Final Manuscript

A new method to identify robust climate analogues

Authors: Carsten Walther[1,], Matthias Lüdeke[1], Ramana Gudipudi[1]*

[1] *Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412 Potsdam, Germany*
*Corresponding author. Tel.: +49 331 288 20710. E-mail address: carsten.walther@pik-potsdam.de

Climate Research 78: 179-187, 2019
published online August 22

Abstract

Climate analogues are a comprehensive approach to learn how to deal with the expected climatic future from current examples. Existing literature on climate analogues struggles with two methodological challenges: how to deal with the unavoidable uncertainty of climate projections and how to define reasonable lower limits of similarity for climate analogues. Here we suggest a new method to identify robust climate analogues (RCAs) which is based on a clustering approach in climate space where each spatial grid element is represented by three points: its current climate, a lower and an upper bound of the climate projections. If the upper and lower bound of the projections for such a grid element share the same cluster and, additionally, this cluster contains also current climate points, then the grid elements related to the latter are defined as RCAs. This definition divides the map of the investigated region into areas with RCAs and uncharted areas where under the current uncertainty range of climate projections such an attribution is not justified. An exemplary application of the algorithm for Europe shows that RCAs can be identified for 37% of the land area and that the new method allows selection of socio-economically reasonable RCAs from climatologically equivalent (given the current uncertainty) grid elements.

Keywords: climate analogue, clustering algorithm, adaptation, uncertainty, regional climate model

1 INTRODUCTION

The development of appropriate strategies and techniques for adaptation to projected climate change is an important (e.g. Pielke et al. 2007; Lobell et al. 2008; Reckien et al. 2015) but challenging task. It comprises comprehensive climate impact assessments, the development of catalogues of related adaptation options and the choice of the most efficient options from these catalogues. Despite large and successful efforts towards this objective (e.g. Nassopoulos et al. 2012; Rosenzweig et al. 2017; Krysanova & Hattermann 2017), there is no guarantee that all relevant climate-sensitive sectors and all possible adaptation options have been considered so far. This motivates the development of complementary and more holistic perspectives. A promising approach in this context uses the spatial variability of the current climate to study how socio-ecological systems (SESs, e.g. urban agglomerations or rural areas) are affected by and deal with different climatic situations (e.g. arid or tropical climates). Given the projected climate for a specific location A and following the space-for-time idea (e.g. Pickett 1989; Rastetter 1996), one may ask where the current climate equals this projection. Such locations are then called climate analogues (CAs) of location A. In the best case such CAs exist and at least one of them also resembles the expected (future) socio-ecological situation of location A – then successful strategies and techniques currently applied in the analogue area can be considered as useful adaptation options for location A. This approach has been followed in several studies in recent decades (e.g. Parry & Carter 1989; Dehn 1999; Dawson et al. 2009; Pugh et al. 2016) and requires an adequate choice of climate variables (depending on the considered SES), the definition of a distance measure in the climate space spanned by these variables and – based on this – an explicit definition of climate analogues.

Here we present a new method for the identification of CAs which addresses two major, interlinked methodological problems: how to deal with the unavoidable uncertainty in climate projections and how to determine the degree of similarity justifying the notion of climate analogues. In earlier work the dependence of the location of CAs on the chosen climate projection was recognized (e.g. Kopf et al. 2008) and partly used to discuss consequences of projection uncertainty (Hallegatte et al. 2007). However, these studies did not address the existence and location of robust climate analogues (RCAs) which are valid despite this uncertainty. More recent literature attempted to address this methodological issue either qualitatively (Arnbjerg-Nielsen et al. 2015) or quantitatively (Hibino et al. 2015). Arnbjerg-Nielsen et al. (2015) projected future rainfall for Denmark by taking measured historical intensity-duration curves (IDFs) from future Denmark's CAs. They used only IDFs from regions which appeared to be good analogues (measured by their ranking) under different climate projections. A more explicit approach to include projection uncertainty was recently presented by Hibino et al. (2015), which included the local uncertainty of climate projections in the parametrization of their "similarity score". This measure compares the standard deviation of different projections for grid element A with the (Euclidian) distance in climate space between this grid element and the potential CA for all considered climate variables. Each climate variable contributes to the similarity score only if the distance is smaller than the standard deviation, i.e. it is possible that the projection of grid element A is equal to the present climate of grid element B. This is a promising perspective on projection uncertainty and therefore we will follow the basic idea behind this approach in our CA definition.

Regarding the second methodological problem of necessary minimum similarity for CAs, to our knowledge, the existing literature identifies a CA by looking for a grid cell which is most similar in climate space and is thus essentially a relative measure. Formally, these approaches result in maps which show the degree of similarity of all grid elements relative to a fixed grid element A for which the CA is sought. The peak of the similarity landscape (or the minimum in the dissimilarity landscape) denotes the location of the CA. Sometimes a more or less subjective threshold is introduced to define a minimum similarity in climate space for a climate analogue, which potentially results in locations without a CA (e.g. Hallegatte et al. 2007). We consider that looking at single maps of the aforementioned kind is only a first step and we advocate analyzing the whole climate space, including the present climate and the different climate projections to systematically identify reasonable requirements for CAs.

To solve the two above-mentioned problems we suggest characterizing projection uncertainty for each grid element by a lower and upper bound for each climate variable. Each grid element thus generates three points in climate space: the present, the lower and the upper bound of projected climates. Then we perform an appropriate cluster analysis in climate space and define grid element B as a RCA of grid element A if and only if the present climate of B shares the same cluster with the lower and upper bound of climate projections for A. Thereby, the deviation in the projections is typically about the same magnitude as their distances to the current climate point, which takes up the basic idea on projection uncertainty consideration by Hibino et al. (2015). Furthermore, the subjective distance threshold definition is now substituted by the cluster partition which is determined by the data structure. Usually data structure excludes several cluster numbers, but some choice remains regarding this parameter. This choice is guided by the tradeoff between the need for sufficient differentiation, necessary for interpretation (pro large cluster number), and the extent of areas without RCAs (pro small cluster number).For a more comprehensive explanation and the interpretation of non-RCA point constellations please refer to the methods section. The proposed cluster method has some similarities with classical climate classifications (e.g. Geiger 1961). However, the major difference is that it generates a classification of the combined current *and* projected climate space - therefore a direct comparison of the resulting maps is not possible.

In this paper we will illustrate the suggested method by analyzing climate analogues in Europe with regard to monthly mean values of precipitation and temperature. These climate variables and the study region were mainly chosen to allow for a comparison of our new method's results with existing CA studies. Regarding lower and upper bound of the climate projections, we used the model ensemble results of EURO-CORDEX for the period 2071-2100 under RCP8.5 (e.g. Jacob et al. 2014).For the present climate we took the European Climate Assessment Data (ECAD) for the period 1971-2000 (for more details please refer to the methods section).

In the next section we introduce the method in more detail and describe the data used for the illustrative example. In the following sections we present the robust climate analogues (RCAs) for Europe, discuss their peculiarities and compare them with the results of Hallegatte et al. (2007) which use the same climate variables.

2 METHODS

2.1 Definition of climate analogues

Our approach considers a gridded area described by *N* grid elements, *i* = 1, …, *N*. We assume climate is characterized by *V* appropriate normalized climate variables (e.g. monthly precipitation, total annual precipitation, monthly temperature) *j* = 1, …, *V*. Each grid element *i* is described by the climate variables for its current climate (*C*) as well as for a lower (*L*) and upper bound of the climate projections (*U*) spanning the projection uncertainty. So each grid element *i* constitutes three data points in the *V*-dimensional climate space: $C_i = \{x^c_{i,1}, \dots, x^c_{i,j}, \dots, x^c_{i,V}\}$, $L_i = \{x^l_{i,1}, \dots, x^l_{i,j}, \dots, x^l_{i,V}\}$ and $U_i = \{x^u_{i,1}, \dots, x^u_{i,j}, \dots, x^u_{i,V}\}$, where *x* denotes the value of the climate variable j in grid element i of the current, lower and upper bound of the projected climates, $d \in \{c, l, u\}$. We then perform a cluster analysis on these 3**N* objects in climate space. For further details on the choice of clustering method used in this analysis refer to appendix (App. 1) as well as Janssen et al. (2012). To illustrate the interpretation of the resulting clusters in terms of robust analogues we show in Fig. 1 a stylized example for only five grid elements and two climate variables. Cluster I includes the lower and upper bound of the projections for grid element 1, as well as the current climate of grid elements 2 and 3. Therefore, in this example, grid elements 2 and 3 are RCAs to grid element 1. As the current climate of grid elements 1, 4 and 5 belongs to other clusters these are not CAs for grid element 1. Since they do not share the lower and upper bound of the projections of a grid element within their respective clusters, they are not robust CAs for any grid element.

The blue lines in Fig. 1 illustrate how our approach relates to Hibino et al. (2015). In cluster I, the difference between L_1 and U_1 is in the same order of magnitude as the distance between the average climate projection and the current climates of grid elements 2 & 3. In contrast, C_4 and C_5 show a significantly larger distance than the uncertainty in climate projection for grid element 1 as they belong to other clusters. This takes up the basic idea of the recent suggestion by Hibino et al. (2015) to include projection uncertainty into CA definition where only distances that are smaller than or equal to the variance of projections contribute to a "similarity index". Another aspect of our method relates to the question of what degree of similarity is necessary to justify the notion of a climate analogue. Here cluster membership is the decisive property, i.e. instead of somewhat subjective thresholds (e.g. Hallegatte et al. 2007) the global data structure in climate space plays a large role in determining which climates can be defined as similar. This notion also results naturally in the identification of climates which disappear (cluster III in Figure 1) and grid elements without (robust) climate analogues "novel climates" after Williams et al. 2007, represented by cluster IV in Figure 1).

Finally, we want to discuss a specific variant of RCAs as defined above. Assume cluster I would also include C_1, i.e. the projection uncertainty for grid element 1 does not allow one to distinguish between current and future climate. This formally fulfills our definition of a RCA (grid element 1 would be its own RCA) but is not interpretable regarding the transfer of adaptation options and we therefore exclude these grid elements from our final RCA results and show them separately (see App. 5).

2.2 Determining lower and upper climate projections

Climate projections entail different sources of uncertainty: forcing, model response, internal variability and the downscaling process (Tebaldi and Knutti 2007; Kopf et al. 2008; Hawkins and Sutton 2011; Taylor et al. 2012; Knutti and Sedláček 2012). We address this by analyzing the outputs of a climate model ensemble (e.g. from EURO-CORDEX) for a given forcing: for each grid element and climate variable we consider the distribution of the climate variable over all climate models and calculate the value for the lower and upper bound of the projections according to fixed percentiles, in our case the 15%- and 85%-percentile, to get rid of outliers. Then, the 15% (85%) percentiles of the model distributions of all climate variables constitute the lower (upper) bound of the climate projections for a grid element. This characterization of the uncertainty range implies a simplification. The hyper cuboid in climate space, spanned by all combinations of upper and lower values of the climate variables for one grid element (2^V vertices), is approximated by one of its main diagonals. As the applied cluster algorithm tends to identify spherical clusters (see App. 1) this is an acceptable approximation which should not affect the identification of robust analogues significantly. Furthermore it should be noted that these upper and lower vertices describe the uncertainty ranges spanned by all available models and are not necessarily physically consistent projections themselves.

While developing the presented algorithm we also considered the possibility to first apply the clustering to each single model projection and to identify afterwards communalities in the large number of resulting maps. One major problem of this procedure is that each single result will have a different optimum cluster number which makes the comparison almost impossible. This led us to the decision to first integrate the projections by using the upper/lower bound or envelope concept described above.

2.3 Applied data

Many existing climate analogue studies are based on monthly mean temperature, monthly mean and sometimes yearly total precipitation (e.g. Hallegatte et al. 2007). As this study focusses on a new method, we limit our analysis to these climate variables to allow for comparison of the results. We selected Europe, a region with a broad range of climatic, cultural and socio-economic regions, as an appropriate study area. Since a large fraction (75%) of the European population resides in urban areas, mainly cities were used to present and compare our results.

The observed climate of the recent past (current climate) is collected from the gridded observational E-OBS data provided by the European Climate Assessment & Data (ECA&D) (Haylock et al. 2008). This data set covers

the whole of Europe with a comparably high spatial resolution of 0.22 degrees and daily temporal resolution. For the projections of climate, ten regional climate model (RCM) outputs taken from the EURO-CORDEX project (Jacob et al. 2014) have been considered in this study. These regional climate projections with a spatial resolution of 0.11 degrees represent downscaled projections from CMIP5 GCMs (IPCC 2013). The RCM data is available only in a rotated grid and was remapped and aggregated to the regular E-OBS grid. The extent of the study area is -13° west to 41° east and 34° to 72° north – this correlates to 32832 grid cells out of which 16844 are land cells.

The E-OBS and the EURO-CORDEX RCM data are the basis for the calculation of the twelve monthly mean temperatures, twelve monthly precipitation sums and the yearly total precipitation under climate change. To account for potential biases of the models the difference of the temperature variables gained from the climate model outputs of the future period (2071-2100) and the historical time period (1971-2000) were added to the observed baseline data of E-OBS for the period 1971-2000 (Delta method; e.g. Mote & Salathé 2010). For precipitation related variables the ratio and multiplication was used instead. To make the different temperature and precipitation variables comparable with each other we mapped them linearly onto the interval [0, 1] (unity based normalization). This allows for the calculation of the Euclidian distance in climate space which is necessary for the cluster algorithm. This choice of the normalization method uses the observed range of a climate variable within the study area as unit. It is applicable as there are no extreme outliers.

3 RESULTS

3.1 Robust Climate Analogue Cluster

The choice of an adequate number of clusters strongly influences the subsequent RCAs. After determining the set of numbers of clusters which are adequate to the data structure (for details see App. 2 and Sietz et al. 2011) we selected ten clusters for further analysis. Once the algorithm delivered the allocation of the data points to the ten clusters the data set was redistributed to three maps of the cluster membership in position space for the current (Fig. 2, right), the upper and the lower bound of the projections. Then we identified the intersection of the two future maps which denotes regions where all EURO-CORDEX projections belong to the same cluster (Fig. 2, left). In our stylized illustration in climate space (Fig. 1) this intersection corresponds to (L_1, U_1) in cluster I and (L_4, U_4,) in cluster IV. As the current climates of grid elements 2 and 3 (C_2 and C_3) are also members of cluster I they are robust climate analogues for grid element 1. In terms of the maps in Fig. 2 this means that for a region A taken from the robust climate projections map (Fig. 2, left), the corresponding cluster (color) of the current climate map (Fig. 2, right) represents the climate analogue region for the future of A.

Taking the example of Berlin, we find that it is allocated to the violet cluster (5) in the projection map (Fig. 2, left). The points in climate space that belong to cluster 5 and describe the current climate (see Fig. 2, right) are located in the region north of Madrid (Spain), in the river Po valley (Italy) and many grid cells in eastern Europe (e.g. Hungary, North Serbia, Macedonia, large parts of the coast of the Black Sea and the higher altitudes of Turkey). With regard to the changes in climate, these grid cells can be seen as the RCAs to Berlin's climatological future. Depending on probable future socioeconomic or environmental features of Berlin, one can now select the most appropriate city within the RCA grid cells. In general this would rather be cities like Bologna or Toulouse than, e.g., Varna in Bulgaria. But the detailed criteria for this choice depend on the focus of the envisaged interpretation. This topic is of relevance in the context the further interpretation of climate analogues but beyond the scope of this paper which focusses only on the climatic aspects.

A look at the other climate analogue clusters discloses more interesting examples of robust climate analogues. We select one city or region from each of the clusters exemplarily and present respective RCAs (see Table 1). For the climatic properties of the clusters please refer to the feature plots (Kok et al. 2016) in App. 3.

Helsinki (Finland) finds its future climate in, e.g., Copenhagen (Denmark) which is located in the red cluster (1) for the current climate. Copenhagen's future climate is represented by the cyan cluster (6), covering France, northern Spain and parts of the Adriatic Sea region in the current climate. Appropriate options within the RCAs

of Copenhagen might be Nantes (France) or Bilbao (Spain). The future climate of Paris is situated in the orange cluster (3), a climate pattern which is the second most warm and second driest in summer among the climate change patterns. The currently comparable regions defined by this cluster are the southern parts of the Mediterranean. Plausible RCAs could therefore be Madrid (Spain) or Rome (Italy).

In the projections for the southern parts of the Mediterranean, we find the climate of the yellow cluster (8), with the driest summers and all-year hottest temperatures including the city of Rome (Italy). In the current climate this cluster comprises regions in Algeria, Tunisia, northern Syria and the region north-east of Huelva in Spain. So Tunis might be the most appropriate RCA for Rome. The turquoise cluster (7) is situated in the far north (see Fig. 2, right hand side) and shows a distinct colder current climate than all other clusters. Due to increasing temperatures, climate patterns typically are moving further north or to higher altitude, which leaves only little space for this climate pattern. Accordingly, it is identified as a "disappearing climate" within Europe. No grid elements that are robust regarding projection uncertainty and that switched cluster membership in the course of climate change remain by the end of the century. As no turquoise pixels occur in Fig. 2, left hand side, there is no location for which the turquoise pixels on the right hand side serve as RCA. However, in Fig. A5 occur some "no change" pixels in the North of Sweden around Abisko, depicting that these pixels stay in the same cluster during climate change – in so far, in climatological terms, it is not exactly a disappearing climate.

The ocean blue cluster (2) is the second coldest region according to summer temperatures but not yet under so much risk of being categorized as a disappearing climate. Large regions of northern Scandinavia will exhibit the former climate of the northeastern Baltic Sea countries. The Russian city of Murmansk can find its RCA in, e.g., the capital of Latvia, Riga, which is also a port city with a Soviet history and similar population size.

The blue (9) and the magenta (10) cluster are strongly influenced by the sea and represent the wet climate of the western shorelines. The blue cluster shows robust projections in the mountainous regions of Norway, starting around Hardangervidda and extending northwards. Villages in this sparsely populated region find their current analogues either closer to or at the Norwegian shoreline (e.g. Kristansand) or, e.g., in the Scottish Highlands. Pixels with robust projections for the magenta cluster which change cluster membership under climate change include only a narrow strip west of the blue cluster in Norway, also sparsely populated. Villages in this region find their RCAs at the Norwegian, Scottish and Irish West Cost.

An example for the green cluster (4) is the city of Lviv in Ukraine. RCAs for that town can be found in the lower Alps and northern parts of former Yugoslavia, e.g. Zagreb in Croatia.

Having in mind the above RCAs one may ask if the degree of similarity of the mentioned city pairs is comparable. As the size of a cluster in climate space is a measure for this similarity, we show the cluster variance in Table 1 which is measured by the sum of the squared distances between the objects (grid cells) and their corresponding cluster centers. Summarizing these results, we obtain the "sharpest" RCAs for Murmansk, Helsinki, Lviv, and Berlin. Intermediate similarity is observed for Copenhagen, Rome and Paris while the identification of RCAs using cluster 9 and 10 have to be taken with care. On the other hand, no better robust analogues can be found here, given the current climate model uncertainty.

For some cities it was possible to do a comparison between RCAs and the study results of Hallegatte et al. (2007) which used two different climate projections for Europe, namely the HadRM3H model (stronger climate response) and the ARPEGE model (weaker). With regard to the climate indices considered, on average the values of these models lie within the range of the EURO-CORDEX models used in our study. Therefore, with respect to the applied data, our study is comparable to that of Hallegatte et al. (2007). Further, Hallegatte et al. (2007) used a minimum distance approach for each of the models including externally determined and variable specific thresholds. For Berlin, this study presented two options for a climate analogue – the region of Rome (Italy) for a weaker climate response (ARPEGE) and the region of Salamanca (Spain) for a stronger response (HadRM3H). In our study, the violet cluster includes Salamanca, but as regards Rome only the mountainous regions nearby are part of the climate cluster. For Copenhagen, the climate analogues are identical with the region of Paris and Albania. For Helsinki, Hallegatte et al. (2007) could not find an analogue fulfilling both the temperature and precipitation distance threshold. The regions of Bordeaux (France) and Cordoba (Spain) are climate analogues for Paris according to Hallegatte et al. (2007). This corresponds to our results for Cordoba but not for Bordeaux. For Rome Hallegatte et al. (2007) did not find an analogue for both models while in our

case Tunis was identified. The results of this comparison are summarized in App. 4 and will be discussed in the next section.

4 DISCUSSION AND CONCLUSION

The exemplary comparison between analogues found by Hallegatte et al. (2007) and by the method proposed in this paper illustrates relevant differences and commonalities. For Helsinki, the presented method identifies Copenhagen as a robust climate analogue while Hallegatte et al. (2007) did not find a CA because the least distances in climate space transgress the rather subjectively determined thresholds for both climate projections. In our method the cluster result based on data structure and reasonably selected cluster number relates Helsinki to cluster 2 (the red cluster in Fig. 2) which has a relatively small extent in climate space (second smallest variance, see Table 1). This example shows the role of the difference in the determination of minimum distances. Contrasting to this, the determination of RCAs of Copenhagen shows agreement for both methods, i.e. the CAs for both climate projections lie within the RCA region identified by our algorithm. But in the method of Hallegatte et al. (2007) the climatic equivalence under uncertainty of the two cities is not recognized and one has to follow both cases for all further interpretations and conclusions. Furthermore, under the existing climate projection, we can choose, e.g., Bilbao or Nantes which are certainly closer to the probable socioeconomic future of Copenhagen. The case of Rome shows that due to new climatic conditions the pointwise (e.g. Hallegatte et al. 2007) method failed to produce an analogue. This might be related to the spatial boundaries of the study area: a spatial extension further south could open up analogues also for regions like Rome in case of the existence of an adequate temperature-precipitation relation. Our algorithm found an RCA within the spatial boundaries (Tunis) but one has to consider that this was within the third-largest cluster in climate space (see Table 1) and therefore less significant than e.g. Murmansk-Riga case (half the cluster size). In the case of Berlin and Paris the comparison with the weaker climate signal analogue shows disagreement which means that the respective analogues do not lie within the RCAs. The reason must be that for these two cities the weaker climate response model used by Hallegatte et al. is outside the (L, U) range defined by the EURO-CORDEX ensemble (see Fig. 1). Therefore this disagreement is caused rather by differences in the input data than the applied method.

In earlier approaches the degree of similarity is determined by the global minimum in distance space together with rather subjective thresholds for maximum distances which are still acceptable (Hallegatte et al. 2007). These procedures suffer from the difficult justification of these thresholds and the point character of the analogue which usually do not lead to an urban region with adequate socioeconomic properties. The approach presented here refers to the data structure of the combined climate data set. However, the choice of the adequate number of clusters is often not unique (2, 3, 5 and 10 in our example) but influences the subsequent RCAs regarding the variance within the clusters and, related to that, the size of the regions without a robust climate analogue ('uncharted regions' in Figure 3, left). Choosing a smaller number of clusters (e.g. 5) will lead to smaller 'uncharted areas' but larger variance within the climate analogue clusters. The first means that robust statements can be made for a larger fraction of European cities while the latter means that these statements are weaker. Since climate analogues can only generate a useful guide to adaptation options with an adequate level of differentiation, we chose ten clusters in the analysis presented here. Besides this, the detection of disappearing climates, due to northward-moving climate patterns (cluster 7) is in line with earlier studies (e.g. Williams et al. 2007).

Climate projections will always be loaded with uncertainty to a certain extent (Meehl et al. 2007; Hawkins & Sutton 2011). But due to ongoing changes in climate, adaptation efforts are necessary even under this unfavorable data situation. It is therefore important to find and interpret climate analogues while simultaneously addressing the uncertainties, which is possible with our approach. In the case that our algorithm finds no analogue we can still identify the reason: if the upper and lower bound of a pixel's climate projections belong to different clusters, projection uncertainty is too large. If both bounds lie within one cluster

which has no current climate members, the city's future climate currently does not exist (within the considered study region).

A further general advantage of the method that has to be acknowledged is that it produces a full and comprehensible picture. While other methodologies always lack the ability to deal with multiple regions of interest (Hallegatte et al. 2007; Williams et al. 2007; Arnbjerg-Nielsen et al. 2015; Hibino et al. 2015), we automatically deliver RCAs for larger areas. Essentially this means that we can display the whole RCA information in two maps (see Fig. 2). This enables another application of the method: the reverse usage of an analogue. As the method delivers a comprehensive picture of the study region, a city in the current climate (Fig. 2, right) could identify a region in the robust projection map (Fig. 2, left) as a potential demand region for a supplier of today's local climate-related applications. E.g., a supplier of air humidifiers for public spaces could identify future sales markets. In other words, Helsinki is a robust climate analogue of Murmansk but Murmansk is a potential sales market of adaptation methods for Helsinki. Here, of course, the time scale of climate change has to be considered.

It is necessary to extend the set of climate variables when applying this method more specifically to urban areas. In particular, extreme weather event indices have to be included due to the disruptive impact of such weather events on various critical infrastructure and public health. Furthermore, the multi-grid-cell character of many urban areas has to be considered. To better meet urban challenges, nearer time horizons and if possible higher spatial resolution data should be considered. A follow-up study should implement other RCPs as this study only focusses on RCP8.5. This reflects the stakeholder's perspective which does not distinguish between different kinds of uncertainty and will deliver climate analogues which are robust towards both model and scenario uncertainty. These different concentration pathways can easily be integrated into the treatment of uncertainty by the method proposed here.

Acknowledgements

This work was funded in parts by the Senate of Berlin (AFOK) and the City of Potsdam ("Adaptation to climate change in the State Capital of Potsdam"). We thank Boris Prahl for helpful comments and Alison Schlums for proofreading the manuscript.

5 REFERENCES

Arnbjerg-Nielsen, K., Funder, S.G. & Madsen, H., 2015. Identifying climate analogues for precipitation extremes for Denmark based on RCM simulations from the ENSEMBLES database. *Water Science and Technology*, 71(3), pp.418–425.

Carpenter, K.E. et al., 2008. One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science (New York, N.Y.)*, 321(5888), pp.560–563.

Dawson, J., Scott, D. & Mcboyle, G., 2009. Climate change analogue analysis of ski tourism in the northeastern USA. *Climate Research*.

Dehn, M., 1999. Application of an analog downscaling technique to the assessment of future landslide activity - a case study in the Italian Alps. *Climate Research*.

Geiger, G., 1961. *Überarbeitete Neuausgabe von Geiger, R.: Koppen-Geiger / Klima der Erde. Wandkarte 1:16 Mill.*, Gotha.

Hallegatte, S., Hourcade, J.-C. & Ambrosi, P., 2007. Using climate analogues for assessing climate change economic impacts in urban areas. *Climatic Change*, 82(1-2), pp.47–60. Available at: http://link.springer.com/10.1007/s10584-006-9161-z.

Hartigan, J.A. & Wong, M.A., 1979. A K-Means Clustering Algorithm. *Applied Statistics*, 28(1), pp.100–108. Available at:

353 http://linkinghub.elsevier.com/retrieve/pii/S0167865503001466.

354 Hawkins, E. & Sutton, R., 2011. The potential to narrow uncertainty in projections of regional precipitation change. *Climate*
355 *Dynamics*, 37(1), pp.407–418.

356 Haylock, M.R. et al., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for
357 1950-2006. *Journal of Geophysical Research: Atmospheres*, 113(20).

358 Hibino, K., Takayabu, I. & Nakaegawa, T., 2015. Objective estimate of future climate analogues projected by an ensemble
359 AGCM experiment under the SRES A1B scenario. *Climatic Change*, 131(4), pp.677–689.

360 Hubert, L. & Arabie, P., 1985. Comparing partitions. *Journal of Classification*, 2(1), pp.193–218. Available at:
361 http://www.springerlink.com/index/x64124718341j1j0.pdf.

362 IPCC, 2013. *Working Group I Contribution to the IPCC Fifth Assessment Report, Climate Change 2013: The Physical Science*
363 *Basis* D. L. Hartmann, A. M. G. K. Tank, & M. Rusticucci, eds., Cambridge University Press.

364 Jacob, D. et al., 2014. EURO-CORDEX: New high-resolution climate change projections for European impact research.
365 *Regional Environmental Change*, 14(2), pp.563–578.

366 Janssen, P., Walther, C. & Lüdeke, M., 2012. Cluster analysis to understand socio-ecological systems: A guideline. *PIK*
367 *Report*, (126).

368 Knutti, R. & Sedláček, J., 2012. Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate*
369 *Change*, 3(4), pp.369–373. Available at: http://www.nature.com/doifinder/10.1038/nclimate1716 [Accessed April 29,
370 2014].

371 Kohonen, T., 1998. The Self-Organizing Map. *Neurocomputing*, 21, pp.1–6.

372 Kok, M. et al., 2016. A new method for analysing socio-ecological patterns of vulnerability. *Regional Environmental Change*,
373 16(6), p.pp 229–243.

374 Kopf, S., Ha-Duong, M. & Hallegatte, S., 2008. Using maps of city analogues to display and interpret climate change
375 scenarios and their uncertainty. *Natural Hazards and Earth System Science*, 8(4), pp.905–918.

376 Krysanova, V. & Hattermann, F., 2017. Hydrological Models Intercomparison for Climate Impact Assessment. *Climatic*
377 *Change*, 141(3).

378 Lobell, D.B. et al., 2008. Prioritizing climate change adaptation needs for food security in 2030. *Science*.

379 MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth*
380 *Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 231–297.

381 Meehl, G. et al., 2007. Global climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of*
382 *Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK:
383 Cambridge University Press,.

384 Mote, P.W. & Salathé, E.P., 2010. Future climate in the Pacific Northwest. *Climatic Change*, 102(1-2), pp.29–50.

385 Nassopoulos, H., Dumas, P. & Hallegatte, S., 2012. Adaptation to an uncertain climate change: Cost benefit analysis and
386 robust decision making for dam dimensioning. *Climatic Change*.

387 Parry, M.L. & Carter, T.R., 1989. An assessment of the effects of climatic change on agriculture. *Climatic Change*.

388 Pickett, S.T.A., 1989. Space-for-Time Substitution as an Alternative to Long-Term Studies. In *Long-Term Studies in Ecology*.

389 Pielke, R. et al., 2007. Lifting the taboo on adaptation. *Nature*.

390 Pugh, T.A.M. et al., 2016. Climate analogues suggest limited potential for intensification of production on current croplands
391 under climate change. *Nature Communications*, 7, p.12608. Available at:
392 http://www.nature.com/doifinder/10.1038/ncomms12608.

393 Rastetter, E.B., 1996. Validating Models of Ecosystem Response to Global Change. *BioScience*.

Reckien, D. et al., 2015. The influence of drivers and barriers on urban adaptation and mitigation plans-an empirical analysis of European Cities. *PLoS ONE*.

Rosenzweig, C. et al., 2017. Assessing inter-sectoral climate change risks: The role of ISIMIP. *Environmental Research Letters*.

Sietz, D., Lüdeke, M.K.B. & Walther, C., 2011. Categorisation of typical vulnerability patterns in global drylands. *Global Environmental Change*, 21(2), pp.431–440. Available at: http://www.sciencedirect.com/science/article/pii/S0959378010001081.

Steinley, D., 2006. K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), pp.1–34.

Taylor, K.E., Stouffer, R.J. & Meehl, G.A., 2012. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4), pp.485–498.

Tebaldi, C. & Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 365(1857), pp.2053–2075.

Williams, J.W., Jackson, S.T. & Kutzbach, J.E., 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), pp.5738–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1851561&tool=pmcentrez&rendertype=abstract.

Cluster number	Color	City/Region of interest – example for each cluster	Optional robust climate analogue (RCA)	Variance
1	Red	Helsinki (Finland)	Copenhagen (Denmark)	0.045
2	ocean blue	Murmansk (Russia)	Riga (Latvia)	0.043
3	Orange	Paris (France)	Madrid (Spain), Rome (Italy)	0.101
4	Green	Lviv (Ukraine)	Zagreb (Croatia)	0.063
5	Violet	Berlin (Germany)	Bologna (Italy), Toulouse (France)	0.059
6	Cyan	Copenhagen (Denmark)	Nantes (France), Bilbao (Spain)	0.081
7	turquoise	NA	NA	0.060
8	Yellow	Rome (Italy)	Tunis (Tunesia)	0.083
9	Blue	Hardangervidda (Norway)	Kristiansand (Norway)	0.155
10	magenta	Strip parallel West Coast (Norway)	Oban (Scottland)	0.342

Table 1: Overview of the ten climate analogue clusters, color, example cities for "City of interest" and "RCA" as well as the variance of each cluster calculated by the mean within cluster sum.

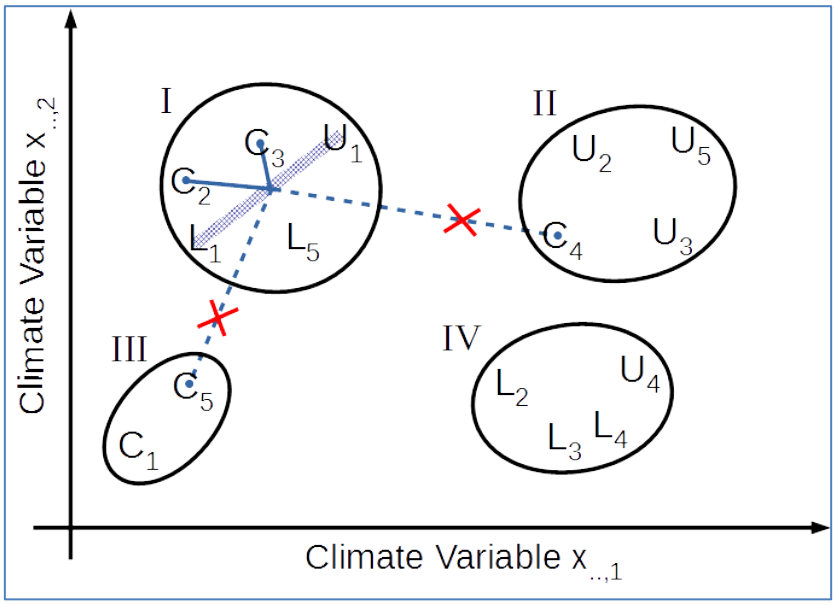

Fig. 1: Stylized illustration of the proposed approach for five grid elements *i* = 1, …, 5 and two climate variables. C_i: current climate of grid element *i*; L_i: lower bound of the climate projections of grid element *i*; U_i: upper bound of the climate projections of grid element i. The 5x3 points in the 2-dimensional climate space form four clusters I to IV. Cluster I: grid elements 2 and3 are robust CAs for grid element 1; Cluster II: grid element 4 is no robust CA to grid elements 2, 3 & 5 (it is only similar to their upper bound of the projections); Cluster III: grid elements 1 & 5 are not CAs for any grid element (disappearing climate of Williams et al., 2009); Cluster IV: grid element 4 has no present analogue (novel climate of Williams et al., 2009). For grid elements 2, 3 & 5 upper and lower bound of the projections belong to different clusters, i.e. they have no robust CAs. The blue lines illustrate differences to and communalities with the approach of Hibino et al. (2015) - see text.

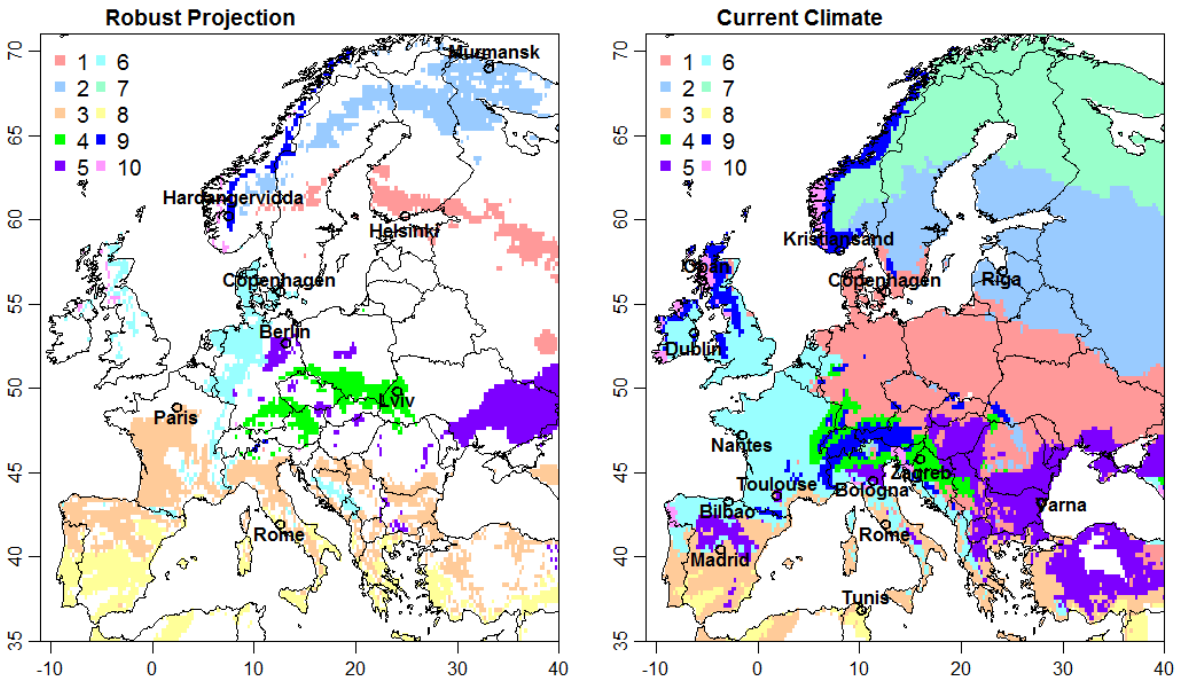

Fig. 2: The figure on the left shows the 10 climate analogue clusters, represented by their robust projections gained by overlapping lower and upper limits of the projections. Here for each cluster one representative city is shown. The right-hand figure shows the current climate with a selection of possible robust climate analogue cities in each of the clusters. White grid cells mark regions with deviating cluster allocations, overlapping, missing data or oceans.

Appendix

App. 1: Choice of clustering method

The definition of robust climate analogues implies that the appropriate clustering method should emphasize the cluster properties of compactness and distance over connectedness (Janssen et al. 2012). Therefore methods which tend to generate spherical clusters seem adequate. As the analysis of the topological structure and potential dimension reduction (by SOM-based methods, e.g. Kohonen 1998) are not of major importance in our case the partitioning cluster method k-means (MacQueen 1967; Hartigan and Wong 1979) has been applied. This algorithm minimizes the total within-cluster sum-of-squares (*TSS*) criterion (Steinley 2006). If the data set consists of *V* variables and the number of groups is chosen to be *K*, the criterion is defined by:

$$TSS = \sum_{j=1}^{V} \sum_{k=1}^{K} \sum_{o \in Q_k} (x_{oj} - \bar{x}_j^{(k)})^2, \qquad \text{(A1)}$$

where x_{oj} is the value of the variable *j* of object *o* and $\bar{x}_j^{(k)}$ is the mean value of variable *j* in cluster *k*. In our case *o* counts all 3**N* objects in climate space defined by grid element *i* and index $d \in \{c, l, u\}$while Q_k denotes the objects belonging to cluster *k*. A hierarchical clustering is used to initialize the partitioning method. The objects are assigned to the given *k* initial cluster centers followed by a calculation of new centers as the average of all objects within each cluster. In an iterative process each object is assigned to a cluster in the way that *TSS* is minimized again followed by a calculation of the centers. This process is repeated until a breakup criterion is reached.

App. 2: Selection of number of clusters

A challenge in applying k-means is the question of the appropriate number of clusters. Here a consistency measure is applied (Sietz et al. 2011) which identifies the most robust partition. This is done by repeating the clustering algorithm many times with different initial conditions for a sequence of fixed cluster numbers and analyzing the variation in the partitioning results for a given cluster number. The cluster number which – independent of the initial condition – repeatedly generates a similar cluster partition (measured by the Rand-index; Hubert & Arabie 1985) is assumed to be the most appropriate to the given data structure. The resulting consistency measure may show local maxima for different cluster numbers. Amongst these, the appropriate cluster number can be chosen according to a further criteria, e.g. the expected detail of the analysis (Janssen et al. 2012).

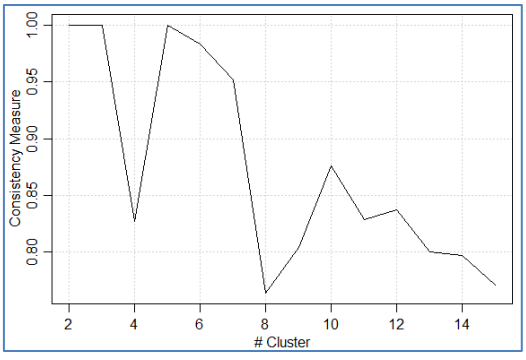

Fig. A2: Result of the consistency measure calculation for the cluster numbers 2 to 15. Pronounced relative maxima denote cluster numbers which reflect the data structure.

For this study we assembled the climate variables in each grid cell from the three climate data sets: the current, the lower and the upper bound of the projections. Then the consistency measure calculation for a varying number of clusters (starting from two clusters and ending at fifteen, see Figure A2) was done. The measure shows pronounced relative maxima at 2, 3, 5 and 10 clusters which reflect the underlying structure of the point

cloud in climate space. Climate analogues can only fulfill their purpose with an adequate level of differentiation – therefore cluster numbers 2, 3 and 5 are too small and the partition with ten climate analogue clusters will be used for further analysis.

After performing the selection of the most appropriate cluster number *K* and the clustering into the *K* clusters in climate space the results can be transformed into spatial maps of cluster membership. To further interpret and discuss these clusters, graphs of the distribution of the climate variables in each cluster were developed.

App. 3: Feature plot of the ten climate analogue clusters

For further characterization of the clusters we show in the following Figure A3 the cluster-specific distribution of some selected climate variables.

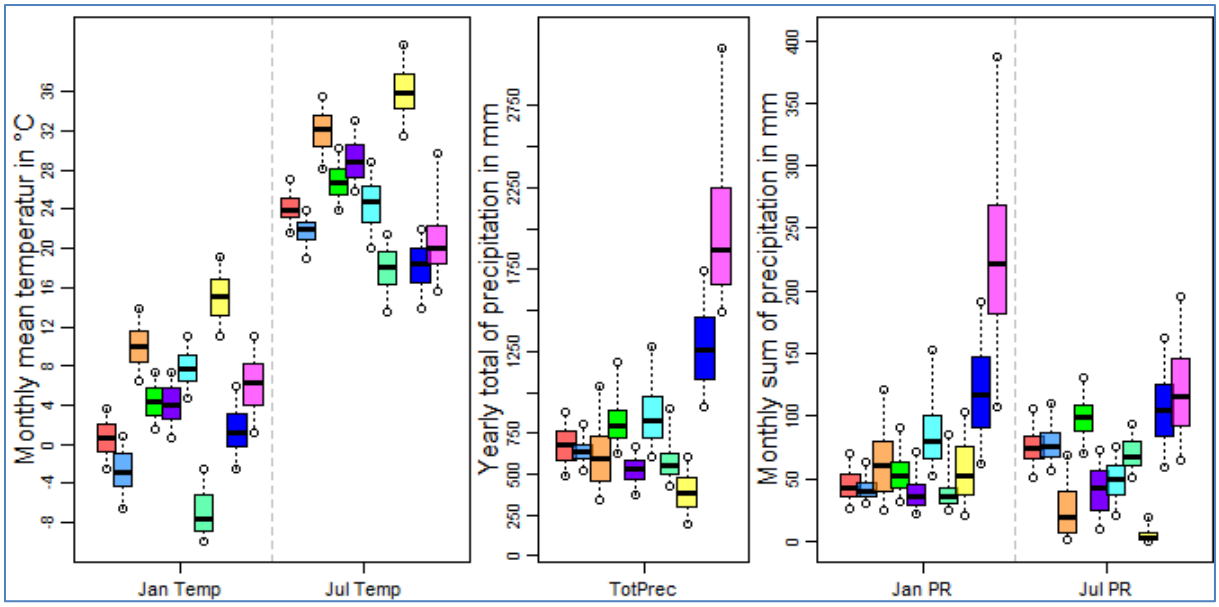

Fig. A3: Feature plot of the ten climate analogue clusters (see e.g. Kok et al. 2016). Examples are shown for monthly mean temperature in the left plot, for total precipitation in the middle (Yearly total – TotPrec) and monthly sum of precipitation in the right plot (January – Jan PR, July – Jul PR). The distribution of the grid cells in the clusters is illustrated by box plots (median, box: 25th and 75th percentile, whiskers: 5th and 95th percentile; the cluster number corresponds to the order of the boxes in the plot from left to right).

App. 4: Comparison of RCAs with Hallegatte et al. (2007)

In the following Table A4 we summarize the results of the comparison between our results and climate analogues as identified by a minimum distance approach (Hallegatte et al., 2007) for some European cities. For each city the latter identifies two analogues, depending on the climate projection. We then check if they lie within the identified RCA region. The results are interpreted in the discussion section of the main article.

City	Climate Cluster	Climate Analogue Region in Hallegatte (2007)		Agree-ment A	Agree-ment B
		Weaker climate response - A	Stronger climate response – B		
Berlin	5	Italy, Reg. Rome	Spain, Reg. Salamanca	No	Yes
Copenhagen	6	France, Reg. Paris	Tirana	Yes	Yes
Helsinki	1	-	-	No	No
Paris	3	France, Reg. Bordeaux	Spain, Reg. Cordoba	No	Yes
Rome	8	-	Not on the map	No	No

Table A4: CAs from Hallegatte et al. (2007) compared with CAs gained by our cluster method.

App. 5:

In Figure A5 we show the pixels where the climate change signal is small compared to the uncertainty range of the projections. In terms of our algorithm this means that the current climate of a location shares a cluster with the lower and upper bound of its climate projections.

Pixels from clusters one and two (red and ocean blue) don't show "no climate change" as defined by our cluster-oriented metrics. These are also the clusters with the smallest extend in climate space (see Table 1, variance ranks 10 and 9). Relatively large areas are found for clusters 6, 8 and 10 (cyan, yellow and magenta) in accordance with their larger variance (ranks 1, 3 and 4).

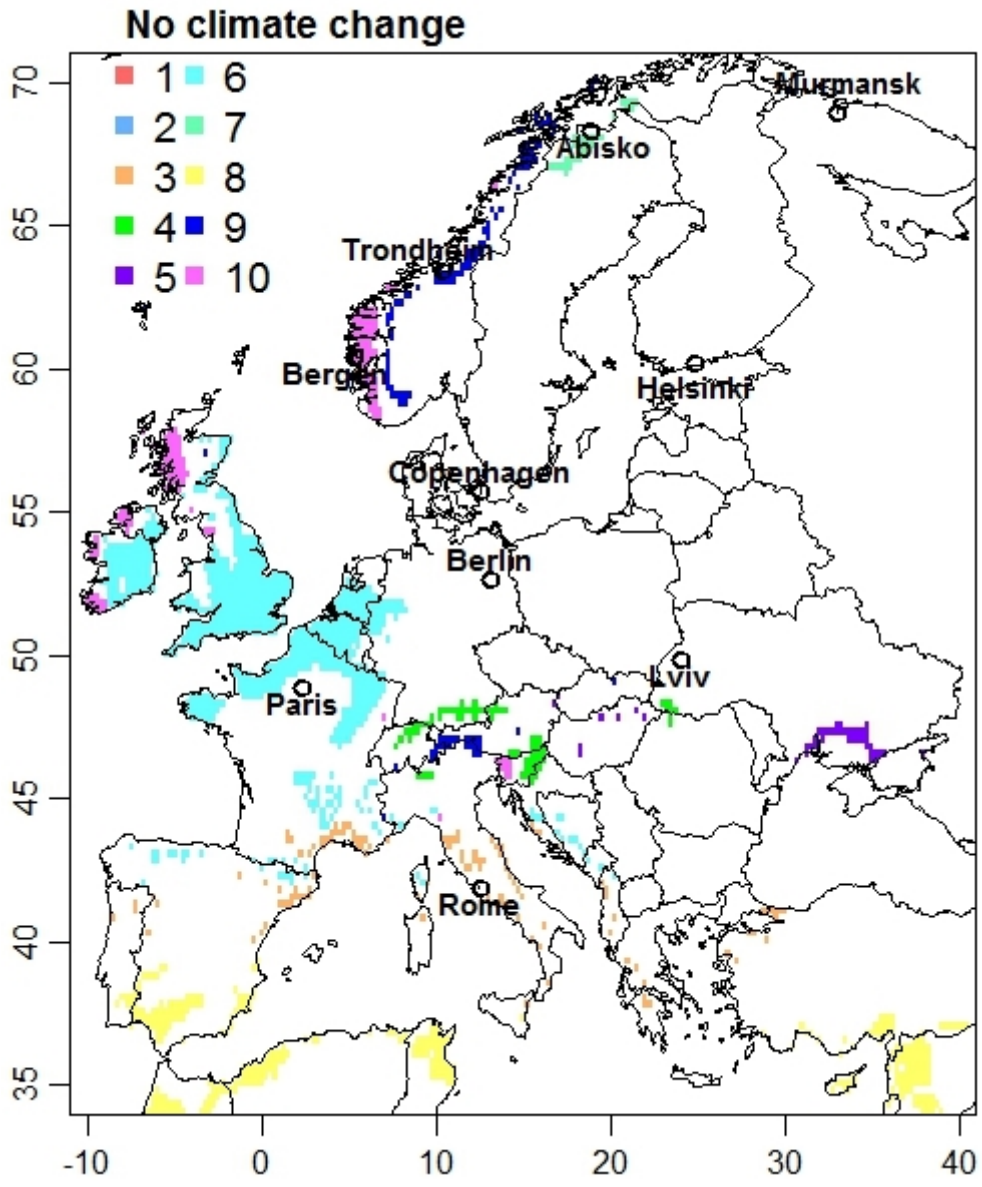

Figure A5: Pixels *i* where C_i, L_i and U_i share the same cluster (Colors as in Figure 2).