

## Journal Pre-proof

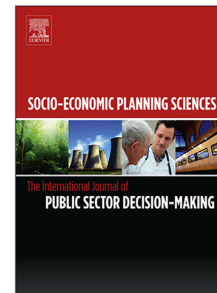
Mapping climate change awareness through spatial hierarchical clustering

Gianpaolo Zammarchi, Paolo Maranzano

PII: S0038-0121(26)00096-0  
DOI: <https://doi.org/10.1016/j.seps.2026.102509>  
Reference: SEPS 102509

To appear in: *Socio-Economic Planning Sciences*

Received date: 10 April 2025  
Revised date: 6 May 2026  
Accepted date: 9 May 2026



Please cite this article as: G. Zammarchi and P. Maranzano, Mapping climate change awareness through spatial hierarchical clustering. *Socio-Economic Planning Sciences* (2026), doi: <https://doi.org/10.1016/j.seps.2026.102509>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier Ltd.

Mapping climate change awareness through spatial  
hierarchical clustering

Gianpaolo Zammarchi<sup>1</sup> and Paolo Maranzano<sup>2,3\*</sup>

<sup>1</sup>Department of Economics and Business Sciences, University of  
Cagliari, Viale Fra Ignazio, 17, Cagliari, 09131, Italy.

<sup>2</sup>Department of Economics, Management and Statistics, University of  
Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milano, 20126, Italy.

<sup>3</sup>Fondazione Eni Enrico Mattei (FEEM), Corso Magenta, 63, Milano,  
20123, Italy.

\*Corresponding author(s). E-mail(s): [paolo.maranzano@unimib.it](mailto:paolo.maranzano@unimib.it);

**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.

001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046

047 **Keywords:** Climate change awareness, Socio-economic and climate-related features,  
048 2022 International Public Opinion on Climate Change survey, Spatial hierarchical  
049 clustering, Data-driven hyperparameters tuning

050  
051  
052  
053  
054  
055

## 1 Climate change and public awareness

056 Climate change represents one of the greatest challenges that this and the future  
057 generations will face. Climate change poses serious risks to the planet's future and  
058 is a cause of environment's degradation. Historically, these changes have happened  
059 during a large part of Earth's history, but their evolution has always been rather  
060 slow if compared to the current trend. In many cases, the changes took hundreds if  
061 not millions of years to manifest themselves in a geologically detectable way (Loarie  
062 et al. 2009). Natural and anthropogenic emissions are both responsible of the current  
063 situation, but the scientific community agrees that humans have a key role in the extent  
064 and speed of these changes. Basically, the main drivers of these negative changes are  
065 associated to human behavior, especially when the rising global population became  
066 focused with the growth at all costs, which frequently leads to disregard the severe  
067 repercussions in natural systems and the consequences that follow from this (Loarie  
068 et al. 2009).

077 Using fossil fuels for power, heating, and transportation has significantly raised  
078 greenhouse gas emissions and changed patterns of precipitations and temperature  
079 throughout the world. In 2022, the average worldwide temperature was approximately  
080 0.86°C higher than the 20th century average and the last 50 years (almost) there was a  
081 continuous exceeding of the average values of previous years (Wang et al. 2023). About  
082 weather patterns, and in particular precipitations, the effects of climate change are  
083 becoming more and more evident globally, as evidenced by the extreme weather events  
084 and related disasters that occur all over the world. Some of the most well-known events  
085 are the forest fires in Australia (like the devastating fires that occurred in Australia  
086  
087  
088  
089  
090  
091  
092

in 2019-2020, [Canadell et al. 2021](#)) and in the United States (especially in California, [Goss et al. 2020](#)) or the intense rainfall in China ([Sun et al. 2022](#)), the droughts in South Africa ([Meza et al. 2021](#)). 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 30% ([Román-Palacios and Wiens 2020](#)), which means millions of animals and plants to disappear ([Andy 2019](#)). 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 levees ([Titus 1986](#)). 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 ([Zahran et al. 2007](#); [Edmonds and Sands 2003](#)). In fact, as stated in the Global Climate Risk Index 2021 issued by the Germanwatch

139 observatory, developing nations are less able to adapt, making them more vulnerable  
140 to the consequences of climate change (Eckstein et al. 2021). In 2019, eight of the ten  
141 most severely impacted countries in terms of deaths and economic losses caused by  
142 extreme weather events (such as storms, floods, heatwaves, etc.) were, low- to lower-  
143 middle-income countries. Bahamas, Zimbabwe and Mozambique were three of the most  
144 impacted countries. Since countermeasures are not so easy and fast to implement,  
145 it is vital that governments of all countries take action to mitigate climate change.  
146 One of the possible pushes to act should come from the populations. For this to hap-  
147 pen, it is necessary for individuals to become aware of the importance and urgency of  
148 interventions. Increasing public awareness on the causes and effects of climate change  
149 is of crucial importance because it can motivate policymakers to take action to cut  
150 greenhouse gas emissions and encourage individual behavioral adjustments. Given the  
151 different levels of vulnerability of countries around the world to the effects of climate  
152 change, the aim of this work was to examine how nations differ in terms of their  
153 degree of knowledge of the problem. Next, we present the findings of an analysis based  
154 on a spatially-constrained hierarchical clustering. To this aim, we compare different  
155 clustering scenarios in which combinations of socio-economic, climate, and geographic  
156 data were tested, while maintaining the settings of number of clusters and the mixing  
157 parameter<sup>1</sup>  $\alpha$  (for more information about  $\alpha$  see Section 2.2).  
158

159 Climate change awareness varies significantly across nations, partly due to the  
160 influence of cultural and environmental factors, as well as their interactions. However,  
161 socioeconomic factors also play their role. Culture and environment can determine  
162 the priority given to addressing climate change, while environmental factors influence  
163 the severity of climate change impacts on the population. Early cross-national stud-  
164 ies have established that wealth, education and unemployment levels, are predictors  
165

166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184

---

<sup>1</sup>In the framework proposed by Chavent et al. (2018), the mixing parameter  $\alpha$  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.

of climate awareness globally [Knight \(2016\)](#); [Kvaløy et al. \(2012\)](#). [Kim and Wolinsky-Nahmias \(2014\)](#) 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 [Lee et al. \(2015\)](#). 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. \(2019\)](#), 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 [Weber \(2006\)](#); [Gifford \(2011\)](#). Additionally, several authors have examined the effects of weather on climate change awareness and found that relationship with personal observation of local weather [Li et al. \(2011\)](#); [Howe et al. \(2013\)](#); [Myers et al. \(2013\)](#) and weather extremes [Leiserowitz et al. \(2018\)](#), 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

231 region, or on a few or single determinants. In our study, however, we aimed at collect-  
232 ing data from over one hundred countries worldwide and jointly assessing numerous  
233 socioeconomic factors and the effects of spatial interdependence.  
234  
235

### 237 **1.1 Related works on geographical clustering and our** 238 **contribution** 239

241 In recent years, many works have shown the usefulness of performing spatial clustering  
242 with data of different nature. [Hincks et al. \(2023\)](#) developed a new spatial typology of  
243 climate change risk for European cities and regions based on k-means clustering. They  
244 used 49 variables at the NUTS-3 level for 33 European countries and showed a sub-  
245 stantial spatial heterogeneity in climate risk across Europe. Their findings show that  
246 climate change risk present patterns which can be addressed only with an approach  
247 that takes into account the transboundary nature of the problem whose fight may be  
248 less efficient because of different political commitment, resource availability, and risk  
249 awareness. In their work, [Abdelgadir et al. \(2025\)](#) showed that if spatial clustering is  
250 combined with predictive modeling it can become a powerful tool for identifying con-  
251 servation areas for endangered species. By combining spatial clustering and ensemble  
252 species distribution modeling they achieved 88-95% prediction accuracy in mapping  
253 the White-eyed gull biogeography. The analysis revealed that climate change will likely  
254 significantly impact the spatial distribution of these seabirds, reducing current suit-  
255 able habitats and shifting it toward the Mediterranean Sea. The spatial clustering  
256 approach can successfully identify geographically coherent areas requiring protection  
257 and can be generalized to other species and with appropriate changes it might also be  
258 possible to adapt it to other contexts. [Yang et al. \(2025\)](#) analyzed the usefulness of  
259 the spatial clustering analysis to discover how organisms' spatial organization affects  
260 climate-related activities like carbon sequestration. They investigated how tree size  
261 and spatial clustering affected soil organic carbon storage in forests in Northeast China,  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276

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 spa-  
tial 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.

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., [Gordon 1996](#); [Ambroise et al. 1997](#); [Liao and Peng 2012](#);  
[Miele et al. 2014](#); [Pawitan and Huang 2003](#)).

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 contin-  
uous 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  $i$ -th element is contiguous to the  $j$ -th ele-  
ment 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

323 separation. In accordance with this logic, some methods proposed to merge the dissim-  
324 ilarity matrix derived from non-geographical features and the matrix of geographical  
325 distances (Oliver and Webster 1989; Bourgault et al. 1992). In this combination, the  
326 weights assigned to the geographical dissimilarities will provide more or less geograph-  
327 ically contiguous clusters, but in this type of approach the issue is shifted to choosing  
328 the right weights.  
329  
330

331 This paper pursues several complementary goals, spanning methodological devel-  
332 opment, comparative assessment, and empirical application. From an applied per-  
333 spective, the study aims to identify and interpret global patterns of climate change  
334 awareness, with particular emphasis on their spatial heterogeneity and their relation-  
335 ship with broader socio-economic and environmental conditions. A second objective  
336 is to evaluate the extent to which incorporating geographical (spatial) information  
337 improves the identification of homogeneous groups of countries, and how spatial  
338 constraints reshape clustering outcomes relative to purely feature-based approaches.  
339 Moreover, the proposed research framework is designed to support policy-relevant seg-  
340 mentation of countries, providing a structured basis for understanding cross-country  
341 differences in climate awareness and for informing geographically targeted communi-  
342 cation strategies. We address these goals by building upon the Ward-like hierarchical  
343 clustering technique with geographical (or spatial) constraints introduced by Chavent  
344 et al. (2018) and further extended to the spatio-temporal setting by Morelli et al.  
345 (2025) and ?. This method employs a convex combination of two dissimilarity matrices,  
346 namely  $D_0$  and  $D_1$ , which contain, respectively, information about a set of clustering  
347 features (in our case climate awareness, climate-related and socio-economic variables)  
348 and the geography of the area of interest. The two dissimilarities are linearly combined  
349 through a mixing hyperparameter  $\alpha$ , enabling a direct comparison between spatially  
350 constrained and unconstrained partitions. The method is suitable with both Euclidean  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368

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  $\alpha$  and the number of clusters  $K$ , remains an open issue in the literature, often addressed through ad hoc or sequential procedures (Mattera and Franses 2023). Therefore, we contribute to this gap by proposing a data-driven procedure for the joint selection of  $\alpha$  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

415 robustness checks aimed at assessing the sensitivity of the clustering results to alter-  
416 native specifications. These include: (i) an extended specification that incorporates a  
417 broad set of climate-related and socio-economic variables, in addition to awareness  
418 and spatial information, to evaluate how these factors affect clusters' composition;  
419 and (ii) a comparison between an unweighted clustering approach, where all obser-  
420 vations receive equal weight, and a population-weighted version that accounts for  
421 cross-country differences in population size. The results of all robustness exercises  
422 are reported in detail in the [Online Supplementary Material](#). Finally, Section 4 con-  
423 cludes the paper by summarizing the main findings and outlining directions for future  
424 research.  
425  
426  
427  
428  
429  
430  
431  
432

## 433 2 Data and methods

### 434 2.1 Survey data on climate change awareness

435 To quantify the level of awareness of each country we used the “International public  
436 opinion on climate change” survey. The most complete currently available edition is  
437 the 2022 edition which includes responses from more than 108 thousands Facebook  
438 users located in 192 countries worldwide ([Leiserowitz et al. 2022](#)). The survey was  
439 carried out between March 25th and April 14th, 2025, in collaboration with “Data for  
440 Good at Meta” and the main aim was to examine people’s knowledge, beliefs, atti-  
441 tudes, policy preferences, and behavior related to climate change. The data collected  
442 is representative of both actual Facebook users and the overall resident population  
443 in a given country<sup>2</sup>. Indeed, for each country surveyed, the survey consists of sam-  
444 ples in proportion to publicly available age and gender benchmarks. Specifically, data  
445 were weighted separately for each country using a multi-stage, pre-and-post-survey  
446 weighting process based on census and nationally representative survey benchmarks,  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457

458 <sup>2</sup>We refer the readers to the Survey Method section in Appendix I “Survey Method” of [Leiserowitz et al.](#)  
459 [\(2022\)](#) for details on the data collection strategy, including the full list of surveyed countries, country-specific  
460 adjustments and the sample distribution across the countries.

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 Face-  
book 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 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 geo-  
graphic or polygonal coordinates. The final count of countries considered in the rest of  
the paper is hundred-three ( $N = 103$ )<sup>3</sup>. Figure B1 in Appendix B shows a map repre-  
senting the unweighted number of respondents to the 2022 Climate Change Opinion  
Survey in the 103 countries considered here. In addition, Figure 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).

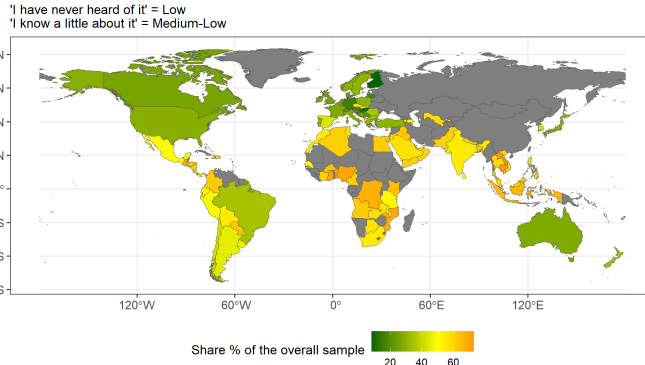
While the complete survey includes answers to numerous questions regarding the  
degree of knowledge about climate change, its main causes, perception of the phe-  
nomenon, and the degree of concern about it<sup>4</sup>, we collected data pertaining to “Overall

<sup>3</sup>Tables B2, B3, and B4 of the Appendix reports the number of unweighted respondents for the  $n = 103$  countries considered in the empirical analysis (taken from Appendix I of Leiserowitz et al. 2022) supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender (source: World Population Review 2026).

<sup>4</sup>We refer the reader to Appendix II “Results by Region” of Leiserowitz et al. (2022) for an extensive descriptive analysis of the responses recorded for each question in the survey based on region of residence.

507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552

### Share of people with medium-low or low climate awareness (2022)



**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”).

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 [Leiserowitz et al. \(2022\)](#), 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. 2018](#)) without having to adopt methodological extensions, such as the multi-matrix case developed by [Morelli et al. \(2025\)](#), which would significantly increase statistical complexity at the expense of interpretative simplicity<sup>5</sup>.

## 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. \(2018\)](#), 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

---

<sup>5</sup>In the scenario of a dissimilarity matrix derived from a socio-economic feature ([Chavent et al. 2018](#)), we would only need to optimize two hyperparameters of the clustering model (the weight of the geographical component  $\alpha$  and the number of clusters  $K^*$ ), while in the case of multiple variables, and therefore multiple matrices ([Morelli et al. 2025](#)), the scalar weight would become a vector with  $P$  elements, where  $P$  is the number of clustering variables to be optimized.

599 feature space in our application will comprehend a set of climate change awareness,  
600 socio-economic, and climate-related variables at the national level for 2022.

602 Notice that, to model geographic proximity between countries, we employ the  
603 Euclidean distance between the centroids of national polygons, computed in a  
604 projected coordinate system and expressed in kilometres. This choice follows a  
605 well-established practice in spatial econometrics and geostatistical clustering litera-  
606 ture (Mattera 2022; Basu and Sen 2023; Mattera and Franses 2023) and provides  
607 several advantages. In particular, by employing the Euclidean distance for both socio-  
608 economic data and space allow us to remain within a Euclidean framework even in  
609 the multi-matrix case, as the two matrices are combined using a convex linear combi-  
610 nation. It follows that all the clustering metrics that could be calculated in the case  
611 of single-matrix Euclidean clustering, can also be used in this case of geographically-  
612 informed clustering without restrictions. Moreover, we stress that several applications  
613 demonstrate that the Euclidean distance provides a reasonable approximation to more  
614 realistic travel or geodetic (i.e., Great Circle) distances. For instance, Jones et al. (2010)  
615 report a correlation of 0.99 between road-network driving distance and Euclidean dis-  
616 tance between ZIP-code centroids, with relatively modest absolute differences<sup>6</sup>. Recent  
617 studies explicitly employ the Euclidean distance between country centroids as a proxy  
618 for travel or trade distance or international spatial proximity (Klein 2020; Nowotny  
619 2019; Czaika 2020; Fenn-Moltu et al. 2023). In line with this literature, we interpret the  
620 Euclidean centroid distance as a parsimonious proxy for the geographical separation  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633

---

634 <sup>6</sup>We replicated this check on our data by computing the linear correlation on the average distance among  
635 each country and the other countries in the sample according to either Euclidean and Geodetic metrics.  
636 According to our finding, the use of Euclidean distance generates values and rankings (to be understood  
637 as the positioning of a country in terms of average distance from other countries in the sample) that are  
638 compatible and consistent with those of geodesic distance. As proof of this, in Figure C2 of Appendix C we  
639 provide the diagrams of the empirical relationship between the average distance of each country from the  
640 rest of the world (i.e., average country-to-World distance) calculated using the Euclidean formula (X axis)  
641 and the geodetic formula (Y axis) (left panel) and the empirical relationship between the ranks calculated  
642 on the same average Euclidean and geodesic distances (right panel). The rankings slightly between  $\pm 20$   
643 positions and the linear correlation is close to 1, showing an high degree of coherence between the two  
644 methods.

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  $\alpha$  (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  $\alpha$  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  $\alpha$  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  $\alpha$  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  $I_\alpha$  denote

691 the mixed pseudo inertia for a given value of  $\alpha$ ; therefore, the mixed pseudo inertia of  
 692 a generic cluster  $C_k^\alpha$  can be defined as follows:  
 693

$$694 \quad I_\alpha(C_k^\alpha) = (1 - \alpha) \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{0,ij}^2 + \alpha \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{1,ij}^2 \quad (1)$$

698 where  $w_i$  is a user-defined weight for the  $i$ -th observation with  $i = 1, \dots, n$ ,  $\mu_k^\alpha =$   
 699  $\sum_{i \in C_k^\alpha} w_i$  is the overall weight of the partition  $C_k^\alpha$ ,  $d_{0,ij}^2$  is the normalized  $n \times n$   
 700 dissimilarity between observations  $i$  and  $j$  in  $D_0$ , and  $d_{1,ij}^2$  is the analogous in  $D_1$ . In  
 701 the mixed setup by [Chavent et al. \(2018\)](#), as the mixed pseudo inertia of cluster  $C_k^\alpha$   
 702 decreases, the units belonging to cluster  $C_k^\alpha$  are more homogeneous (i.e.,  $\downarrow I_\alpha(C_k^\alpha)$   
 703 then  $\uparrow$  homogeneity within  $C_k^\alpha$ ). When considering a partitioning into  $K$  clusters,  
 704 the overall homogeneity is given by the mixed within-cluster pseudo inertia, namely  
 705  $W(\mathcal{P}_K)$ , which is computed as

$$706 \quad W_\alpha(\mathcal{P}_K^\alpha) = \sum_{k=1}^K I_\alpha(C_k^\alpha) \quad (2)$$

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

709 We stress that the unit-level weights  $w_i$  (for  $i = 1, \dots, n$ ) are pre-defined by the  
 710 user and can assume either a single value common to all the observations or a unit-  
 711 specific value (see Section 2 of [Chavent et al. 2018](#), for a formal discussion on the  
 712 topic). In the former case, also referred to as the uniform weighting design, we assume  
 713 that  $w_i = \frac{1}{n} \quad \forall i$ ; conversely, in the latter, that is, the non-uniform weighting design,  
 714 the generic value  $w_i$  can be computed based on other information, such as using the  
 715 population count or socio-economic variables like the GDP. Confident in the fact that  
 716 the questionnaire responses have already been weighted and adjusted for the demo-  
 717 graphic composition of the countries and respondents, in the empirical application  
 718

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 $\alpha$ 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 2020](#)) or model-based indicators based on the likelihood function (e.g., likelihood criteria as in [Rossbroich et al. 2022](#)). 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  $\alpha$ , that is,  $\alpha^*$ . Also, we recall that there exists a mutual relationship among  $\alpha$  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  $\alpha^*$ , we recall the proposal by [Chavent et al. \(2018\)](#) and [Morelli et al. \(2025\)](#), both based on a sequential approach in which  $\alpha$  is determined by conditioning on the number of clusters  $K$ . The former proposal defines  $\alpha$  that best compromises between loss of feature-based and loss of geographical homogeneity. In other words, [Chavent et al. \(2018\)](#) determine  $\alpha$  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

783 clustering based only on the geographical distance), namely  $\tilde{Q}_{D_1}(\mathcal{P}_K^\alpha)$ . Conversely,  
 784  
 785 [Morelli et al. \(2025\)](#) propose to select the mixing parameter  $\alpha$  which jointly maximizes  
 786 the amount of pseudo inertia explained from both the socio-economic features and  
 787 the geographical information, weighted by the cumulated spatial and socio-economic  
 788 pseudo inertia embedded the data. A comprehensive discussion about the computa-  
 789 tion and the interpretation of the two methods, as well as their relationship with the  
 790 clustering-related metrics, is provided in Section 4 of [Morelli et al. \(2025\)](#) and Section  
 791 3 of [Chavent et al. \(2018\)](#).

796 In [Morelli et al. \(2025\)](#), the authors propose several sequential algorithms to show  
 797 how the selection of both  $K$  and  $\alpha$  can be performed; in particular, conditioning on an  
 798 initial value of  $K$ , the mixing parameter  $\alpha^*$  is find according to the criterion above, and  
 799 then the number of clusters  $K^*$  is re-optimized using one or more clustering criteria  
 800 defined by the user. Here, we propose a procedure that combines the [Morelli et al.](#)  
 801 [\(2025\)](#) and [Chavent et al. \(2018\)](#) criteria and extends the set of clustering indicators  
 802 used to select a suitable number of clusters and the mixing parameters. The proposed  
 803 procedure is described in Algorithm 1. In practice, Algorithm 1 is implemented setting  
 804 a maximum number of clusters  $K_{max} = 10$  and a grid of candidate  $\alpha$  values ranging  
 805 from 0 to 1 with step size  $\Delta\alpha = 0.01^7$ . Furthermore, since distances depend on the  
 806 unit of measurement and the scale of the data (in particular, when the Euclidean  
 807 distance is calculated between countries scattered across the globe, the values can be  
 808 very high), both the spatial dissimilarity matrix and the feature dissimilarity matrix  
 809 were normalized. The spatial matrix was normalized with respect to its maximum  
 810 value as  $\frac{X_i}{\max(X)}$ , while the climate change awareness matrix was normalized using a  
 811 robust scaling that accounts for potential outliers, namely  $\frac{X_i - \text{median}(X)}{q_{0.75}(X) - q_{0.25}(X)}$ , where  $X_i$   
 812 is the feature's value associated with a country  $i$  and  $q_{0.75}(X)$  and  $q_{0.25}(X)$  are the  
 813 75° and 25° quantiles, respectively.

826  
 827 <sup>7</sup>For a given setting (i.e., weighted or unweighted clustering), the total number of hyperparameters combina-  
 828 tions is  $15 \times 101 = 1515$  and requires a computing time lower than two minutes to be executed without  
 specific computational and hardware requirements.

<b>Algorithm 1</b> Spatial Hierarchical Clustering: choice of the hyperparameters $\alpha$ and $K$	829
<b>Step 1: define the algorithm's inputs</b>	830
Define as $D_0 = [d_{0,ij}]_{i,j=1,\dots,n}$ the (normalized) feature dissimilarity matrix	831
Define as $D_1 = [d_{1,ij}]_{i,j=1,\dots,n}$ the (normalized) spatial dissimilarity matrix	832
Define as $K_{max}$ the maximum number of clusters	833
Define as $\underline{\alpha} \in [0, 1]$ a sequence of mixing parameters	834
	835
<b>Step 2: identify the set of potential optimal mixing parameters</b>	836
<b>for</b> $K = 2, \dots, K_{max}$ <b>do</b>	837
<b>for</b> $\alpha \in \underline{\alpha}$ <b>do</b>	838
Compute the linear combination of the two dissimilarity matrices $D(\alpha) =$	839
$(1 - \alpha)D_0 + \alpha D_1$ ;	840
Compute the $\mathcal{P}_K^\alpha =$ partition in $K$ clusters according to Ward's hierarchical	841
algorithm on $D$ ;	842
Compute the weighted average of the explained mixed pseudo inertia $\bar{Q}(\mathcal{P}_K^\alpha)$	843
as in <a href="#">Morelli et al. (2025)</a> and the two normalized pseudo inertia $\bar{Q}_{D_0}(\mathcal{P}_K^\alpha)$ and	844
$\bar{Q}_{D_1}(\mathcal{P}_K^\alpha)$ as in <a href="#">Chavent et al. (2018)</a> ;	845
<b>end for</b>	846
Select the best $\alpha$ for each $K$ such that $\alpha_{K,max}^* = \operatorname{argmax}_\alpha \bar{Q}(\mathcal{P}_K^\alpha)$ (that is,	847
according to the criterion by <a href="#">Morelli et al. (2025)</a> ) and the best $\alpha$ for each $K$ such	848
that $\alpha_{K,min}^* = \operatorname{argmin}_\alpha  \bar{Q}_{D_0}(\mathcal{P}_K^\alpha) - \bar{Q}_{D_1}(\mathcal{P}_K^\alpha) $ (that is, according to the criterion	849
by <a href="#">Chavent et al. (2018)</a> )	850
<b>end for</b>	851
	852
<b>Step 3: define the optimal combination of number of clusters and mixing parameters</b>	853
Choose $K^*$ (evaluated at the corresponding $\alpha_{K,max}^*$ and $\alpha_{K,min}^*$ ) based on a set of	854
hierarchical clustering criteria, such as the Silhouette index, the Dunn's index, the	855
C-index, the Calinski-Harabasz's index, and the McClain-Rao's index. In particular,	856
consider the combination of three potential rule-of-thumb: (1) graphical analysis of	857
the indices for varying $\alpha$ 's and $K$ 's; (2) majority voting rule on the absolute values	858
of the clustering criteria; (3) majority voting rule on the gain/loss of the indices	859
evaluated at $\alpha_{K,max}^*$ and $\alpha_{K,min}^*$ (i.e., geographically-informed clustering) compared	860
to the case of the indices at $\alpha = 0$ (i.e., non-spatial clustering).	861
	862
	863
The algorithm works in a three-stage setting. In the first stage, the user has to	864
define the hierarchical clustering's inputs (i.e., the feature dissimilarity and spatial dis-	865
similarity matrices, the maximum number of clusters to be considered, and a sequence	866
of candidate mixing parameters to weight the matrices. In the second step we identify	867
the set of potential optimal mixing parameters based on both the <a href="#">Morelli et al. (2025)</a>	868
(denoted as $\alpha_{K,max}^*$ ) and <a href="#">Chavent et al. (2018)</a> (denoted as $\alpha_{K,min}^*$ ) criteria for all	869
	870
	871
	872
	873
	874

875 the considered  $K$ 's. In the third step we use a combination of three potential rule-of-  
 876 thumb to identify the definitive number of clusters  $K^*$ , that is, graphical analysis of  
 877 the indices for varying  $\alpha$ 's and  $K$ 's; majority voting rule (Charrad et al. 2014) on the  
 878 absolute values of a set of clustering criteria (i.e., Silhouette index Rousseeuw (1987),  
 880 the Dunn's index Dunn (1974), the C-index Hubert and Levin (1976), the Calinski-  
 881 Harabasz's index Caliński and Harabasz (1974), and the McClain-Rao's index McClain  
 882 and Rao (1975)); majority voting rule on the gain/loss of the previous set of indices  
 883 evaluated at  $\alpha_{K,max}^*$  and  $\alpha_{K,min}^*$  (i.e., geographically-informed clustering) compared  
 884 to the case in which we consider only the feature space and ignore the geographical  
 885 information (i.e.,  $\alpha = 0$ ). Specifically, the latter rule-of-thumb is based on the com-  
 886 putation of the percentage gain or loss obtained by using a geographically-informed  
 887 clustering approach instead of the clustering without geographical constraints, that  
 888 is when fixing  $\alpha = 0$ . For a given index, the gain/loss is obtained as the following  
 889 percentage variation:

$$898 \quad GL = \frac{Index_{\alpha_K^*} - Index_{\alpha_K=0}}{Index_{\alpha_K=0}} \times 100. \quad (3)$$

903 Notice that the definition of gain or loss depends on how each index is optimized.  
 904 Indeed, when considering the C-index and the McClain-Rao index, the optimal number  
 905 of clusters is the minimizer, while for Silhouette, Calinski-Harabasz, and Dunn indices  
 906 the optimal value is  $K$  such that the indicators are maximized. Thus, for the former  
 907 two a gain (loss) is detected when the indicator associated with the geographically-  
 908 informed clustering returns a lower (greater) value compared to the non-spatial case.  
 909 Conversely, for the latter three a gain (loss) is detected when the indicator associated  
 910 with the geographically-informed clustering returns a greater (lower) value compared  
 911 to the non-spatial case. In other words, for Silhouette, Calinski-Harabasz, and Dunn  
 912 the spatial information induces improvements for positive variations, while for the  
 913  
 914  
 915  
 916  
 917  
 918  
 919  
 920

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 2.3. 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.

**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  $(\alpha_K^*, K^*)$  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 Akhanli and Hennig 2020), 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.,

967  $w_i = \frac{1}{n} \quad \forall i = 1, \dots, 103$ ) are used, whereas in Section 3.2 we provide the results for a  
 968  
 969 robustness check in which the feature dissimilarity matrix is computed by combing the  
 970 share of respondents to the survey with socio-economic and climate-related country-  
 971 specific variables.

972  
 973 In the [Online Supplementary Material](#) we report the results obtained from other  
 974  
 975 robustness experiments, such as the spatial clustering in which the country-specific  
 976 awareness is weighted by the total resident population (i.e.,  $w_i = Pop_i \quad \forall i =$   
 977  $1, \dots, 103$ ) and the clustering performed using the share of respondents declaring an  
 978 high or medium-high climate change awareness. Overall, the results obtained without  
 979 weighting the observations and those obtained by weighting by total resident popula-  
 980 tion are mutually consistent. First, the optimal number of clusters is typically found to  
 981 lie between two and four (see, e.g., Figures 37–39 in the [Online Supplementary Mate-](#)  
 982 [rial](#)). Second, the relative importance of the spatial component, as captured by the  
 983 parameter  $\alpha$ , ranges between 0.21 and 0.58, with optimal values equal to  $\alpha = 0.21$  when  
 984  $K = 4$  and  $\alpha = 0.43$  when  $K = 3$ . Third, the resulting partitions continue to reflect  
 985 the main socio-economic, political, and historical blocks identified in the unweighted  
 986 clustering (see Figures 40–43 in the [Online Supplementary Material](#)). Finally, the inclu-  
 987 sion of additional climate-related and socio-economic covariates yields results that are  
 988 fully consistent and qualitatively very similar to those obtained in the unweighted case  
 989 (Figure 45 in the [Online Supplementary Material](#)) presented below.

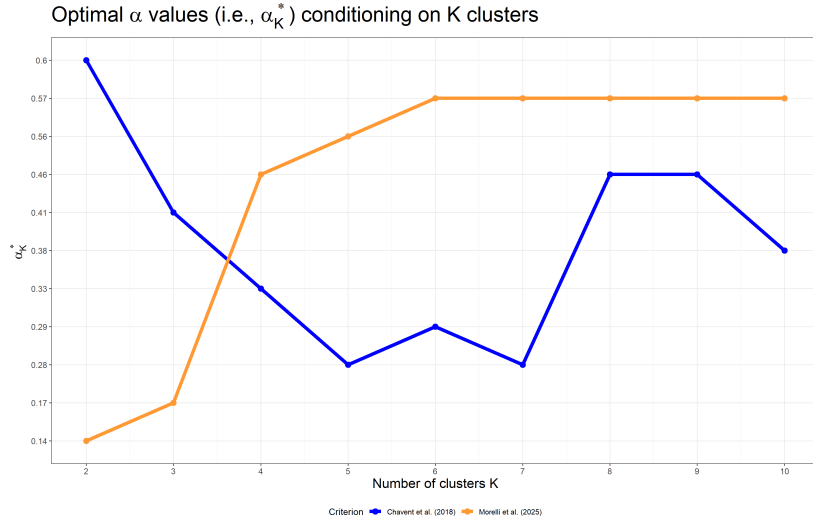
1000

### 1001 **3.1 Spatial hierarchical clustering using climate change**

#### 1002 **awareness information and geographical distance**

1003  
 1004  
 1005 In Figure 2 we show the optimal  $\alpha$  values (i.e.  $\alpha_K^*$ ) computed for each value of  $K$   
 1006 from 2 to 10. The  $(K_{max} - 1) \alpha_K^*$  values were obtained using the criteria proposed  
 1007 by [Morelli et al. \(2025\)](#) (orange line) and [Chavent et al. \(2018\)](#) (blue line). From the  
 1008 figure we can easily infer that while the [Chavent et al. \(2018\)](#)'s criterion generally  
 1009  
 1010  
 1011  
 1012

assigns a lower weight to the geographical information (i.e.,  $\alpha_{K,min}^*$  lies between 0.28 1013  
and 0.46 when  $K \geq 4$ , with higher values for a fewer number of clusters), the criterion 1014  
by Morelli et al. (2025) suggest a more relevant role for the spatial constraint (i.e., 1015  
above 0.46 for  $K \geq 4$ ). 1016  
1017  
1018  
1019



**Fig. 2** Optimal values of the mixing parameter  $\alpha$  conditioning on  $K$  clusters (that is,  $\alpha_K^*$ ) obtained 1042  
using the criterion by Morelli et al. (2025) ( $\alpha_{K,max}^*$  in orange) and by Chavent et al. (2018) ( $\alpha_{K,min}^*$  1043  
in blue). 1044  
1045

As shown in Figure 3, for a given value of  $K$  and  $\alpha_K^*$  it is possible to obtain a specific 1046  
index value for each of the five proposed metrics. These metrics were the Silhouette 1047  
index, the Dunn's index, the C-index, the Calinski-Harabasz's index, and the McClain- 1048  
Rao's index. However, not all the metrics work in the same manner. In fact, for the 1049  
C-index and the McClain-Rao index the optimal value is the one corresponding to the 1050  
minimizer, while for Silhouette, Dunn and Calinski-Harabasz indices indicate that the 1051  
optimal value is the maximum. Taking these facts into account, Figure 3 can be used 1052  
to establish which value of  $K$  corresponds to the optimal value, that we will indicate 1053  
1054  
1055  
1056  
1057  
1058

1059 as  $K^*$ . Potentially, any combination of metrics (i.e., the indices) and methods (i.e.,  
 1060  
 1061 Chavent et al. (2018) and Morelli et al. (2025)) could return a different value for  $K^*$ ,  
 1062 thus leading to a maximum of ten optimal  $K^*$ . In our analysis only seven different  
 1063 values were selected as a potential optimal value for  $K$ . These values were 2 (four  
 1064 times), 3 (one time), 4 (one time), 5 (one time), 6 (one time), 7 (one time), and 8 (one  
 1065 time).  
 1066 times), 3 (one time), 4 (one time), 5 (one time), 6 (one time), 7 (one time), and 8 (one  
 1067 time).  
 1068

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

1084

1085

1086

1087

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

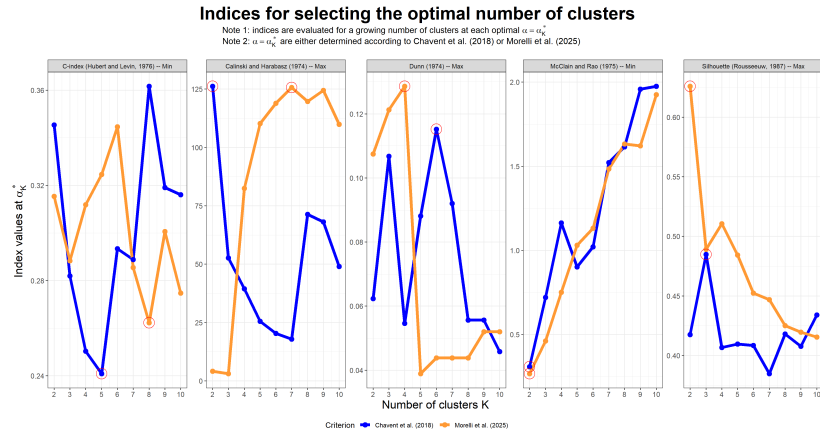
1100

1101

1102

1103

1104

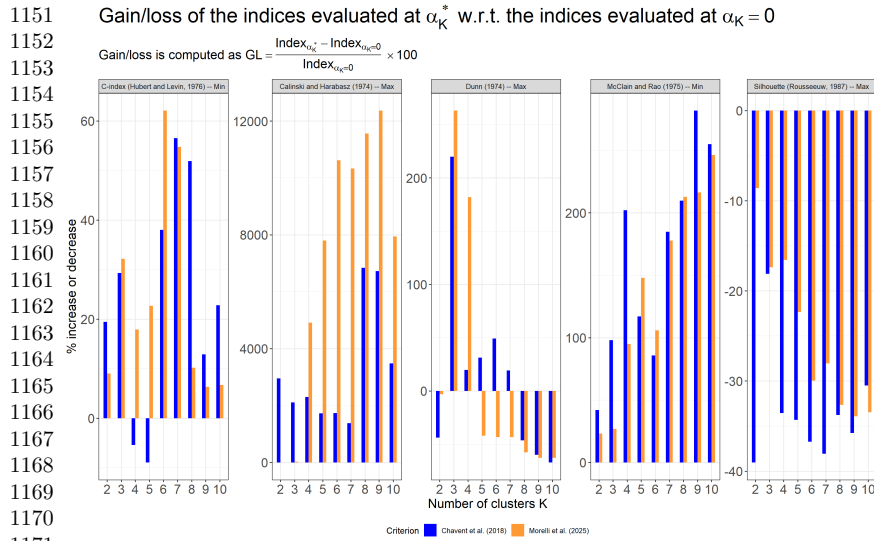


1088 **Fig. 3** Estimates of the Silhouette index, the Dunn's index, the C-index, the Calinski-Harabasz's  
 1089 index, and the McClain-Rao's index for  $K$  clusters evaluated at the corresponding  $\alpha_{K,max}^*$  (orange)  
 1090 and  $\alpha_{K,min}^*$  (blue). Red circles identifies the optimal values, that is, the minimizer for C-index and  
 1091 McClain-Rao's indices and the maximizer for Silhouette, Dunn's and Calinski-Harabasz's indices.

1092  
 1093 We previously stated that one of the objectives of the analysis is to understand if  
 1094 the geographical information can improve the clustering findings, and thus the inter-  
 1095 pretation of the climate change awareness phenomenon. Then, we carried on a direct  
 1096 comparison between the five validation indices obtained including and excluding the  
 1097 geographical information. In Figure 4 we show the estimated percentage gain or loss of  
 1098 the five metrics when using the geographical constraint (i.e., indices are evaluated at  
 1099  $\alpha = \alpha_{K,min}^*$  or  $\alpha = \alpha_{K,max}^*$ ) compared to the case of clustering without geographical

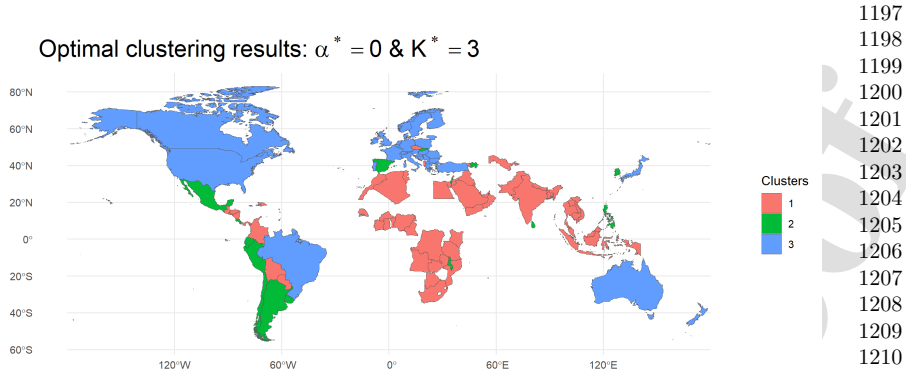
constraints (i.e., indices are evaluated at  $\alpha = 0$ ). For all the indices except the Silhouette, we observe that the optimized  $\alpha_K^*$  produce larger values in the absolute scale, leading to positive percentage increments. The magnitude of the gain varies based on the metrics and the specific  $K$  values. The Silhouette index is the only one showing negative variations for all the considered combinations of  $K$  and  $\alpha$ . 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  $\alpha_K^*$  values computed using the method proposed by Morelli et al. (2025) 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. (2025) 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. (2025). Below we present the results related to these two cases when considering the geographical constraint and when ignoring the spatial dimension.

Figures 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. Figures 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 all over the World. When geography is included as a clustering variable (even with

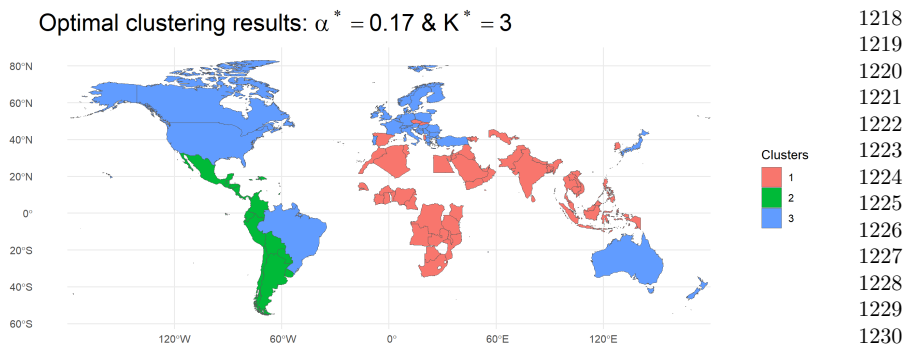


**Fig. 4** Estimated percentage gain (positive values) or loss (negative values) of the Silhouette, Dunn's, C-index, Calinski-Harabasz's index, and McClain-Rao's indices evaluated at the corresponding  $\alpha_{K,max}^*$  (orange) and  $\alpha_{K,min}^*$  (blue) with respect to the indices evaluated 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  $\alpha = 0$ .

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 (Figures 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



**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.

are merged with Europe-North America without geography, whereas they split off and moved to an independent cluster in Figure 8. Similarly, when space is used South and Central America appear to be more compact with respect to the case of awareness-only clustering.

1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260

1261

1262

1263

1264

1265

1266

1267

1268

1269

1270

1271

1272

1273

1274

1275

1276

1277

1278

1279

1280

1281

1282

1283

1284

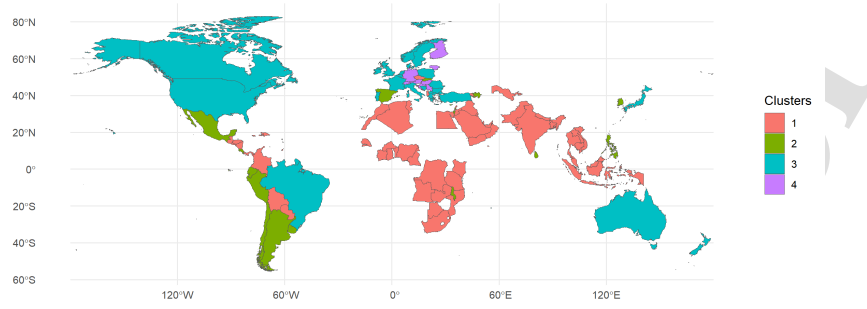
1285

1286

1287

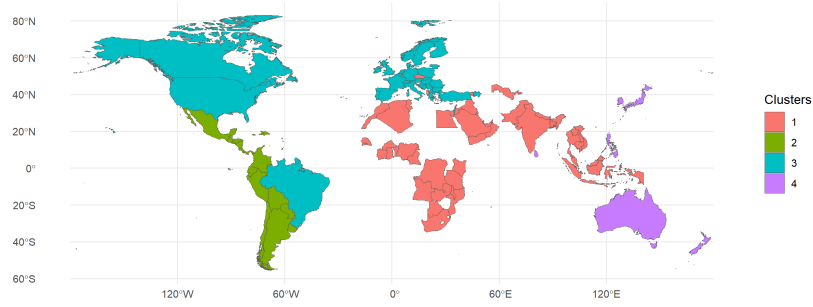
1288

Optimal clustering results:  $\alpha^* = 0$  &  $K^* = 4$



**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.

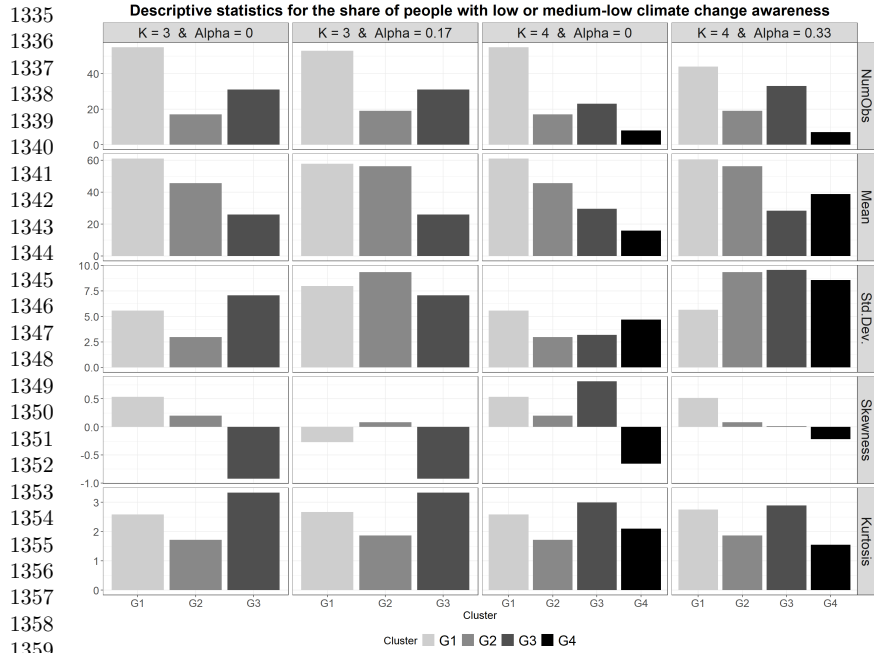
Optimal clustering results:  $\alpha^* = 0.33$  &  $K^* = 4$



**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.

Figure 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., 1289  
the clustering ignoring geographical proximity), the three clusters are clearly ordered 1290  
by their means: G1 (i.e., African and Asian countries) exhibits the highest mean (worst 1291  
awareness) and is also the largest group, G2 (i.e., mixed European and South Ameri- 1292  
can countries) is intermediate, and G3 (i.e., Western countries) shows the lowest mean, 1293  
hence the best awareness, with comparatively higher within-cluster heterogeneity and 1294  
a markedly negative skewness, consistent with mass at relatively high shares and a tail 1295  
toward lower shares. When spatial information is incorporated ( $K = 3$ ,  $\alpha = 0.17$ ), the 1296  
mean separation partly changes: G1 and G2 become closer in average levels (both char- 1297  
acterized by relatively high shares of low-awareness respondents), while G3 remains 1298  
the group with the highest awareness; at the same time, dispersion increases notably in 1299  
G1 and G2, suggesting that imposing spatial coherence can group together geograph- 1300  
ically proximate units that remain more heterogeneous in the awareness indicator. 1301  
Moving to  $K = 4$ , the non-spatial partition (i.e.,  $\alpha = 0$ ) essentially refines the aware- 1302  
ness gradient by carving out an additional, small cluster (G4) with the lowest mean 1303  
(best awareness) relative to the other groups, while G1 remains the worst awareness 1304  
and largest cluster; higher-order moments indicate that the newly formed group dif- 1305  
fers not only in location but also in shape, pointing to a distinct distributional profile 1306  
rather than a mere mean shift. Under the spatial specification ( $\alpha = 0.33$ ), cluster sizes 1307  
and moments shift more visibly: the high-mean cluster remains sizable, but the added 1308  
cluster G4 (i.e., Pacific and Australian countries) no longer sits at the extreme low 1309  
mean and instead appears as a smaller, spatially coherent group with intermediate 1310  
mean and substantial dispersion, indicating that the inclusion of space can reallocate 1311  
extreme-awareness observations across nearby areas and alter within-cluster variabil- 1312  
ity and tail behavior without eliminating the broader pattern of differentiation by 1313  
awareness levels. 1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334



**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  $\alpha$  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.

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 A1 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 variables is described in [Figure 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.

1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472

Linear correlation among awareness, socio-economic and climate-related variables (average 2019-2022)



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

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*)

3. Low and medium-low climate change awareness + climate-related features + geographical distance (later on denoted as *Robustness: Awareness + Climate*) 1473  
1474  
1475
4. Low and medium-low climate change awareness + socio-economic features + climate-related features + geographical distance (later on denoted as *Robustness: Awareness + Climate + Socioeconomic*) 1476  
1477  
1478  
1479  
1480  
1481

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  $\alpha$ . In particular, all alternative clustering exercises are based on the values of  $K$  and  $\alpha$  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<sup>8</sup>. 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*. 1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497

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 [Gordon 1999](#)). 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. 1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511

However, it is important to note that fixing  $(K, \alpha)$  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 1512  
1513  
1514  
1515

---

<sup>8</sup>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 1516  
1517  
1518

1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564

**Adjusted Rand Index for the main clusterings and the robustness clustering**



**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.

(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.

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;
  - 1565
  - 1566
  - 1567
  - 1568
  - 1569
- 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  $ARI = 0.34$ );
  - 1570
  - 1571
  - 1572
  - 1573
  - 1574
  - 1575
  - 1576
  - 1577
- 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;
  - 1578
  - 1579
  - 1580
  - 1581
  - 1582
  - 1583
  - 1584
  - 1585
  - 1586
  - 1587
  - 1588
  - 1589
- 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  $K = \{3, 4\}$ ;
  - 1590
  - 1591
  - 1592
  - 1593
  - 1594
  - 1595
  - 1596
  - 1597
  - 1598
  - 1599
  - 1600
  - 1601
  - 1602
- 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).
  - 1603
  - 1604
  - 1605
  - 1606
  - 1607
  - 1608
  - 1609
  - 1610

1611 When geography is omitted some of the blocks remain (e.g., stability of European  
1612 countries or the North America-Europe-Japan cluster) while blocks such as South  
1613 America and Africa are scattered and more heterogeneous. These considerations  
1614 apply across the various scenarios considered;

1615  
1616  
1617 In general, the above points suggest greater stability of partitions generated by includ-  
1618 ing geographic information than the purely-feature clustering case. More specifically,  
1619 the geographically-informed clustering leads to more geographically-compact aggrega-  
1620 tions (clusters reflect continental or cultural blocks); reduces the potential variations  
1621 generated as the chosen number of clusters varies and thus to more stable and less  
1622 volatile partitions; and generates consistent partitions even when information sets of  
1623 different nature and varying numbers of features are considered.

## 1630 1631 **4 Conclusions**

1632  
1633 In this paper, we investigated global patterns of climate change awareness through  
1634 a geographically-informed hierarchical clustering framework that explicitly accounts  
1635 for spatial constraints. To this end, we combined awareness indicators with a broad  
1636 set of climate-related and socio-economic variables, allowing for the identification of  
1637 cross-country patterns and their spatial heterogeneity. Our approach builds upon the  
1638 *hclustgeo* clustering framework proposed by Chavent et al. (2018), where a mixing  
1639 parameter  $\alpha$  is used to linearly combine the dissimilarity matrix in the feature space  
1640 with geographical distances. In this context, we addressed the open issue of hyperpa-  
1641 rameter tuning by introducing a data-driven procedure for the joint selection of the  
1642 mixing parameter  $\alpha$  and the number of clusters  $K$ . The proposed approach combines  
1643 multiple criteria (including within-cluster homogeneity, between-cluster separation,  
1644 and the comparison between spatially constrained and unconstrained solutions) to  
1645 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 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

1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702

1703 underrepresented population groups; nevertheless, robustness analyses based on alter-  
1704 native weighting schemes and population-adjusted specifications confirm the overall  
1705 stability of the results. Second, the identified spatial partitions are inherently sen-  
1706 sitive to the choice and calibration of clustering hyperparameters (e.g.,  $\alpha$  and  $K$ ).  
1707  
1708 In the robustness analysis, these parameters are intentionally held fixed to ensure  
1709 comparability with the main specification. While this choice may lead to suboptimal  
1710 partitions for alternative feature sets, additional tuning exercises indicate that the  
1711 resulting deviations are moderate and do not materially affect the main conclusions,  
1712 although some degree of model dependence cannot be fully excluded. Third, uncer-  
1713 tainty in country-level estimates limits the strength of the generalizations that can  
1714 be drawn. Fourth, geodetic distances are computed using country centroids, which  
1715 may provide only a coarse proxy of actual human proximity, particularly for large  
1716 or spatially discontinuous countries (e.g., Russia, US, Indonesia), where population  
1717 distributions are highly uneven. This limitation was partially mitigated through the  
1718 integration of socio-economic, climatic, and awareness-related variables within the  
1719 multivariate distance framework, as well as through robustness analyses based on aug-  
1720 mented specifications and population-weighted schemes, which together contribute to  
1721 a more realistic representation of cross-country proximity.  
1722

1723 Despite these limitations, the analysis offers a coherent and globally comparable  
1724 framework for assessing climate change awareness and its spatial heterogeneity. By  
1725 integrating behavioral, socio-economic, and environmental dimensions within a unified  
1726 clustering approach, the study provides a flexible basis for comparative analysis and  
1727 policy-oriented interpretation. Future research can build on this framework to extend  
1728 coverage to underrepresented contexts and to develop predictive tools capable of infer-  
1729 ring awareness levels from observable structural characteristics, thereby supporting  
1730 more targeted and geographically nuanced climate communication strategies.  
1731  
1732

1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748

**Declarations**

- Conflict of interest/Competing interests: The authors have no competing interests to declare that are relevant to the content of this article;
- Funding: The authors did not receive support from any organization for the submitted work;
- Data availability and codes: All results presented in this paper can be reproduced using the R statistical software. The codes were developed entirely by the authors. For the reproducibility purposes, all scripts and the data are made available for the public on the following GitHub folder: [https://github.com/PaoloMaranzano/GPZ\\_PM\\_HierSpatClusterClimateAwareness.git](https://github.com/PaoloMaranzano/GPZ_PM_HierSpatClusterClimateAwareness.git)
- Figures: All images included in the paper were created by the authors and do not require any publication permission.

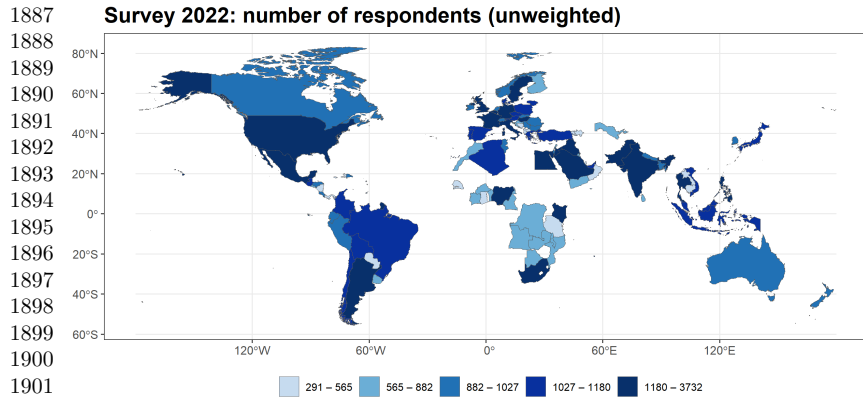
1795 **Appendix A** list of climate-related and  
1796  
1797 socio-economic variables used in the  
1798  
1799 robustness analysis  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840

**Table A1** 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 <sub>2</sub> -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 <sub>2</sub> emissions ( $\mu g$ ) 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.

1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886



**Fig. B1** World map representing the unweighted number of respondents to the 2022 Climate Change Opinion Survey in the 103 countries used in this empirical analysis.

**Appendix B** Descriptive statistics and maps of the unweighted number of respondents to the 2022 Climate Change Opinion Survey at the country level.

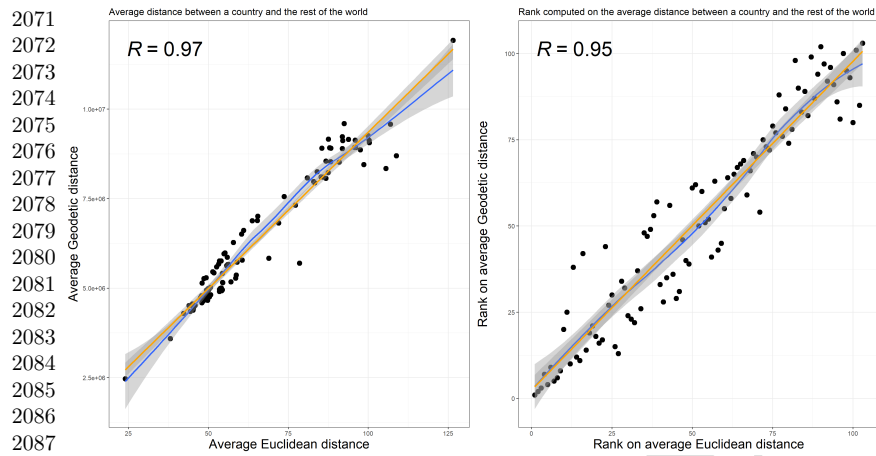
**Table B2** Synthetic statistics on the 2022 Climate Change Opinion Survey in the 103 countries used in the empirical analysis supplemented with information on the national population and the number of Facebook users registered from 2023 to 2025, broken down by gender (source: WorldPopulationReview). (Part 1).

ISO Code	Country Name	Number of respondents to the CCOS 2022 survey	Share (%) of people with med-low or low clim. aware.		Share (%) of people with med-high or high clim. aware.	Pop. (mln)	Facebook users (mln)		% female users		% male users	
			2023	2024			2023	2024	2023	2024	2023	2024
1	AL	329	53	44	2.88	1.87	1.87	1.10	45	55	55	55
2	DZ	1066	61	37	43.05	32.63	32.63	20.80	43	57	57	57
3	AO	727	62	35	31.83	5.82	5.82	3.50	42	58	58	58
4	AR	1257	46	54	44.78	41.70	41.70	27.40	54	46	46	46
5	AM	306	61	37	2.96	2.27	2.27	1.40	54	46	46	46
6	AU	1012	28	72	25.20	22.70	22.70	14.90	52	48	48	48
7	AT	1120	18	81	8.96	5.41	5.41	3.10	52	48	48	48
8	AZ	541	49	49	10.05	5.08	5.08	1.50	45	55	55	55
9	BD	910	61	35	163.05	67.18	67.18	43.30	35	65	65	65
10	BE	1201	28	72	11.54	8.68	8.68	6.30	48	52	52	48
11	BJ	665	71	25	11.80	2.89	2.89	1.10	38	62	62	62
12	BO	1070	55	42	11.51	8.68	8.68	7.10	48	52	52	52
13	BA	291	27	73	3.30	2.02	2.02	1.40	51	49	49	49
14	BW	690	57	39	2.30	1.42	1.42	1.00	50	50	50	50
15	BR	1048	34	65	211.05	175.11	175.11	0.11	55	45	45	45
16	BG	1012	32	67	7.00	4.43	4.43	3.40	53	47	47	47
17	BF	581	64	32	20.32	3.47	3.47	1.90	30	70	70	70
18	KH	445	73	23	16.49	13.86	13.86	10.40	46	54	54	54
19	CM	606	61	32	25.88	6.17	6.17	3.60	42	58	58	58
20	CA	1026	26	74	37.41	32.00	32.00	20.60	53	47	47	47
21	CL	1094	42	57	18.95	18.68	18.68	12.20	52	48	48	48
22	HK	417	58	41	7.44	6.40	6.40	4.20	58	42	42	42
23	CO	1082	62	36	50.34	43.60	43.60	33.50	52	48	48	48
24	CR	1012	43	56	5.05	4.41	4.41	3.10	51	49	49	49
25	HR	635	17	83	4.13	2.57	2.57	1.80	52	48	48	48
26	CY	494	44	56	0.87	1.28	1.28	0.79	49	51	51	51
27	CZ	1049	57	43	10.69	8.76	8.76	5.70	38	62	62	62
28	CI	646	61	31	25.72	8.07	8.07	4.70	37	63	63	63
29	CD	587	68	27	86.79	8.07	8.07	4.70	37	63	63	63
30	DK	1101	25	75	5.77	4.85	4.85	3.50	54	46	46	46
31	DO	884	60	35	10.74	0.05	8.31	5.20	52	48	48	50
32	EC	1963	1966	1951	19.48	19.45	19.45	12.90	19.39	19.37	19.37	1933
33	EG	1963	1966	1952	19.48	19.45	19.45	12.90	19.39	19.37	19.37	1933
34	SV	817	64	33	6.45	4.85	4.85	3.60	53	47	47	47



**Table B4** 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 Code	Country Name	Number of respondents to the CCOS 2022 survey	Share (%) of people with		Share (%) of people with med-high or high clim. aware.	Pop. (mln)	Facebook users (mln)		% female users		% male users	
			med-low or low clim. aware.	med-high or high clim. aware.			2024	2023	2025	2024	2025	2024
71	PK	1181	61	31	216.57	60.42	37.30	24	24	76	76	
72	PA	541	54	44	4.25	3.43	1.80	51	51	49	49	
73	PY	550	62	34	7.04	4.81	3.50	50	50	50	50	
74	PE	952	49	50	32.51	29.05	22.90	48	48	52	52	
75	PH	1249	44	51	108.12	102.32	80.30	53	53	47	47	
76	PL	1069	30	70	37.89	24.75	17.90	55	55	45	45	
77	PT	1029	29	70	10.23	9.09	5.90	53	53	47	47	
78	QA	498	61	33	2.83	3.46	1.90	30	30	70	70	
79	KR	921	44	56	51.23	27.15	9.60	53	53	47	47	
80	RO	907	35	62	19.36	12.34	9.60	52	52	48	48	
81	SA	1252	57	38	34.27	29.18	10.80	34	34	66	66	
82	SN	502	51	42	16.30	4.66	2.60	36	36	64	64	
83	RS	949	21	78	6.99	4.83	2.90	52	52	48	48	
84	SG	369	43	56	5.80	5.32	3.30	51	51	49	49	
85	SK	985	45	55	5.46	3.47	2.50	53	53	47	47	
86	ZA	1369	55	43	58.56	31.97	22.10	51	51	49	49	
87	ES	1163	43	56	46.74	35.90	19.40	53	53	47	47	
88	LK	632	50	49	21.32	9.41	6.50	41	41	59	59	
89	SE	1338	30	69	10.04	8.75	5.80	53	53	47	47	
90	CH	990	21	79	8.59	5.65	3.00	51	51	49	49	
91	TH	1220	58	41	69.63	58.30	48.10	52	52	48	48	
92	TT	518	56	44	1.39	1.14	0.75	53	53	47	47	
93	TN	992	55	41	11.69	8.84	6.50	48	48	52	52	
94	TR	1158	30	70	83.43	70.30	32.80	44	44	56	56	
95	TZ	505	52	39	56.38	8.38	3.80	41	41	59	59	
96	AE	1096	57	41	9.77	13.86	7.30	31	31	69	69	
97	GB	1951	26	74	67.53	55.90	34.40	53	53	47	47	
98	US	3096	29	71	329.06	279.81	175.00	54	54	46	46	
99	UY	580	46	53	3.46	3.27	2.00	55	55	45	45	
100	UZ	710	58	42	32.98	10.10	1.60	37	37	63	63	
101	VN	1096	60	36	96.46	86.11	66.20	51	51	49	49	
102	RU	1511	60	38	147.86	135.45	103.45	51	51	49	49	
103	ZM	692	57	41	17.86	20.37	2.80	43	43	57	57	



2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089 **Fig. C2** Diagrams of the empirical relationship between the average distance of each country from  
2090 the rest of the world. Left panel: scatterplot of the average country-to-World distance computed using  
2091 the Euclidean formula (X axis) and the geodesic formula (Y axis). Right panel: empirical relationship  
2092 between the ranked distances calculated on the average Euclidean and geodesic distances.

2093 **Appendix C Robustness of the spatial distance:**  
2094  
2095 **Euclidean distance versus Geodesic**  
2096 **distance**  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116

<b>References</b>	2117
	2118
Abdelgadir M, AlRashidi M, Alharbi R, et al (2025) Spatial clustering analysis com-	2119
bined with ensemble modeling identified potential coastal conservation hotspots of	2120
white-eyed gulls in the red sea. <i>Global Ecology and Conservation</i> 59:e03581	2121
	2122
	2123
	2124
Akhanli SE, Hennig C (2020) Comparing clusterings and numbers of clusters by	2125
aggregation of calibrated clustering validity indexes. <i>Statistics and Computing</i>	2126
30(5):1523–1544. <a href="https://doi.org/10.1007/s11222-020-09958-2">https://doi.org/10.1007/s11222-020-09958-2</a> , URL <a href="https://doi.org/10.1007/s11222-020-09958-2">https://doi.org/10.1007/s11222-020-09958-2</a>	2127
	2128
	2129
	2130
	2131
	2132
Ambroise C, Dang M, Govaert G (1997) Clustering of spatial data by the em algorithm.	2133
In: <i>geoENV I—Geostatistics for Environmental Applications: Proceedings of the</i>	2134
<i>Geostatistics for Environmental Applications Workshop, Lisbon, Portugal, 18–19</i>	2135
November 1996, Springer, pp 493–504	2136
	2137
	2138
	2139
Andy P (2019) A million threatened species? thirteen questions and answers. <a href="https://www.ipbes.net/news/million-threatened-species-thirteen-questions-answers">https:</a>	2140
<a href="https://www.ipbes.net/news/million-threatened-species-thirteen-questions-answers">//www.ipbes.net/news/million-threatened-species-thirteen-questions-answers</a> ,	2141
Last accessed on August 31 2024	2142
	2143
	2144
	2145
Basu S, Sen S (2023) Covid 19 pandemic, socio-economic behaviour and infection	2146
characteristics: An inter-country predictive study using deep learning. <i>Computa-</i>	2147
<i>tional Economics</i> 61(2):645–676. <a href="https://doi.org/10.1007/s10614-021-10223-5">https://doi.org/10.1007/s10614-021-10223-5</a> , URL	2148
<a href="https://doi.org/10.1007/s10614-021-10223-5">https://doi.org/10.1007/s10614-021-10223-5</a>	2149
	2150
	2151
	2152
	2153
Bourgault G, Marcotte D, Legendre P (1992) The multivariate (co) variogram as a	2154
spatial weighting function in classification methods. <i>Mathematical Geology</i> 24:463–	2155
478	2156
	2157
	2158
	2159
Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. <i>Communications</i>	2160
in <i>Statistics-theory and Methods</i> 3(1):1–27	2161
	2162

- 2163 Canadell JG, Meyer C, Cook GD, et al (2021) Multi-decadal increase of forest burned  
2164 area in australia is linked to climate change. *Nature communications* 12(1):6921  
2165  
2166
- 2167 Charrad M, Ghazzali N, Boiteau V, et al (2014) Nbclust: An r package for determining  
2168 the relevant number of clusters in a data set. *Journal of Statistical Software* 61(6):1  
2169 – 36. <https://doi.org/10.18637/jss.v061.i06>, URL [https://www.jstatsoft.org/index.  
2170 php/jss/article/view/v061i06](https://www.jstatsoft.org/index.php/jss/article/view/v061i06)  
2171  
2172  
2173
- 2174 Chavent M, Kuentz-Simonet V, Labenne A, et al (2018) Clustgeo: an r pack-  
2175 age for hierarchical clustering with spatial constraints. *Computational Statistics*  
2176 33(4):1799–1822  
2177  
2178  
2179
- 2180 Czaika M (2020) On the negative impact of time zone differences on tourist arrivals.  
2181 Current Issues in Tourism Employs Euclidean (“as the crow flies”) distance between  
2182 countries as spatial control variable.  
2183  
2184  
2185
- 2186 Dunn JC (1974) Well-separated clusters and optimal fuzzy partitions. *Journal of*  
2187 *cybernetics* 4(1):95–104  
2188  
2189
- 2190 Eckstein D, Künzel V, Laura S (2021) Global climate risk index. [https://reliefweb.  
2191 int/report/world/global-climate-risk-index-2021](https://reliefweb.int/report/world/global-climate-risk-index-2021), Last accessed on August 31 2024  
2192  
2193
- 2194 Edmonds JA, Sands RD (2003) What are the costs of limiting co2 concentrations.  
2195 Global climate change: The science, economics, and politics pp 140–86  
2196  
2197
- 2198 Fenn-Moltu G, Ollier S, Bates OK, et al (2023) Global flows of insect transport and  
2199 establishment: The role of biogeography, trade and regulations. *Diversity and Distri-  
2200 butions* 29(11):1478–1491. <https://doi.org/https://doi.org/10.1111/ddi.13772>, URL  
2201 <https://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.13772>  
2202  
2203  
2204  
2205  
2206  
2207  
2208

Gifford R (2011) The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. <i>American psychologist</i> 66(4):290	2209 2210 2211 2212
Gordon A (1996) A survey of constrained classification. <i>Computational Statistics &amp; Data Analysis</i> 21(1):17–29	2213 2214 2215 2216
Gordon AD (1999) <i>Classification</i> . CRC Press	2217 2218 2219
Goss M, Swain DL, Abatzoglou JT, et al (2020) Climate change is increasing the likelihood of extreme autumn wildfire conditions across california. <i>Environmental Research Letters</i> 15(9):094016	2220 2221 2222 2223 2224 2225
Hincks S, Carter J, Connelly A (2023) A new typology of climate change risk for european cities and regions: Principles and applications. <i>Global Environmental Change</i> 83:102767	2226 2227 2228 2229 2230 2231
Howe PD, Markowitz EM, Lee TM, et al (2013) Global perceptions of local temperature change. <i>Nature climate change</i> 3(4):352–356	2232 2233 2234 2235
Hubert LJ, Levin JR (1976) A general statistical framework for assessing categorical clustering in free recall. <i>Psychological bulletin</i> 83(6):1072	2236 2237 2238 2239
Jones SG, Ashby AJ, Momin SR, et al (2010) Spatial implications associated with using euclidean distance measurements and geographic centroid imputation in health care research. <i>Health services research</i> 45(1):316–327	2240 2241 2242 2243 2244
Kim SY, Wolinsky-Nahmias Y (2014) Cross-national public opinion on climate change: The effects of affluence and vulnerability. <i>Global Environmental Politics</i> 14(1):79–106	2245 2246 2247 2248 2249 2250 2251 2252 2253 2254

- 2255 Klein A (2020) The determinants of international migration in early modern europe:  
2256  
2257 Evidence from the maritime sector, c. 1700–1800. *Social Science History* 44(1):143–  
2258 167. Uses straight-line Euclidean distance between country centroids (“as the crow  
2259 flies”) to model bilateral migration distances.  
2260  
2261  
2262 Knight KW (2016) Public awareness and perception of climate change: A quantitative  
2263 cross-national study. *Environmental Sociology* 2(1):101–113  
2264  
2265  
2266 Kvaløy B, Finseraas H, Listhaug O (2012) The publics’ concern for global warming:  
2267 A cross-national study of 47 countries. *Journal of Peace Research* 49(1):11–22  
2268  
2269  
2270 Lee TM, Markowitz EM, Howe PD, et al (2015) Predictors of public climate change  
2271 awareness and risk perception around the world. *Nature climate change* 5(11):1014–  
2272 1020  
2273  
2274  
2275  
2276 Leiserowitz A, Carman J, Buttermore N, et al (2022) International public opin-  
2277 ion on climate change 2022. New Haven, CT: Yale Program on Climate Change  
2278 Communication and Facebook Data for Good  
2279  
2280  
2281  
2282 Leiserowitz AA, Maibach E, Roser-Renouf C, et al (2018) Climate change in the  
2283 American mind. University of Washington Seattle, WA, USA  
2284  
2285  
2286 Li Y, Johnson EJ, Zaval L (2011) Local warming: Daily temperature change influences  
2287 belief in global warming. *Psychological science* 22(4):454–459  
2288  
2289  
2290 Liao ZX, Peng WC (2012) Clustering spatial data with a geographic constraint:  
2291 exploring local search. *Knowledge and information systems* 31:153–170  
2292  
2293  
2294 Loarie SR, Duffy PB, Hamilton H, et al (2009) The velocity of climate change. *Nature*  
2295 462(7276):1052–1055  
2296  
2297  
2298  
2299  
2300

- Mattera R (2022) A weighted approach for spatio-temporal clustering of covid-19 spread in italy. *Spatial and Spatio-temporal Epidemiology* 41:100500. <https://doi.org/https://doi.org/10.1016/j.sste.2022.100500>, URL <https://www.sciencedirect.com/science/article/pii/S1877584522000247>
- Mattera R, Franses PH (2023) Are african business cycles synchronized? evidence from spatio-temporal modeling. *Economic Modelling* 128:106485. <https://doi.org/https://doi.org/10.1016/j.econmod.2023.106485>, URL <https://www.sciencedirect.com/science/article/pii/S0264999323002973>
- McClain JO, Rao VR (1975) Clustisz: A program to test for the quality of clustering of a set of objects. *Journal of Marketing Research* pp 456–460
- Meza I, Rezaei EE, Siebert S, et al (2021) Drought risk for agricultural systems in south africa: Drivers, spatial patterns, and implications for drought risk management. *Science of the Total Environment* 799:149505
- Miele V, Picard F, Dray S (2014) Spatially constrained clustering of ecological networks. *Methods in Ecology and Evolution* 5(8):771–779
- Morelli C, Boccaletti S, Maranzano P, et al (2025) Multidimensional spatiotemporal clustering – an application to environmental sustainability scores in europe. *Environmetrics* 36(2):e2893. <https://doi.org/https://doi.org/10.1002/env.2893>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2893>
- Myers TA, Maibach EW, Roser-Renouf C, et al (2013) The relationship between personal experience and belief in the reality of global warming. *Nature climate change* 3(4):343–347
- Nowotny K (2019) Network migration: Do neighbouring regions matter? *Regional Studies* Distance between regions measured in km “as the crow flies”.

- 2347 Oliver M, Webster R (1989) A geostatistical basis for spatial weighting in multivariate  
2348  
2349 classification. *Mathematical geology* 21:15–35  
2350
- 2351 Pawitan Y, Huang J (2003) Constrained clustering of irregularly sampled spatial data.  
2352  
2353 *Journal of Statistical Computation and Simulation* 73(12):853–865  
2354
- 2355 Poortinga W, Whitmarsh L, Steg L, et al (2019) Climate change perceptions and their  
2356  
2357 individual-level determinants: A cross-european analysis. *Global environmental*  
2358  
2359 *change* 55:25–35  
2360
- 2361 Román-Palacios C, Wiens JJ (2020) Recent responses to climate change reveal the  
2362  
2363 drivers of species extinction and survival. *Proceedings of the National Academy of*  
2364  
2365 *Sciences* 117(8):4211–4217  
2366
- 2367 Rossbroich J, Durieux J, Wilderjans TF (2022) Model selection strategies for  
2368  
2369 determining the optimal number of overlapping clusters in additive overlapping par-  
2370  
2371 titional clustering. *Journal of Classification* 39(2):264–301. [https://doi.org/10.1007/](https://doi.org/10.1007/s00357-021-09409-1)  
2372 [s00357-021-09409-1](https://doi.org/10.1007/s00357-021-09409-1), URL <https://doi.org/10.1007/s00357-021-09409-1>  
2373
- 2374 Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation  
2375  
2376 of cluster analysis. *Journal of computational and applied mathematics* 20:53–65  
2377
- 2378 Sun Y, Zhang X, Ding Y, et al (2022) Understanding human influence on climate  
2379  
2380 change in china. *National science review* 9(3):nwab113  
2381
- 2382 Titus JG (1986) Greenhouse effect, sea level rise, and coastal zone management.  
2383  
2384 *Coastal Management* 14(3):147–171  
2385
- 2386 Wang F, Harindintwali JD, Wei K, et al (2023) Climate change: Strategies for  
2387  
2388 mitigation and adaptation. *The Innovation Geoscience* 1(1):100015–61  
2389  
2390  
2391  
2392

Weber EU (2006) Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). Climatic change 77(1):103–120	2393 2394 2395 2396
World Population Review (2026) Facebook users by country 2026, retrieved January 15, 2026, from. URL <a href="https://worldpopulationreview.com/country-rankings/facebook-users-by-country">https://worldpopulationreview.com/country-rankings/facebook-users-by-country</a>	2397 2398 2399 2400 2401 2402
Yang S, She D, Wang L, et al (2025) Tree size and spatial clustering drive contrasting carbon sequestration in soil aggregates in temperate forests in china. Urban Forestry & Urban Greening p 128998	2403 2404 2405 2406 2407 2408
Zahran S, Kim E, Chen X, et al (2007) Ecological development and global climate change: A cross-national study of kyoto protocol ratification. Society and Natural Resources 20(1):37–55	2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438

### **Highlights**

- A geographically-informed hierarchical clustering approach is used to analyze cross-country climate change awareness.
- Spatial information is integrated with awareness, socio-economic, and climate-related indicators through a Ward-type algorithm with spatial constraints.
- A novel procedure is proposed to tune clustering hyperparameters and compare spatial and non-spatial partitions.
- Incorporating geography improves cluster stability and yields more interpretable, spatially compact groupings.
- Western countries exhibit high and homogeneous climate change awareness, while Asia, Africa, and the Middle East show lower and more heterogeneous awareness levels.

**Paolo Maranzano**

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.

**Gianpaolo Zammarchi**

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

Dear *Socio-Economic Planning Sciences* Editor,

Dear Guest Editors of the Special Issue “Statistical Measurement of Economic Analysis: Methods and Approaches for Sustainable, Innovative, and Digital Planning”

Prof.ssa Tiziana Laureti, prof.ssa Giorgia Riveccio, prof. Gennaro Punzo e prof. Luca Secondi

It is our pleasure to submit the revised version of the paper

G. Zammarchi & P. Maranzano

*Mapping climate change awareness through spatial hierarchical clustering*

Declaration of competing interests

At the end of the paper, right before the references, you will find a statement in which the authors declare that they have no known competing interests or personal relationships that could have appeared to influence the work reported in this paper.

Please receive our best regards,

Gianpaolo Zammarchi & Paolo Maranzano

Milano, Italy, January 28th 2026