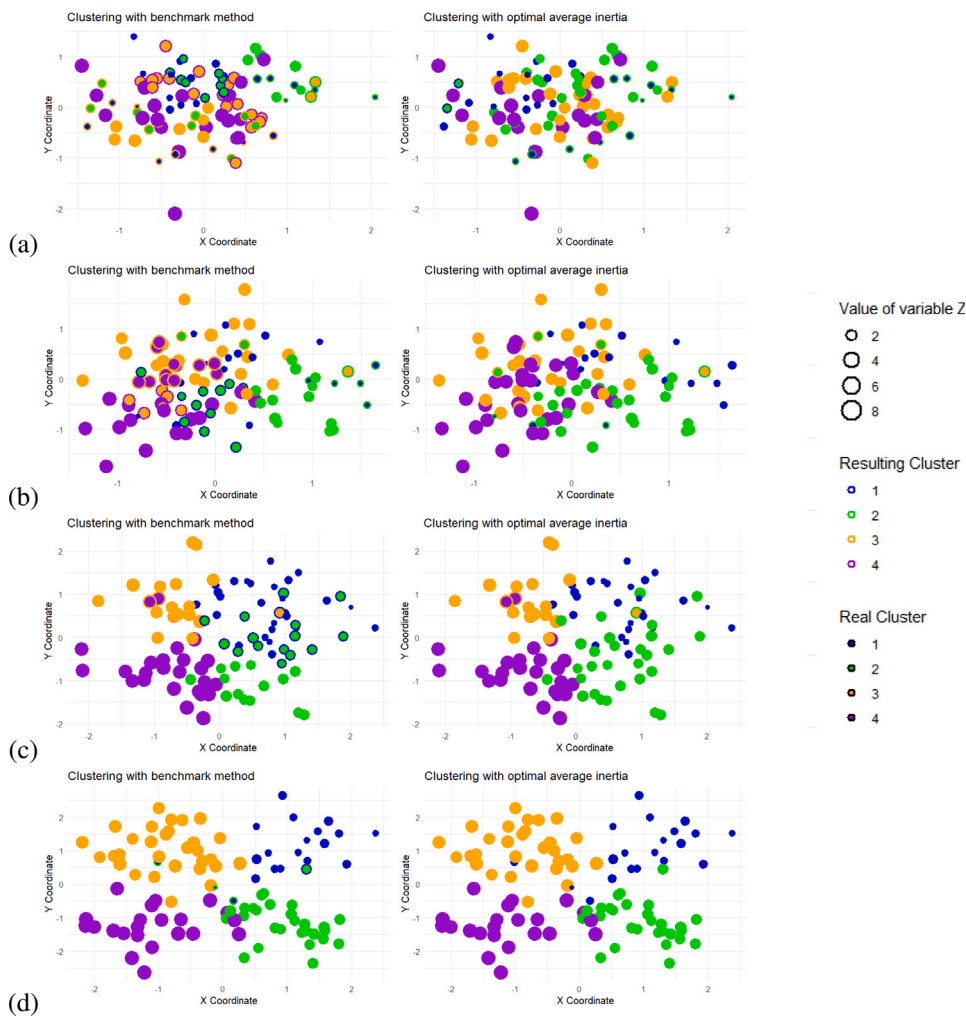


Fig. 2. Clustering partition in four simulations with different overlapping parameters $d \in \{0, 1/3, 2/3, 1\}$ respectively in panel (a), (b), (c), and (d). The color of the points refers to the true cluster from which the observation is drawn, while the color of the border refers to the resulting clustering partition obtained by applying the benchmark algorithm (Chavent et al., 2018) (left column) and the optimal solution for choosing α in Morelli et al. (2025) (right column).

Table 2

Comparison of the performance resulting from clustering partition according to the benchmark procedure from Chavent et al. (2018) and the optimal solution in Morelli et al. (2025), with Accuracy and Adjusted Rand Index and the resulting Joint Inertia to assess the role of the spatial component in the partition.

d	Benchmark			Optimal			Joint inertia
	α	Accuracy	Adj. rand index	α	Accuracy	Adj. rand index	
0	0.75	0.43	0.27	0.35	0.87	0.74	0.17
1/3	0.65	0.63	0.34	0.50	0.93	0.83	0.49
2/3	0.85	0.82	0.61	0.45	0.97	0.93	0.86
1	0.75	0.96	0.91	0	1.00	1.00	0.93

Moreover, we confirm that the parameter α cannot be interpreted as a measure of assessment of the role of the spatial component as the optimized values do not reflect the degree of overlapping of the data-generating process. On the other hand, the Joint Inertia index appears effective in assessing the relevance of the spatial component in the resulting cluster distribution.

In Algorithm 2, we provide a scheme of the simulation, from the data generating process, to the choice of α according to Chavent et al. (2018) and Morelli et al. (2025), and the computation of the validity measures and the Joint Inertia.

Below, we present the results from the Monte Carlo experiment and random assignment for parameters $\pi_k \sim U(0.15, 0.35)$, $v \sim U(0.1, 0.6)$, $v_{sp} = 0.4$ and $d \sim U(0, 1)$. From Table 3, we can clearly observe that maximum average criterion by Morelli et al.

Algorithm 2 Simulation of Spatial Hierarchical Clustering

Simulate data from $K = 4$, with probabilities $\pi_1, \dots, \pi_4$, with variable $Z_k \sim N(\mu_k, \sigma)$ and spatial coordinates $(X, Y)_k \sim N(\mu_{sp,k}, \Sigma_{sp})$.
 Set the weight of each unit $w_i = 1/n$
 Compute $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ the feature dissimilarity matrix as Euclidean distance across values of the variable.
 Compute $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$ the spatial dissimilarity matrix as distance across the coordinates.
 Compute $A_0 = [\delta_{0,ij}]_{i,j=1,\dots,n}$ and $A_1 = [\delta_{1,ij}]_{i,j=1,\dots,n}$ as $\delta_{ij} = \frac{w_i w_j}{w_i + w_j} d_{ij}^2$
 Define as $\Delta\alpha$ the increment of α
for $\alpha \in [0, 1]$, by $\Delta\alpha$ **do**
 Compute the aggregation matrix $\Delta(\alpha) = (1 - \alpha)A_0 + \alpha A_1$
 Find the $\mathcal{P}_{K=4}^\alpha$ partition using Ward algorithm on the combined matrix $\Delta(\alpha)$, which correspond to apply the Ward algorithm on the combination of inertia
 Compute the proportion of explained inertia $\mathcal{Q}_{D_0,\alpha}$ and $\mathcal{Q}_{D_1,\alpha}$
 Compute the weighted average of the explained mixed pseudo inertia $\bar{\mathcal{Q}}(\mathcal{P}_K^\alpha)$
end for
 Select $\alpha_K^c = \operatorname{argmin}_\alpha (\operatorname{abs}(\mathcal{Q}_{D_0,\alpha} - \mathcal{Q}_{D_1,\alpha}))$
 Select $\alpha_K^* = \operatorname{argmax}_\alpha \bar{\mathcal{Q}}(\mathcal{P}_K^\alpha)$
 Compute Accuracy, Precision Sensitivity and Adjusted Rand Index, for $\mathcal{P}_{K=4}^{\alpha^c}$ and $\mathcal{P}_{K=4}^{\alpha^*}$
 Compute Joint Inertia according to equation (8).

Table 3

Comparison of the performance resulting from clustering partition according to the benchmark procedure from [Chavent et al. \(2018\)](#) and the optimal solution in [Morelli et al. \(2025\)](#), using a Monte Carlo simulation, in terms of Accuracy, Precision, Sensitivity and Adjusted Rand Index.

	Accuracy		Precision		Sensitivity		Adj. rand index	
	Benchmark	Optimal	Benchmark	Optimal	Benchmark	Optimal	Benchmark	Optimal
Mean	0.74	0.85	0.75	0.85	0.74	0.84	0.55	0.71
Median	0.76	0.91	0.78	0.91	0.76	0.90	0.52	0.77
SD	0.16	0.14	0.17	0.15	0.16	0.15	0.23	0.22

(2025) is able to find clustering partitions that are very similar to the true partitions and performs remarkably better in identifying the actual mixing parameter α with respect to the ([Chavent et al., 2018](#)) criterion.

In [Fig. 3](#) we provide a comparison of the resulting clustering performance with respect to the overlapping parameter d used in the simulation. It is evident that [Morelli et al. \(2025\)](#) algorithm performs generally better, in particular when the overlapping parameter is low and clusters are partially overlapping in space. As expected, when the overlapping parameter d is close to 1, both algorithms are able to find exactly the true partition because, as explained in the previous section, when the feature dissimilarity and the spatial dissimilarity lead to the same partition, any combination of them will lead to it.

In [Fig. 4](#), we provide several insights into the relationship between the spatial distance among centroids, determined by the overlapping parameter d , the mixing parameter α , and the Joint Inertia index. In the left panel, as the overlapping parameter grows (from left to right), the mixing parameter selected by both methods seems to be unaffected, confirming that the weight assigned to the spatial dissimilarity matrix does not provide information about the actual contribution of the spatial component in determining the clusters. Conversely, the right panel remarkably supports the use of the Joint Inertia index in quantifying the importance of the spatial component. Indeed, a strong and positive correlation between the overlapping parameter d and the Joint Inertia is detected, both using α_{chavent} and α_{morelli} . The positive correlation is definitely consistent with the interpretation of the overlapping parameter d given above, that is, as d grows from 0 to 1, the distance between the centroids of the groups tends to increase, increasing the informative power induced by the spatial/geographical compared to the non-spatial distance, exactly as described by the Joint Inertia.

5. Dynamic and regional clustering of GHGs emissions

In this section, we present an application of spatiotemporal clustering on environmental data concerning the annual GHGs emissions at the regional level for Europe. Our aim is to explore regional patterns in the environmental impact caused by human activities on air quality, which may also depend on the geographical and morphological characteristics of the area as well as on the productive specialization of the regions.

We analyze the annual average emissions per km^2 of several GHGs, that is, methane, nitrous oxide, fluorinated gases (F-gases), and carbon dioxide produced by the main economic sectors, namely agriculture, buildings, energy, industry, transport, and waste. Data are measured in tons per km^2 of CO_2 equivalent.

Greenhouse gases in Europe originate from different activities, each of them contributing to climate change in unique ways ([Crippa et al., 2024](#)). Methane (CH_4) emissions primarily arise from agriculture, especially livestock farming and waste

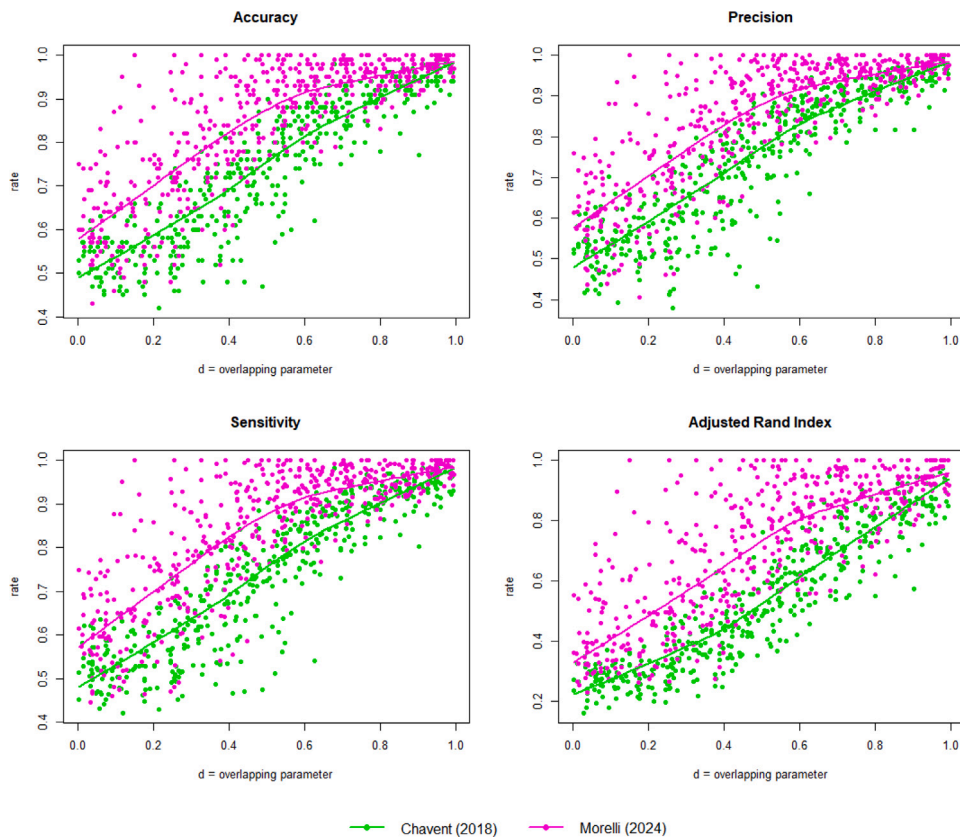


Fig. 3. Clustering performance comparison between Chavent et al. (2018) and Morelli et al. (2025) in terms of Accuracy, Sensitivity, Precision, and Adjusted Rand Index, with respect to the overlapping parameter d .

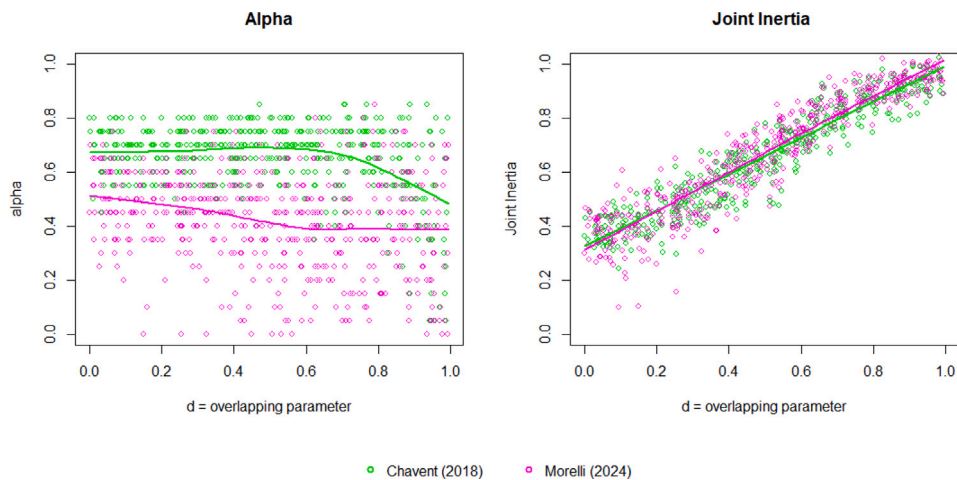


Fig. 4. In panel (a), on the left, we can observe the resulting α according to Chavent et al. (2018) and Morelli et al. (2025) when considering clusters with different overlapping degrees (i.e., when the overlapping parameter d varies between 0 and 1). In panel (b), on the right, we report the resulting Joint Inertia obtained in the simulation experiment with respect to the overlapping parameter d .

management, as well as fossil fuel extraction and processing. Nitrous oxide (N_2O) is largely emitted through agricultural practices, such as the use of nitrogen-based fertilizers, with smaller contributions from industrial processes and fossil fuel combustion. F-gases, synthetic in nature, are used in refrigeration, air conditioning, and industrial processes like aluminum production. Carbon dioxide

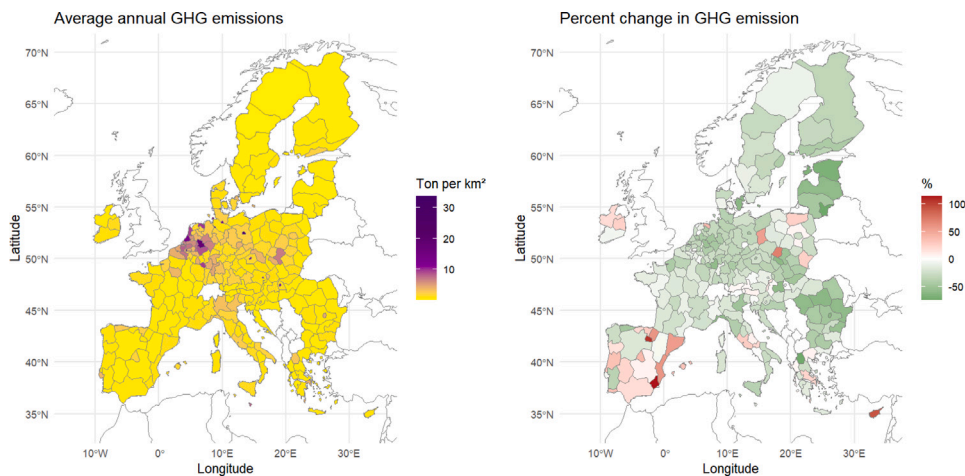


Fig. 5. Descriptive statistics of GHG emission in each region at NUTS2 level: annual average tons CO₂ equivalent per km² between 1990–2022 (on the left) and percent change from 1990 to 2022 (on the right).

(CO₂), the most abundant GHGs, is predominantly released through fossil fuel combustion for energy, transportation, and industrial activities, as well as deforestation and cement production.

We believe that the identification of patterns and dynamics in regional GHGs emissions within a spatiotemporal framework is a valuable contribution that extends the current knowledge on the status of environmental degradation in Europe, enhancing the awareness of citizens and institutions regarding the policies to be planned to contrast climate change and global warming.

The analysis can be carried out using the SCDA package (Cerqueti et al., 2025), written in the R language (R Core Team, 2025), which provides the HSTC functions to identify the clustering partition, optimize the number of clusters, and compute the joint inertia⁴

5.1. Dataset description

The data of interest relate to atmospheric emissions of sector-specific greenhouse gases measured at the regional level provided by the Emissions Database for Global Atmospheric Research (EDGAR) database (Crippa et al., 2024) within the Annual Regional Database of the European Commission (ARDECO) (Bucciarelli, 2024) project.

The regional emission data are taken from EDGAR v8.0 and are modeled estimates produced with a technology-based emission-factor approach under an IPCC-consistent methodology. National and sectoral totals are downscaled to a 0.1° × 0.1° grid using spatial proxies (e.g. facility locations, road networks, population) and then aggregated to NUTS2 polygons (NUTS 2021). Emissions are reported in kton CO₂-eq (GWP-100, IPCC AR5). Further methodological details and discussion of assumptions and uncertainty are given in Crippa et al. (2024). In particular, we consider yearly measurements of GHGs emissions from 1990 to 2022 (i.e., $T = 33$ time stamps) for the $N = 234$ regions⁵ in the European Union, in 27 countries, identified according to the Nomenclature of territorial units for statistics - Level 2 (NUTS-2) classification (European Commission, 2016). We also collect the area of NUTS-2 regions in 2022 from Eurostat⁶ dataset in order to obtain the annual emissions in tons per km² of CO₂ equivalent, by gases and by sector.

In Fig. 5 we provide descriptive statistics about our dataset, showing the long-run emission intensity with its evolution: the left panel highlights persistent high-intensity hotspots concentrated in major metropolitan/industrial corridors, whereas the right panel shows broad declines alongside localized increases from 1990 to 2022, underscoring marked spatial heterogeneity in both levels and trends.

The dataset contains annual time series of GHG emissions for four gases (i.e., CH₄, F-gases, fossil-CO₂, and N₂O) and for six emitting sectors: Agriculture, Buildings, Energy, Industry, Transport, and Waste. Averaged across all years in the sample, the global Moran's I is approximately 0.28 for CH₄, 0.13 for F-gases, 0.17 for fossil-CO₂, and 0.24 for N₂O. By sector, the mean Moran's I is highest for Agriculture and Transport which are respectively 0.48 and 0.44, followed by Industry (≈ 0.17), Waste, Energy and Buildings with the lowest annual average Moran's I , respectively 0.09, 0.09 and 0.08. Complete year-by-year values and plots are reported in Appendix B.

We run the spatiotemporal cluster analysis twice, considering two different datasets. In the first case, we implement the clustering by aggregating GHGs emissions according to the type of gas (i.e., by summing the emissions at the sectoral level for each gas). In

⁴ <https://cran.r-project.org/web/packages/SCDA/index.html> (Accessed on January 21st, 2026).

⁵ Notice that we excluded regions classified as extra-continental territories and those located in non-European continents, ensuring that the extremely high distances involving these regions do not render the spatial dissimilarities among the remaining regions insignificant.

⁶ https://ec.europa.eu/eurostat/databrowser/view/REG_AREA3/default/table (Accessed on October 9th, 2025)

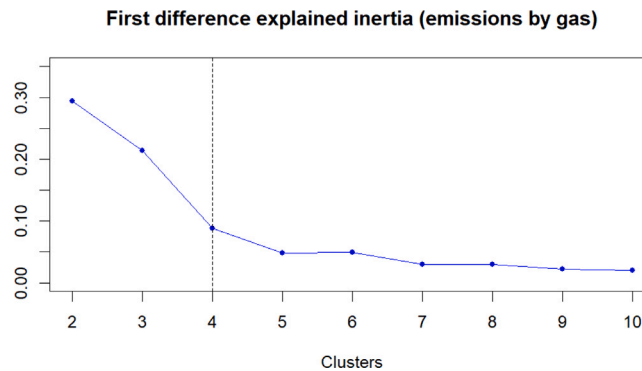


Fig. 6. Results: Increment in the weighted average proportion of explained inertia generated by a unitary increase in the number of clusters. Recall that, for each K , we considered the best combination of inertia of the dissimilarity matrices according to the α_K^* found using the maximization inertia criterion.

Table 4

Summary of the α^* weights and the inertia (proportion and normalized) returned by the spatiotemporal clustering of emissions by gases at the optimal solution $K^* = 4$. $Q_{D_p}(\mathcal{P}_K^a)$ is the proportion of explained inertia in D_p ; $\tilde{Q}_{D_p}(\mathcal{P}_K^a)$ is the normalized proportion of inertia in D_p ; $\tilde{Q}_{D_{-p}}(\mathcal{P}_K^a)$ is the normalized proportion of weighted average inertia excluding D_p ; and $J I_p(\mathcal{P}_K^a)$ is the Joint Inertia relative to D_p .

	D_{CH_4}	$D_{F-gases}$	D_{CO_2}	D_{N_2O}	D_{sp}
α_p^*	0.20	0.20	0.05	0.15	0.40
$Q_{D_p}(\mathcal{P}_K^a)$	0.67	0.68	0.63	0.38	0.59
$\tilde{Q}_{D_p}(\mathcal{P}_K^a)$	0.72	0.80	0.66	0.47	0.82
$\tilde{Q}_{D_{-p}}(\mathcal{P}_K^a)$	0.97	0.96	1.00	0.99	0.81
$J I_p(\mathcal{P}_K^a)$	0.69	0.76	0.66	0.46	0.64

the second case, we aggregate the emissions with respect to the sectors, thus by summing the gas-specific emissions for each sector. In both cases, we utilize the Dynamic Time Warping (DTW) strategy described in Section 2.2 to build the temporal dissimilarity matrices, whereas we compute the spatial dissimilarity matrix as the geodetic distance across the coordinates of the centroids of each region.

5.2. Spatiotemporal clusters for emissions by gas

We run the algorithm described in Morelli et al. (2025) setting $4\alpha = 0.05$, for $K = 2, \dots, 10$, combining the inertia from $P = 5$ dissimilarity matrices referring to the CH_4 , F-Gases, Fossil- CO_2 , N_2O emission time series and the spatial component. We select $K^* = 4$ according to the elbow rule, observing the weighted average proportion of explained inertia, which can be interpreted as the gain in explained variability due to an additional cluster. In Fig. 6 we represent the first difference in the weighted average proportion of explained inertia, highlighting our choice of $K^* = 4$.

Consequently, we selected $\alpha_{4,p}^* = (0.20, 0.20, 0.15, 0.5, 0.40)$ as the optimal weighting vector for the inertia from each dissimilarity matrix. Table 4 provides a summary of the weights α_p^* , the explained inertia (both proportion and normalized), and the Joint Inertia associated with the resulting clustering partition for each dissimilarity matrix included in the algorithm.

From Table 4, it is possible to observe that the partition obtained with $K^* = 4$ clusters allows us to explain almost 40% of the DTW dissimilarities across N_2O time series and about 60% of the dissimilarities across time series of CH_4 , F-gases, CO_2 and the spatial dissimilarity. The latter has a Joint Inertia of 0.64, which allows us to conclude that spatial information is crucial when determining the clusters composition, and therefore, the spatial distance is useful in the identification of patterns to explain well the data under consideration. In particular, through the resulting clustering partition $\mathcal{P}_K^a$ we are able to incorporate 82% of the explained inertia in the spatial component, relative to the case in which we consider only D_{sp} and, considering the non-spatial dissimilarities, we are still able to include 81% of the weighted average explained inertia with respect to the partition obtained excluding the spatial component. All the dissimilarity matrices have a weight greater than zero and they all seem to be relevant in the composition of the clusters, in fact they have a rather high Joint Inertia. It is possible to note that the weights assigned to the dissimilarity matrices do not always reflect the relevance of the matrix itself in relation to the clustering partition obtained, for example the matrix D_{N_2O} is assigned a weight of 0.15, greater than the 0.05 assigned to matrices D_{CO_2} , yet the proportion of inertia explained in the first, which is 0.46, is less than the inertia explained in the other 0.66.

In Fig. 7 we present a map of the regions belonging to different clusters and in Fig. 8 we report the average annual emission in each cluster by gas.

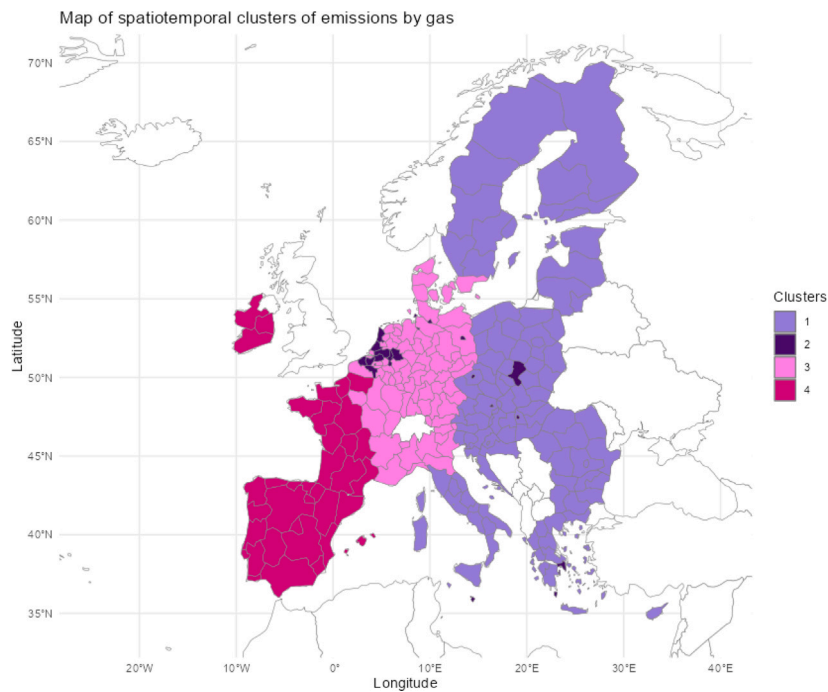


Fig. 7. Results: Map of spatiotemporal clusters obtained combining the inertia from the dissimilarity matrices of the time series of emissions of different gases between 1990 and 2022, that is D_{CH_4} , $D_{F-gases}$, D_{CO_2} and D_{N_2O} , and the geodetic distances across centroids of NUTS2 level regions D_{sp} .

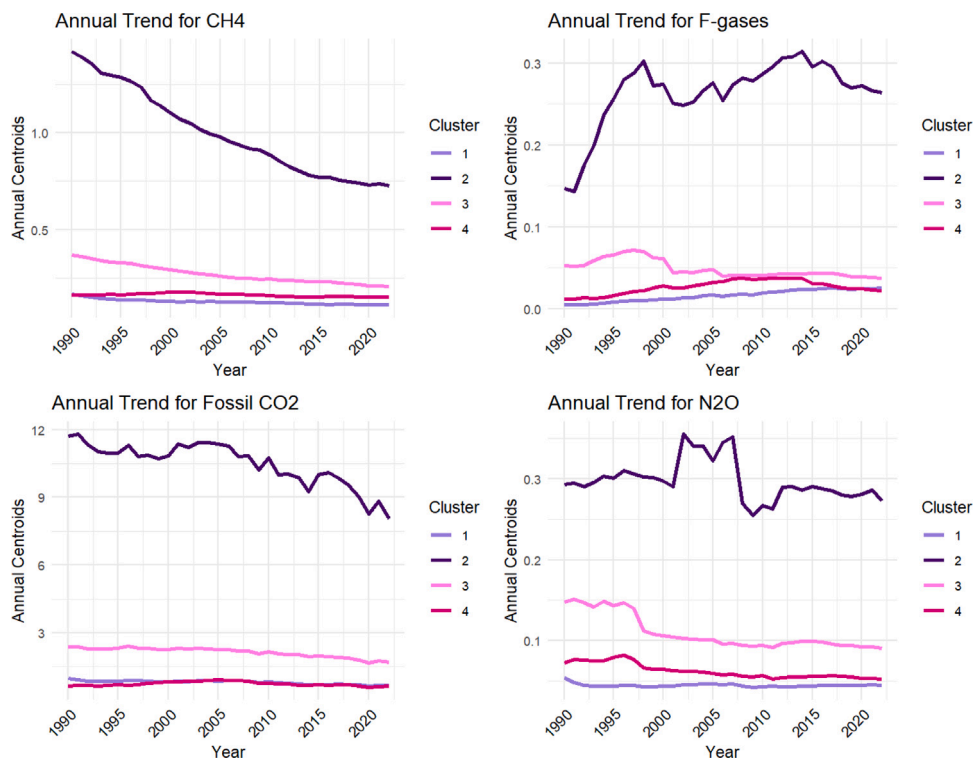


Fig. 8. Results: Annual average of emission per km^2 of gases for spatiotemporal clusters obtained combining the inertia of the dissimilarity matrices of the time series of emissions of different gases between 1990 and 2022 that is D_{CH_4} , $D_{F-gases}$, D_{CO_2} and D_{N_2O} , and the geodetic distances across centroids of NUTS2 level regions D_{sp} .

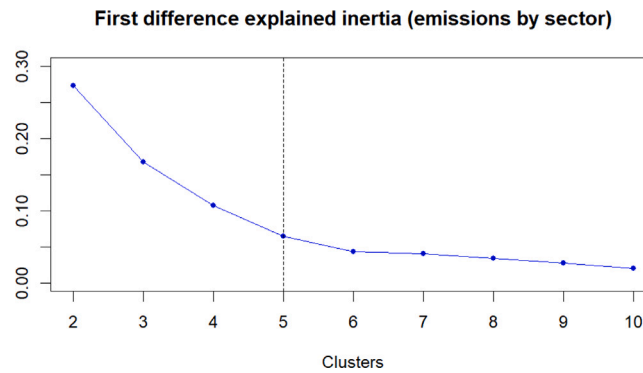


Fig. 9. Results: Increment in the weighted average proportion of explained inertia generated by a unitary increase in the number of clusters. Recall that, for each K , we considered the best combination of inertia from the dissimilarity matrices according to the α_K^* found using the maximization inertia criterion.

Cluster 1 includes regions located in East Europe, from the Scandinavian peninsula to the Balkan countries, the central and southern areas of Italy and the eastern area of Austria. It shows the lowest level of CH_4 , F-gases, CO_2 and N_2O emissions compared to the other clusters and it appears to be quite stable over time, except for F-gases emissions that are slowly increasing.

Regions of the Netherlands, Malta, and several metropolitan areas in East-Central Europe belong to Cluster 2, which is characterized by emission levels that are remarkably higher — by a significant margin — than those observed in any other cluster. As for the dynamics of CH_4 emissions, these exhibit a decreasing trend, nearly halving between 1990 and 2022. Similarly, CO_2 emissions show a decline over the years, dropping from approximately 12 tons per km^2 to about 8 tons per km^2 . In contrast, emissions of F-gases experience a sharp increase during the 1990s, stabilizing thereafter at a level of approximately 0.25 tons per km^2 . Meanwhile, emissions of NO_2 display fluctuations over time without any discernible trend.

Moving to Cluster 3, we can see from the map that it incorporates regions from North-central Europe, mainly Denmark, Germany, east part of France, and Northern Italy. It appears to be the second worst cluster in term of emission per km^2 , showing a slowly decreasing trend in CH_4 emissions, quite stable level of CO_2 and F-gases emissions and a drop in NO_2 emissions in the last '90s.

Finally, cluster 4 consists of regions from Ireland, Spain, Portugal, and the West part of France. Regions in this cluster exhibit very low levels of emissions, comparable to those in Cluster 1. These emissions are characterized by a slightly increasing trend for F-gases and stable levels for CH_4 , CO_2 , and N_2O .

5.3. Spatiotemporal clusters for emissions by sector

The second spatiotemporal clustering considers the GHGs emissions by sector. We compute the dissimilarity matrices of the total GHGs emissions in agricultural, buildings, energy, industry, transport, and waste sectors, and we combine them with the spatial dissimilarity matrix, thus having $P = 7$. We set $\Delta\alpha = 0.1$, for $K = 2, \dots, 10$, and we select $K^* = 5$ according to the elbow rule and $\alpha_{s,p}^* = (0.2, 0, 0, 0.1, 0, 0.3, 0.4)$ as the weight of the inertia from each dissimilarity matrix. In Fig. 9, we represent the first difference in the weighted average proportion of explained inertia, highlighting our choice of $K^* = 5$.

In Table 5, we report a summary of the weights $\underline{\alpha}^*$, the explained inertia (proportion and normalized), and the Joint Inertia in the resulting clustering partition for each dissimilarity matrix included in the algorithm.

From Table 5, it is possible to observe that the partition obtained with $K^* = 5$ clusters allows to explain at least 35% of the proportion of inertia in each dissimilarity matrix and the 36% of the inertia with respect to the cases in which clusters are computed only relying on one dissimilarity matrix. Each dissimilarity matrix seems to be quite helpful in obtaining the resulting partition. In particular, the spatial component presents a Joint Inertia equal to 0.62, and it incorporates 0.82 of the inertia normalized to the explained inertia in the partition obtained from the purely spatial case. On the other hand, including the spatial component, we are still able to capture 0.80 of the inertia in the non-spatial component. Although all the dissimilarity matrix D_{Tran} has weights equal to 0, thus it is not included, the resulting partition incorporates 0.61 of the inertia, which corresponds to the 67% normalized to the explained inertia from the partition obtain only considering D_{Tran} .

In Fig. 10 we present a map of the regions belonging to different clusters and in Fig. 11 we report the average annual emission in each cluster, for each sector.

Cluster 1 consists primarily of regions in Eastern Europe, including Finland, Sweden, Latvia, Lithuania, Poland, Slovakia, Hungary, Romania, Bulgaria, and Greece. This cluster exhibits the overall lowest emission levels, with stable trends in the Buildings, Energy, Industry, and Waste sectors. Emissions in the Agricultural sector show a slight decrease during the 1990s, while emissions in the Transport sector display a gradual increase over time.

Cluster 2 includes the smallest number of regions, located in Belgium and the Netherlands and some metropolitan areas in Germany and Eastern Europe. As highly industrialized and densely populated areas, it is unsurprising that this group exhibits the highest levels of emissions across all sectors except Agriculture. In particular, emissions from the Buildings and Industry sectors

Table 5

Summary of the α^* weights and the inertia (proportion and normalized) returned by the spatiotemporal clustering of emissions by gases at the optimal solution $K^* = 5$. $Q_{D_p}(P_K^{\alpha})$ is the proportion of explained inertia in D_p ; $\tilde{Q}_{D_p}(P_K^{\alpha})$ is the normalized proportion of inertia in D_p ; $\tilde{Q}_{D_{-p}}(P_K^{\alpha})$ is the normalized proportion of weighted average inertia excluding D_p ; and $JI_p(P_K^{\alpha})$ is the Joint Inertia relative to D_p .

	D_{Agri}	D_{Build}	D_{Energy}	D_{Ind}	D_{Tran}	D_{Waste}	D_{sp}
α_p^*	0.20	0.00	0.00	0.10	0.00	0.30	0.40
$Q_{D_p}(P_K^{\alpha})$	0.69	0.57	0.35	0.58	0.61	0.61	0.65
$\tilde{Q}_{D_p}(P_K^{\alpha})$	0.71	0.58	0.36	0.59	0.64	0.63	0.82
$\tilde{Q}_{D_{-p}}(P_K^{\alpha})$	0.94	1.00	1.00	0.99	1.00	1.00	0.80
$JI_p(P_K^{\alpha})$	0.65	0.58	0.36	0.58	0.64	0.63	0.62

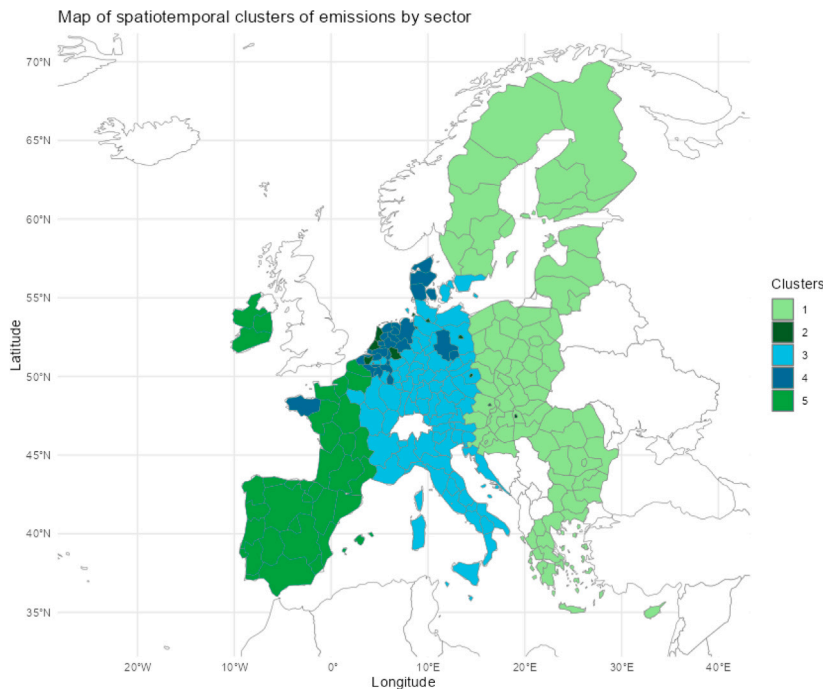


Fig. 10. Results: Map of spatiotemporal clusters obtained combining the inertia of dissimilarity matrices of the time series of emissions of different sectors between 1990 and 2022 that is D_{Agri} , D_{Build} , D_{Energy} , D_{Ind} , D_{Tran} and D_{Waste} , and the geodetic distances across centroids of nuts2 level regions D_{sp} .

display a decreasing trend, declining from 6 tonnes per km^2 to less than 4 tonnes per km^2 . Emissions from the Energy sector demonstrate high variability over time, with a clear reduction from 6 to 3.5 tonnes per km^2 in the last five years. Waste emissions also follow a generally decreasing trend but appear to have stabilized in recent years. Notably, only the Transport sector shows an increase in emissions over time.

Cluster 3 primarily includes regions located in Central Europe, such as Germany, the Czech Republic, Austria, Italy, Croatia, Slovenia, along with a few regions of France. This cluster exhibits steady levels of emissions, which are neither the highest nor the lowest compared to other clusters across all sectors.

Considering Cluster 4, it includes regions located in Belgium and the Netherlands, Northern France and Northern Germany and Denmark. This group is characterized by particularly high levels of emissions in the Agricultural sector, which exhibit a slowly decreasing trend. For other sectors, it ranks as the second worst cluster, following Cluster 3. In the Transport and Waste sectors, emissions follow a dynamic similar to Cluster 3, showing increasing and decreasing trends, respectively, while no clear trend is observed in the Buildings, Energy, and Industry sectors.

Finally, cluster 5 comprises regions located in Spain, Portugal, Ireland, and certain areas of France. This group is characterized by relatively low emissions across all sectors. Notably, emission levels in this group do not exhibit pronounced rising or falling trends. However, a slight upward trend is observed in the Transport sector, while the Energy sector shows a modest decline over time.

Overall, we can conclude that a relatively small number of regions, encompassing a limited portion of the European territory, are responsible for exceptionally high GHG emissions. Within each breakdown, a group with particularly elevated emission levels



Fig. 11. Results: Annual average of emission per km^2 of sectors for spatiotemporal clusters obtained combining the inertia of the dissimilarity matrices of the time series of emissions of different sectors between 1990 and 2022 that is D_{Agri} , D_{Build} , D_{Energy} , D_{Ind} , D_{Tran} and D_{Waste} , and the geodetic distances across centroids of nuts2 level regions D_{sp} .

could be distinctly identified. While it is clear that this small group contributes disproportionately to total emissions, it is equally evident that these regions also exhibit decreasing trends over time. This suggests ongoing improvements, albeit at varying rates, across individual sectors and for gases with higher emission levels.

5.4. Clustering comparison and internal validation

To further investigate the role of the spatial component in cluster formation, we performed non-spatial clustering as a robustness check for emissions by gas and by sector. The resulting clusters show greater overlap compared to the spatiotemporal clustering approach, in both cases. Nevertheless, regions in Central-Northern Europe and metropolitan areas remain identifiable as those with the highest levels of emissions per km^2 across multiple sectors. A detailed discussion of this robustness check and the associated results can be found in [Appendix C](#).

In addition, [Appendix D](#) reports an internal validation of the partitions. For each gas and sector we compute standard indices (Average Silhouette, Calinski–Harabasz, Within/Between ratio, Pearson’s Gamma) and compare three solutions: spatiotemporal, temporal-only, and spatial-only. Results show the expected pattern — temporal-only scores higher on feature-based dissimilarities, spatial-only on the spatial one — while the spatiotemporal partition consistently lies between them, indicating a balanced trade-off. Silhouettes are generally modest (as typical with high-dimensional series), but the overall evidence supports combining temporal and spatial information.

Moreover, to assess cluster stability across environmental dimensions and to examine gas- and sector-specific patterns in greater detail, we provide in [Appendix E](#) a comparative set of maps showing spatiotemporal clusters by gas (CH_4 , F-gases, CO_2 , N_2O) and

by sector (Agriculture, Buildings, Energy, Industry, Transport, Waste). Furthermore, in order to evaluate their interaction with the spatial component, we show for each partition the resulting Joint Inertia.

6. Conclusion

Spatiotemporal hierarchical clustering methodology is a versatile and flexible approach to analyze patterns and dynamics of complex phenomena. In our study, we applied this methodology to greenhouse gas emissions across European regions, combining temporal emission patterns and spatial proximity into a unified analysis. Notably, we sought to assess whether the geographical information is helpful in explaining emission dynamics, providing both methodological advancements and empirical insights.

To this end, we introduced a novel measure, the Joint Inertia, to evaluate the role of the spatial component in the clustering process. This measure captures the contribution of spatial distances in generating the final clustering partition. Through a simulation experiment, we demonstrated the advantages of our methodology and highlighted how the Joint Inertia effectively quantifies the influence of spatial information on the clustering results. These findings underscore the robustness of our approach and its capacity to reveal spatial-temporal structures in the data that might otherwise remain hidden.

Our application provides compelling evidence that spatial context meaningfully enhances the analysis of emission trends. The Joint Inertia values indicate that including spatial distances enriches the clustering process, revealing region-specific patterns and dynamics that reflect Europe's inherent heterogeneity in economic, industrial, and social characteristics.

By separately analyzing emissions by gas type and sector, we uncovered distinct trends and dynamics within each cluster. This analysis highlighted how regions with high emissions are often associated with specific sectors, such as agriculture or industry, while others demonstrate relatively low emissions with stable or slowly evolving patterns. The interplay of gas-specific and sector-specific emissions varies significantly across regions, reflecting the diverse economic, industrial, and social landscapes of Europe. This detailed approach provides a more nuanced understanding of emission patterns, identifying both hotspots of high emissions and regions with notable improvements or stability. Such insights are invaluable for tailoring regional environmental policies and strategies to address sector-specific challenges effectively.

The methodological contributions and empirical findings of this study hold significant utility for both statisticians and environmental researchers. From a methodological perspective, our approach demonstrates the versatility of spatiotemporal clustering for high-dimensional data, offering a framework that can be adapted to other environmental and socioeconomic phenomena. For environmental research, this study delivers actionable insights into the spatiotemporal dynamics of greenhouse gas emissions, offering a foundation for targeted mitigation policies and resource allocation.

Future work could extend these contributions by linking cluster dynamics to socioeconomic, industrial, or policy variables, further enriching the understanding of emission drivers, exploiting proper spatiotemporal models. This would enable a more comprehensive framework for analyzing greenhouse gas emissions in a holistic spatiotemporal context, paving the way for interdisciplinary research and evidence-based policy development.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT, a generative AI tool developed by OpenAI, to assist with drafting, rephrasing, and refining the text. After utilizing this tool, the authors carefully reviewed, edited, and revised the content as necessary to ensure accuracy, clarity, and alignment with the objectives of the publication. The authors take full responsibility for the content of this manuscript, including any ideas, analyses, and conclusions presented.

Funding

This study was partly funded by the University of Milano-Bicocca, Italy in the framework of the Sequestering CARbon through Forests, AgriCulture, and land use (SCARFACE) research project (2024-ATEQC-0048).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Appendix A. Technical details of joint inertia

In this appendix we focus on the properties of Joint Inertia, in particular we explain why its value varies between 0 and 1 and what are the conditions that guarantee it, starting from the simplest case where we have only two matrices of dissimilarity and then extending the reasoning to the case where we want to combine more than two matrices.

A.1. Joint inertia for the case $P = 2$ dissimilarity matrices

Let $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ and $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$ refer respectively to any distances matrix of variables and the spatial distances matrix considering a sample of n units and let w_i with $i = 1, \dots, n$ be the weight of the i th cross-sectional unit to be clustered. Let us consider a partition $\mathcal{P}_K^\alpha = (C_1^\alpha, \dots, C_K^\alpha)$ in K clusters obtained mixing the dissimilarity matrices D_0 and D_1 with the parameter α . Also, let us denote its within-clusters mixed inertia as $W(\mathcal{P}_K^\alpha)$ and the corresponding proportion of the total pseudo inertia explained as

$$Q_m(\mathcal{P}_K^\alpha) = 1 - \frac{W_m(\mathcal{P}_K^\alpha)}{W_m(\mathcal{P}_1)} = 1 - \frac{\sum_{k=1}^K \frac{1}{2\mu_k} \sum_{i \in C_k} \sum_{j \in C_k} w_i w_j d_{mij}^2}{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{mij}^2} \quad m = 0, 1.$$

where $\mu_k = \sum_{i \in C_k} w_i$ and $W_0(\mathcal{P}_1)$ and $W_1(\mathcal{P}_1)$ are the total pseudo inertia under dissimilarity matrix D_0 and under dissimilarity matrix D_1 , respectively. It is easy to observe that the possible values of this ratio vary between 0 and 1 because the sum of the square dissimilarities between the units belonging to the same clusters will not be greater than the total of the square dissimilarities within a matrix, and being quantities square, will all have positive values, for each possible value of α .

The combination in within mixed pseudo inertia will capture more the dissimilarities in D_0 (or D_1), when the value of α is closer to 0 (or close to 1), thus the resulting partition $\mathcal{P}_K^\alpha$ will be able to capture an higher proportion of explained inertia in D_0 (or D_1). Therefore, when α increases, $Q_{D_0}(\mathcal{P}_K^\alpha)$ will increase and $Q_{D_1}(\mathcal{P}_K^\alpha)$ will decrease. This is also explained in details in Section 3.2 of [Chavent et al. \(2018\)](#). Notice that the proportion of explained inertia is equal to 0 only when $K = 1$ and it is 1 only when $K = n$, for each value of α . Following we summarize and formalize the concept which hold for $1 < K < n$

$$0 = Q_{D_0}(\mathcal{P}_{K=1}) < Q_{D_0}(\mathcal{P}_K^{\alpha=1}) \leq Q_{D_0}(\mathcal{P}_K^\alpha) \leq Q_{D_0}(\mathcal{P}_K^{\alpha=0}) < Q_{D_0}(\mathcal{P}_{K=n}) = 1$$

$$0 = Q_{D_1}(\mathcal{P}_{K=1}) < Q_{D_1}(\mathcal{P}_K^{\alpha=0}) \leq Q_{D_1}(\mathcal{P}_K^\alpha) \leq Q_{D_1}(\mathcal{P}_K^{\alpha=1}) < Q_{D_1}(\mathcal{P}_{K=n}) = 1.$$

Now, we consider the normalized proportion of explained inertia for dissimilarity matrix D_0 (or D_1), which correspond to the ratio between the proportion of explained inertia in the partition obtained using the combined dissimilarity matrix with parameter α , and the proportion of inertia explained using only the features (or spatial) dissimilarities, thus the maximum proportion of explained inertia.

$$\tilde{Q}_{D_0}(\mathcal{P}_K^\alpha) = \frac{Q_{D_0}(\mathcal{P}_K^\alpha)}{Q_{D_0}(\mathcal{P}_K^0)} \quad \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha) = \frac{Q_{D_1}(\mathcal{P}_K^\alpha)}{Q_{D_1}(\mathcal{P}_K^1)}.$$

These quantities are strictly positive and no greater than 1. In particular, the normalized proportion of explained inertia in D_0 (or D_1) reaches its minimum value when the feature (or spatial) dissimilarity matrix is not included in the combination, thus $\alpha = 1$ (or $\alpha = 0$), and it reaches its maximum when the clustering partition is obtained from the purely no-spatial (or spatial) case, thus in $\mathcal{P}_K^0 = (C_1^0, \dots, C_K^0)$ (or $\mathcal{P}_K^1 = (C_1^1, \dots, C_K^1)$).

$$0 < \tilde{Q}_{D_0}(\mathcal{P}_K^{\alpha=1}) \leq \tilde{Q}_{D_0}(\mathcal{P}_K^\alpha) \leq \tilde{Q}_{D_0}(\mathcal{P}_K^{\alpha=0}) = 1$$

$$0 < \tilde{Q}_{D_1}(\mathcal{P}_K^{\alpha=0}) \leq \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha) \leq \tilde{Q}_{D_1}(\mathcal{P}_K^{\alpha=1}) = 1$$

The Joint Inertia between two dissimilarity matrices is computed using the following formula

$$JI(\mathcal{P}_K^\alpha) = \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha) - (1 - \tilde{Q}_{D_0}(\mathcal{P}_K^\alpha)).$$

It captures the net gain in explained inertia from jointly incorporating spatial and feature dissimilarities. The indicator quantifies the added explanatory power from combining both sources — i.e., the synergy between spatial and non-spatial components — revealing patterns that would be missed if either were used alone. It can be expressed in term of dissimilarities as follows:

$$\begin{aligned} JI(\mathcal{P}_K^\alpha) &= \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha) - (1 - \tilde{Q}_{D_0}(\mathcal{P}_K^\alpha)) \\ &= \frac{Q_{D_1}(\mathcal{P}_K^\alpha)}{Q_{D_1}(\mathcal{P}_K^1)} - \left(1 - \frac{Q_{D_0}(\mathcal{P}_K^\alpha)}{Q_{D_0}(\mathcal{P}_K^0)} \right) \\ &= \frac{1 - \frac{W_{D_1}(\mathcal{P}_K^\alpha)}{W_{D_1}(\mathcal{P}_1)}}{1 - \frac{W_{D_1}(\mathcal{P}_K^1)}{W_{D_1}(\mathcal{P}_1)}} - \left(1 - \frac{1 - \frac{W_{D_0}(\mathcal{P}_K^\alpha)}{W_{D_0}(\mathcal{P}_1)}}{1 - \frac{W_{D_0}(\mathcal{P}_K^0)}{W_{D_0}(\mathcal{P}_1)}} \right) \\ &= \frac{W_{D_1}(\mathcal{P}_1) - W_{D_1}(\mathcal{P}_K^\alpha)}{W_{D_1}(\mathcal{P}_1) - W_{D_1}(\mathcal{P}_K^1)} - \left(1 - \frac{W_{D_0}(\mathcal{P}_1) - W_{D_0}(\mathcal{P}_K^\alpha)}{W_{D_0}(\mathcal{P}_1) - W_{D_0}(\mathcal{P}_K^0)} \right) \\ &= \frac{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{1,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{1,ij}^2}{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{1,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^1} \sum_{j \in C_k^1} w_i w_j d_{1,ij}^2} \\ &\quad - \left(1 - \frac{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{0,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{0,ij}^2}{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{0,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^0} \sum_{j \in C_k^0} w_i w_j d_{0,ij}^2} \right). \end{aligned}$$

It is strictly positive because the partition obtained from the mixed dissimilarity matrix allows at least to explain as much as the partition obtained from purely non-spatial case in D_0 or the partition obtained from purely spatial case in D_1 . Moreover, it is no greater than 1 because the partition $\mathcal{P}_K^0 = (C_1^0, \dots, C_K^0)$ is the one that minimize the within cluster non-spatial dissimilarity with respect to other partitions, such that $\frac{1}{2\mu_k} \sum_{i \in C_k^0} \sum_{j \in C_k^0} w_i w_j d_{0,ij}^2 \leq \frac{1}{2\mu_k} \sum_{i \in C_k} \sum_{j \in C_k} w_i w_j d_{0,ij}^2$ and thus $\frac{1}{2\mu_k} \sum_{i \in C_k^0} \sum_{j \in C_k^0} w_i w_j d_{0,ij}^2 \leq \frac{1}{2\mu_k} \sum_{i \in C_k^0} \sum_{j \in C_k^0} w_i w_j d_{0,ij}^2$. Similarly for the spatial dissimilarities, partition $\mathcal{P}_K^1 = (C_1^1, \dots, C_K^1)$ is the one that minimize the within cluster dissimilarities in the spatial component with respect to other partitions, such that $\frac{1}{2\mu_k} \sum_{i \in C_k^1} \sum_{j \in C_k^1} w_i w_j d_{1,ij}^2 \leq \frac{1}{2\mu_k} \sum_{i \in C_k} \sum_{j \in C_k} w_i w_j d_{1,ij}^2$ and thus $\frac{1}{2\mu_k} \sum_{i \in C_k^1} \sum_{j \in C_k^1} w_i w_j d_{1,ij}^2 \leq \frac{1}{2\mu_k} \sum_{i \in C_k^0} \sum_{j \in C_k^0} w_i w_j d_{1,ij}^2$.

A.2. Joint inertia for the case $P > 2$ dissimilarity matrices

The generalization to the case of multiple dissimilarity matrices, one of which relates to the spatial component, is straightforward. Let us suppose that we are interested in computing the Joint Inertia for the generic variable m (with m being the generic index for the P variables of interests) compared to the combination of all the other variables used to cluster the time series.⁷

Recalling the notation previously introduced, let us consider a set of P variables observed over time at n spatial locations (or regions) and let $\underline{\alpha} = [\alpha_p]_{p=1, \dots, P} = (\alpha_1, \dots, \alpha_m, \dots, \alpha_P)$ on the $(P-1)$ -simplex Δ^{P-1} be the vector of mixing weights⁸ used to compute the linear combination of the mixed within pseudo inertia from dissimilarity matrices $D_p = [d_{p,ij}]_{p=1, \dots, P-1; i,j=1, \dots, n}$, such that $I(C_K^\alpha) = \sum_{p=1}^P \alpha_p \sum_{i \in C_K} \sum_{j \in C_K} \frac{w_i w_j}{2 \sum_{i \in C_K} w_i} d_{p,ij}^2$.

By applying the Ward-like hierarchical clustering algorithm presented in the manuscript the resulting partition is denoted as $\mathcal{P}_K^\alpha$. Given the generic partition $\mathcal{P}_K^\alpha$ obtained from the mixed dissimilarity matrix, the proportion of explained inertia for matrix m th is computed as

$$Q_{D_m}(\mathcal{P}_K^\alpha) = 1 - \frac{W_{D_m}(\mathcal{P}_K^\alpha)}{W_{D_m}(\mathcal{P}_1)}$$

and the normalized proportion of explained inertia, with respect to the case in which only variable p is used, is computed as

$$\tilde{Q}_{D_m}(\mathcal{P}_K^\alpha) = \frac{Q_{D_m}(\mathcal{P}_K^\alpha)}{Q_{D_m}(\mathcal{P}_K^{\alpha_p=1})}$$

Recall that these quantities are both included between 0 and 1 for the same reasons and considerations as in the case of two dissimilarity matrices.

To retrieve the equation of the Joint Inertia for variable m we need to focus separately on the inertia explained in the m th dissimilarity matrix and the overall inertia explained in the others dissimilarity matrices different from m . Let $D_m = [d_{m,ij}]_{i,j=1, \dots, n}$ be the dissimilarity matrix for variable m across the n units and let $D_{-m} = [d_{p,ij}]_{p=1, \dots, m-1, m+1, \dots, P; i,j=1, \dots, n}$ be the dissimilarity matrices containing the distances across the $P-1$ time series different from m . Moreover, let us define $\alpha' | \alpha_m = 0$ as the vector of weights that maximizes the weighted average explained inertia from all the dissimilarity matrices excluding m th matrix and the resulting clustering partition as $\mathcal{P}_K^{\alpha' | \alpha_m=0}$. Similarly, define the partition obtained by using only the information contained in the m th matrix (i.e., by fixing $\alpha_m = 1$) as $\mathcal{P}_K^{\alpha_m=1}$.

If we denote the total inertia in D_{-m} as $W_{D_{-m}}(\mathcal{P}_1)$, we can compute the following quantities

$$\bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha) = 1 - \frac{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^\alpha)}{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1)} \quad \bar{Q}_{D_{-m}}(\mathcal{P}_K^{\alpha' | \alpha_m=0}) = 1 - \frac{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^{\alpha' | \alpha_m=0})}{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1)}$$

which represent the weighted average proportion of explained inertia in the dissimilarity matrices excluding the m th matrix, when considering the partition obtained including the m th matrix, thus the partition resulting from the optimal weights $\underline{\alpha}$ (left formula) and the weighted average proportion of explained inertia in the dissimilarity matrices excluding the m th matrix, when considering the partition obtained excluding the m th matrix, thus the partition resulting from the optimal weights $\underline{\alpha}' | \alpha_m = 0$ (right formula).⁹

The normalized proportion of explained inertia in the non-spatial combined dissimilarity matrix can be computed as

$$\tilde{Q}_{D_{-m}}(\mathcal{P}_K^\alpha) = \frac{\bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha)}{\bar{Q}_{D_{-m}}(\mathcal{P}_K^{\alpha' | \alpha_m=0})}$$

By definition, all these quantities are bounded between 0 and 1. Indeed, If we employ one of the procedures in [Chavent et al. \(2018\)](#) or [Morelli et al. \(2025\)](#) to compute the optimal weighting vector for the combination of inertia from all the dissimilarity

⁷ In the main manuscript, we focused on the specific case of the Joint Inertia for the spatial dissimilarity matrix, that we assumed to be the P th variable. Hence, the results in the Appendix can be related to the results in the main manuscript by assuming that $m = P$.

⁸ Notice that, although the grid can be either regular or irregular, here we consider a regular grid with constant step $\Delta\alpha$.

⁹ Notice that, when m is the spatial matrix, the first formula corresponds to the proportion of explained inertia in the non-spatial dissimilarity matrices when including the spatial component, while the second formula is the proportion of explained inertia induced by the non-spatial dissimilarity matrices when excluding the spatial component.

matrices different from m (i.e., $\alpha^*|\alpha_m = 0$ subject to $\sum_{p \neq m} \alpha_p = 1$), then, the corresponding convex combination of inertia from dissimilarity matrices (i.e., $D_{-m}(\alpha^*|\alpha_m = 0)$) will explain as much as possible the dissimilarities in the D_m matrix, thus being $Q_{D_{-m}}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})$ greater than any other potential $Q_{D_{-m}}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})$.

Finally, considering a generic combination of α and K , the Joint Inertia for variable m when the number of dissimilarity matrices is $P > 2$ is given by

$$JI_m(\mathcal{P}_K^\alpha) = \bar{Q}_{D_m}(\mathcal{P}_K^\alpha) - (1 - \bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha)) = \frac{\bar{Q}_{D_m}(\mathcal{P}_K^\alpha)}{\bar{Q}_{D_m}(\mathcal{P}_K^{\alpha_m=1})} - \left(1 - \frac{\bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha)}{\bar{Q}_{D_{-m}}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})}\right),$$

which can be expressed in term of dissimilarities as follows

$$\begin{aligned} JI_m(\mathcal{P}_K^\alpha) &= \bar{Q}_{D_m}(\mathcal{P}_K^\alpha) - (1 - \bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha)) \\ &= \frac{\bar{Q}_{D_m}(\mathcal{P}_K^\alpha)}{\bar{Q}_{D_m}(\mathcal{P}_K^{\alpha_m=1})} - \left(1 - \frac{\bar{Q}_{D_{-m}}(\mathcal{P}_K^\alpha)}{\bar{Q}_{D_{-m}}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})}\right) \\ &= \frac{1 - \frac{W_{D_m}(\mathcal{P}_K^\alpha)}{W_{D_m}(\mathcal{P}_1)}}{1 - \frac{W_{D_m}(\mathcal{P}_K^{\alpha_m=1})}{W_{D_m}(\mathcal{P}_1)}} - \left(1 - \frac{1 - \frac{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^\alpha)}{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1)}}{1 - \frac{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})}{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1)}}\right) \\ &= \frac{W_{D_m}(\mathcal{P}_1) - W_{D_m}(\mathcal{P}_K^\alpha)}{W_{D_m}(\mathcal{P}_1) - W_{D_m}(\mathcal{P}_K^{\alpha_m=1})} - \left(1 - \frac{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1) - \sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^\alpha)}{\sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_1) - \sum_{p=1, p \neq m}^P W_{D_p}(\mathcal{P}_K^{\alpha^*|\alpha_m=0})}\right) \\ &= \frac{W_{D_m}(\mathcal{P}_1) - W_{D_m}(\mathcal{P}_K^\alpha)}{W_{D_m}(\mathcal{P}_1) - W_{D_m}(\mathcal{P}_K^{\alpha_m=1})} - \left(1 - \frac{\sum_{p=1, p \neq m}^P (W_{D_p}(\mathcal{P}_1) - W_{D_p}(\mathcal{P}_K^\alpha))}{\sum_{p=1, p \neq m}^P (W_{D_p}(\mathcal{P}_1) - W_{D_p}(\mathcal{P}_K^{\alpha^*|\alpha_m=0}))}\right) \\ &= \frac{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{m,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{m,ij}^2}{\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{m,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^{\alpha_m=1}} \sum_{j \in C_k^{\alpha_m=1}} w_i w_j d_{m,ij}^2} \\ &\quad - \left(1 - \frac{\sum_{p=1, p \neq m}^P (\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{p,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} w_i w_j d_{p,ij}^2)}{\sum_{p=1, p \neq m}^P (\frac{1}{2\mu} \sum_i \sum_j w_i w_j d_{p,ij}^2 - \frac{1}{2\mu_k} \sum_{i \in C_k^{\alpha'}} \sum_{j \in C_k^{\alpha'}} w_i w_j d_{p,ij}^2)}\right). \end{aligned}$$

As in the two-matrix case, if the contribution of D_m is not useful, the resulting partition captures inertia from the other dissimilarity matrices but not from D_m (or vice versa). Consequently, the normalized proportion of explained inertia for D_m will be close to zero (though strictly positive), while the normalized proportion for the weighted average of all other matrices will be close to one. Conversely, if the partition from D_m alone and the partition from the combination that excludes D_m both coincide with the partition from the full combination, thus $\mathcal{P}_K^{\alpha_m=1} = \mathcal{P}_K^{\alpha^*|\alpha_m=0} = \mathcal{P}_K^\alpha$, then the Joint Inertia equals 1 (its maximum), indicating that D_m alone suffices to reproduce the same sum of squared dissimilarities captured by the full set of matrices.

Appendix B. Spatial dependence over time

In this appendix, we provide a brief overview of spatial dependence over time via annual Moran's I ; although the main paper operates on dissimilarities (not correlations), which are conceptually distinct, we consider this descriptive analysis complementary and helpful for readers to gauge the strength and evolution of spatial structure in the data. We compute the global Moran's I annually (1990–2022) for NUTS-2 emissions, separately by each gas and by each sector. We used as spatial-weights matrix $\mathbf{W}$ the k -nearest-neighbor. Significance is assessed via 999 permutations. We provide the results with $k = 6$ and, as a robustness check, we repeated the analysis with $k = 10$ and $k = 15$ obtaining very similar values over time (results are available upon request to the authors).

In Fig. B.1, N_2O shows a marked increase in spatial autocorrelation in the mid-2000s (approximately 2005–2013), while F-gases decline toward near-zero spatial dependence after the early 2010s.

Fig. B.2 Industry and Agriculture exhibit persistent, relatively high spatial autocorrelation; Energy is moderate; Buildings, Transport, and Waste are lower, with Transport showing a mild uptick in the late 2010s.

These diagnostics document sustained — but heterogeneous — spatial dependence across regional GHG emissions. They motivate including a spatial component in our clustering framework and help interpret the partitions.

Appendix C. Non-spatial clustering

With the aim of better illustrating the utility of the spatiotemporal clustering methodology employed and emphasizing the role of the spatial component — central to this paper — in this appendix, we present the results obtained from performing the cluster analysis without incorporating the spatial component. This comparison underscores the added value of integrating spatial information in identifying meaningful patterns and dynamics in the data.

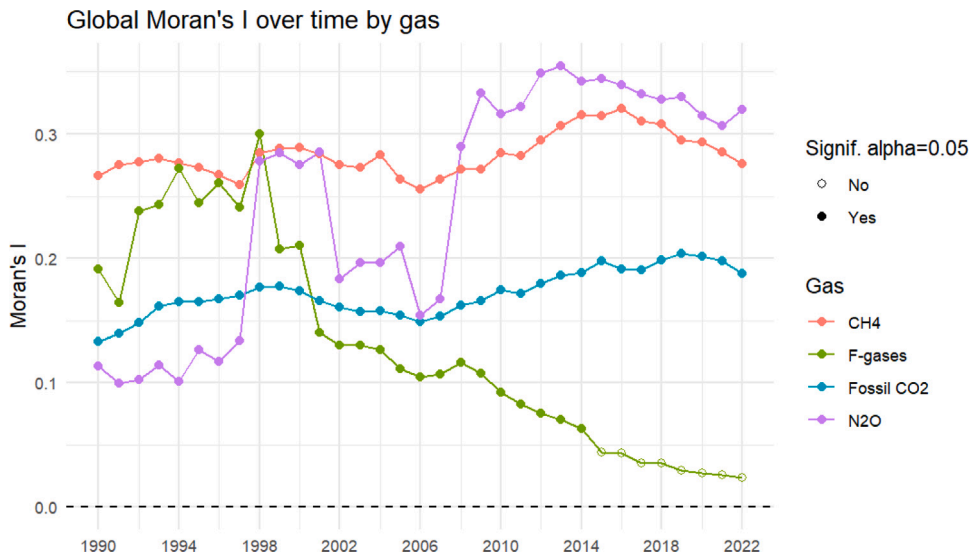


Fig. B.1. Global Moran's I for NUTS-2 emissions by gas (CH_4 , F-gases, Fossil CO_2 , N_2O), computed for each year from 1990 to 2022. Spatial weights W are k -nearest-neighbor with $k = 6$. Filled dots indicate values significantly different from zero at the 95% level (permutation test, $p < 0.05$).

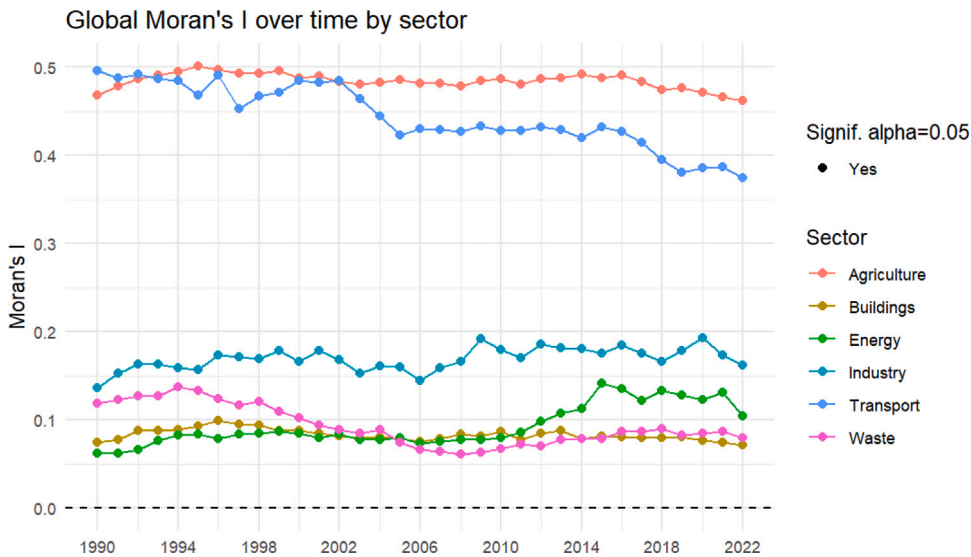


Fig. B.2. Global Moran's I for NUTS-2 emissions by sector (Agriculture, Buildings, Energy, Industry, Transport, Waste), computed for each year from 1990 to 2022. Spatial weights W are k -nearest-neighbor with $k = 6$. Filled dots indicate values significantly different from zero at the 95% level (permutation test, $p < 0.05$).

C.1. Temporal clustering for emissions by gas

Considering dissimilarity matrices D_{CH_4} , $D_{\text{F-gases}}$, D_{CO_2} and $D_{\text{N}_2\text{O}}$, as defined in Section 5.2, we set $\Delta\alpha = 0.05$ and $K = 4$ and we obtain $\alpha_{4,p}^* = (0.1, 0.35, 0.15, 0.4)$ as the optimal weighting vector for the mixed inertia from the dissimilarity matrices. In Fig. C.3 we show a map of the regions belonging to different clusters and in Fig. C.4 we provide the average annual emission in each cluster by gas.

The clustering results exhibit a high degree of spatial overlap, yet discernible spatial patterns remain evident. Notably, regions in clusters 1 and 2, which demonstrate the poorest performance, include Belgium, the Netherlands, and certain metropolitan areas in Northern Europe. These regions display a decreasing trend in emissions for the gases with the highest levels, namely D_{CH_4} and D_{CO_2} . In contrast, cluster 3 consists of sparsely populated regions in Central Europe with significantly lower emissions than the

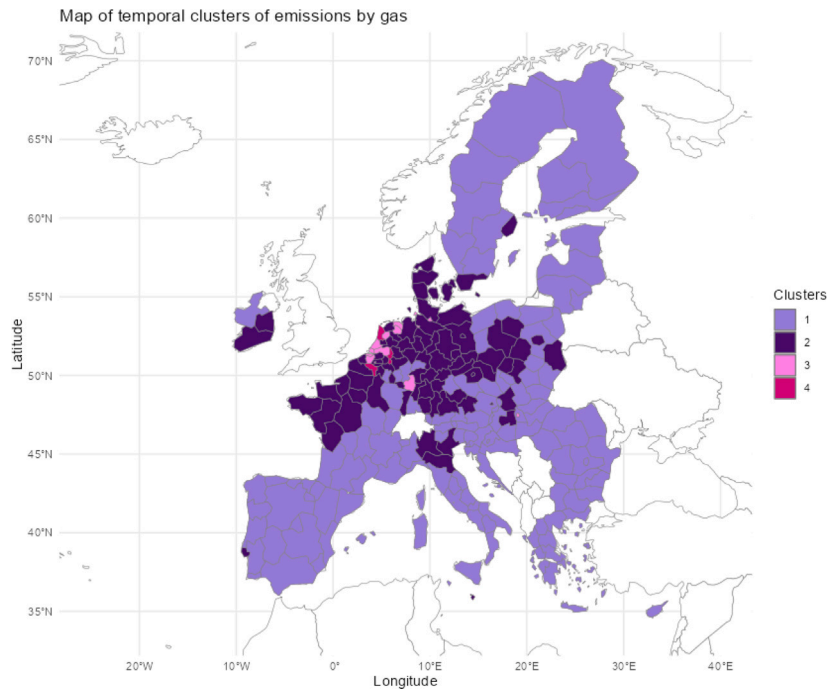


Fig. C.3. Map of temporal clusters obtained combining the inertia from the dissimilarity matrices of the time series of emissions of different gases between 1990 and 2022 that is D_{CH_4} , $D_{F-gases}$, D_{CO_2} and D_{N_2O} .

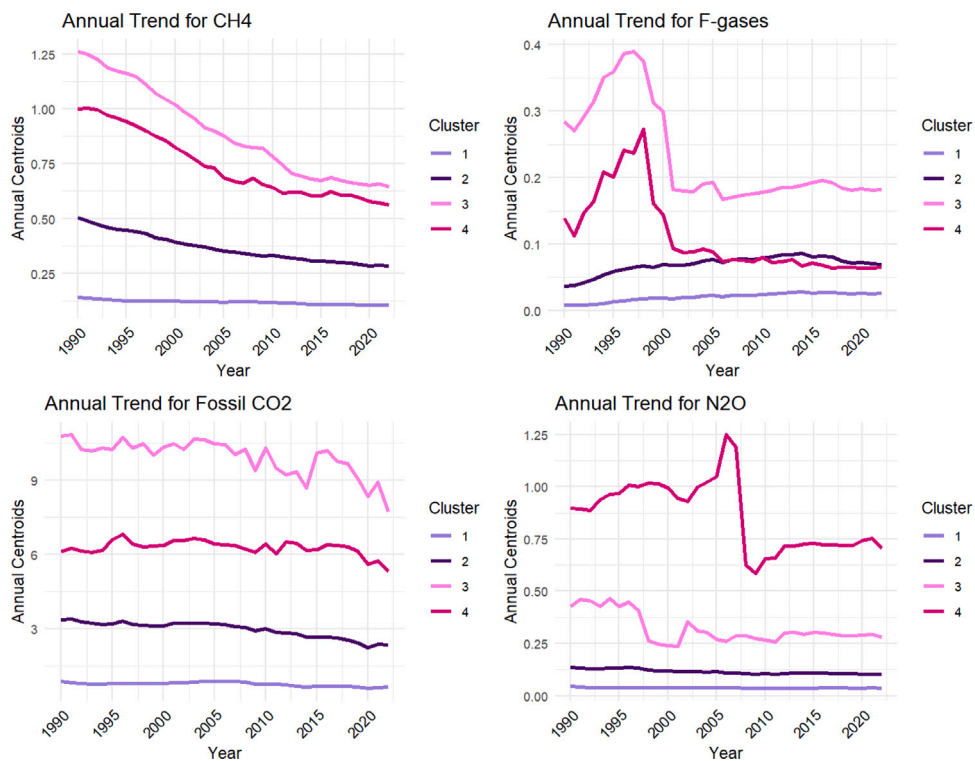


Fig. C.4. Annual average of emission per km² of sectors for temporal clusters obtained combining the inertia from the dissimilarity matrices of the time series of emissions of different gases between 1990 and 2022 that is D_{CH_4} , $D_{F-gases}$, D_{CO_2} and D_{N_2O} .

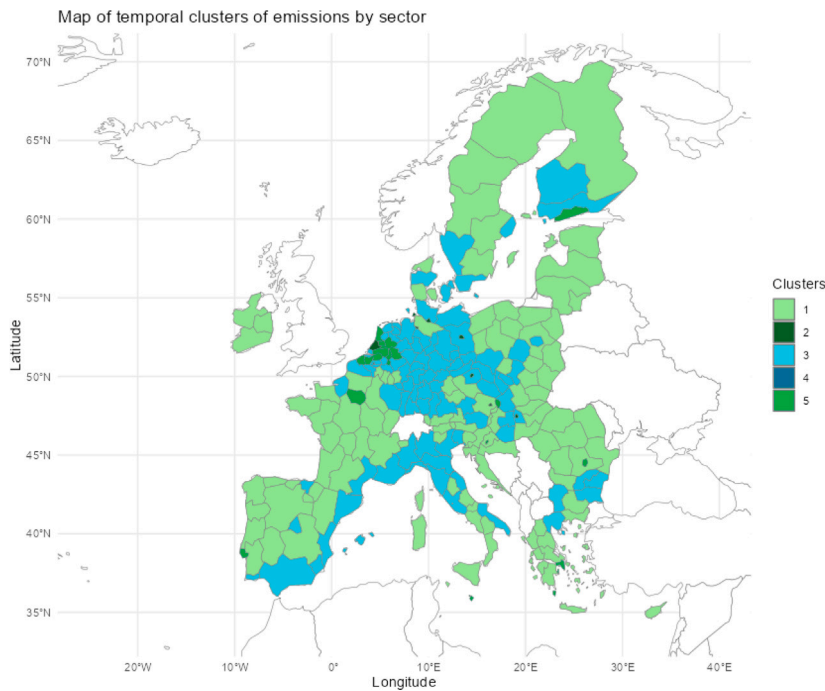


Fig. C.5. Map of temporal clusters obtained combining the inertia from the dissimilarity matrices of the time series of emissions of different sectors between 1990 and 2022, that is D_{Agri} , D_{Build} , D_{Energy} , D_{Ind} , D_{Tran} and D_{Waste} .

previous clusters. Lastly, the cluster encompassing the largest number of regions, covering over two-thirds of the analyzed area, is characterized by remarkably low emissions and appears relatively stable over time.

C.2. Temporal clustering for emissions by sector

Moving to the emissions by sectors, we consider dissimilarity matrices D_{Agri} , D_{Build} , D_{Energy} , D_{Ind} , D_{Tran} and D_{Waste} , as defined in Section 5.3, we set $\Delta\alpha = 0.1$ and $K = 5$ and we obtain $\alpha_{5,p}^* = (0.3, 0, 0.2, 0, 0.3, 0.2)$ as the optimal weighting vector for the mixed inertia from the dissimilarity matrices. In Fig. C.5 we provide a map of the regions belonging to different clusters and in Fig. C.6 we describe the average annual emission in each cluster by gas.

From the results, we can immediately see that, even without incorporating the spatial component, there is still a degree of spatial overlap in the clustering outcomes. Notably, regions with particularly low performance are concentrated within the European Union. Specifically, Cluster 1, the largest in terms of geographic area, exhibits very low emission levels across all sectors, with no discernible increasing or decreasing trends. Cluster 2 encompasses industrial regions in Northern Europe, characterized by high emissions in the Energy and Industry sectors, which show a declining trend over time. Cluster 3 is associated with low, though not minimal, emission levels, accompanied by a slight downward trend. Cluster 4 includes several metropolitan regions and displays the highest emission levels in the Buildings, Transport, and Waste sectors. Lastly, Cluster 5, primarily located in Northern Europe and Northern Italy, stands out as the largest emitter in the Agricultural sector.

Overall, our analysis confirms the significant relevance of the spatial component. Regions that are geographically close tend to exhibit similar emission patterns, both by gas type and sector. Incorporating the spatial component explicitly into the clustering algorithm produces groups that are more interpretable in terms of geographic proximity. This leads to results that are particularly valuable in environmental statistics, where spatiotemporal approaches and methodologies are essential for understanding and addressing complex phenomena.

Appendix D. Internal validation of clustering results

To strengthen both the methodological grounding and the practical evaluation, we provide an internal validation of the clustering partitions. For each set of dissimilarities (by gas and by sector), we compare the partition obtained with the proposed spatiotemporal approach to those produced in the purely temporal and purely spatial cases. We compute four standard indices with `fpc::cluster.stats` in R: (i) the average silhouette, which summarizes how well each unit fits its assigned cluster relative to the nearest alternative (Kaufman and Rousseeuw, 1990; Rousseeuw, 1987); (ii) the Calinski–Harabasz (CH) index (Caliński and Harabasz, 1974; Hennig, 2014), which weighs between-cluster dispersion against within-cluster dispersion; (iii) the within/between

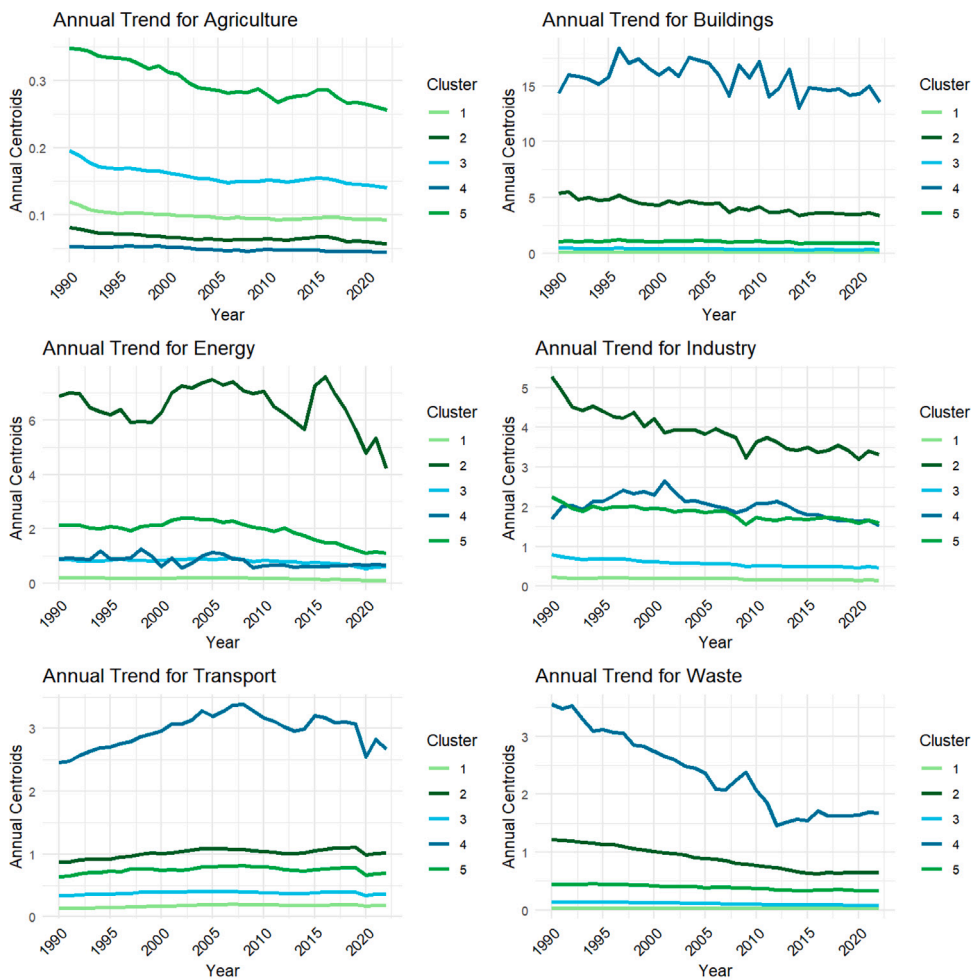


Fig. C.6. Annual average of emission per km^2 of sectors for temporal clusters obtained combining the inertia from the dissimilarity matrices of the time series of emissions of different sectors between 1990 and 2022, that is D_{Agri} , D_{Build} , D_{Energy} , D_{Ind} , D_{Tran} and D_{Waste} .

(WB) ratio of average distances, a compact measure of cluster separation versus compactness; and (iv) Pearson's gamma, a rank correlation between pairwise dissimilarities and a same/different-cluster indicator that gauges the monotonic association between the clustering and the dissimilarity structure (Halkidi et al., 2001).

The average silhouette quantifies, for each unit, how close it is to its own cluster compared with the nearest alternative; values lie in $[-1, 1]$, with larger values indicating better cohesion and separation (Kaufman and Rousseeuw, 1990; Rousseeuw, 1987). In Fig. D.7a and Fig. D.8a, silhouettes are generally modest — as expected for high-dimensional time-series dissimilarities — but the blue bars (spatiotemporal partition) typically fall between the orange (temporal-only) and green (spatial-only) bars. This confirms that the spatiotemporal solution strikes a compromise: it inherits part of the cohesion visible when clustering on temporal features while retaining some spatial structure, whereas the purely spatial solution yields good cohesion only with respect to the spatial dissimilarity.

The Calinski-Harabasz (CH) index balances between-cluster dispersion against within-cluster dispersion; higher values imply tighter, well-separated clusters. Panels (b) of Figs. D.7 and D.8 show that temporal-only partitions tend to score higher when evaluated on feature-based dissimilarities, while the spatial-only solution dominates when the spatial dissimilarity is used. The spatiotemporal partition lies in between, reinforcing its role as a balanced allocation.

The within/between (WB) ratio summarizes compactness versus separation as the ratio of average within-cluster to average between-cluster dissimilarities (smaller values are preferable). In Figs. D.7c and D.8c, the blue bars again occupy an intermediate position between the temporal-only and spatial-only cases, consistent with a partition that trades off feature cohesion and geographic proximity.

Finally, Pearson's gamma measures the monotone association between pairwise dissimilarities and the same/different-cluster indicator (Halkidi et al., 2001). Larger values indicate that pairs declared “in different clusters” are, in rank terms, more dissimilar

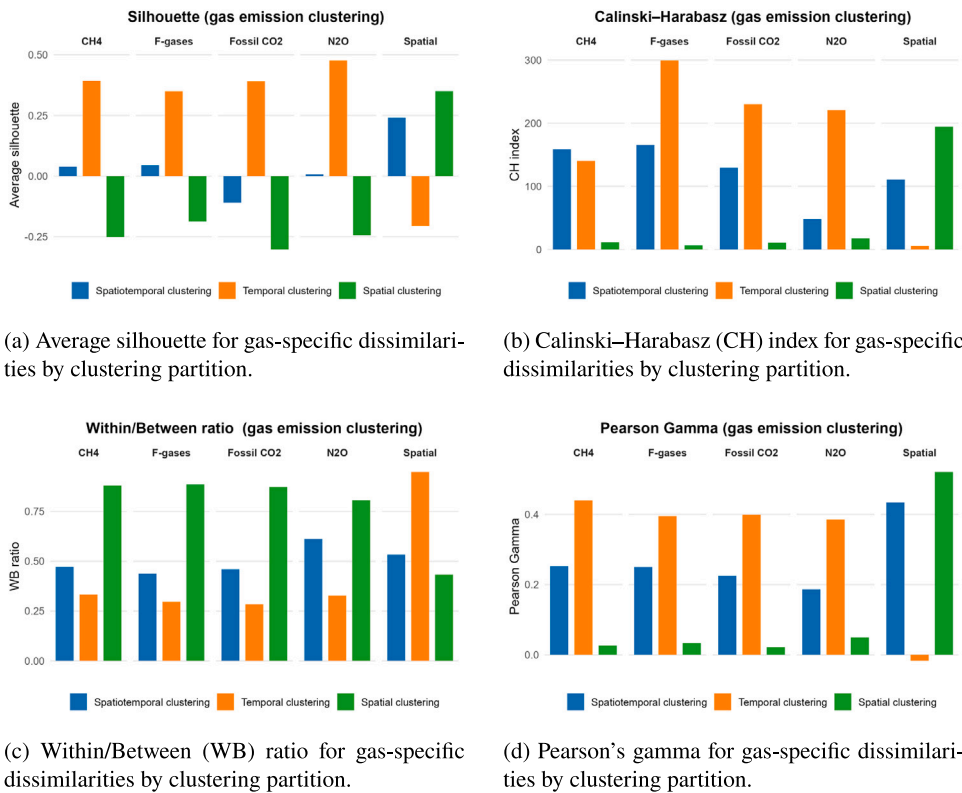


Fig. D.7. Internal validation indices for gas-specific dissimilarities by clustering partition (panels a–d).

than pairs “in the same cluster”. Panels (d) show the same qualitative pattern observed for CH and WB: the temporal-only clustering aligns best with feature-based dissimilarities, the spatial-only clustering aligns best with the spatial dissimilarity, and the spatiotemporal partition achieves a reasonable balance across both information sources.

Taken together, these diagnostics (Figs. D.7–D.8) corroborate the main message of the paper: the spatiotemporal partition provides a principled middle ground between purely temporal and purely spatial solutions, preserving meaningful geographic structure without sacrificing too much feature-based cohesion.

Appendix E. Spatiotemporal clustering of single time series

In this appendix we provide the clustering partitions obtained by combining the inertia from dissimilarity matrix of the emission time series for each gas with the spatial dissimilarity, and the dissimilarity matrix of the emission time series for each sector with the spatial dissimilarity, thus combining only two dissimilarity matrices each time. These results help assess how stable the spatial grouping is across environmental dimensions and where gas- or sector-specific patterns emerge.

In Fig. E.9 CO₂, F-gases, and CH₄ display a similar backbone: three large, spatially compact clusters that roughly separate Western/South-Western, Central, and Eastern Europe, plus a fourth, partially overlapping cluster concentrated in the Northern/Central belt that gathers cross-country regions with comparable temporal trajectories. By contrast, N₂O departs from this pattern and yields four compact, contiguous clusters with much less overlap, producing a sharper spatial segmentation.

According to Table E.1, F-gases, CH₄, and Fossil CO₂ exhibit high Joint Inertia (0.70, 0.60, and 0.60, respectively), indicating that the resulting partitions capture a large proportion of the dissimilarity from the time-series and spatial dissimilarities. By contrast, N₂O shows a low Joint Inertia (0.25), meaning the partition is not able to capture the inertia from both the spatial dissimilarity and the N₂O time series. The selected α values differ across gases, underscoring that the trade-off between temporal features and space is gas-specific.

In Fig. E.10, sectoral partitions show a common backbone: four compact groups that are largely similar across sectors, plus a fifth, overlapping cluster that typically sits in Northern/Central Europe and often captures metropolitan/high-density corridors. Despite this shared structure, boundaries and memberships differ slightly by sector—most visibly for Energy and Industry, which deviate more from the common pattern. Transport, Buildings, and Waste each display four markedly contiguous blocks plus a distinct cluster that typically includes metropolitan and other high-density regions in Central Europe. Agriculture shows a similar pattern—four compact clusters and another grouping that includes agricultural regions in Northern Europe. By contrast, Energy and Industry

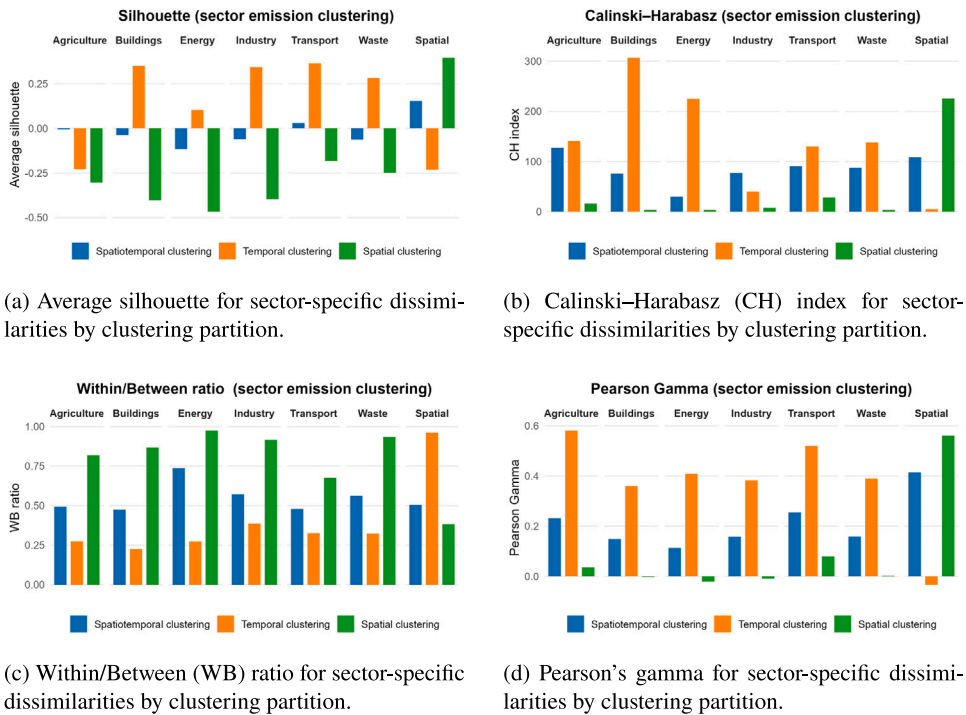


Fig. D.8. Internal validation indices for sector-specific dissimilarities by clustering partition (panels a–d).

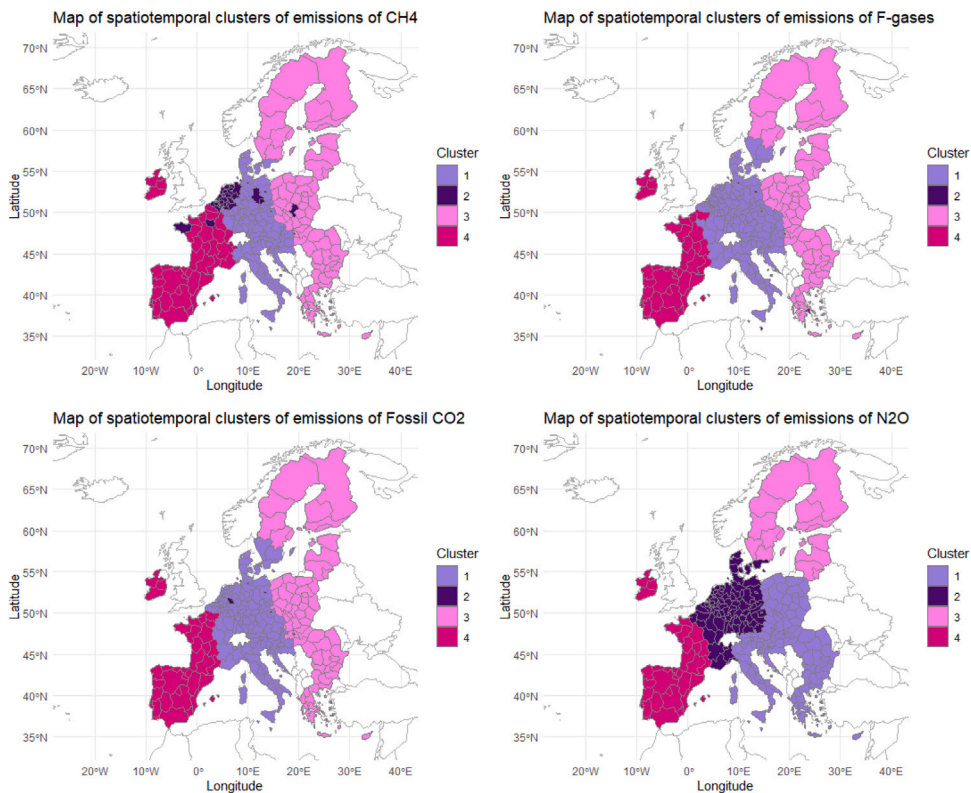


Fig. E.9. Map of spatiotemporal clusters obtained combining the inertia from dissimilarity matrices of the time series of emissions of each gas between 1990 and 2022 with the spatial dissimilarities. We compute the optimal α of each combination according to [Morelli et al. \(2025\)](#), setting $K = 4$.

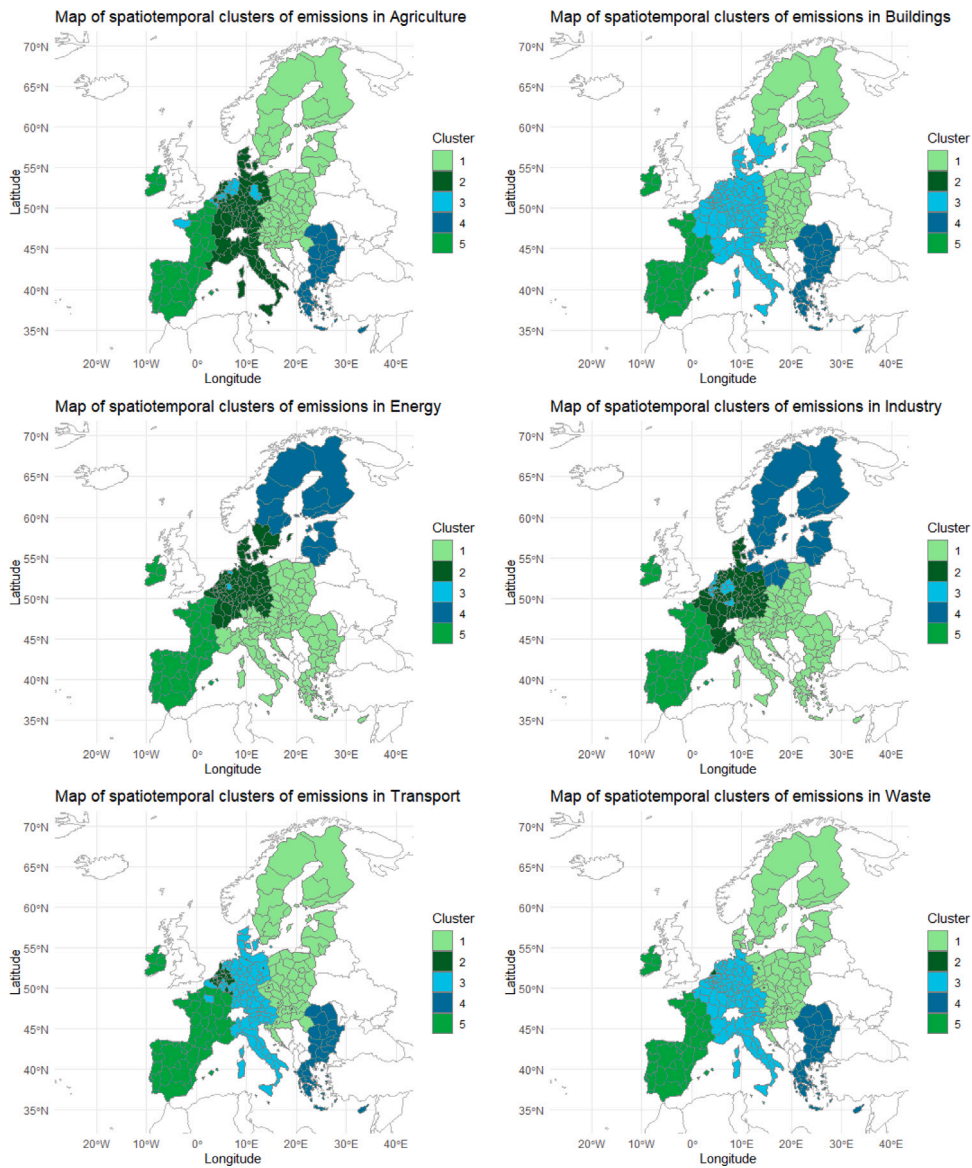


Fig. E.10. Map of spatiotemporal clusters obtained combining the inertia from dissimilarity matrix of the time series of emissions by each sector between 1990 and 2022 with the inertia from spatial dissimilarity matrix. We compute the optimal α of each combination according to [Morelli et al. \(2025\)](#), setting $K = 5$.

Table E.1

Summary of the α^* weights of the feature dissimilarities in the spatiotemporal clustering of the different partitions obtained combining the inertia from dissimilarity matrix of the time series of emissions by each gas between 1990 and 2022 with the inertia from spatial dissimilarity matrix, setting $K = 4$. $J I(P_K^{\alpha^*})$ is the Joint Inertia resulting from each partition. Note that the joint inertia when considering only two dissimilarity matrices, remains the same whether it is calculated with respect to one matrix or the other dissimilarity matrix.

	D_{CH_4}	$D_{\text{F-gases}}$	D_{CO_2}	$D_{\text{N}_2\text{O}}$
α	0.60	0.65	0.40	0.40
$J I(P_K^{\alpha^*})$	0.58	0.70	0.60	0.25

Table E.2

Summary of the α^* weights of the feature dissimilarities in the spatiotemporal clustering of the different partitions obtained combining the inertia from the dissimilarity matrix of emissions in each sector and the inertia from spatial dissimilarity matrix, setting $K = 5$. $JI(P_K^{\alpha^*})$ is the Joint Inertia resulting from each partition. Note that the joint inertia when considering only two dissimilarity matrices, remains the same whether it is calculated with respect to one matrix or the other dissimilarity matrix.

	D_{Agri}	D_{Build}	D_{Energy}	D_{Ind}	D_{Tran}	D_{Waste}
α	0.50	0.50	0.45	0.70	0.65	0.70
$JI(P_K^{\alpha^*})$	0.65	0.63	0.54	0.55	0.73	0.60

present a different structure, with a compact cluster spanning Germany, the Netherlands, Denmark, and eastern France, alongside a smaller cluster of regions that spatially overlaps the former.

In Table E.2 Transport shows the highest Joint Inertia (0.73), pointing to a strong role of geography with respect to the dissimilarity in the time series of Transport emission, followed by Agriculture, Buildings and Waste, respectively 0.65, 0.63, 0.60. Energy and Industry display more moderate interplay between spatial information and feature information (0.54–0.55). The selected α values differ across sectors, indicating sector-specific balances between temporal features and spatial proximity.

Overall, the partitions are not identical across gases and sectors but they share a common macro-regional backbone. Moreover, the Joint Inertia values and confirm that the relevance of spatial information is dimension-specific.

Data and code availability

Data used in the paper are public (source JRC-ARDECO). Since we use only public data, no special permission is need to use copyrighted material from other sources (including the Internet). All results presented in this paper can be reproduced using the R statistical software. The codes were developed entirely by the authors. For reproducibility purposes, code and data are made public through the following GitHub folder: https://github.com/PaoloMaranzano/CM_PM_PO_STClusterGHG.git.

References

- Ansari, M.Y., Ahmad, A., Khan, S.S., Bhushan, G., Mainuddin, 2020. Spatiotemporal clustering: a review. *Artif. Intell. Rev.* 53 (4), 2381–2423. <http://dx.doi.org/10.1007/s10462-019-09736-1>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068956432&doi=10.1007%2fs10462-019-09736-1&partnerID=40&md5=e6852a2faff860d4938e522e3ebd9554>.
- Banja, M., Crippa, M., Guizzardi, D., Muntean, M., Pagani, F., Pisoni, E., 2025. A comparative analysis of EDGAR and UNFCCC GHG emissions inventories: insights on trends, methodology and data discrepancies. *Earth Syst. Sci. Data Discuss.* 2025, 1–39. <http://dx.doi.org/10.5194/essd-2025-385>, URL <https://essd.copernicus.org/preprints/essd-2025-385/>.
- Benevento, A., Durante, F., 2024. Correlation-based hierarchical clustering of time series with spatial constraints. *Spat. Stat.* 59, <http://dx.doi.org/10.1016/j.spasta.2023.100797>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179118911&doi=10.1016%2fs1016j.spasta.2023.100797&partnerID=40&md5=4ff3edcc09735e57aed7a404e27f932>.
- Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: *AAAIWS'94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. pp. 359–370.
- Birant, D., Kut, A., 2007. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* 60 (1), 208–221. <http://dx.doi.org/10.1016/j.datak.2006.01.013>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33846024749&doi=10.1016%2fs1016j.datak.2006.01.013&partnerID=40&md5=8594c7a23a768f20115490fc091be556>.
- Bourgault, G., Marcotte, D., Legendre, P., 1992. The multivariate (co)variogram as a spatial weighting function in classification methods. *Math. Geol.* 24 (5), 463–478. <http://dx.doi.org/10.1007/BF00890530>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0026482068&doi=10.1007%2fbf00890530&partnerID=40&md5=3976146f3363dd6cf4447a474bb7d0f>.
- Bucci, A., Ippoliti, L., Valentini, P., 2023. Analysing spatiotemporal patterns of COVID-19 confirmed deaths at the NUTS-2 regional level. *Reg. Stat.* 13 (2), 214–239. <http://dx.doi.org/10.15196/RS130202>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85153956522&doi=10.15196%2fRS130202&partnerID=40&md5=e0caf54c37f055fba4f7c0586f2ae93>.
- Bucciarelli, G., 2024. ARDECO: Annual regional database of the European commission (ARDECO). URL <https://CRAN.R-project.org/package=ARDECO>. R package version 2.1.0.
- Calculi, C., Fassò, A., Finazzi, F., Pollice, A., Turnone, A., 2015. Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy. *Environmetrics* 26 (6), 406–417. <http://dx.doi.org/10.1002/env.2345>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2345>.
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat.* 3 (1), 1–27. <http://dx.doi.org/10.1080/03610927408827101>, URL <https://www.tandfonline.com/doi/abs/10.1080/03610927408827101>.
- Cerqueti, R., Maranzano, P., Mattera, R., 2025. Spatially-clustered spatial autoregressive models with application to agricultural market concentration in Europe. *J. Agric. Biol. Environ. Stat.* <http://dx.doi.org/10.1007/s13253-024-00672-4>.
- Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J., 2018. ClustGeo: an R package for hierarchical clustering with spatial constraints. *Comput. Statist.* 33, 1–24. <http://dx.doi.org/10.1007/s00180-018-0791-1>.
- Chen, X.C., Faghmous, J.H., Khandelwal, A., Kumar, V., 2015. Clustering dynamic spatio-temporal patterns in the presence of noise and missing data. In: *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2015-January, pp. 2575–2581, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84949815953&partnerID=40&md5=82c802d091285adb0ac115c82d0caa44>.
- Crippa, M., Guizzardi, D., Solazzo, E., Muntean, M., Schaaf, E., Monforti-Ferrario, F., Banja, M., Olivier, J., Grassi, G., Rossi, S., 2024. GHG Emissions of All World Countries - 2024. Report, Publications Office of the European Union, Luxembourg, http://dx.doi.org/10.2760/40028977_JRC138862, URL <https://publications.jrc.ec.europa.eu/repository/handle/JRC138862>.

- Deb, S., Karmakar, S., 2023. A novel spatio-temporal clustering algorithm with applications on COVID-19 data from the United States. *Comput. Statist. Data Anal.* 188, <http://dx.doi.org/10.1016/j.csda.2023.107810>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85162867311&doi=10.1016%2fj.csda.2023.107810&partnerID=40&md5=d7a1715870a1ad5cbf14818729e88f39>.
- D'Urso, P., Vitale, V., 2020. A robust hierarchical clustering for georeferenced data. *Spat. Stat.* 35, <http://dx.doi.org/10.1016/j.spasta.2020.100407>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077915270&doi=10.1016%2fj.spasta.2020.100407&partnerID=40&md5=9cd3820014642ad9a4d5a3c0b4896221>.
- Ester, M., Kriegl, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings - 2nd International Conference on Knowledge Discovery and Data Mining. KDD 1996*, pp. 226–231, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85170282443&partnerID=40&md5=bbefa1633035b0779ae320a55ac5323a>.
- European Commission, 2016. NUTS – nomenclature of territorial units for statistics. <https://ec.europa.eu/eurostat/web/nuts>.
- Fouedjio, F., 2016. A hierarchical clustering method for multivariate geostatistical data. *Spat. Stat.* 18, 333–351. <http://dx.doi.org/10.1016/j.spasta.2016.07.003>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84995642372&doi=10.1016%2fj.spasta.2016.07.003&partnerID=40&md5=6894eeb07594509a3a2488f606c63f5c>.
- Georgoulas, G., Konstantaras, A., Katsifarakis, E., Stylios, C., Maravelakis, E., Vachtsevanos, G., 2013. "Seismic-mass" density-based algorithm for spatio-temporal clustering. *Expert Syst. Appl.* 40 (10), 4183–4189. <http://dx.doi.org/10.1016/j.eswa.2013.01.028>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84875370729&doi=10.1016%2fj.eswa.2013.01.028&partnerID=40&md5=1264bd9bbf5f98ef8dd97a4bd069cf9e>.
- Gordon, A.D., 1999. *Classification*. CRC Press.
- Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. *J. Intell. Inf. Syst.* 17 (2–3), 107–145. <http://dx.doi.org/10.1023/A:1012801612483>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035676057&doi=10.1023%2fA%3a1012801612483&partnerID=40&md5=39cbe04ae9f020af717bd2dcf23c32c9>.
- Hennig, C., 2014. How many bee species? A case study in determining the number of clusters. *Stud. Classification Data Anal. Knowledge Organ.* 47, 41–49. http://dx.doi.org/10.1007/978-3-319-01595-8_5, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84951805221&doi=10.1007%2f978-3-319-01595-8_5&partnerID=40&md5=a5b6e129245d7f69d416d4617ebef073.
- Jaya, I.G.N.M., Folmer, H., 2020. Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia. *J. Geogr. Syst.* 22 (1), 105–142. <http://dx.doi.org/10.1007/s10109-019-00311-4>.
- Jaya, I.G.N.M., Folmer, H., 2021. Identifying spatiotemporal clusters by means of agglomerative hierarchical clustering and Bayesian regression analysis with spatiotemporally varying coefficients: Methodology and application to dengue disease in Bandung, Indonesia. *Geogr. Anal.* 53 (4), 767–817. <http://dx.doi.org/10.1111/gean.12264>.
- Jaya, I., Ruchjana, B., Andriyana, Y., Agata, R., 2019. Clustering with spatial constraints: The case of diarrhea in Bandung city, Indonesia. *J. Phys.: Conf. Ser.* 1397 (1), <http://dx.doi.org/10.1088/1742-6596/1397/1/012068>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078401440&doi=10.1088%2f1742-6596%2f1397%2f1%2f012068&partnerID=40&md5=ff7709958e63de22a0a3fa07a0711055>.
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, United States.
- Maranzano, P., Fassó, A., 2022. The Impact of the Lockdown Restrictions on Air Quality During COVID-19 Pandemic in Lombardy, Italy. Springer, pp. 343–374. http://dx.doi.org/10.1007/978-3-031-07155-3_15, URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85150925207&doi=10.1007%2f978-3-031-07155-3_15&partnerID=40&md5=596ef5a3bd958120c5e17c129fe5941e.
- Maranzano, P., Fassó, A., Pelagatti, M., Modelsee, M., 2020. Statistical modeling of the early-stage impact of a new traffic policy in Milan, Italy. *Int. J. Environ. Res. Public Health* 17 (3), <http://dx.doi.org/10.3390/ijerph17031088>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079309659&doi=10.3390%2fijerph17031088&partnerID=40&md5=df18f8e1769457e5d4cd6f9df25200b>.
- Maranzano, P., Otto, P., Fassó, A., 2023. Adaptive LASSO estimation for functional hidden dynamic geostatistical models. *Stoch. Environ. Res. Risk Assess.* <http://dx.doi.org/10.1007/s00477-023-02466-5>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85169242030&doi=10.1007%2fs00477-023-02466-5&partnerID=40&md5=b0290bc5b274de3e33e24f49ad2c3e16>.
- Maranzano, P., Pelagatti, M., 2024. Spatiotemporal event studies for environmental data under cross-sectional dependence: An application to air quality assessment in Lombardy. *J. Agric. Biol. Environ. Stat.* 29 (1), 147–168. <http://dx.doi.org/10.1007/s13253-023-00564-z>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85167788841&doi=10.1007%2fs13253-023-00564-z&partnerID=40&md5=382c633708b835c3f80aed8c521166eb>.
- Marongiu, A., Collalto, A.G., Distefano, G.G., Angelino, E., 2024. Application of machine learning to estimate ammonia atmospheric emissions and concentrations. *Air 2* (1), 38–60. <http://dx.doi.org/10.3390/air2010003>, URL <https://www.mdpi.com/2813-4168/2/1/3>.
- Marongiu, A., Colombo, L., Collalto, A.G., 2025. Dynamic emission inventory of ammonia in northern Italy by machine learning. *Air Qual. Atmos. Health* <http://dx.doi.org/10.1007/s11869-025-01779-4>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105008884040&doi=10.1007%2fs11869-025-01779-4&partnerID=40&md5=998fc68a8477c2b602ad98111c53cb87>.
- Mattera, R., Frances, P.H., 2023. Are African business cycles synchronized? Evidence from spatio-temporal modeling. *Econ. Model.* 128, <http://dx.doi.org/10.1016/j.econmod.2023.106485>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85169829326&doi=10.1016%2fj.econmod.2023.106485&partnerID=40&md5=2a9dbcc5024b0a4cad6b73062c9b3393>.
- Montero, P., Vilar, J.A., 2014. TScust: An R package for time series clustering. *J. Stat. Softw.* 62 (1), 1–43. <http://dx.doi.org/10.18637/jss.v062.i01>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84920439414&doi=10.18637%2fjss.v062.i01&partnerID=40&md5=505b736bcdfe34c4d63d0739b1d401>.
- Morelli, C., Boccaletti, S., Maranzano, P., Otto, P., 2025. Multidimensional spatiotemporal clustering - An application to environmental sustainability scores in Europe. *Environmetrics* 36 (2), <http://dx.doi.org/10.1002/env.2893>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85216969489&doi=10.1002%2fenv.2893&partnerID=40&md5=a350a71d40b156c892cf7e52579bfff25>.
- Mutascu, M., 2022. CO2 emissions in the USA: new insights based on ANN approach. *Environ. Sci. Pollut. Res.* 29 (45), 68332–68356. <http://dx.doi.org/10.1007/s11356-022-20615-1>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129771831&doi=10.1007%2fs11356-022-20615-1&partnerID=40&md5=306f9052ceca0bc0f6370352f8fe3cd4>.
- Naidu, G., Zuva, T., Sibanda, E.M., 2023. A review of evaluation metrics in machine learning algorithms. In: *Artificial Intelligence Application in Networks and Systems*. Springer International Publishing, pp. 15–25.
- Najafabadi, A.M., Mahaki, B., Hajizadeh, Y., 2020. Spatiotemporal modeling of airborne fine particulate matter distribution in Isfahan. *Int. J. Environ. Health Eng. (IJEHE)* 2020 (July), 1–7. http://dx.doi.org/10.4103/ijehe.ijehe_6_20.
- Oliver, M., Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. *Math. Geol.* 21 (1), 15–35. <http://dx.doi.org/10.1007/BF00897238>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0005852342&doi=10.1007%2fBF00897238&partnerID=40&md5=1f84b2077cd38e299e540a6d4e148a1f>.
- Otto, P., Fusta Moro, A., Rodeschini, J., Shaboviq, Q., Ignaccolo, R., Golini, N., Cameletti, M., Maranzano, P., Finazzi, F., Fassó, A., 2024. Spatiotemporal modelling of PM 2.5 concentrations in Lombardy (Italy): a comparative study. *Environ. Ecol. Stat.* 31 (2), 245–272. <http://dx.doi.org/10.1007/s10651-023-00589-0>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85183928401&doi=10.1007%2fs10651-023-00589-0&partnerID=40&md5=7287355023438b6b6da56eeb4b8cf969>.
- Pértega-Díaz, S., Vilar, J.A., 2010. Comparing several parametric and nonparametric approaches to time series clustering: A simulation study. *J. Classification* 27 (3), 333–362. <http://dx.doi.org/10.1007/s00357-010-9064-6>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871904170&doi=10.1007%2fs00357-010-9064-6&partnerID=40&md5=04e799db318bee7727784b44b52b79a6>.

- Piccolo, D., 1990. A distance measure for classifying arima models. *J. Time Series Anal.* 11 (2), 153–164. <http://dx.doi.org/10.1111/j.1467-9892.1990.tb00048.x>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84981470151&doi=10.1111%2fj.1467-9892.1990.tb00048.x&partnerID=40&md5=fec7891c50b735b7fc531aeba883b87c>.
- Pisoni, E., Zauli-Sajani, S., Belis, C.A., Khomenko, S., Thunis, P., Motta, C., Van Dingenen, R., Bessagnet, B., Monforti-Ferrario, F., Maes, J., Feyen, L., 2025. High resolution assessment of air quality and health in Europe under different climate mitigation scenarios. *Nat. Commun.* 16 (1), <http://dx.doi.org/10.1038/s41467-025-60449-2>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105007185788&doi=10.1038%2fs41467-025-60449-2&partnerID=40&md5=0e12d55e732355e69b53a7e452185bed>.
- R Core Team, 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org/>.
- Rocha, J.A.M., Times, V.C., Oliveira, G., Alvares, L.O., Bogorny, V., 2010. DB-SMoT: A direction-based spatio-temporal clustering method. In: 2010 IEEE International Conference on Intelligent Systems, IS 2010 - Proceedings. pp. 114–119. <http://dx.doi.org/10.1109/IS.2010.5548396>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-77957843307&doi=10.1109%2fis.2010.5548396&partnerID=40&md5=5b971c9788f4b2b97ee31908772d3134>.
- Rodeschini, J., Fassò, A., Finazzi, F., Fusta Moro, A., 2024. Scenario analysis of livestock-related PM2.5 pollution based on a new heteroskedastic spatiotemporal model. *Socio-Economic Plan. Sci.* 96, <http://dx.doi.org/10.1016/j.seps.2024.102053>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85203506126&doi=10.1016%2fj.seps.2024.102053&partnerID=40&md5=3a04c644cdcd15dc2c29d2e45bc3eb33>.
- Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (C), 53–65. [http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7), URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0023453329&doi=10.1016%2f0377-0427%2887%2990125-7&partnerID=40&md5=0b64d7ff4409fc3b6d7d88108ab2d07c>.
- Rowland, O.E., 2024. Comparative analysis of meteorological parameters and their relationship with NO2, PM10, PM2.5 and O3 concentrations at selected urban air quality monitoring stations in Krakow, Paris, and Milan. *Discov. Environ.* 2 (1), <http://dx.doi.org/10.1007/s44274-024-00060-2>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105004913809&doi=10.1007%2fs44274-024-00060-2&partnerID=40&md5=0e1b48ca5a0f7d02c7236acc949b1db5>.
- Sakoe, H., Chiba, S., 1971. A dynamic programming approach to continuous speech recognition. *Comput. Sci. Linguist.* URL <https://api.semanticscholar.org/CorpusID:107516844>.
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* 26 (1), 43–49. <http://dx.doi.org/10.1109/TASSP.1978.1163055>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0017930815&doi=10.1109%2ftassp.1978.1163055&partnerID=40&md5=214a08cd9889c7763a7eb3235d1718c6>.
- Taghavi-Shahri, S.M., Fassò, A., Mahaki, B., Amini, H., 2020. Concurrent spatiotemporal daily land use regression modeling and missing data imputation of fine particulate matter using distributed space-time expectation maximization. *Atmos. Environ.* 224, <http://dx.doi.org/10.1016/j.atmosenv.2019.117202>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079163547&doi=10.1016%2fj.atmosenv.2019.117202&partnerID=40&md5=302b22fc3aad92569abc95ec7d76ade>.
- Tietbohl, A., Bogorny, V., Kuijpers, B., Alvares, L.O., 2008. A clustering-based approach for discovering interesting places in trajectories. In: Proceedings of the ACM Symposium on Applied Computing. pp. 863–868. <http://dx.doi.org/10.1145/1363686.1363886>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-56749150055&doi=10.1145%2f1363686.1363886&partnerID=40&md5=d2a6a24e973393a25daf0fd75145141f>.
- Yu, C., Tan, J., Cheng, Y., Mi, X., 2024. Data analysis and preprocessing techniques for air quality prediction: a survey. *Stoch. Environ. Res. Risk Assess.* 38 (6), 2095–2117. <http://dx.doi.org/10.1007/s00477-024-02693-4>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85187939827&doi=10.1007%2fs00477-024-02693-4&partnerID=40&md5=54de4eac2727ec9c8b5e3d9a32018f5c>.
- Zammarchi, G., Maranzano, P., 2024. Mapping climate change awareness through spatial hierarchical clustering. 2409.13760. URL <https://arxiv.org/pdf/2409.13760>.
- Zhou, J., Liu, T., Zhu, J., 2019. Weighted adjacent matrix for K-means clustering. *Multimedia Tools Appl.* 78 (23), 33415–33434. <http://dx.doi.org/10.1007/s11042-019-08009-x>, URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069733662&doi=10.1007%2fs11042-019-08009-x&partnerID=40&md5=b40514c48f94623390de55d70caca427>.