Socio-Economic Planning Sciences 105 (2026) 102509

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

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

Mapping climate change awareness through spatial hierarchical clustering ☆

Gianpaolo Zammarchi [a], Paolo Maranzano [b,c,*]

[a] Department of Economics and Business Sciences, University of Cagliari, Viale Fra Ignazio, 17, Cagliari, 09131, Italy
[b] Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milano, 20126, Italy
[c] Fondazione Eni Enrico Mattei (FEEM), Corso Magenta, 63, Milano, 20123, Italy

ARTICLE INFO

Dataset link: https://github.com/PaoloMaranzano/GPZ_PM_HierSpatClusterClimateAwareness.git, GPZ_PM_HierSpatClusterClimateAwareness (Reference data)

Keywords:
Climate change awareness
Socio-economic and climate-related features
2022 international public opinion on climate change survey
Spatial hierarchical clustering
Data-driven hyperparameters tuning

ABSTRACT

Climate change is a critical issue that is expected to remain high on the political agenda for decades to come. While policy discussions at the international and national levels are essential, public awareness of climate change plays an equally crucial role in shaping effective responses. Since countries differ markedly in their exposure to climate-related risks, understanding cross-country differences in climate change awareness is particularly important. In this paper, we present a geographically-informed hierarchical clustering analysis aimed at identifying groups of countries with similar levels of climate change awareness. We employ a Ward-type clustering algorithm that integrates information on climate change awareness, socio-economic conditions, climate-related characteristics, and geographical distances between countries. To select appropriate values for the clustering hyperparameters, we propose a data-driven selection procedure that jointly accounts for within-cluster homogeneity, between-cluster separation, and explicit comparisons between geographically informed and non-spatial partitions. Findings reveal that incorporating spatial information leads to more stable clustering solutions and yields interpretable, geographically compact groupings relative to clustering approaches that ignore geography. In particular, we identify a clear contrast between Western countries, characterized by high and spatially cohesive levels of climate change awareness, and countries in Asia, Africa, and the Middle East, which exhibit lower average awareness and greater internal heterogeneity.

1. Climate change and public awareness

Climate change represents one of the greatest challenges that this and the future generations will face. Climate change poses serious risks to the planet's future and is a cause of environment's degradation. Historically, these changes have happened during a large part of Earth's history, but their evolution has always been rather slow if compared to the current trend. In many cases, the changes took hundreds if not millions of years to manifest themselves in a geologically detectable way [1]. Natural and anthropogenic emissions are both responsible of the current situation, but the scientific community agrees that humans have a key role in the extent and speed of these changes. Basically, the main drivers of these negative changes are associated to human behavior, especially when the rising global population became focused with the growth at all costs, which frequently leads to disregard the severe repercussions in natural systems and the consequences that follow from this [1].

Using fossil fuels for power, heating, and transportation has significantly raised greenhouse gas emissions and changed patterns of precipitations and temperature throughout the world. In 2022, the average worldwide temperature was approximately 0.86 °C higher than the 20th century average and the last 50 years (almost) there was a continuous exceeding of the average values of previous years [2]. About weather patterns, and in particular precipitations, the effects of climate change are becoming more and more evident globally, as evidenced by the extreme weather events and related disasters that occur all over the world. Some of the most well-known events are the forest fires in Australia (like the devastating fires that occurred in Australia in 2019–2020, [3]) and in the United States (especially in California, [4]) or the intense rainfall in China [5], the droughts in South Africa [6]. These facts are now becoming common nowadays, but unfortunately they are not the only catastrophic events that are happening, just consider the melting of the glaciers and the reduction of snow on mountain tops, the rising sea levels, the changes to river flow patterns, and the risk of extinction of many species worldwide. In relation to this last point in fact, it has been estimated that, as worst case scenario, it is possible to witness a species loss between 16% and

☆ This article is part of a Special issue entitled: 'Statistical Measurement of Economic Analysis' published in Socio-Economic Planning Sciences.

* Corresponding author at: Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milano, 20126, Italy.
E-mail addresses: gianpaolo.zammarchi@unica.it (G. Zammarchi), paolo.maranzano@unimib.it (P. Maranzano).

https://doi.org/10.1016/j.seps.2026.102509
Received 10 April 2025; Received in revised form 6 May 2026; Accepted 9 May 2026
Available online 16 May 2026
0038-0121/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

30% [7], which means millions of animals and plants to disappear [8]. Each one of these events will have several repercussions on the life of human beings and the animal species that live, hunt or base their life cycle on their environment.

As previously stated, all countries are impacted by climate change, however certain areas may be more exposed to particular effects. Given the seriousness of the problem, it is necessary to find solutions which, however, will probably require a long term to be implemented. For instance, reducing the amount of carbon dioxide in the atmosphere should be the primary goal to mitigate and carry out adaptation strategies for climate change. Other strategies include limiting development in floodplains, safeguarding naturally occurring wetlands and barrier islands, and fortifying vulnerable coastal communities with sea walls and levels [9]. The goal of these policy tools is to make ecosystems and people more resilient to the fluctuations and changes in climate. Nevertheless, these initiatives may worsen disparities about the effects of climate change and provide challenges for regional policy coordination. For example, carbon intensive societies have a bigger absolute burden from attempts to reduce greenhouse gas emissions, since the costs associated with abatement, transition, and compliance are higher [10,11]. In fact, as stated in the Global Climate Risk Index 2021 issued by the Germanwatch observatory, developing nations are less able to adapt, making them more vulnerable to the consequences of climate change [12]. In 2019, eight of the ten most severely impacted countries in terms of deaths and economic losses caused by extreme weather events (such as storms, floods, heatwaves, etc.) were, low- to lower-middle-income countries. Bahamas, Zimbabwe and Mozambique were three of the most impacted countries. Since countermeasures are not so easy and fast to implement, it is vital that governments of all countries take action to mitigate climate change. One of the possible pushes to act should come from the populations. For this to happen, it is necessary for individuals to become aware of the importance and urgency of interventions. Increasing public awareness on the causes and effects of climate change is of crucial importance because it can motivate policymakers to take action to cut greenhouse gas emissions and encourage individual behavioral adjustments. Given the different levels of vulnerability of countries around the world to the effects of climate change, the aim of this work was to examine how nations differ in terms of their degree of knowledge of the problem. Next, we present the findings of an analysis based on a spatially-constrained hierarchical clustering. To this aim, we compare different clustering scenarios in which combinations of socio-economic, climate, and geographic data were tested, while maintaining the settings of number of clusters and the mixing parameter[1] α (for more information about α see Section 2.2).

Climate change awareness varies significantly across nations, partly due to the influence of cultural and environmental factors, as well as their interactions. However, socioeconomic factors also play their role. Culture and environment can determine the priority given to addressing climate change, while environmental factors influence the severity of climate change impacts on the population. Early cross-national studies have established that wealth, education and unemployment levels, are predictors of climate awareness globally [14,15]. Kim and Wolinsky-Nahmias [16] showed that vulnerability to climate impacts and national wealth interact in complex ways, with wealthier nations showing greater baseline awareness, however, less wealthy nations frequently show more support for climate policy. Recent research examined the influence on climate change awareness and risk perceptions of socio-demographic information, geography, perceived well-being and beliefs [17]. The authors found that educational levels are the most reliable indicator of climate change awareness and they suggest improving basic education and climate knowledge. In a more recent study by Poortinga et al. [18], the authors explained the role of political orientation and human values as well as the well-known socio-demographic and educational aspects, stressing the importance of collecting data on people's opinions. This work is based on data from 23 countries (mostly European) and showed that in Central and Eastern European regions many predictors tend to have a weaker effect, while in Northern and Western Europe the relationship between these predictors and climate awareness is stronger. Importantly, higher awareness does not automatically translate into proportional risk perception or behavioral change, because these are influenced by personal experience and in many regions of the world people never personally witnessed the consequences of climate change [19,20]. Additionally, several authors have examined the effects of weather on climate change awareness and found that relationship with personal observation of local weather [21–23] and weather extremes [24], underlining the importance of also taking these data into consideration in the analyses. Despite the growing body of research, significant gaps remain in our understanding of the spatial structure of awareness patterns and how geographic aggregation can inform targeted policy interventions. Most existing studies focus on one or a few countries at a time, often from the same geographic region, or on a few or single determinants. In our study, however, we aimed at collecting data from over one hundred countries worldwide and jointly assessing numerous socioeconomic factors and the effects of spatial interdependence.

1.1. Related works on geographical clustering and our contribution

In recent years, many works have shown the usefulness of performing spatial clustering with data of different nature. Hincks et al. [25] developed a new spatial typology of climate change risk for European cities and regions based on k-means clustering. They used 49 variables at the NUTS-3 level for 33 European countries and showed a substantial spatial heterogeneity in climate risk across Europe. Their findings show that climate change risk present patterns which can be addressed only with an approach that takes into account the transboundary nature of the problem whose fight may be less efficient because of different political commitment, resource availability, and risk awareness. In their work, Abdelgadir et al. [26] showed that if spatial clustering is combined with predictive modeling it can become a powerful tool for identifying conservation areas for endangered species. By combining spatial clustering and ensemble species distribution modeling they achieved 88%–95% prediction accuracy in mapping the White-eyed gull biogeography. The analysis revealed that climate change will likely significantly impact the spatial distribution of these seabirds, reducing current suitable habitats and shifting it toward the Mediterranean Sea. The spatial clustering approach can successfully identify geographically coherent areas requiring protection and can be generalized to other species and with appropriate changes it might also be possible to adapt it to other contexts. Yang et al. [27] analyzed the usefulness of the spatial clustering analysis to discover how organisms' spatial organization affects climate-related activities like carbon sequestration. They investigated how tree size and spatial clustering affected soil organic carbon storage in forests in Northeast China, analyzing 720 plots across a 7.2 hectare in a forest. Their findings showed that soil with different types of trees was positively associated with the carbon sequestration, partially because they improved the soil physical conditions and aggregate stability. This work demonstrates that the spatial structure of organisms that are distributed and clustered in space can be of critical importance, suggesting that discovering spatial patterns could represent a nature-based but underutilized climate solution. These works reinforce the importance of discovery structures in the data and therefore the idea that strategies to adapt to climate change cannot follow a uniform approach, as the uneven distribution of risks, resources or awareness across regions necessitates different spatially-informed strategies.

[1] In the framework proposed by Chavent et al. [13], the mixing parameter α governs the contribution (i.e., the relative importance) of each dissimilarity matrix to the clustering criterion, effectively tuning the trade-off (i.e., the strength) between socio-economic similarity and spatial proximity.

With respect to the clustering methods, several solutions have been proposed in the literature to determine the optimal partition based on a homogeneity criteria based on dissimilarity [see e.g., 28–32].

In some cases, it makes sense to limit the range of feasible solutions, for instance by imposing contiguity constraints. Contiguity constraints can be both in time or in space and are the most prevalent kind. These constraints arise when an object in a cluster must be both similar to every other elements of a group and part of a continuous group of elements. Basically, if there is a path connecting each pair of elements in a cluster, then the cluster is said to be contiguous. Let $\mathbf{C}$ be the contiguity matrix where each entry c_{ij} can be equal to 1 if the ith element is contiguous to the jth element and equal to 0 otherwise, then we can consider two clusters to be contiguous if the contiguity matrix shows a relationship ($c_{ij} = 1$) between two elements (one from each cluster). However, many of the proposed methods work by considering that, if two elements are very similar but are located in different areas, the criterion based on spatial proximity will separate them into two different clusters. A possible solution is to consider soft constraints, so that the spatial contiguity does not create a sharp separation. In accordance with this logic, some methods proposed to merge the dissimilarity matrix derived from non-geographical features and the matrix of geographical distances [33,34]. In this combination, the weights assigned to the geographical dissimilarities will provide more or less geographically contiguous clusters, but in this type of approach the issue is shifted to choosing the right weights.

This paper pursues several complementary goals, spanning methodological development, comparative assessment, and empirical application. From an applied perspective, the study aims to identify and interpret global patterns of climate change awareness, with particular emphasis on their spatial heterogeneity and their relationship with broader socio-economic and environmental conditions. A second objective is to evaluate the extent to which incorporating geographical (spatial) information improves the identification of homogeneous groups of countries, and how spatial constraints reshape clustering outcomes relative to purely feature-based approaches. Moreover, the proposed research framework is designed to support policy-relevant segmentation of countries, providing a structured basis for understanding cross-country differences in climate awareness and for informing geographically targeted communication strategies. We address these goals by building upon the Ward-like hierarchical clustering technique with geographical (or spatial) constraints introduced by Chavent et al. [13] and further extended to the spatio-temporal setting by Morelli et al. [35] and Morelli et al. [36]. This method employs a convex combination of two dissimilarity matrices, namely D_0 and D_1, which contain, respectively, information about a set of clustering features (in our case climate awareness, climate-related and socio-economic variables) and the geography of the area of interest. The two dissimilarities are linearly combined through a mixing hyperparameter α, enabling a direct comparison between spatially constrained and unconstrained partitions. The method is suitable with both Euclidean and non-Euclidean distances, being the latter the situation when considering geographical distances. The general idea under the algorithm is to generate geographically compact clusters (i.e., with a marked spatial contiguity) of countries without excessively deteriorating the quality of the solution based on the set of the available features. However, under this setting, the selection of the key hyperparameters, namely the mixing parameter α and the number of clusters K, remains an open issue in the literature, often addressed through ad hoc or sequential procedures [37]. Therefore, we contribute to this gap by proposing a data-driven procedure for the joint selection of α and K, enabling a principled trade-off between spatial cohesion and feature-based similarity, while improving the stability and interpretability of the resulting partitions. Specifically, the proposed selection algorithm automatically combines multiple criteria, including within-cluster homogeneity, between-cluster separation, and the comparison between geographically-informed and unconstrained solutions. The proposed approach provides a systematic and transparent way to balance the contribution of spatial and feature-based information, identifying configurations that preserve clustering quality while ensuring an adequate degree of spatial coherence.

The remainder of the paper is organized as follows. Section 2 describes, on the one hand, the climate change awareness data provided by the 2022 International Public Opinion on Climate Change survey and the related statistical challenges, and, on the other hand, the geographically-informed hierarchical clustering algorithm employed in the empirical analysis. In this section, we also introduce a new, computationally feasible procedure for tuning the clustering hyperparameters that jointly accounts for within-cluster homogeneity, between-cluster separation and extends existing approaches in the literature. Section 3 presents the empirical findings. We first discuss a core clustering analysis in which countries are grouped using only climate change awareness information and geographical proximity. We then report a set of robustness checks aimed at assessing the sensitivity of the clustering results to alternative specifications. These include: (i) an extended specification that incorporates a broad set of climate-related and socio-economic variables, in addition to awareness and spatial information, to evaluate how these factors affect clusters' composition; and (ii) a comparison between an unweighted clustering approach, where all observations receive equal weight, and a population-weighted version that accounts for cross-country differences in population size. The results of all robustness exercises are reported in detail in the Online Supplementary Material. Finally, Section 4 concludes the paper by summarizing the main findings and outlining directions for future research.

2. Data and methods

2.1. Survey data on climate change awareness

To quantify the level of awareness of each country we used the "International public opinion on climate change" survey. The most complete currently available edition is the 2022 edition which includes responses from more than 108 thousands Facebook users located in 192 countries worldwide [38]. The survey was carried out between March 25th and April 14th, 2025, in collaboration with "Data for Good at Meta" and the main aim was to examine people's knowledge, beliefs, attitudes, policy preferences, and behavior related to climate change. The data collected is representative of both actual Facebook users and the overall resident population in a given country.[2] Indeed, for each country surveyed, the survey consists of samples in proportion to publicly available age and gender benchmarks. Specifically, data were weighted separately for each country using a multi-stage, pre-and-post-survey weighting process based on census and nationally representative survey benchmarks, Facebook demographics, and Facebook engagement metrics, balanced to the total number of survey completions.

We stated that the survey covers a total of 192 countries worldwide; however, data for 81 countries and territories with small populations and/or limited numbers of Facebook users were combined into groups or areas, thereby forming a single representative sample for each area. The five areas are Caribbean, Asian and Pacific Islands, and Sub-Saharan African countries. Since the aim of the paper is to investigate the role of geographic/spatial dimensions in determining homogeneous groups of countries based on climate change awareness, defining the notion of spatial "centroid" for countries that are already aggregated, and potentially scattered across the globe, could lead to inconsistencies in the results and subsequent interpretations. Therefore, we chose to

[2] We refer the readers to the Survey Method section in Appendix I "Survey Method" of Leiserowitz et al. [38] for details on the data collection strategy, including the full list of surveyed countries, country-specific adjustments and the sample distribution across the countries.

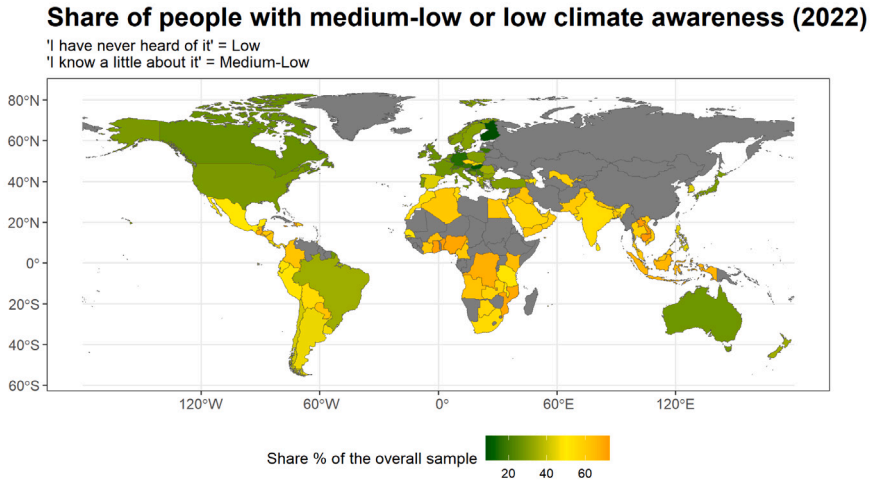

Fig. 1. World map representing the share of people in the 2022 survey with low and medium-low climate change awareness at the country level. The share of people with low and medium-low climate change awareness decreases toward green and increases toward red. The share was created by aggregating the participants declaring little or no awareness (i.e., "*I have never heard of it*" and "*I know a little about it*"), while the complementary was created by aggregating the participants who declared moderate or high awareness of climate change (i.e.,"*I know a lot of it*" and "*I know a moderate amount about it*").

exclude from the analysis the pre-grouped countries, which represent 5.62% of the Facebook sample, and to consider only countries that can be matched to precise geographic or polygonal coordinates. The final count of countries considered in the rest of the paper is hundred-three ($N = 103$).[3] reports the number of unweighted respondents for the $n = 103$ countries considered in the empirical analysis (taken from Appendix I of Leiserowitz et al. [38]) supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender [source: 39]. Fig. B.1 in Appendix B shows a map representing the unweighted number of respondents to the 2022 Climate Change Opinion Survey in the 103 countries considered here. In addition, Fig. 1 shows the world map based on the awareness levels of each of the 103 selected countries. The map shows that the US, Canada, Brazil, Australia, New Zealand, and almost all Europe have high awareness values, while African countries and Asian countries have rather low values of awareness (with some exceptions, e.g. Japan).

While the complete survey includes answers to numerous questions regarding the degree of knowledge about climate change, its main causes, perception of the phenomenon, and the degree of concern about it,[4] we collected data pertaining to "Overall Climate Awareness", summarized in the question "How much do you know about climate change?". Users could respond to the question with five possible answers: "*I have never heard of it*", "*I know a little about it*", "*I know a moderate amount about it*" and "*I know a lot about it*", plus an option to decline to answer. To pursue our goals, in line with [38], we reclassified the previous statements into a binary response, namely the degree of climate change awareness, in which people answering "*I have never heard of it*" and "*I know a little about it*" were merged into the class "*Low or medium-low climate change awareness*", whereas people answering "*I know a lot of it*" and "*I know a moderate amount about it*" were merged into "*High or medium-high climate change awareness*". Then, for each country, we computed the share of respondents associated with the two reclassified answers. We stress that, while dichotomizing may result in a loss of information due to the compression of the number of categories, it also simplifies analysis and interpretation and is immediately compatible with the hierarchical geographical clustering algorithm that will be introduced in Section 2.2. In fact, given the perfect collinearity between the two responses (i.e., adding the two classes gives 1 by definition), it is only necessary to use the dissimilarity matrix generated by one of the two categories (combined with the geographic distance matrix) to partition the countries (following the approach of Chavent et al. [13])without having to adopt methodological extensions, such as the multi-matrix case developed by Morelli et al. [35], which would significantly increase statistical complexity at the expense of interpretative simplicity.[5]

2.2. Hierarchical clustering with geographical constraints

The survey data described above are employed as input of a hierarchical clustering algorithm embedding geographical constraints to explore the climate change awareness patterns. To do so, we considered the spatial hierarchical clustering algorithm proposed by Chavent et al. [13], which relies on a linear combination of the dissimilarity in the feature space and the geographical distances (i.e., distance in the spatial coordinates) to obtain the partitioning of the units under proximity constraint. Specifically, this method makes use of two dissimilarity matrices, that is, a feature-based distance matrix $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ and a geographical distances matrix $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$. The interpretation of the two matrices is straightforward: while the D_0 matrix provides information on dissimilarity in the "feature space", the D_1 distance matrix embeds information in the geographical "constraint space". Notice that to compute the geographical distance we used the standard Euclidean distance function, which consider the shortest path between two points on a curved surface, such as an ellipsoid or sphere, which in this case is the Earth's surface. Also, as we will see later, the feature space in our application will comprehend a set of climate change awareness, socio-economic, and climate-related variables at the national level for 2022.

Notice that, to model geographic proximity between countries, we employ the Euclidean distance between the centroids of national polygons, computed in a projected coordinate system and expressed in

[3] Tables B.2, B.3, and B.4 of the Appendix

[4] We refer the reader to Appendix II "Results by Region" of Leiserowitz et al. [38] for an extensive descriptive analysis of the responses recorded for each question in the survey based on region of residence.

[5] In the scenario of a dissimilarity matrix derived from a socio-economic feature [13], we would only need to optimize two hyperparameters of the clustering model (the weight of the geographical component α and the number of clusters K^*), while in the case of multiple variables, and therefore multiple matrices [35], the scalar weight would become a vector with P elements, where P is the number of clustering variables to be optimized.

kilometers. This choice follows a well-established practice in spatial econometrics and geostatistical clustering literature [37,40,41] and provides several advantages. In particular, by employing the Euclidean distance for both socio-economic data and space allow us to remain within a Euclidean framework even in the multi-matrix case, as the two matrices are combined using a convex linear combination. It follows that all the clustering metrics that could be calculated in the case of single-matrix Euclidean clustering, can also be used in this case of geographically-informed clustering without restrictions. Moreover, we stress that several applications demonstrate that the Euclidean distance provides a reasonable approximation to more realistic travel or geodetic (i.e., Great Circle) distances. For instance, Jones et al. [42] report a correlation of 0.99 between road-network driving distance and Euclidean distance between ZIP-code centroids, with relatively modest absolute differences.[6] Recent studies explicitly employ the Euclidean distance between country centroids as a proxy for travel or trade distance or international spatial proximity [43–46]. In line with this literature, we interpret the Euclidean centroid distance as a parsimonious proxy for the geographical separation between countries. Overall, this simple robustness analysis confirms that the qualitative features of the resulting spatial clusters remain stable when Euclidean distance is replaced by the geodetic distance and vice versa.

The clustering is performed following a Ward-like hierarchical strategy in which the dissimilarity matrix is given by a convex linear combination of D_0 and D_1, linearly related by a mixing parameter α (with $0 \leq \alpha \leq 1$), which controls the importance (i.e., the strength) given to each dissimilarity matrix employed to cluster the units. When the mixing parameter α approaches 1, contiguous units will be forced to cluster together as the weight of the features dissimilarity becomes negligible and the only relevant dimension is the geographical distance. In the extreme case of $\alpha = 1$, the clustering outcome is purely geographic, which can result in areas united only by their proximity although being extremely dissimilar. On the contrary, for values of α approaching 0, the spatial constraint will become progressively negligible and the feature-based dissimilarities will prevail. When $\alpha = 0$, the clustering outcome is driven exclusively by features, potentially creating spatially fragmented clusters consisting of regions/states located at great distances.

As in the classical Ward-like hierarchical procedure, the scope of the algorithm is to minimize the convex combination of an homogeneity criterion calculated with D_0 (i.e., using only non-geographical features, such as the country-specific climate change awareness) and the homogeneity criterion calculated with D_1 (i.e., computed using only geographical distances among countries' centroids). To this extent, recall that when the mixing parameter α increases, the homogeneity computed with D_0 decreases while the homogeneity computed with D_1 increases. Also, in the same spirit of Ward's hierarchical clustering, we recall that the degree of homogeneity associated with a partition is measured through its pseudo-inertia, that is a generalization of inertia – defined as the total dispersion of a set of statistical units with respect to the Euclidean distance – to the case of non-Euclidean distances. In the mixed approach, let I_α denote the mixed pseudo inertia for a given value of α; therefore, the mixed pseudo inertia of a generic cluster C_k^α can be defined as follows:

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

where w_i is a user-defined weight for the ith observation with $i = 1,\ldots,n$, $\mu_k^\alpha = \sum_{i\in C_k^\alpha} w_i$ is the overall weight of the partition C_k^α, $d_{0,ij}^2$ is the normalized $n \times n$ dissimilarity between observations i and j in D_0, and $d_{1,ij}^2$ is the analogous in D_1. In the mixed setup by Chavent et al. [13], as the mixed pseudo inertia of cluster C_k^α decreases, the units belonging to cluster C_k^α are more homogeneous (i.e., $\downarrow I_\alpha\left(C_k^\alpha\right)$ then $\uparrow$ homogeneity within C_k^α). When considering a partitioning into K clusters, the overall homogeneity is given by the mixed within-cluster pseudo inertia, namely $W(\mathcal{P}_K)$, which is computed as

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

In this framework, in order to obtain a high degree of homogeneity of the partition $\mathcal{P}_K$, the clusters are formed by minimizing the mixed within-cluster pseudo inertia.

We stress that the unit-level weights w_i (for $i = 1,\ldots,n$) are predefined by the user and can assume either a single value common to all the observations or a unit-specific value [see Section 2 of 13, for a formal discussion on the topic]. In the former case, also referred to as the uniform weighting design, we assume that $w_i = \frac{1}{n} \quad \forall i$; conversely, in the latter, that is, the non-uniform weighting design, the generic value w_i can be computed based on other information, such as using the population count or socio-economic variables like the GDP. Confident in the fact that the questionnaire responses have already been weighted and adjusted for the demographic composition of the countries and respondents, in the empirical application we will provide insights from the uniform weighting setting. However, in addition, we will leverage on information about the countries' population to generate to weigh the survey responses and obtain results from the non-uniform clustering.

2.3. Tuning of the hyperparameters α and K

In the typical context of hierarchical clustering a crucial issue is how to determine the optimal number of clusters, that is, K^*. Several criteria are available in the literature, and ranging from graphical methods, such as the elbow method, to analytical tools, such as distance-based indicators (e.g., the Calinski–Harabasz index or the silhouette width as in Akhanli and Hennig [47]) or model-based indicators based on the likelihood function (e.g., likelihood criteria as in Rossbroich et al. [48]). In our framework, the complexity is even higher as in addition to the number of clusters we have to choose a suitable value for the mixing parameter α, that is, α^*. Also, we recall that there exists a mutual relationship among α and K as the mixing parameter depends on the number of clusters, but also the mixing parameter can influence the number of clusters.

Among the available ways to define α^*, we recall the proposal by Chavent et al. [13], Morelli et al. [35], both based on a sequential approach in which α is determined by conditioning on the number of clusters K. The former proposal defines α that best compromises between loss of feature-based and loss of geographical homogeneity. In other words, Chavent et al. [13] determine α such that it increases the spatial contiguity without deteriorating too much the quality of the solution based on the feature space. Such objective is obtained by minimizing the distance among the proportion of the total mixed pseudo inertia explained by the partition $\mathcal{P}_K$ in K clusters normalized with respect to the D_0 matrix (that is, compared to the case of clustering based only on the non-spatial features), namely $\tilde{Q}_{D_0}(\mathcal{P}_K^\alpha)$, and the proportion normalized with respect to the D_1 matrix (that is, compared to the case of clustering based only on the geographical distance), namely $\tilde{Q}_{D_1}(\mathcal{P}_K^\alpha)$. Conversely, Morelli et al. [35] propose to select the mixing parameter α which jointly maximizes the amount of pseudo inertia

[6] We replicated this check on our data by computing the linear correlation on the average distance among each country and the other countries in the sample according to either Euclidean and Geodetic metrics. According to our finding, the use of Euclidean distance generates values and rankings (to be understood as the positioning of a country in terms of average distance from other countries in the sample) that are compatible and consistent with those of geodesic distance. As proof of this, in Fig. C.2 of Appendix C we provide the diagrams of the empirical relationship between the average distance of each country from the rest of the world (i.e., average country-to-World distance) calculated using the Euclidean formula (X axis) and the geodetic formula (Y axis) (left panel) and the empirical relationship between the ranks calculated on the same average Euclidean and geodesic distances (right panel). The rankings slightly between ±20 positions and the linear correlation is close to 1, showing an high degree of coherence between the two methods.

Algorithm 1 Spatial Hierarchical Clustering: choice of the hyperparameters α and K

Step 1: define the algorithm's inputs
Define as $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ the (normalized) feature dissimilarity matrix
Define as $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$ the (normalized) spatial dissimilarity matrix
Define as K_{max} the maximum number of clusters
Define as $\underline{\alpha} \in [0,1]$ a sequence of mixing parameters

Step 2: identify the set of potential optimal mixing parameters
for $K = 2, \dots, K_{max}$ **do**
  for $\alpha \in \underline{\alpha}$ **do**
    Compute the linear combination of the two dissimilarity matrices $D(\alpha) = (1-\alpha)D_0 + \alpha D_1$;
    Compute the $\mathcal{P}_K^\alpha$ = partition in K clusters according to Ward's hierarchical algorithm on D;
    Compute the weighted average of the explained mixed pseudo inertia $\tilde{Q}(\mathcal{P}_K^\alpha)$ as in Morelli et al. [35] and the two normalized pseudo inertia $\tilde{Q}_{D_0}(\mathcal{P}_K^\alpha)$ and $\tilde{Q}_{D_1}(\mathcal{P}_K^\alpha)$ as in Chavent et al. [13];
  end for
  Select the best α for each K such that $\alpha^*_{K,max} = argmax_\alpha \tilde{Q}(\mathcal{P}_K^\alpha)$ (that is, according to the criterion by Morelli et al. [35]) and the best α for each K such that $\alpha^*_{K,min} = argmin_\alpha |\tilde{Q}_{D_0}(\mathcal{P}_K^\alpha) - \tilde{Q}_{D_1}(\mathcal{P}_K^\alpha)|$ (that is, according to the criterion by Chavent et al. [13])
end for

Step 3: define the optimal combination of number of clusters and mixing parameters
Choose K^* (evaluated at the corresponding $\alpha^*_{K,max}$ and $\alpha^*_{K,min}$) based on a set of hierarchical clustering criteria, such as the Silhouette index, the Dunn's index, the C-index, the Calinski–Harabasz's index, and the McClain–Rao's index. In particular, consider the combination of three potential rule-of-thumb: (1) graphical analysis of the indices for varying α's and K's; (2) majority voting rule on the absolute values of the clustering criteria; (3) majority voting rule on the gain/loss of the indices evaluated at $\alpha^*_{K,max}$ and $\alpha^*_{K,min}$ (i.e., geographically-informed clustering) compared to the case of the indices at $\alpha = 0$ (i.e., non-spatial clustering).

explained from both the socio-economic features and the geographical information, weighted by the cumulated spatial and socio-economic pseudo inertia embedded the data. A comprehensive discussion about the computation and the interpretation of the two methods, as well as their relationship with the clustering-related metrics, is provided in Section 4 of Morelli et al. [35] and Section 3 of Chavent et al. [13].

In [35], the authors propose several sequential algorithms to show how the selection of both K and α can be performed; in particular, conditioning on an initial value of K, the mixing parameter α^* is find according to the criterion above, and then the number of clusters K^* is re-optimized using one or more clustering criteria defined by the user. Here, we propose a procedure that combines the Morelli et al. [35], Chavent et al. [13] criteria and extends the set of clustering indicators used to select a suitable number of clusters and the mixing parameters. The proposed procedure is described in Algorithm 1. In practice, Algorithm 1 is implemented setting a maximum number of clusters $K_{max} = 10$ and a grid of candidate α values ranging from 0 to 1 with step size $\Delta\alpha = 0.01$.[7] Furthermore, since distances depend on the unit of measurement and the scale of the data (in particular, when the Euclidean distance is calculated between countries scattered across the globe, the values can be very high), both the spatial dissimilarity matrix and the feature dissimilarity matrix were normalized. The spatial matrix was normalized with respect to its maximum value as $\frac{X_i}{max(X)}$, while the climate change awareness matrix was normalized using a robust scaling that accounts for potential outliers, namely $\frac{X_i - median(X)}{q_{0.75}(X) - q_{0.25}(X)}$, where X_i is the feature's value associated with a country i and $q_{0.75}(X)$ and $q_{0.25}(X)$ are the 75° and 25° quantiles, respectively.

The algorithm works in a three-stage setting. In the first stage, the user has to define the hierarchical clustering's inputs (i.e., the feature dissimilarity and spatial dissimilarity matrices, the maximum number of clusters to be considered, and a sequence of candidate mixing parameters to weight the matrices. In the second step we identify the set of potential optimal mixing parameters based on both the Morelli et al. [35] (denoted as $\alpha^*_{K,max}$) and Chavent et al. [13] (denoted as $\alpha^*_{K,min}$) criteria for all the considered K's. In the third step we use a combination of three potential rule-of-thumb to identify the definitive number of clusters K^*, that is, graphical analysis of the indices for varying α's and K's; majority voting rule [49] on the absolute values of a set of clustering criteria (i.e., Silhouette index [50], the Dunn's index [51], the C-index [52], the Calinski–Harabasz's index [53], and the McClain–Rao's index [54]); majority voting rule on the gain/loss of the previous set of indices evaluated at $\alpha^*_{K,max}$ and $\alpha^*_{K,min}$ (i.e., geographically-informed clustering) compared to the case in which we consider only the feature space and ignore the geographical information (i.e., $\alpha = 0$). Specifically, the latter rule-of-thumb is based on the computation of the percentage gain or loss obtained by using a geographically-informed clustering approach instead of the clustering without geographical constraints, that is when fixing $\alpha = 0$. For a given index, the gain/loss is obtained as the following percentage variation:

$$GL = \frac{Index_{\alpha^*_K} - Index_{\alpha_K=0}}{Index_{\alpha_K=0}} \times 100. \tag{3}$$

Notice that the definition of gain or loss depends on how each index is optimized. Indeed, when considering the C-index and the McClain–Rao index, the optimal number of clusters is the minimizer, while for Silhouette, Calinski–Harabasz, and Dunn indices the optimal value is K such that the indicators are maximized. Thus, for the former two a gain (loss) is detected when the indicator associated with the geographically-informed clustering returns a lower (greater) value compared to the non-spatial case. Conversely, for the latter three a gain (loss) is detected when the indicator associated with the geographically-informed clustering returns a greater (lower) value compared to the non-spatial case. In other words, for Silhouette, Calinski–Harabasz, and Dunn the spatial information induces improvements for positive variations, while for the C-index and the McClain–Rao index an improvement is obtained when negative variations occur. A summary of the interpretation criteria for each index is reported in Table 1. We also remark that by expressing the improvements with respect to the case without geographical constraints, in the unlucky situation of losses, we can both quantify the amount of such losses and eventually choose a set of hyperparameters that guarantees performances similar to the non-spatial clustering.

[7] For a given setting (i.e., weighted or unweighted clustering), the total number of hyperparameters combinations is $15 \times 101 = 1515$ and requires a computing time lower than two minutes to be executed without specific computational and hardware requirements.

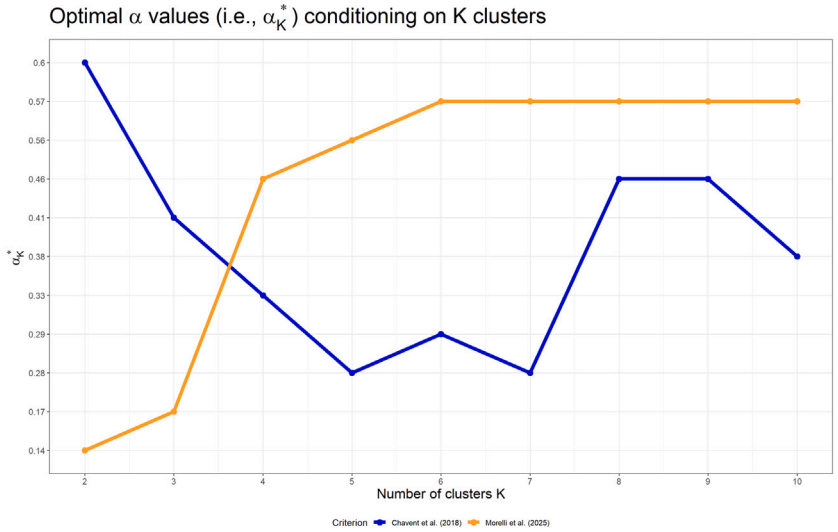

Fig. 2. Optimal values of the mixing parameter α conditioning on K clusters (that is, α^*_K) obtained using the criterion by Morelli et al. [35] ($\alpha^*_{K,max}$ in orange) and by Chavent et al. [13] ($\alpha^*_{K,min}$ in blue).

Table 1
Gain and loss interpretation with respect to non-spatial clustering.

Index	Optimality	↑	↓
C-index	Min	Loss ×	Gain ✓
McClain–Rao	Min	Loss ×	Gain ✓
Silhouette	Max	Gain ✓	Loss ×
Calinski–Harabasz	Max	Gain ✓	Loss ×
Dunn	Max	Gain ✓	Loss ×

The symbol ↑ (↓) indicates an increase (decrease) of the index value when moving from the non-spatial clustering to the geographically-informed clustering.

To sum up, the optimal combination (α^*_K, K^*) is selected such that it simultaneously takes into account the within-clusters homogeneity (both when selecting the mixing parameter using the pseudo inertia and when computing the clustering criteria), between-clusters separation (e.g., the Calinski–Harabasz's index consider both within-cluster and between cluster variability [47]), and an explicit comparison between the geographically-informed and non-geographical partitions (when computing the gain/loss of the indicators).

3. Empirical results

In this section we report and comment on the results obtained from the application of the spatial clustering procedure that employs as key feature the share of respondents declaring a low or medium-low climate change awareness. Specifically, in Section 3.1 we show the results from a spatial clustering when equally-weighted observations (i.e, $w_i = \frac{1}{n} \quad \forall i = 1, \dots, 103$) are used, whereas in Section 3.2 we provide the results for a robustness check in which the feature dissimilarity matrix is computed by combining the share of respondents to the survey with socio-economic and climate-related country-specific variables.

In the Online Supplementary Material we report the results obtained from other robustness experiments, such as the spatial clustering in which the country-specific awareness is weighted by the total resident population (i.e, $w_i = Pop_i \quad \forall i = 1, \dots, 103$) and the clustering performed using the share of respondents declaring an high or medium-high climate change awareness. Overall, the results obtained without weighting the observations and those obtained by weighting by total resident population are mutually consistent. First, the optimal number of clusters is typically found to lie between two and four (see, e.g., Figures 37–39 in the Online Supplementary Material). Second, the relative importance of the spatial component, as captured by the parameter α, ranges between 0.21 and 0.58, with optimal values equal to $\alpha = 0.21$ when $K = 4$ and $\alpha = 0.43$ when $K = 3$. Third, the resulting partitions continue to reflect the main socio-economic, political, and historical blocks identified in the unweighted clustering (see Figures 40–43 in the Online Supplementary Material). Finally, the inclusion of additional climate-related and socio-economic covariates yields results that are fully consistent and qualitatively very similar to those obtained in the unweighted case (Figure 45 in the Online Supplementary Material) presented below.

3.1. Spatial hierarchical clustering using climate change awareness information and geographical distance

In Fig. 2 we show the optimal α values (i.e. α^*_K) computed for each value of K from 2 to 10. The $(K_{max} - 1)$ α^*_K values were obtained using the criteria proposed by Morelli et al. [35] (orange line) and Chavent et al. [13] (blue line). From the figure we can easily infer that while the Chavent et al. [13]'s criterion generally assigns a lower weight to the geographical information (i.e., $\alpha^*_{K,min}$ lies between 0.28 and 0.46 when $K \geq 4$, with higher values for a fewer number of clusters), the criterion by Morelli et al. [35] suggest a more relevant role for the spatial constraint (i.e., above 0.46 for $K \geq 4$).

As shown in Fig. 3, for a given value of K and α^*_K it is possible to obtain a specific index value for each of the five proposed metrics. These metrics were the Silhouette index, the Dunn's index, the C-index, the Calinski–Harabasz's index, and the McClain–Rao's index. However, not all the metrics work in the same manner. In fact, for the C-index and the McClain–Rao index the optimal value is the one corresponding to the minimizer, while for Silhouette, Dunn and Calinski–Harabasz indices indicate that the optimal value is the maximum. Taking these facts into account, Fig. 3 can be used to establish which value of K corresponds to the optimal value, that we will indicate as K^*. Potentially, any combination of metrics (i.e., the indices) and methods (i.e., [13,35]) could return a different value for K^*, thus leading to a maximum of ten optimal K^*. In our analysis only seven different values were selected as a potential optimal value for K. These values were 2 (four times), 3 (one time), 4 (one time), 5 (one time), 6 (one time), 7 (one time), and 8 (one time).

We previously stated that one of the objectives of the analysis is to understand if the geographical information can improve the clustering findings, and thus the interpretation of the climate change awareness phenomenon. Then, we carried on a direct comparison between the five validation indices obtained including and excluding the geographical

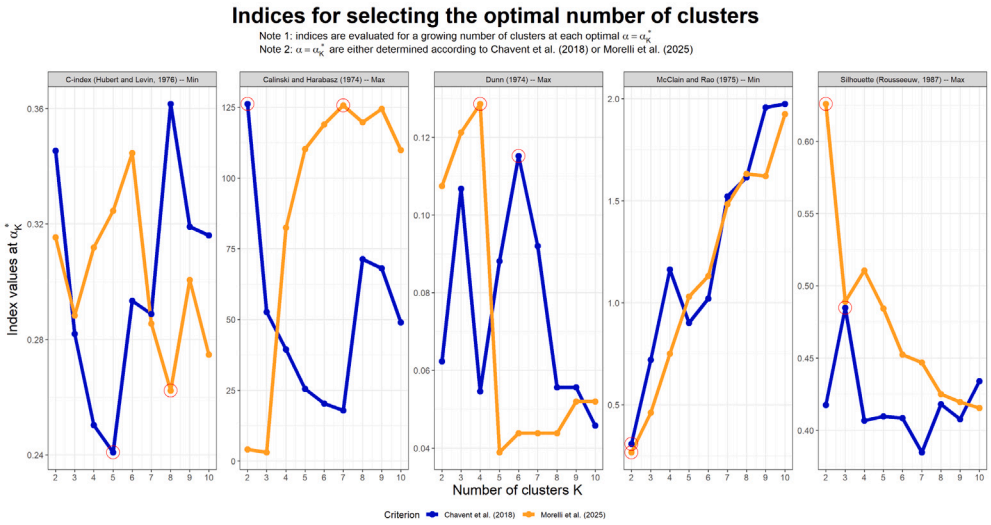

Fig. 3. Estimates of the Silhouette index, the Dunn's index, the C-index, the Calinski–Harabasz's index, and the McClain–Rao's index for K clusters evaluated at the corresponding $\alpha^*_{K,max}$ (orange) and $\alpha^*_{K,min}$ (blue). Red circles identifies the optimal values, that is, the minimizer for C-index and McClain–Rao's indices and the maximizer for Silhouette, Dunn's and Calinski–Harabasz's indices.

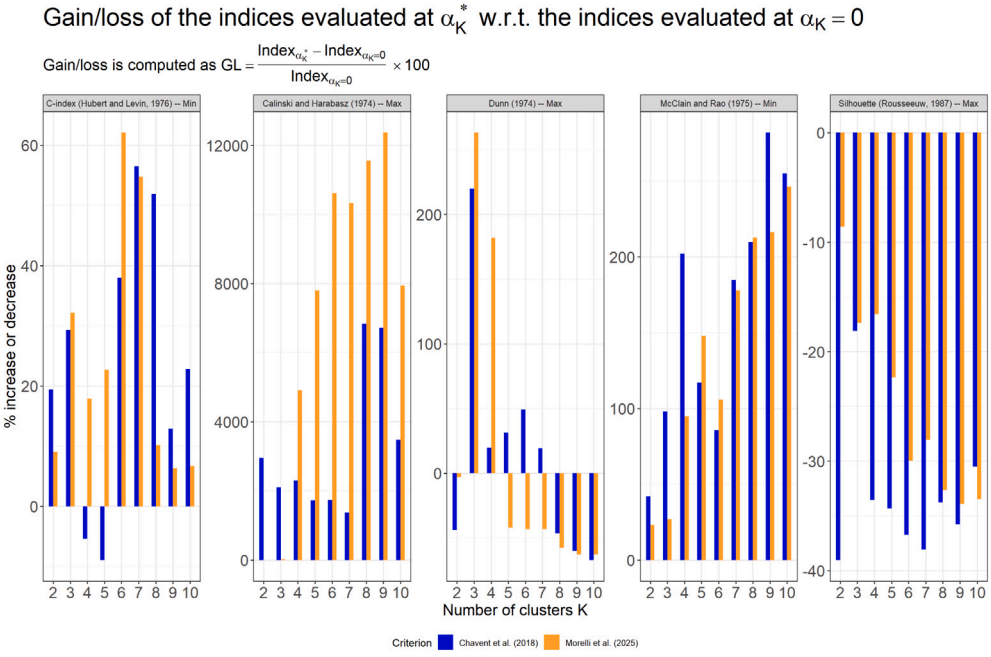

Fig. 4. Estimated percentage gain (positive values) or loss (negative values) of the Silhouette, Dunn's, C-index, Calinski–Harabasz's index, and McClain–Rao's indices evaluated at the corresponding $\alpha^*_{K,max}$ (orange) and $\alpha^*_{K,min}$ (blue) with respect to the indices evaluated at $\alpha = 0$. Percentage gain/loss is computed as $GL = \frac{Index_{\alpha_K^*} - Index_{\alpha_K=0}}{Index_{\alpha_K=0}} \times 100$; thus, positive values have to be read as the percentage increase of the indices w.r.t. the baseline case of index at $\alpha = 0$, while negative values have to be read as the percentage decrease of the indices w.r.t. the baseline case of index at $\alpha = 0$.

information. In Fig. 4 we show the estimated percentage gain or loss of the five metrics when using the geographical constraint (i.e., indices are evaluated at $\alpha = \alpha^*_{K,min}$ or $\alpha = \alpha^*_{K,max}$) compared to the case of clustering without geographical constraints (i.e., indices are evaluated at $\alpha = 0$). For all the indices except the Silhouette, we observe that the optimized α^*_K produce larger values in the absolute scale, leading to positive percentage increments. The magnitude of the gain varies based on the metrics and the specific K values. The Silhouette index is the only one showing negative variations for all the considered combinations of K and α. In many cases, $K = 3$ and $K = 4$ lead to marked gains or moderate losses. For instance, regarding the Calinski–Harabasz's index, using the α^*_K values computed using the method proposed by Morelli et al. [35] allows to obtain a particularly high gain for $K \geq 5$, while moderate improvements are detected when considering $K = 4$ regardless the selected method. Also, $K = 3$ or $K = 4$ evaluated at $\alpha^*_{K,max}$ provide very minor losses for C-index and Silhouette index, while for $K = 4$ evaluated at $\alpha^*_{K,min}$ the C-index improves and for either $K = 3$ or $K = 4$ the Dunn's index improves substantially (especially when the Morelli et al. [35] criterion is adopted). Considering the partial concordance among the optimal solution, we believe that a good compromise can be achieved by choosing $K^* = 3$ or $K^* = 4$ evaluated using the methods by Morelli et al. [35]. Below we present the results related to these two cases when considering the geographical constraint and when ignoring the spatial dimension.

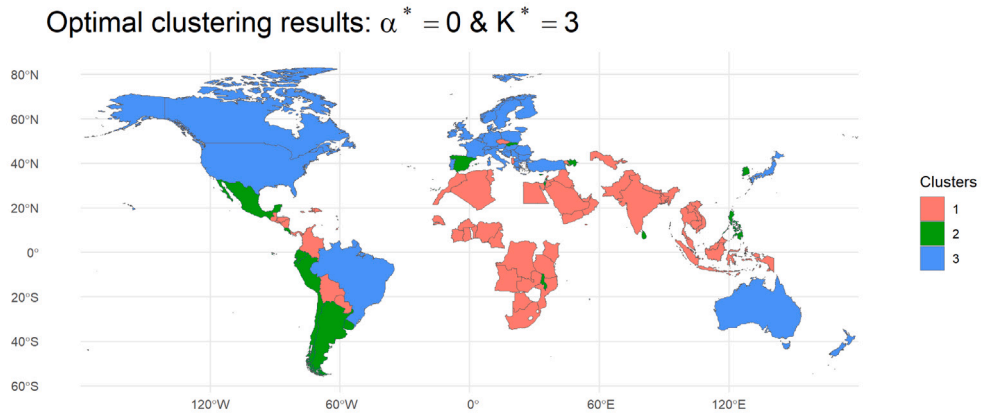

Fig. 5. Map of clustering partitions obtained by setting $\alpha = 0$ and $K = 3$ and using as unique feature the share of people in the 2022 survey with medium-low and low climate change awareness.

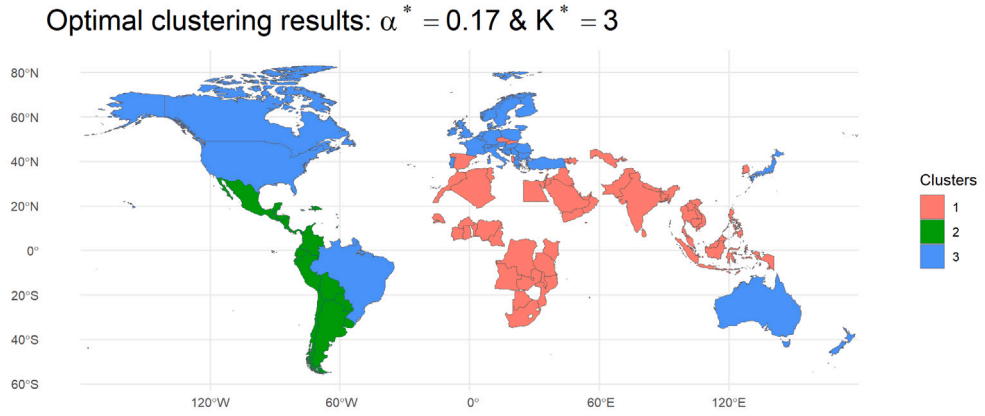

Fig. 6. Map of clustering partitions obtained by setting $\alpha = 0.17$ and $K = 3$ and using as unique feature the share of people in the 2022 survey with medium-low and low climate change awareness.

Figs. 5 through 8 show the maps of the partitions generated using $K = 3$ or $K = 4$ clusters with $\alpha = 0$ or $\alpha^*_{K,max} = 0.17$ (for $K = 3$) or $\alpha^*_{K,max} = 0.33$ (for $K = 4$), allowing a graphical comparison between the output of clustering with both features and geographical components and that from the clustering ignoring the geographic information. Figs. 5 and 6 show the effect of the spatial restriction when $K = 3$ clusters are considered. In absence of geographical distance, the climate change awareness is able to identify historically-and-economic relevant partitions, with (1) European countries unified with Japan, Brazil, Oceania and North America; (2) African and Asian countries merged together; (3) a third group with a mix of countries from allover the World. When geography is included as a clustering variable (even with a moderate weight, that is, $\alpha^* = 0.17$) the previous blocks are consistently retrieved, while preserving some interesting exceptions, such as Czech Republic, Spain, Slovakia and Albania, that, instead of being allocated to the European cluster, are placed in the cluster largely populated by African and Asian countries. Similarly, when the pair $\{K = 3, \alpha = 0.17\}$ is considered, Spain and Slovakia are grouped with Albania and Czechia in the African–Asian cluster, while Japan and Oceania are separated from Europe and North America and put in an independent group. With the addition of the fourth cluster (Figs. 7 and 8), while the European–North American and African–Asian clusters are detected, the allocation of some countries depends on the presence or the absence of spatial information. For instance, Japan, New Zealand and Oceania are merged with Europe–North America without geography, whereas they split off and moved to an independent cluster in Fig. 8. Similarly, when space is used South and Central America appear to be more compact with respect to the case of awareness-only clustering.

Fig. 9 reports descriptive statistics for the share of respondents declaring low or medium-low climate change awareness across clusters under several settings. Recall that, higher values of this indicator imply a larger fraction of low-awareness respondents (i.e., worse climate-change awareness overall), whereas lower values correspond to improved awareness. For the non-spatial specification with $K = 3$ and $\alpha = 0$ (i.e., the clustering ignoring geographical proximity), the three clusters are clearly ordered by their means: G1 (i.e., African and Asian countries) exhibits the highest mean (worst awareness) and is also the largest group, G2 (i.e., mixed European and South American countries) is intermediate, and G3 (i.e., Western countries) shows the lowest mean, hence the best awareness, with comparatively higher within-cluster heterogeneity and a markedly negative skewness, consistent with mass at relatively high shares and a tail toward lower shares. When spatial information is incorporated ($K = 3$, $\alpha = 0.17$), the mean separation partly changes: G1 and G2 become closer in average levels (both characterized by relatively high shares of low-awareness respondents), while G3 remains the group with the highest awareness; at the same time, dispersion increases notably in G1 and G2, suggesting that imposing spatial coherence can group together geographically proximate units that remain more heterogeneous in the awareness indicator. Moving to $K = 4$, the non-spatial partition (i.e., $\alpha = 0$) essentially refines the awareness gradient by carving out an additional, small cluster (G4) with the lowest mean (best awareness) relative to the other groups, while G1 remains the worst awareness and largest cluster; higher-order moments indicate that the newly formed group differs not only in location but also in shape, pointing to a distinct distributional profile rather than a mere mean shift. Under the spatial specification

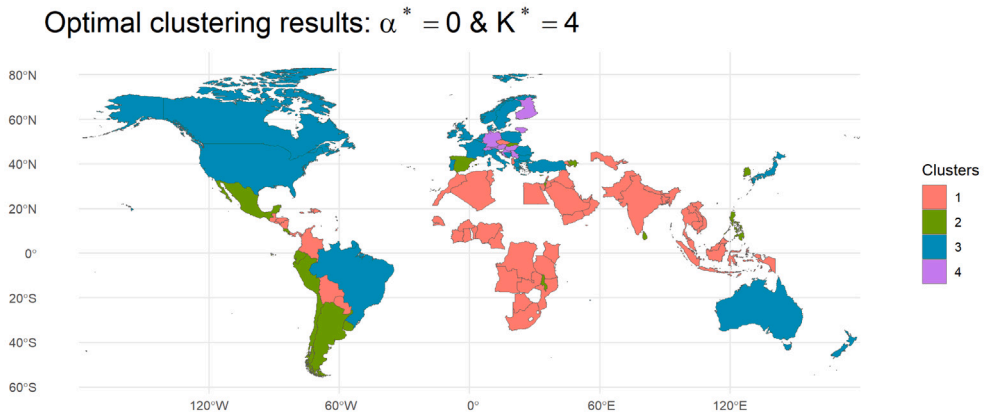

Fig. 7. Map of clustering partitions obtained by setting $\alpha = 0$ and $K = 4$ and using as unique feature the share of people in the 2022 survey with medium-low and low climate change awareness.

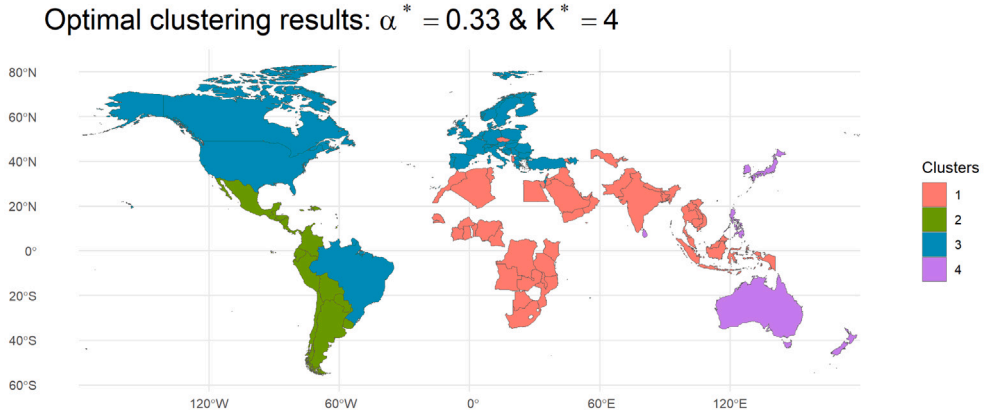

Fig. 8. Map of clustering partitions obtained by setting $\alpha = 0.33$ and $K = 4$ and using as unique feature the share of people in the 2022 survey with medium-low and low climate change awareness.

($\alpha = 0.33$), cluster sizes and moments shift more visibly: the high-mean cluster remains sizable, but the added cluster G4 (i.e., Pacific and Australian countries) no longer sits at the extreme low mean and instead appears as a smaller, spatially coherent group with intermediate mean and substantial dispersion, indicating that the inclusion of space can reallocate extreme-awareness observations across nearby areas and alter within-cluster variability and tail behavior without eliminating the broader pattern of differentiation by awareness levels.

In summary, the four partitions agree in joining European and North American countries and those with advanced economies, as well as in joining African and Middle Eastern countries. Without particular surprise, these two blocs are well evidenced by the descriptive statistics, which show that awareness is significantly higher (on average) and more compact (reduced variability) in the Western block than in the other groups.

3.2. Robustness analysis: spatial hierarchical clustering using climate change awareness, climate-related and socio-economic information

In the main analysis we built the feature dissimilarity matrix D_0 considering only the information provided the climate change awareness survey. In particular, we used the national share of interviewed people declaring a medium-low or low knowledge about climate change.

To expand the empirical study, we enriched the survey data with a set of climate-related and socio-economic variables collected from the United Nations Data Portal, the World Bank Climate Change Knowledge Portal, the Penn World Table 11.0, the Global Carbon Atlas, and the Our World in Data web portal. To account for the presence of missing values, for every country and variable we considered the average value recorded in the period from 2019 to 2022. In Table A.1 of Appendix A we present the complete list of variables gathered, the source from which they were obtained, a synthetic description, and their classification into socio-economic or climate-related variables. In the same table, we also report the essential descriptive statistics for the period 2019–2022. The relationship between climate change awareness data and the additional variables is described in Fig. 10, in which we report the heatmap of the pairwise Pearson's linear correlation index. The plot shows that the medium-high and high awareness is positively correlated with the Human Development Index, the share of government expenditure in GDP, the per capita real GDP at constant national prices, the employment rate, the sum of imported and exported merchandise over GDP, and excess mortality risk (tx84rr). The medium-high and high awareness is also negatively correlated mainly only with the mean surface air temperature and the warm spell duration index, while all the other negative correlations are relatively weak.

To strengthen the main findings discussed in Section 3.1 we performed several robustness check in which we modified the set of features used to generate the clustering while letting unchanged the spatial dissimilarity matrix D_1. We considered four alternative scenarios:

1. High and medium-high climate change awareness + geographical distance (later on denoted as *Robustness: High awareness only*)
2. Low and medium-low climate change awareness + socio-economic features + geographical distance (later on denoted as *Robustness: Awareness + Socioeconomic*)

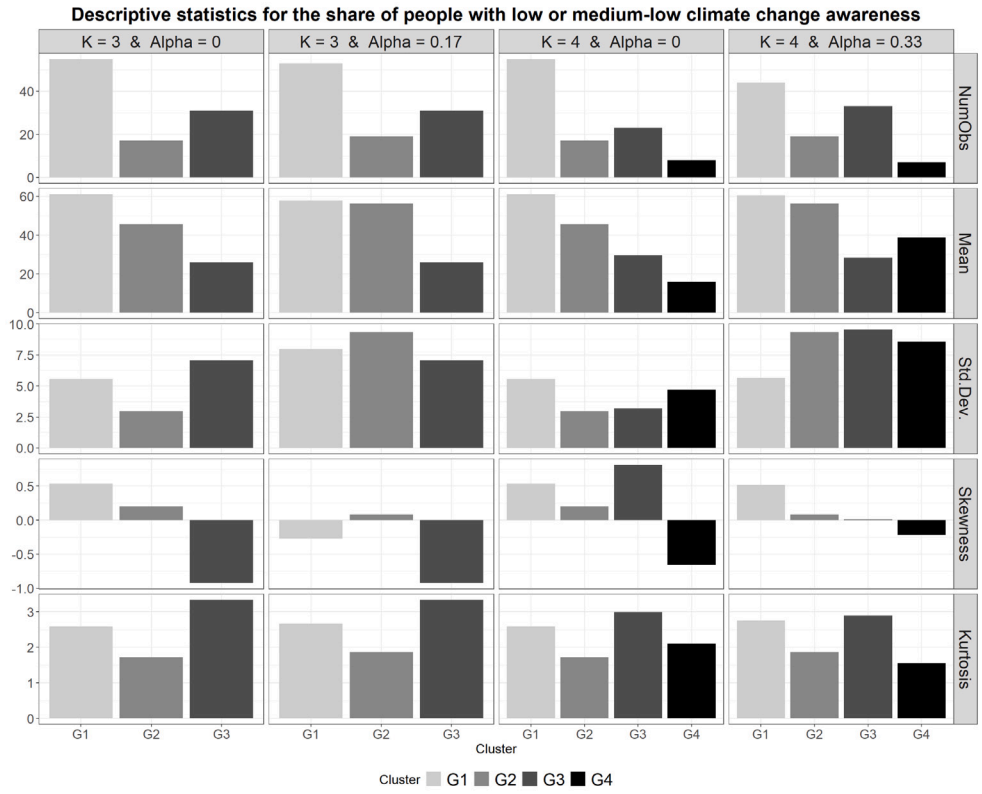

Fig. 9. Descriptive statistics by clustering setting for the share of interviewed people declaring a low or medium-low climate change awareness. Statistics (i.e., cluster size, mean, standard deviation, skewness and kurtosis) are computed by cluster and setting (i.e., combination of optimal α and K). Colors allow to compare the statistics across clusters for different settings, while side-by-side bars allow the comparison of the statistics across the settings for a specific cluster.

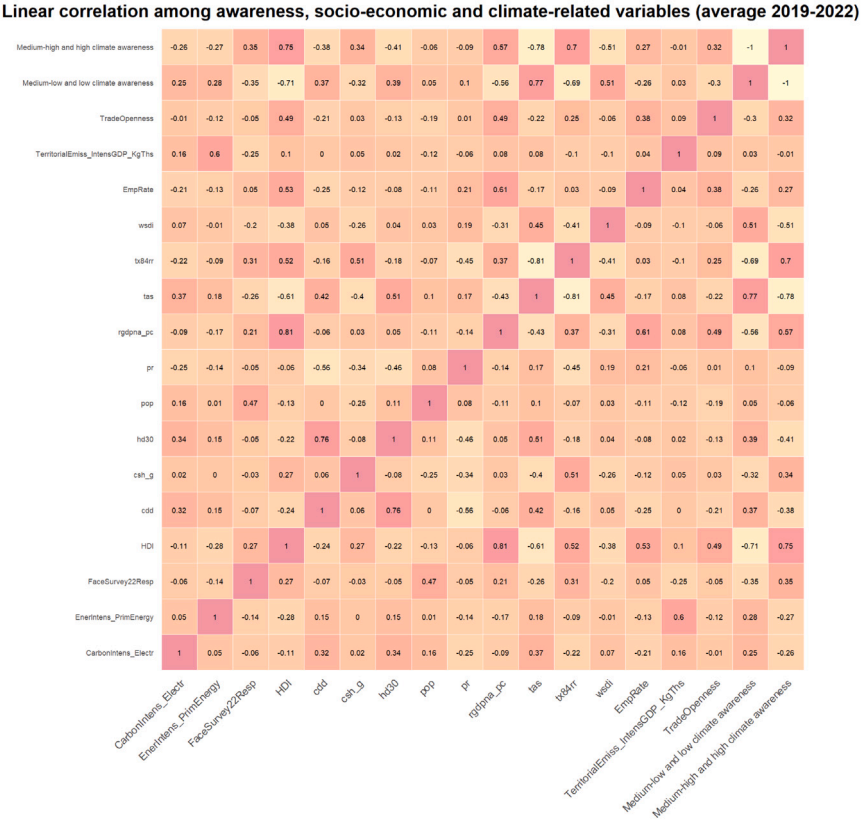

Fig. 10. Heatmap of pairwise Pearson's linear correlation index for the full set of climate awareness, climate-related and socio-economic features.

3. Low and medium-low climate change awareness + climate-related features + geographical distance (later on denoted as *Robustness: Awareness + Climate*)
4. Low and medium-low climate change awareness + socio-economic features + climate-related features + geographical distance (later on denoted as *Robustness: Awareness + Climate + Socioeconomic*)

The robustness analysis aims to assess the coherence between the partitions obtained in the awareness-only case (main analysis) and those derived under alternative scenarios, while keeping fixed the number of clusters K and the mixing parameter α. In particular, all alternative clustering exercises are based on the values of K and α identified in the main analysis, namely $(K = 3, \alpha = 0)$, $(K = 3, \alpha = 0.43)$, $(K = 4, \alpha = 0)$, and $(K = 4, \alpha = 0.21)$. The resulting partitions are then compared across different feature sets.[8] The combination of these values yields twelve additional partitions, which are considered alongside the four from the main analysis and the four obtained from the robustness exercise based on *high or medium-high climate change awareness*.

Given the large number of possible outputs, the full set of results is reported in the Online Supplementary Material, where maps corresponding to each scenario and combination of hyperparameters are provided. In the main text, we summarize the comparison of partitions using the Adjusted Rand Index (ARI) (e.g., see Section 7.2 of [55]). Among others, the ARI is commonly used to assess the degree of agreement between two independently derived partitions of the same set of objects; specifically, values close to 1 indicate strong agreement, whereas values near 0 or negative suggest little or no concordance.

However, it is important to note that fixing (K, α) across different feature sets may lead to suboptimal partitions in the alternative scenarios, as the optimal configuration of these hyperparameters is data-dependent. Therefore, the comparison of partitions (e.g., via the ARI) should be interpreted as conditional on the baseline specification and primarily aimed at assessing the coherence and robustness of clustering outcomes, rather than their absolute optimality (see Fig. 11).

The main findings from the robustness analysis can be summarized as follows:

- The highest similarity is obtained when comparing the partitions generated using awareness alone (both high and low awareness). This can be seen by observing the 8x8 quadrant in the top right, where some of the highest ARI values are present;
- The use of $K = 3$ or $K = 4$ groups generates highly-coherent partitions both in the case of geographically-informed and purely-feature clustering: this is highlighted from the cells around the northeast-southwest diagonal, which show several ARIs above 0.80, except for the cases including the socio-economic features without spatial component (i.e., ARI = 0.24 and ARI = 0.34);
- The coherence between awareness-only and covariates-augmented partitions is low or moderate: in fact, the ARI varies between 0 and 0.38 (with a clear prevalence of values close to zero) in the first eight rows of the matrix, highlighting that the inclusion of further socio-economic or climate information can strongly change the partitions generated by the proposed algorithm. However, when the geographical information is used (see, for example, rows 3–4 and 7–8) the partitions tend to be more similar;
- The results obtained when using awareness and/or socio-economic and climate features in addition to the geographic information (i.e., all the *Robustness* rows/column with $\alpha \neq 0$) are robust to the inclusion of socio-economic and climate change-related information (see, for instance, the intersection of *Awareness+Socio* and *Awareness+Socio+Climate*), while they remarkably deviate from the results obtained when ignoring the geographical information (i.e., all the *Robustness* rows/column with $\alpha = 0$); see, for example, the cells at the intersection between *Awareness+Socio*, *Awareness+Climate* and *Awareness+Socio+Climate* when $\alpha = 0$ and $K = \{3, 4\}$;
- Comparison of the maps in the Online Supplementary Material shows that when geography is included, the clusters reflect (consistently) continental blocks (e.g., European countries are always clustered together, as is the Asian–Pacific block) or cultural–historical blocks (e.g., European countries, North America, and Australia). When geography is omitted some of the blocks remain (e.g., stability of European countries or the North America–Europe–Japan cluster) while blocks such as South America and Africa are scattered and more heterogeneous. These considerations apply across the various scenarios considered;

In general, the above points suggest greater stability of partitions generated by including geographic information than the purely-feature clustering case. More specifically, the geographically-informed clustering leads to more geographically-compact aggregations (clusters reflect continental or cultural blocks); reduces the potential variations generated as the chosen number of clusters varies and thus to more stable and less volatile partitions; and generates consistent partitions even when information sets of different nature and varying numbers of features are considered.

4. Conclusions

In this paper, we investigated global patterns of climate change awareness through a geographically-informed hierarchical clustering framework that explicitly accounts for spatial constraints. To this end, we combined awareness indicators with a broad set of climate-related and socio-economic variables, allowing for the identification of cross-country patterns and their spatial heterogeneity. Our approach builds upon the *hclustgeo* clustering framework proposed by Chavent et al. [13], where a mixing parameter α is used to linearly combine the dissimilarity matrix in the feature space with geographical distances. In this context, we addressed the open issue of hyperparameter tuning by introducing a data-driven procedure for the joint selection of the mixing parameter α and the number of clusters K. The proposed approach combines multiple criteria (including within-cluster homogeneity, between-cluster separation, and the comparison between spatially constrained and unconstrained solutions) to identify clustering configurations that balance feature similarity and spatial coherence.

The empirical findings show that incorporating geographical information substantially reshapes clustering outcomes relative to purely feature-based approaches. When clustering is based exclusively on feature similarity, the resulting partitions tend to be geographically fragmented, with countries belonging to the same cluster often scattered across distant regions. Although such configurations capture similarities in awareness levels, they may generate spatially disconnected groupings that are less informative from an interpretative and policy-oriented perspective. In contrast, the inclusion of geographical information generally produces more spatially coherent partitions, often revealing regional or culturally homogeneous aggregation patterns that are less evident when spatial proximity is ignored. Similar evidence is consistently observed across the alternative specifications considered in the robustness analysis.

Robustness analyses based on alternative sets of socio-economic and climate-related variables further support the consistency of the main findings across different specifications. Overall, the results reveal a marked spatial heterogeneity in climate change awareness, with Western countries (Europe, the Americas, and Oceania) generally exhibiting higher and more homogeneous awareness levels, while countries in

[8] To maintain coherence with the main analysis, all socio-economic and climate-related features were normalized using the same robust scaling procedure described in Section 2.3

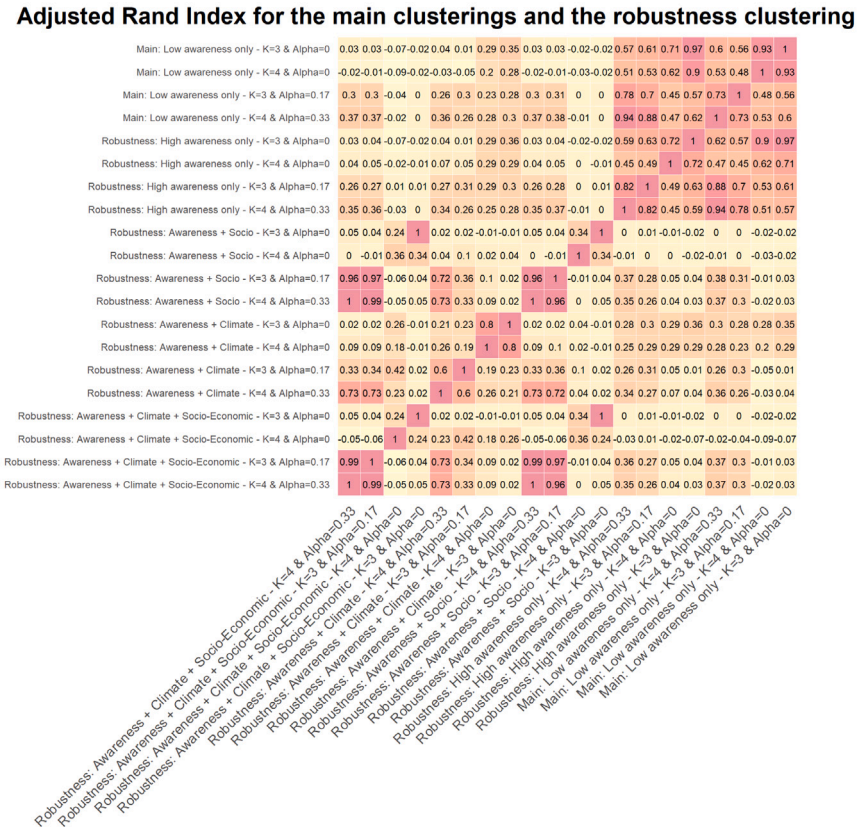

Fig. 11. Heatmap of pairwise Adjusted Rand Index for the partitions from the main analysis (marked as *main*) and the partitions generated by the robustness checks. Values near to 1 indicate high similarity among the two partitions, while negative or null values (expected value is 0) is associated with partitions selected at random or completely different. Columns or rows marked as *Robustness: Awareness + Climate + Socioeconomic* refer to the clustering jointly considering climate awareness, climate-related and socio-economic features; Columns or rows marked as *Robustness: Awareness + Climate* refer to the clustering jointly considering climate awareness and climate-related features; Columns or rows marked as *Robustness: Awareness + Socioeconomic* refer to the clustering jointly considering climate awareness and socio-economic features.

Asia, Africa, and the Middle East display lower and more heterogeneous profiles. These findings highlight the relevance of geographically-informed clustering as a flexible tool for comparative analysis and policy-oriented segmentation in the study of global climate awareness.

To conclude, the results should be interpreted in light of several limitations. First, the analysis relies on Facebook-based survey data which, despite the application of country-specific survey weights calibrated on census benchmarks, demographic characteristics, and platform usage patterns, may still be affected by residual sampling and selection biases related to differential internet access and social media penetration. Although the weighting procedure adopted in this paper improves representativeness, some discrepancies may persist, particularly in countries with uneven digital access or underrepresented population groups; nevertheless, robustness analyses based on alternative weighting schemes and population-adjusted specifications confirm the overall stability of the results. Second, the identified spatial partitions are inherently sensitive to the choice and calibration of clustering hyperparameters (e.g., α and K). In the robustness analysis, these parameters are intentionally held fixed to ensure comparability with the main specification. While this choice may lead to suboptimal partitions for alternative feature sets, additional tuning exercises indicate that the resulting deviations are moderate and do not materially affect the main conclusions, although some degree of model dependence cannot be fully excluded. Third, uncertainty in country-level estimates limits the strength of the generalizations that can be drawn. Fourth, geodetic distances are computed using country centroids, which may provide only a coarse proxy of actual human proximity, particularly for large or spatially discontinuous countries (e.g., Russia, US, Indonesia), where population distributions are highly uneven. This limitation was partially mitigated through the integration of socio-economic, climatic, and awareness-related variables within the multivariate distance framework, as well as through robustness analyses based on augmented specifications and population-weighted schemes, which together contribute to a more realistic representation of cross-country proximity.

Despite these limitations, the analysis offers a coherent and globally comparable framework for assessing climate change awareness and its spatial heterogeneity. By integrating behavioral, socio-economic, and environmental dimensions within a unified clustering approach, the study provides a flexible basis for comparative analysis and policy-oriented interpretation. Future research can build on this framework to extend coverage to underrepresented contexts and to develop predictive tools capable of inferring awareness levels from observable structural characteristics, thereby supporting more targeted and geographically nuanced climate communication strategies.

CRediT authorship contribution statement

Gianpaolo Zammarchi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Paolo Maranzano:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

The authors did not receive support from any organization for the submitted work.

Declaration of competing interest

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

Appendix A. List of climate-related and socio-economic variables used in the robustness analysis

See Table A.1.

Appendix B. Descriptive statistics and maps of the unweighted number of respondents to the 2022 climate change opinion survey at the country level.

See Fig. B.1 and Tables B.2–B.4.

Appendix C. Robustness of the spatial distance: Euclidean distance versus geodetic distance

See Fig. C.2.

Table A.1
Climate change awareness, socio-economic and climate-related variables used in the empirical analysis, data sources and essential descriptive statistics over the period 2019–2022.

Variable name	Variable description	Source	Min	Average	Median	Max	SD	Skew	Kurt
Medium-low and low climate awareness	Share of respondents with no or little awareness about climate change	CCOS	8.12	47.96	54.13	72.77	16.50	−0.53	2.18
Medium-high and high climate awareness	Share of respondents with moderate or high awareness about climate change	CCOS	19.30	49.68	43.08	91.72	18.01	0.48	2.07
CarbonIntens_Electr	Kg of CO_2-equivalents emitted per kWh of electricity generated	OWID	23.71	365.97	377.38	796.01	192.86	−0.13	2.00
EnerIntens_PrimEnergy	Amount of energy needed to produce one unit of economic output (MJ/$)	OWID	1.32	4.20	3.62	19.90	2.46	3.41	19.64
HDI	Human Development Index	UNDP	42.73	76.59	77.88	96.35	14.29	−0.51	2.29
csh_g	Share of government expenditure over GDP	PWT	2.27	17.90	18.22	47.43	6.73	0.62	5.32
rgdpna_pc	Per capita real GDP at constant national prices (in mil. 2021US$)	PWT	1043	24863	16690	107477	22209	1.32	4.63
EmpRate	Employment rate (%)	PWT	18.97	43.80	44.34	73.58	9.36	−0.24	3.63
Pop	Population (mill.)		0.87	51.26	17.10	1366.42	142.88	7.82	71.31
TradeOpenness	Sum of imported and exported merchandise over GDP (%)	PWT	11.82	64.63	50.20	444.56	58.64	3.44	19.90
TerrEmiss_IntensGDP	Ratio between CO_2 emissions (μg) and real GDP	GCA	0.21	27.88	10.98	669.67	71.54	7.33	64.58
cdd	Maximum number of consecutive dry days	CCKP	7.94	60.51	32.86	300.48	60.02	1.69	5.34
hd30	Number of hot days (Tmax >30 °C)	CCKP	0.00	25.49	3.52	180.32	44.38	2.03	6.17
pr	Precipitation	CCKP	16.53	1136.26	943.59	3388.01	788.00	1.05	3.93
tx84rr	Excess Mortality	CCKP	0.11	1.10	0.90	2.53	0.69	0.44	1.91
wsdi	Warm Spell Duration Index	CCKP	4.07	21.04	19.20	97.91	11.97	2.81	17.92
tas	Average Mean Surface Air Temperature	CCKP	−3.38	18.94	21.08	29.19	7.48	−0.55	2.28

Abbreviations: CCOS = Climate Change Opinion Survey (2022); OWID = Our World In Data; UNDP = United Nations Data Portal; GCA = Global Carbon Atlas; PWT = Penn World Table 11.0; CCKP = World Bank Climate Change Knowledge Portal.

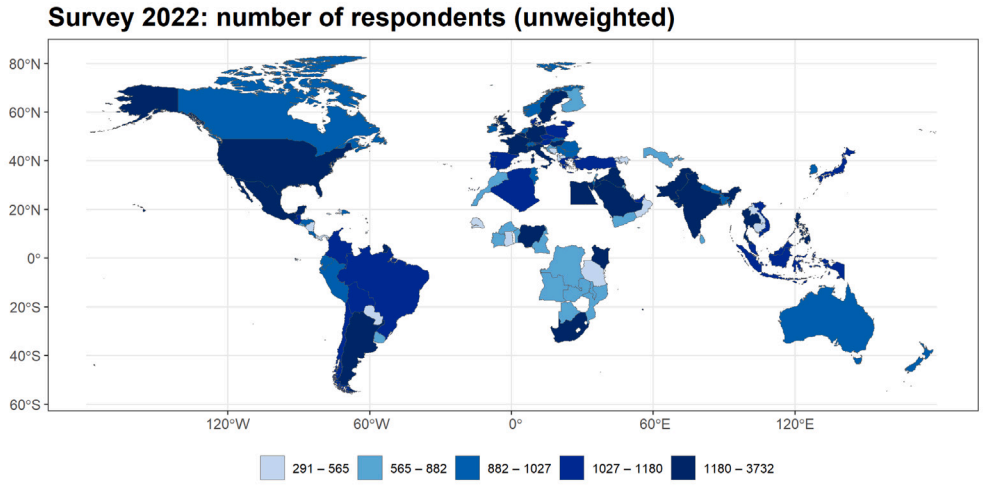

Fig. B.1. World map representing the unweighted number of respondents to the 2022 Climate Change Opinion Survey in the 103 countries used in this empirical analysis.

Table B.2
Synthetic statistics on the 2022 Climate Change Opinion Survey in the 103 countries used in the empirical analysis supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender (source: WorldPopulationReview).(Part 1).

	ISO	Country	Number of respondents to the CCOS	Share (%) of people with med-low	Share (%) of people with med-high	Pop.	Facebook users (mln)			% female users		% male users	
	Code	Name	2022 survey	or low clim. aware.	or high clim. aware.	(mln)	2025	2024	2023	2025	2024	2025	2024
1	AL	Albania	329	53	44	2.88	1.87	1.87	1.10	45	45	55	55
2	DZ	Algeria	1066	61	37	43.05	32.63	32.63	20.80	43	43	57	57
3	AO	Angola	727	62	35	31.83	5.82	5.82	3.50	42	42	58	58
4	AR	Argentina	1257	46	54	44.78	41.70	41.70	27.40	54	54	46	46
5	AM	Armenia	306	61	37	2.96	2.27	2.27	1.40	54	54	46	46
6	AU	Australia	1012	28	72	25.20	22.70	22.70	14.90	52	52	48	48
7	AT	Austria	1120	18	81	8.96	5.41	5.41	3.10	52	52	48	48
8	AZ	Azerbaijan	541	49	49	10.05	5.08	5.08	1.50	45	45	55	55
9	BD	Bangladesh	910	61	35	163.05	67.18	67.18	43.30	35	35	65	65
10	BE	Belgium	1201	28	72	11.54	8.68	8.84	6.30	48	52	52	48
11	BJ	Benin	665	71	25	11.80	2.89	2.89	1.10	38	38	62	62
12	BO	Bolivia	1070	55	42	11.51	8.68	8.68	7.10	48	48	52	52
13	BA	Bosnia–Herzegovina	291	27	73	3.30	2.02	2.02	1.40	51	51	49	49
14	BW	Botswana	690	57	39	2.30	1.42	1.42	1.00	50	50	50	50
15	BR	Brazil	1048	34	65	211.05	175.11	175.11	0.11	55	55	45	45
16	BG	Bulgaria	1012	32	67	7.00	4.43	4.43	3.40	53	53	47	47
17	BF	Burkina Faso	581	64	32	20.32	3.47	3.47	1.90	30	30	70	70
18	KH	Cambodia	445	73	23	16.49	13.86	13.86	10.40	46	46	54	54
19	CM	Cameroon	606	61	32	25.88	6.17	6.17	3.60	42	42	58	58
20	CA	Canada	1026	26	74	37.41	32.00	32.00	20.60	53	53	47	47
21	CL	Chile	1094	42	57	18.95	18.68	18.68	12.20	52	52	48	48
22	HK	Hong Kong	417	58	41	7.44	6.40	6.40	4.20	58	58	42	42
23	CO	Colombia	1082	62	36	50.34	43.60	43.60	33.50	52	52	48	48
24	CR	Costa Rica	1012	43	56	5.05	4.41	4.41	3.10	51	51	49	49
25	HR	Croatia	635	17	83	4.13	2.57	2.57	1.80	52	52	48	48
26	CY	Cyprus	494	44	56	0.87	1.28	1.28	0.79	49	49	51	51
27	CZ	Czech Rep.	1049	57	43	10.69							
28	CI	Côte d'Ivoire	646	61	31	25.72	8.76	8.76	5.70	38	38	62	62
29	CD	D.R. of Congo	587	68	27	86.79	8.07	8.07	4.70	37	37	63	63
30	DK	Denmark	1101	25	75	5.77	4.85	4.85	3.50	54	54	46	46
31	DO	Dominican Rep.	884	60	35	10.74	0.05	8.31	5.20	52	50	48	50
32	EC	Ecuador	1017	48	50	17.37	15.32	15.32	12.00	50	50	50	50
33	EG	Egypt	1206	62	35	100.39	56.40	56.40	42.00	40	40	60	60
34	SV	El Salvador	817	64	33	6.45	4.85	4.85	3.60	53	53	47	47

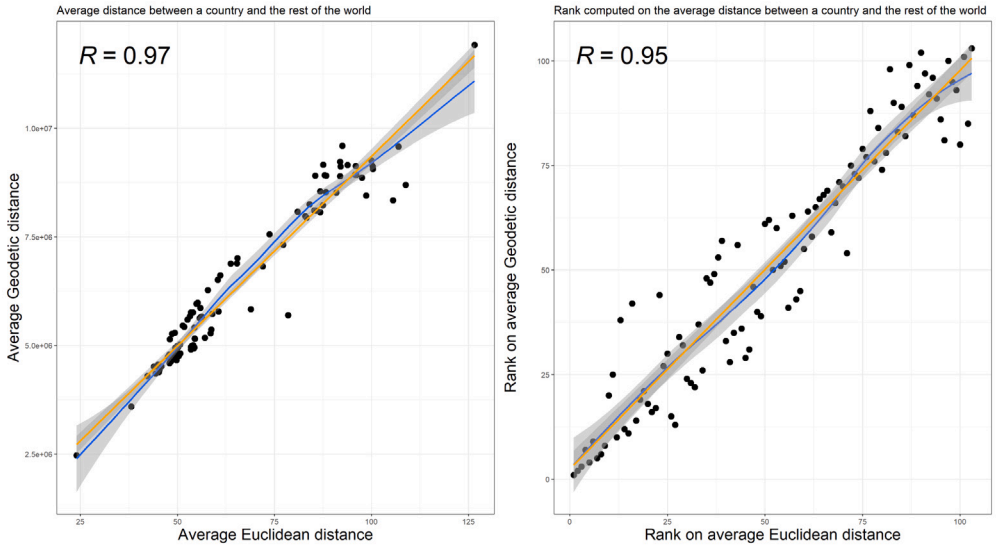

Fig. C.2. Diagrams of the empirical relationship between the average distance of each country from the rest of the world. Left panel: scatterplot of the average country-to-World distance computed using the Euclidean formula (X axis) and the geodetic formula (Y axis). Right panel: empirical relationship between the ranked distances calculated on the average Euclidean and geodesic distances.

Table B.3
Synthetic statistics on the 2022 Climate Change Opinion Survey in the 103 countries used in the empirical analysis supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender (source: WorldPopulationReview).(Part 2).

	ISO	Country	Number of respondents to the CCOS	Share (%) of people with med-low	Share (%) of people with med-high	Pop.	Facebook users (mln)			% female users		% male users	
	Code	Name	2022 survey	or low clim. aware.	or high clim. aware.	(mln)	2025	2024	2023	2025	2024	2025	2024
35	FI	Finland	683	8	92	5.53	3.86	3.86	2.40	55	55	45	45
36	FR	France	2536	27	72	67.35	47.30	47.30	30.40	53	53	47	47
37	DE	Germany	3732	15	84	83.52	46.60	46.60	24.50	52	52	48	48
38	GH	Ghana	563	72	23	30.42	10.09	10.09	5.70	42	42	58	58
39	GR	Greece	1051	32	68	10.47	7.30	7.30	5.00	50	50	50	50
40	GT	Guatemala	1031	68	29	17.58	10.48	10.48	8.00	50	50	50	50
41	HT	Haiti	543	73	19	11.26	3.21	3.21	1.90	47	47	53	53
42	HN	Honduras	1012	61	33	9.75	5.41	5.41	4.00	53	53	47	47
43	HU	Hungary	1280	10	90	9.68	7.11	7.11	5.50	53	53	47	47
44	IN	India	2574	55	42	1366.42	581.58	581.58	314.60	32	32	68	68
45	ID	Indonesia	1178	66	29	270.63	174.00	174.00	119.90	46	46	54	54
46	IQ	Iraq	1403	63	34	39.31	30.93	30.93	17.90	35	35	65	65
47	IE	Ireland	977	28	72	4.88	4.06	4.06	2.50	55	55	45	45
48	IL	Israel	914	41	57	8.52	6.92	6.92	4.50	52	52	48	48
49	IT	Italy	1279	28	71	60.55	43.60	43.60	27.90	51	51	49	49
50	JM	Jamaica	375	70	30	2.95	1.72	1.72	0.85	54	54	46	46
51	JP	Japan	1174	29	71	126.86	61.70	61.70	15.10	56	56	44	44
52	JO	Jordan	1315	57	42	10.10	7.25	7.25	4.90	46	46	54	54
53	KE	Kenya	1423	64	31	52.57	17.04	17.04	9.30	44	44	56	56
54	KW	Kuwait	568	64	32	4.21	4.63	4.63	1.80	38	38	62	62
55	LA	Lao People's DR	323	72	27	7.17	4.64	4.64	3.20	47	47	53	53
56	LB	Lebanon	653	57	40	6.86	4.53	4.53	3.00	47	47	53	53
57	LT	Lithuania	1162	17	82	2.76	2.17	2.17	1.60	55	55	45	45
58	MW	Malawi	874	50	46	18.63	1.67	1.67	0.75	37	37	63	63
59	MY	Malaysia	1153	60	37	31.95	31.00	31.00	20.30	48	48	52	52
60	MX	Mexico	1218	49	50	127.58	111.41	111.41	0.08	52	52	48	48
61	MA	Morocco	788	54	41	36.47	27.50	27.50	17.30	42	42	58	58
62	MZ	Mozambique	756	69	27	30.37	3.91	3.91	2.30	42	42	58	58
63	NP	Nepal	896	61	35	28.61	16.48	16.48	0.01	44	44	56	56
64	NL	Netherlands	978	26	74	17.10	12.87	12.87	8.10	54	54	46	46
65	NZ	New Zealand	971	34	65	4.78	4.47	4.47	3.00	53	53	47	47
66	NI	Nicaragua	519	61	38	6.55	4.45	4.45	3.30	50	50	50	50
67	NG	Nigeria	1382	70	26	200.96	51.23	51.23	21.80	43	43	57	57
68	MK	North Macedonia	464	37	63	2.08	1.37	1.37	0.91	48	48	52	52
69	NO	Norway	1004	29	70	5.38	4.60	4.60	3.30	53	53	47	47
70	OM	Oman	499	60	36	4.97	3.77	3.77	1.20	33	33	67	67

Table B.4
Synthetic statistics on the 2022 Climate Change Opinion Survey in the 103 countries used in the empirical analysis supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender (source: WorldPopulationReview).(Part 2).

	ISO	Country	Number of respondents to the CCOS	Share (%) of people with med-low	Share (%) of people with med-high	Pop.	Facebook users (mln)			% female users		% male users	
	Code	Name	2022 survey	or low clim. aware.	or high clim. aware.	(mln)	2025	2024	2023	2025	2024	2025	2024
71	PK	Pakistan	1181	61	31	216.57	60.42	60.42	37.30	24	24	76	76
72	PA	Panama	541	54	44	4.25	3.43	3.43	1.80	51	51	49	49
73	PY	Paraguay	550	62	34	7.04	4.81	4.81	3.50	50	50	50	50
74	PE	Peru	952	49	50	32.51	29.05	29.05	22.90	48	48	52	52
75	PH	Philippines	1249	44	51	108.12	102.32	102.32	80.30	53	53	47	47
76	PL	Poland	1069	30	70	37.89	24.75	24.75	17.90	55	55	45	45
77	PT	Portugal	1029	29	70	10.23	9.09	9.09	5.90	53	53	47	47
78	QA	Qatar	498	61	33	2.83	3.46	3.46	1.90	30	30	70	70
79	KR	Republic of Korea	921	44	56	51.23	27.15	27.15	9.60	53	53	47	47
80	RO	Romania	907	35	62	19.36	12.34	12.34	9.60	52	52	48	48
81	SA	Saudi Arabia	1252	57	38	34.27	29.18	29.18	10.80	34	34	66	66
82	SN	Senegal	502	51	42	16.30	4.66	4.66	2.60	36	36	64	64
83	RS	Serbia	949	21	78	6.99	4.83	4.83	2.90	52	52	48	48
84	SG	Singapore	369	43	56	5.80	5.32	5.32	3.30	51	51	49	49
85	SK	Slovakia	985	45	55	5.46	3.47	3.47	2.50	53	53	47	47
86	ZA	South Africa	1369	55	43	58.56	31.97	31.97	22.10	51	51	49	49
87	ES	Spain	1163	43	56	46.74	35.90	35.90	19.40	53	53	47	47
88	LK	Sri Lanka	632	50	49	21.32	9.41	9.41	6.50	41	41	59	59
89	SE	Sweden	1338	30	69	10.04	8.75	8.75	5.80	53	53	47	47
90	CH	Switzerland	990	21	79	8.59	5.65	5.65	3.00	51	51	49	49

(*continued on next page*)

Table B.4 (*continued*).

	ISO	Country	Number of respondents to the CCOS	Share (%) of people with med-low	Share (%) of people with med-high	Pop.	Facebook users (mln)			% female users		% male users	
	Code	Name	2022 survey	or low clim. aware.	or high clim. aware.	(mln)	2025	2024	2023	2025	2024	2025	2024
91	TH	Thailand	1220	58	41	69.63	58.30	58.30	48.10	52	52	48	48
92	TT	Trinidad–Tobago	518	56	44	1.39	1.14	1.14	0.75	53	53	47	47
93	TN	Tunisia	992	55	41	11.69	8.84	8.84	6.50	48	48	52	52
94	TR	Turkey	1158	30	70	83.43	70.30	70.30	32.80	44	44	56	56
95	TZ	Tanzania	505	52	39	56.38	8.38	8.38	3.80	41	41	59	59
96	AE	United Arab Emir.	1096	57	41	9.77	13.86	13.86	7.30	31	31	69	69
97	GB	United Kingdom	1951	26	74	67.53	55.90	55.90	34.40	53	53	47	47
98	US	United States	3096	29	71	329.06	279.81	279.81	175.00	54	54	46	46
99	UY	Uruguay	580	46	53	3.46	3.27	3.27	2.00	55	55	45	45
100	UZ	Uzbekistan	710	58	42	32.98	10.10	10.10	1.60	37	37	63	63
101	VN	Viet Nam	1096	60	36	96.46	86.11	86.11	66.20	51	51	49	49
102	YE	Yemen	731	61	33	29.16	5.34	5.34	2.50	17	17	83	83
103	ZM	Zambia	692	57	41	17.86	4.11	4.11	2.60	45	45	55	55

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.seps.2026.102509.

Data availability

For the reproducibility purposes, all scripts and data are made available for the public at the following GitHub folder: https://github.com/PaoloMaranzano/GPZ_PM_HierSpatClusterClimateAwareness.git

GPZ_PM_HierSpatClusterClimateAwareness (Reference data) (GitHub)

References

[1] Loarie SR, Duffy PB, Hamilton H, et al. The velocity of climate change. Nature 2009;462(7276):1052–5.

[2] Wang F, Harindintwali JD, Wei K, et al. Climate change: Strategies for mitigation and adaptation. Innov Geosci 2023;1(1). 100015–61.

[3] Canadell JG, Meyer C, Cook GD, et al. Multi-decadal increase of forest burned area in Australia is linked to climate change. Nat Commun 2021;12(1):6921.

[4] Goss M, Swain DL, Abatzoglou JT, et al. Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. Environ Res Lett 2020;15(9):094016.

[5] Sun Y, Zhang X, Ding Y, et al. Understanding human influence on climate change in China. Natl Sci Rev 2022;9(3):nwab113.

[6] Meza I, Rezaei EE, Siebert S, et al. Drought risk for agricultural systems in South Africa: Drivers, spatial patterns, and implications for drought risk management. Sci Total Environ 2021;799:149505.

[7] Román-Palacios C, Wiens JJ. Recent responses to climate change reveal the drivers of species extinction and survival. Proc Natl Acad Sci 2020;117(8):4211–7.

[8] Andy P. A million threatened species? Thirteen questions and answers. 2019, https://www.ipbes.net/news/million-threatened-species-thirteen-questions-answers. [Accessed 31 August 2024].

[9] Titus JG. Greenhouse effect, sea level rise, and coastal zone management. Coast Manag 1986;14(3):147–71.

[10] Zahran S, Kim E, Chen X, et al. Ecological development and global climate change: A cross-national study of kyoto protocol ratification. Soc Nat Resour 2007;20(1):37–55.

[11] Edmonds JA, Sands RD. What are the costs of limiting CO2 concentrations. Glob Clim Chang: Sci Econ Politics 2003;140–86.

[12] Eckstein D, Künzel V, Laura S. Global climate risk index. 2021, https://reliefweb.int/report/world/global-climate-risk-index-2021. [Accessed 31 August 2024].

[13] Chavent M, Kuentz-Simonet V, Labenne A, et al. ClustGeo: an R package for hierarchical clustering with spatial constraints. Comput Statist 2018;33(4):1799–822.

[14] Knight KW. Public awareness and perception of climate change: A quantitative cross-national study. Environ Sociol 2016;2(1):101–13.

[15] Kvaløy B, Finseraas H, Listhaug O. The publics' concern for global warming: A cross-national study of 47 countries. J Peace Res 2012;49(1):11–22.

[16] Kim SY, Wolinsky-Nahmias Y. Cross-national public opinion on climate change: The effects of affluence and vulnerability. Glob Environ Politics 2014;14(1):79–106.

[17] Lee TM, Markowitz EM, Howe PD, et al. Predictors of public climate change awareness and risk perception around the world. Nat Clim Chang 2015;5(11):1014–20.

[18] Poortinga W, Whitmarsh L, Steg L, et al. Climate change perceptions and their individual-level determinants: A cross-European analysis. Glob Environ Chang 2019;55:25–35.

[19] Weber EU. Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). Clim Change 2006;77(1):103–20.

[20] Gifford R. The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. Am Psychol 2011;66(4):290.

[21] Li Y, Johnson EJ, Zaval L. Local warming: Daily temperature change influences belief in global warming. Psychol Sci 2011;22(4):454–9.

[22] Howe PD, Markowitz EM, Lee TM, et al. Global perceptions of local temperature change. Nat Clim Chang 2013;3(4):352–6.

[23] Myers TA, Maibach EW, Roser-Renouf C, et al. The relationship between personal experience and belief in the reality of global warming. Nat Clim Chang 2013;3(4):343–7.

[24] Leiserowitz AA, Maibach E, Roser-Renouf C, et al. Climate change in the American mind. University of Washington Seattle, WA, USA; 2018.

[25] Hincks S, Carter J, Connelly A. A new typology of climate change risk for European cities and regions: Principles and applications. Glob Environ Chang 2023;83:102767.

[26] Abdelgadir M, AlRashidi M, Alharbi R, et al. Spatial clustering analysis combined with ensemble modeling identified potential coastal conservation hotspots of white-eyed gulls in the red sea. Glob Ecol Conserv 2025;59:e03581.

[27] Yang S, She D, Wang L, et al. Tree size and spatial clustering drive contrasting carbon sequestration in soil aggregates in temperate forests in China. Urban For Urban Green 2025;128998.

[28] Gordon A. A survey of constrained classification. Comput Statist Data Anal 1996;21(1):17–29.

[29] Ambroise C, Dang M, Govaert G. Clustering of spatial data by the EM algorithm. In: GeoENV i—geostatistics for environmental applications: proceedings of the geostatistics for environmental applications workshop, Lisbon, Portugal, 18–19 November 1996. Springer; 1997, p. 493–504.

[30] Liao ZX, Peng WC. Clustering spatial data with a geographic constraint: exploring local search. Knowl Inf Syst 2012;31:153–70.

[31] Miele V, Picard F, Dray S. Spatially constrained clustering of ecological networks. Methods Ecol Evol 2014;5(8):771–9.

[32] Pawitan Y, Huang J. Constrained clustering of irregularly sampled spatial data. J Stat Comput Simul 2003;73(12):853–65.

[33] Oliver M, Webster R. A geostatistical basis for spatial weighting in multivariate classification. Math Geol 1989;21:15–35.

[34] Bourgault G, Marcotte D, Legendre P. The multivariate (co) variogram as a spatial weighting function in classification methods. Math Geol 1992;24:463–78.

[35] Morelli C, Boccaletti S, Maranzano P, et al. Multidimensional spatiotemporal clustering – An application to environmental sustainability scores in europe. Environmetrics 2025;36(2):e2893. http://dx.doi.org/10.1002/env.2893, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2893.

[36] Morelli C, Maranzano P, Otto P. Spatiotemporal clustering of ghgs emissions in europe: Exploring the role of spatial component. Spatial Stat 2026;100960. http://dx.doi.org/10.1016/j.spasta.2026.100960, URL https://www.sciencedirect.com/science/article/pii/S2211675326000084.

[37] Mattera R, Franses PH. Are african business cycles synchronized? Evidence from spatio-temporal modeling. Econ Model 2023;128:106485. http://dx.doi.org/10.1016/j.econmod.2023.106485, URL https://www.sciencedirect.com/science/article/pii/S0264999323002973.

[38] Leiserowitz A, Carman J, Buttermore N, et al. International public opinion on climate change 2022. New Haven, CT: Yale Program on Climate Change Communication and Facebook Data for Good; 2022.

[39] World Population Review. Facebook users by country 2026. 2026, Retrieved January 15, 2026, from, URL https://worldpopulationreview.com/country-rankings/facebook-users-by-country.
[40] Mattera R. A weighted approach for spatio-temporal clustering of COVID-19 spread in Italy. Spat Spatio-Temporal Epidemiology 2022;41:100500. http://dx.doi.org/10.1016/j.sste.2022.100500, URL https://www.sciencedirect.com/science/article/pii/S1877584522000247.
[41] Basu S, Sen S. COVID 19 pandemic, socio-economic behaviour and infection characteristics: An inter-country predictive study using deep learning. Comput Econ 2023;61(2):645–76. http://dx.doi.org/10.1007/s10614-021-10223-5.
[42] Jones SG, Ashby AJ, Momin SR, et al. Spatial implications associated with using euclidean distance measurements and geographic centroid imputation in health care research. Health Serv Res 2010;45(1):316–27.
[43] Klein A. The determinants of international migration in early modern europe: Evidence from the maritime sector, c. 1700–1800. Soc Sci Hist 2020;44(1):143–67, Uses straight-line Euclidean distance between country centroids (“as the crow flies”) to model bilateral migration distances..
[44] Nowotny K. Network migration: Do neighbouring regions matter? Reg Stud 2019. Distance between regions measured in km “as the crow flies”..
[45] Czaika M. On the negative impact of time zone differences on tourist arrivals. Curr Issues Tour 2020. Employs Euclidean (“as the crow flies”) distance between countries as spatial control variable.
[46] Fenn-Moltu G, Ollier S, Bates OK, et al. Global flows of insect transport and establishment: The role of biogeography, trade and regulations. Diversity and Distributions 2023;29(11):1478–91. http://dx.doi.org/10.1111/ddi.13772, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.13772.
[47] Akhanli SE, Hennig C. Comparing clusterings and numbers of clusters by aggregation of calibrated clustering validity indexes. Stat Comput 2020;30(5):1523–44. http://dx.doi.org/10.1007/s11222-020-09958-2.
[48] Rossbroich J, Durieux J, Wilderjans TF. Model selection strategies for determining the optimal number of overlapping clusters in additive overlapping partitional clustering. J Classification 2022;39(2):264–301. http://dx.doi.org/10.1007/s00357-021-09409-1.
[49] Charrad M, Ghazzali N, Boiteau V, et al. Nbclust: An R package for determining the relevant number of clusters in a data set. J Stat Softw 2014;61(6):1–36. http://dx.doi.org/10.18637/jss.v061.i06, URL https://www.jstatsoft.org/index.php/jss/article/view/v061i06.
[50] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[51] Dunn JC. Well-separated clusters and optimal fuzzy partitions. J Cybern 1974;4(1):95–104.
[52] Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free recall. Psychol Bull 1976;83(6):1072.
[53] Caliński T, Harabasz J. A dendrite method for cluster analysis. Comm Statist Theory Methods 1974;3(1):1–27.
[54] McClain JO, Rao VR. Clustisz: A program to test for the quality of clustering of a set of objects. J Mark Res 1975;456–60.
[55] Gordon AD. Classification. CRC Press; 1999.

Gianpaolo Zammarchi is Researcher Tenure Track (Assistant Professor) of Statistics at the Department of Economics and Management of the University of Cagliari, Italy.

Paolo Maranzano is Researcher Tenure Track (Assistant Professor) of Economic Statistics at the Department of Economics, Management and Statistics (DEMS) of the University of Milano-Bicocca. Also, he serves as Senior Researcher within the Econometrics of the Energy Transition research group at Eni Enrico Mattei Foundation in Milano, Italy.