Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of Temperatures and Precipitation

M. J. Carvalho[a,], P. Melo-Gonçalves[a], J. C. Teixeira[a], A. Rocha[a]

[a]*Department of Physics & CESAM, University of Aveiro, Aveiro, Portugal*

Abstract

In order to study climate change on a regional scale using Earth System Models, it is useful to parition the spatial domain into regions according to their climate changes. The aim of this work is to divide the European domain into regions of similar projected climate changes using a simulation of daily total precipitation, minimum and maximum temperatures for the recent-past (1986 – 2005) and long-term future (2081 – 2100) provided by the Coupled Model Intercomparison Project (CMIP5). The difference between the long-term future and recent-past daily climatologies of these three variables is determined. Aiming to objectively identify the grid points with coherent climate changes, a K-Mean Cluster Analysis is applied to these differences. This method is performed for each variable independently (univariate version) and for the aggregation of the three variables (multivariate version). A mathematical approach to determine the optimal number of clusters is pursued. However, due to the method characteristics, a sensitivity test to the number of clusters is performed by analysing the consistency of the results. This is a novel method, allowing for the determination of regions based on the climate change of multiple variables. Results from this method are in accordance with results found in the literature, showing overall similar regions of changes.

Keywords: Climate Change, Surface Temperatures, K-Means Clustering, Precipitation, Europe

1. Introduction

Climate change studies are usually carried out either globally or regionally. Either way, they usually focus on areas with different climate characteristics and large variability. In Europe, temporal variability of daily surface climate variables (such as minimum and maximum temperatures and

*Corresponding author
Email address: mariajcarvalho@ua.pt (M. J. Carvalho)

precipitation) has high spatial gradients. Therefore, statistics of the temporal behaviour of a particular variable or its derived quantities over the target domain must be estimated taking into account these spatial gradients. Some statistics can be displayed over a map; however there are statistics, such as Probability Density Functions at each grid point of the domain, that are impossible to be displayed in a map. Because of this, it is mandatory to reduce the number of degrees of freedom which, in this case, consists of a reduction of the time series representative of the domain. This, together with the large amount of data, adds up to the need to define regions to be analysed using either grid point as a representation of the region or the average behaviour of the grid points.

In an attempt to divide the overland areas of the globe into a manageable number of regions, each with simple shape and representing a different climatic regime, several authors (such as Sillmann et al. (2013, 2014)) have followed the approach of Giorgi and Francisco (2000). When studying the uncertainty in regional climate change prediction using ensemble simulations from a coupled Atmosphere-Ocean Global Climate Model (AOGCM), they proposed a division of the domain, creating rectangle-like overland areas, admitting, however, that this was a subjective approach to the issue. On a regional scale, the simple use of geographic markers has been extensively used in order to define regions. For example, in their study of European heat waves in present-day and future climates, Lau and Nath (2014) simply divided the domain into three regions: Russia, eastern Europe and western Europe. Much like this, when studying record high maximum and low minimum temperatures, Meehl et al. (2009) used the 100 °W meridian to divide the United States of America into eastern and western USA. Fischer and Schär (2009) simply use the Iberian Peninsula, Scandinavia and France as key regions when studying the PRUDENCE regional climate model scenarios for temperature and the driving processes in temperature extremes while Wójcik (2014) divides Poland taking into account some orographic characteristic in order to study the reliability of CMIP5 simulations in reproducing atmospheric circulation. An upgrade to this methodology is the approach of Fischer et al. (2014) who segregate grid points based on their altitude, in order to study projected changes in precipitation intensity and frequency in Switzerland.

Due to the subjectivity of the regionalization methodologies described above, several authors have pursued a more objective approach. For example, Richman and Lamb (1985, 1987) used Principal Component Analysis (PCA) in order to study the spatial distribution of three- and seven-day rainfall events for the USA and Canada, respectively. In a later work, White et al. (1991) applied Rotation Principal Analysis to monthly precipitation in Pennsylvania using different rotation algorithms in order to assess the sensitivity of the regionalization result to rotation. They found that the resulting regions varied widely with the use of different rotation algorithms.

In addition to having been used to identify Weather Types (or Weather Regimes) by Santos et al. (2005), Cluster Analysis has also been used to define climatic regions. DeGaetano and Shulman (1990) applied a flexible clustering technique to the first three principal components of several climatological variables in order to identify regions of coherent plant hardiness. Much like this, Fovell and Fovell (1993) used K-Means Cluster Analysis in order to identify climatic zones of the Conterminous United States based on both temperature and precipitation.

However objective these studies may be, they regionalize the domain using observed (or modeled) data for a given period and therefore obtain regions of coherent climate. In a changing climate, the main interest becomes knowing the regions of coherent changes, instead of the definition of regions with the same climate characteristics, since they can change in time. The main goal of this work was to identify regions with consistent climate changes in precipitation and surface minimum and maximum temperature seasonal cycles in Europe.This is a novel method, allowing for the determination of regions based on the climate change of multiple variables.

2. Method and Data

The data used was provided by the Coupled Model Intercomparison Project Phase 5 (CMIP5), simulated by the MPI-ESM-LR model with the r1i1p1 initialization, with a horizontal resolution of 1.9° horizontal resolution (Giorgetta et al., 2013). As a representation of the recent-past climate, the last 20 years (1986 – 2005) of the historical experiment which runs from 1850 to 2005 were used.

The future climate used was simulated by using the 8.5 Representative Concentration Pathway (RCP8.5) which stabilizes radiative forcing at $8.5\,\mathrm{W \cdot m^{-2}}$ in the year 2100 without exceeding that value (Riahi et al., 2011). From the future climate experiment, which runs from 2006 to 2100, only the long-term future (from 2081 to 2100) was used in this work, since changes for this period relative to the recent-past are expected to be greater than those found for both the near-term (2016 – 2035) and mid-term (2046 – 2065) periods. The variables used were the daily minimum and maximum near-surface air temperatures as well as total daily precipitation, which includes both liquid and solid phases from all types of clouds (both large scale and convective). The simulations are available for the entire globe. However, this study focused on a domain containing only Europe: $25\,°\mathrm{N} - 70\,°\mathrm{N}, 45\,°\mathrm{W} - 65\,°\mathrm{E}$. This domain is presented in Figure 1 where the model grid points corresponding to the aforementioned resolution are also plotted.

Taking each of the two climates – recent-past and long-term future – daily climatologies of each of the variables were determined, for each of the domain grid points, using a 15 day-running window as a low frequency filter. Taking each grid point, the difference between the recent-past and long-term future climatologies were determined, creating a measure of the changes in the seasonal cycle.

The challenge was then identifying grid points where the changes in the seasonal cycle of the variables are similar. To the difference fields, the K-Means Cluster Analysis was applied. This is a non-hierarchical clustering method which starts by computing the centroids for each cluster and then calculates the distances between the current data vector and each of the centroids, assigning the vector to the cluster whose centroid is closest to it. Since this is a dynamic method, meaning that vectors can change cluster after being assigned to it, this process is repeated until all vectors are assigned a cluster and their members are closest to the centroid than to the mean of other clusters (Wilks, 2011). The mathematical condition for the cluster C_k and the k centroids μ_k can be expressed as Equation 1.

$$Minimize \quad \sum_{k=1}^{K} \sum_{x_n \in C_k} \parallel x_n - \mu_k \parallel^2 \quad with\ respect\ to\ C_k,\ \mu_k \tag{1}$$

It is important to note that, unlike when applied to determine weather types, the K-Means Clustering is done in space and not time, resulting in each grid point (instead of a time step) being assigned to a cluster.

The described methodology was applied to the climatology differences of each of the three variables independently (univariate version) and using the daily climatology differences of a synthetic joint variable composed by concatenating the temporal-varying spatial fields of the three variables (multivariate version). Since the goal is to determine one set of regions on which to base the analysis, the univariate version is only used to, once more, analyse the consistency of the multivariate results and ultimately validate them, since this is a novel statistical approach.

In order to determine the number of clusters, the Gap Statistics was used as described by Pham et al. (2005). However, since this is a blind statistic, the sensitivity to the number of clusters was estimated by computing the K-Means for different k's, which allowed for the verification of the results' consistency.

Lastly, the statistical significance of the differences between the regional averaged long-term future and recent-past daily climatologies for each variable was estimated using the non-parametric Rank-sum test (Mann and Whitney, 1947). This test was chosen due to its resistance to wild data or outliers which could otherwise contaminate the results by providing false negatives (Wilks, 2011). Furthermore, since it is a non-parametric test, it does not require data with a normal probability distribution.

3. Results and Discussion

Due to the chosen clustering method, the number of clusters, k, must be chosen a priori. Using the Gap Statistics as described in Section 2, it was determined that $k = 6$ was the optimal choice. As mentioned before, since this is a purely mathematical method, the clustering analysis was also performed with $k = 3$, 10 and 13. For each k, the method was applied to each of the variables individually, followed by the multivariate version. For the sake of conciseness, and since the univariate version serves as a validation of coherence, only the results of the multivariate cluster analysis will be presented.

3.1. Regionalization

When applying the clustering algorithm to each of the variables using $k = 3$ (results not shown), the domain is divided into three regions which are similar for maximum and minimum temperatures, with the exception of the Mediterranean region which is grouped together with the Atlantic region for minimum temperature and with northern Europe for maximum temperature. Precipitation regions show a clear division of Europe into two along the 55 °N latitude, with the third region located northwest of Iceland being considerably smaller. Results obtained by the multivariate version, shown in Figure 2a), are consistent with those obtained by the univariate one. There is a region south of 40/50 °N latitude, which includes north Africa, the Mediterranean, Italy, Greece and most of the Iberian Peninsula. The second (middle) region goes from the 40/50 °N latitude up to 60 °N overland, and encroaches up to the upper boundary of the domain over the Atlantic ocean. The third regions encloses two subregions: northern Europe (north of 60 °N) and Greenland.

When the number of clusters is increased, there are more regions to which each point can be attributed and therefore the regions obtained will provide a more detailed definition of the changes. As for $k = 3$, the $k = 6$ regions obtained for maximum and minimum temperatures are similar and differ from the precipitation ones especially over the Atlantic ocean, in the western most part of the domain. In this area, there is a larger number of regions for precipitation while there is more differentiation over land when temperature is used. The regions based on the precipitation changes follow the approximate layering described for the multivariate $k = 3$ regionalization for the overland area while, for the western side of the 20 °W meridian, there is an increased number of regions from three to four. Results obtained by the multivariate version, shown in Figure 2b), approximately follow the temperature ones; there is a small region over Greenland and a larger region over the northern part of the Atlantic. The southern region of the Atlantic ocean belongs to a distinct region which extends into the Mediterranean and includes countries such as Portugal, southern Spain, western France, and Italy). Overland points in Africa are grouped into another region, while the remaining parts of Europe are divided once more approximately along the 55/60 °N parallel. The fact that so many overland points belong to the southern-Atlantic and Mediterranean region is most likely due to the combined effect of the model's coastline and its low resolution.

The results for $k = 3$ results show that there is a latitudinal layering of the changes affecting precipitation as well as maximum and minimum temperatures in Europe. These results are consistent with the work of Field et al. (2014) who have shown that there is a clear latitudinal differentiation of climate change of precipitation and temperatures, as well as extreme events such as dry spells and heat waves. Also, there is a distinctive area of changes in central Europe which, in this study, appears in the $k = 6$ regionalization.

When applying $k = 10$, the number of regions is higher over the Atlantic for precipitation and over land for the temperatures, even though more balanced than what was found for a lower k. Also, in the univariate versions of the clustering, the regions become less organized and layered when compared to what happened for a lower k. However, the multivariate regionalization shown in Figure 2c) retains the horizontal layering format, even though presenting differences between overland and over-ocean areas. The north African region, obtained for $k = 3$ and 6, become divided into two distinct regions: one between the 30 °N and 40 °N parallels in the overland points, and another below the southern parallel. Once more Europe is divided into three belt regions: the first between the 40 °N and 50 °N parallels, the second up to the 60 °N and the third, north of that. It is noteworthy that some grid points in the southern Iberian Peninsula are attributed to the same region as north Africa. However, this should be considered carefully since the connection between the two parts of the region, as well as the Iberian part of the region itself, is cell-thin. Over the ocean, one of the regions includes the entire Atlantic below 40 °N, encroaching into the Mediterranean (similar with a region found for $k = 6$). Above that, there is another region which is bordered above and to the west by a different one. As seen for $k = 6$, there is also another distinct region over Greenland.

When moving to $k = 13$, the increase in regions is still distributed the same way (more regions over the ocean for precipitation and more regions overland for temperature). However, and as can be seen in Figure 2d) there are several cell-thin regions which leads to the belief that the number of clusters is too high for the spatial resolution of this model.

From the comparison of results obtained from different k's in the K-Means algorithm, it can be concluded that $k = 6$ is, as predicted by the Gap Statistics, the best choice. It could be argued that applying $k = 10$ to the multivariate version of the clustering method yields equally good results with the added value of a more localized regionalization. However, when using the univariate version, there are several cases of cell-thin regions suggesting that this number of clusters is too high for the resolution in use.

3.2. Validation of the k regions

Since both the Gap Statistics and the analysis of the results for different number of clusters points to six being the optimum k, it is worth looking into the differences in climatologies of the variables for each of the different regions. These differences can, not only show the differences between regions, but also serve as an objective characterization of the regions in terms of their climate change patterns in the long-term future.

Since the regions were defined based on climate change in the seasonal cycle of precipitation, maximum and minimum temperature, it is worth comparing the Probability Distribution Functions (PDFs) of two climates (recent-past and long-term future) of each region (Figure 3d) – i)). The Rank-Sum statistical test (see Section 2) was applied to the PDFs of the two climates. The only regions for whih the long-term future climate is considered to be different is R1 (with a confidence level of 95%) and R4 (confidence level 90%) and only for precipitation. The remaining PDFs cannot be considered significantly different. However, and since it was the daily climatology difference that was used to define the clusters, that was also tested using the Rank-Sum test. When testing the region-averaged daily climatologies of the long-term future agains those of the recent-past, the results show that, for all six regions, the long-term future daily climatology is distinct from the recent-past one at a confidence level of 99%. These results point to an interesting development. While there is, on average, no significant climate change detection in the PDFs of the variables, for all the regions, that change is present and statistically significant in the climatologies, showing that, while detecting seasonal changes, those might not be evident in the original PDFs.

On the left column of Figure 3, the mean regional seasonal climatological differences (deviations of the long-term future climatology from the recent-past climatology) for each variable can be seen, for the six regions. For all three variables, it is evident that each region presents different climatological changes, which vary throughout the seasons in distinct ways. That said, it can be seen that the climatological differences are always positive for both minimum and maximum temperatures (top two plots), pointing to a warming of all regions, albeit of different magnitude. Overall, the seasonal differences between the recent-past and long-term future are statistically significant at the 95% confidence level for temperatures and precipitation, for all regions. The exceptions are R2 and R6 in spring and R1 and R6 for autumn (marked with the asterisks).

The regions showing largest climatological differences for both temperatures are R1 (north-Atlantic), R3 (northern Africa) and R6 (Central an eastern Europe) followed by R5 (Greenland). The largest differences found are for R3, with minimum temperature changes reaching an 8 °C increase (6 °C in maximum temperature) in the future during winter, and a 5 °C increase in both temperatures throughout the remaining seasons. On the other hand, R2 (northern Europe) and R4 (southern part of the Atlantic extending to the Mediterranean and IP) show the lowest magnitude of changes, ranging from $\sim$ 2.5°C for the earlier and $\sim$ 3.5 °C for the later. These results point to a stronger warming in areas where temperatures tend to be more extreme, such as north Africa and central Europe which are warmer and Greenland where temperatures tend to be lower.

When compared to temperature, precipitation changes (lower plot) show a different scenario, with a less organized seasonal pattern, which would be expected due to the fact that precipitation (unlike temperature) is neither continuous in time or space. The regions which show higher and therefore significant changes throughout the seasons are R3, R4 and R5. In magnitude, the season showing largest changes is autumn with a decrease of $\sim 0.7\,\mathrm{mm}$ (R4 + R6) and increase of $1.5\,\mathrm{mm}$ (R2 + R3 + R5). R4, or the southern Atlantic and Mediterranean region shows a systematic and significant decrease of precipitation through the seasons while regions such as northern Africa (R3) and Europe (R2) and Greenland (R6) show increase in precipitation. Central Europe shows a redistribution of the rainfall patterns through the seasons, with its decrease in precipitation is mainly due to the summer. As for temperature, it is clear that the largest changes are projected for the most arid area of the domain (north Africa), as well as snow prone regions such as north Atlantic and Greenland.

Table 1: Cross-test of the regions to determine if the daily climatologies of each of the variables are significantly different for each region. Checks represent pairs of regions which have shown to be distinct from each other for all three variables, at the 95% confidence level. The remaining pairs are considered to have the same distributions for the mentioned variable. Dark gray cells show pairs of regions for which the precipitation distributions are considered to be different, but only at the 90% confidence level.

	R1	R2	R3	R4	R5	R6
R1		✓	tasmax	✓	tasmin	pr
R2			✓	✓	pr	✓
R3				✓	pr	✓
R4					✓	✓
R5						tasmax
R6						

In order to verify if the regions are significantly different from each other, the Rank-sum statistical test was applied to each pair of regions, at the 95% confidence level (Table 1). The only region which is considered to be significantly different from all others at a 95% confidence level for all three variables under study is R4. However, if the confidence level is lowered to 90%, R2 also passes the test for all variables and is therefore concluded to be distinct from the other regions. Even though the distribution of one of the variables is not significantly different for some pairs of regions (as happens with R1 and R3 and R5 and R6 for maximum temperature, R1 and R5 for minimum temperature and R3 and R5 for precipitation), it is worth considering these regions as having different characteristics since the clustering is performed on several features (i.e. variables) rather than just one. This outcome shows a shortfall of this methodology which is, however, overcome by the fact that the method produces one set of regions based on the climate change of a group of variables.

4. Concluding Remarks

This work aims to develop a novel approach to the regionalization of a domain, in this case Europe, based on the climate change of a range of variables. The focus was on the long-term changes in precipitation, maximum and minimum temperatures. This was achieved by applying a K-Means clustering analysis to the daily climatological difference for each of the variables independently (univariate – not shown) and, most importantly, using each of the variables as a feature (multivariate version). The result is a map in which each grid point is associated to a cluster (region).

Results show that the multivariate version is congruent with the univariate version, although creating new and more intricate features in the regions. For the optimum determined k (six), there is a clear latitudinal layering of the regions, which is then overrun by the inland-ocean differences. The Atlantic Ocean area is divided into northern and southern part, with the later extending over the Iberian Peninsula and the Mediterranean. Greenland, north Africa, Central to eastern and northern Europe comprise the other four regions.

When analysing the seasonally averaged climatology differences for each region, it becomes clear that these regions have, in fact, different characteristics concerning precipitation, maximum and minimum temperature projected changes, even though these are not statistically significant for every variable when region pairs are compared.

Is is noteworthy that the maximum and minimum temperature changes projected for the long-term are positive and statistically significant in every region, pointing to a clear warming of Europe, for every season. Precipitation changes show a more complex outlook, with the Mediterranean and southern Atlantic showing a systematic and significant decrease in precipitation and regions such as northern Africa and Europe and Greenland showing increase.

The sensitivity of the results to the number of regions was tested by performing the same methodology for $k = 3$, 6, 10, 13. Results show that, when increasing the number of clusters considered there is increased detail in the spatial features obtained. However, due to the rather coarse resolution of the data, when $k = 10$, 13, single grid points of a region engulfed by other regions start to appear. These features may not be geographical and therefore, the number of clusters needs adaptation for different resolutions.

Even though the seasonal climate change detected is not evident on the Probability Distribution Functions of the original variables and that some regions were found to not be significantly different from each other concerning the changes of a variable, this methodology presents a novel way to approach the subject of identifying regions of coherent climate change. Furthermore, it creates the possibility to determine areas based on several variables, rather than just one.

5. Acknowledgments

This study was supported by FEDER funds through the Programa Operacional Factores de Competitividade – COMPETE and by Portuguese national funds through FCT – Fundação para a Ciência e a Tecnologia, within the framework of the following projects: CLIPE Project Reference PTDC/AAC-CLI/111733/2009; CLICURB EXCL/AAG-MAA/0383/2012. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the Max Planck Institute for Meteorology for producing and making available their model output as described in Section 2. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals

6. References

DeGaetano, A. T., Shulman, M. D., 1990. A climatic classification of plant hardiness in the United States and Canada. Agricultural and forest meteorology 51 (3), 333–351.

Field, C. B., Barros, V. R., Mach, K., Mastrandrea, M., 2014. Climate change 2014: impacts, adaptation, and vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Fischer, A., Keller, D., Liniger, M., Rajczak, J., Schär, C., Appenzeller, C., 2014. Projected changes in precipitation intensity and frequency in Switzerland: a multi-model perspective. International Journal of Climatology.

Fischer, E. M., Schär, C., 2009. Future changes in daily summer temperature variability: driving processes and role for temperature extremes. Climate Dynamics 33 (7-8), 917–935.

Fovell, R. G., Fovell, M.-Y. C., 1993. Climate zones of the conterminous United States defined using cluster analysis. Journal of Climate 6 (11), 2103–2135.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al., 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. Journal of Advances in Modeling Earth Systems 5 (3), 572–597.

Giorgi, F., Francisco, R., 2000. Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. Climate Dynamics 16 (2-3), 169–182.

Lau, N.-C., Nath, M. J., 2014. Model simulation and projection of European heat waves in present-day and future climates. Journal of Climate 27 (10), 3713–3730.

Mann, H. B., Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50–60.

Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D., McDaniel, L., 2009. Relative increase of record high maximum temperatures compared to record low minimum temperatures in the US. Geophysical Research Letters 36 (23).

Pham, D. T., Dimov, S. S., Nguyen, C., 2005. Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219 (1), 103–119.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. RCP 8.5 – A scenario of comparatively high greenhouse gas emissions. Climatic Change 109 (1-2), 33–57.

Richman, M. B., Lamb, P. J., 1985. Climatic pattern analysis of three-and seven-day summer rainfall in the central United States: Some methodological considerations and a regionalization. Journal of Climate and Applied Meteorology 24 (12), 1325–1343.

Richman, M. B., Lamb, P. J., 1987. Pattern analysis of growing season precipitation in southern Canada. Atmosphere-Ocean 25 (2), 137–158.

Santos, J., Corte-Real, J., Leite, S., 2005. Weather regimes and their connection to the winter rainfall in Portugal. International Journal of Climatology 25 (1), 33–50.

Sillmann, J., Donat, M. G., Fyfe, J. C., Zwiers, F. W., 2014. Observed and simulated temperature extremes during the recent warming hiatus. Environmental Research Letters 9 (6), 064023.

Sillmann, J., Kharin, V., Zhang, X., Zwiers, F., Bronaugh, D., 2013. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. Journal of Geophysical Research: Atmospheres 118 (4), 1716–1733.

White, D., Richman, M., Yarnal, B., 1991. Climate regionalization and rotation of principal components. International Journal of Climatology 11 (1), 1–25.

Wilks, D. S., 2011. Statistical methods in the atmospheric sciences. Vol. 100. Academic press.

Wójcik, R., 2014. Reliability of CMIP5 GCM simulations in reproducing atmospheric circulation over Europe and the North Atlantic: a statistical downscaling perspective. International Journal of Climatology.

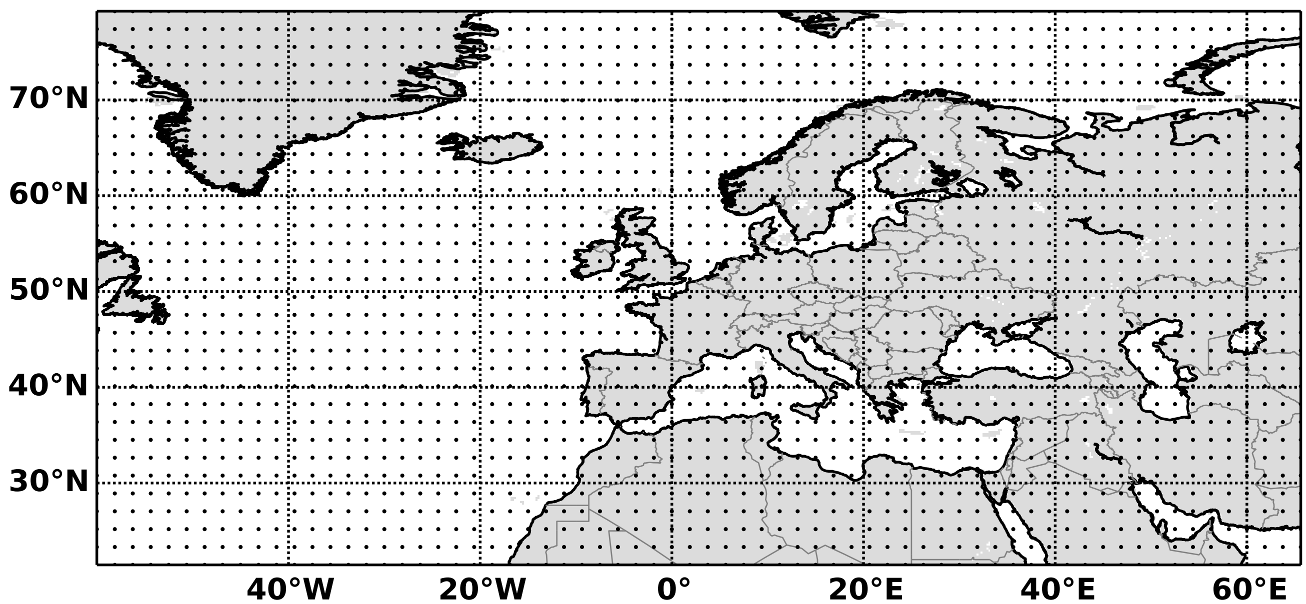

Figure 1: Study domain to which the MPI-ESM-LR Earth System Model data was cut and model grid points (1.9° horizontal resolution.

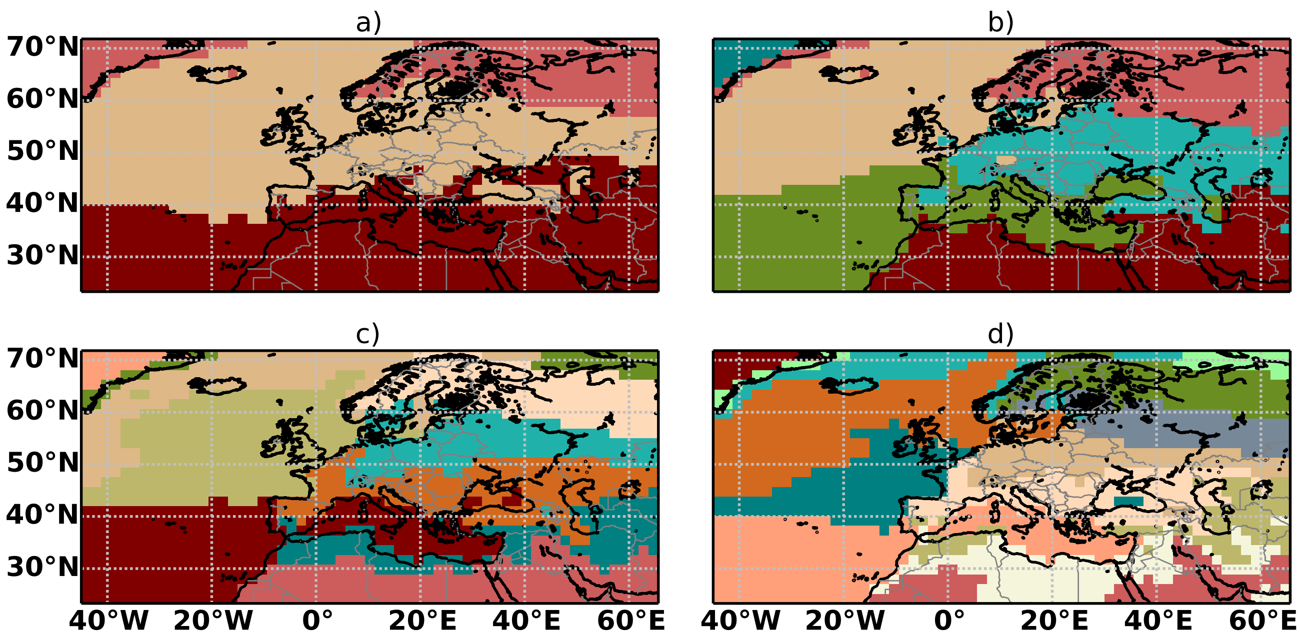

Figure 2: Regions in Europe as determined using K-Means Clustering Analysis applied to the climatology difference between the recent-past (1986 – 2005) and long-term future (2081 – 2100) using a) $k = 3$, b) $k = 6$, c) $k = 10$ and d) $k = 13$. Note that the colors are not correspondent to each other from panel to panel and that they only serving only to allow better differentiatiation between regions inside each panel.

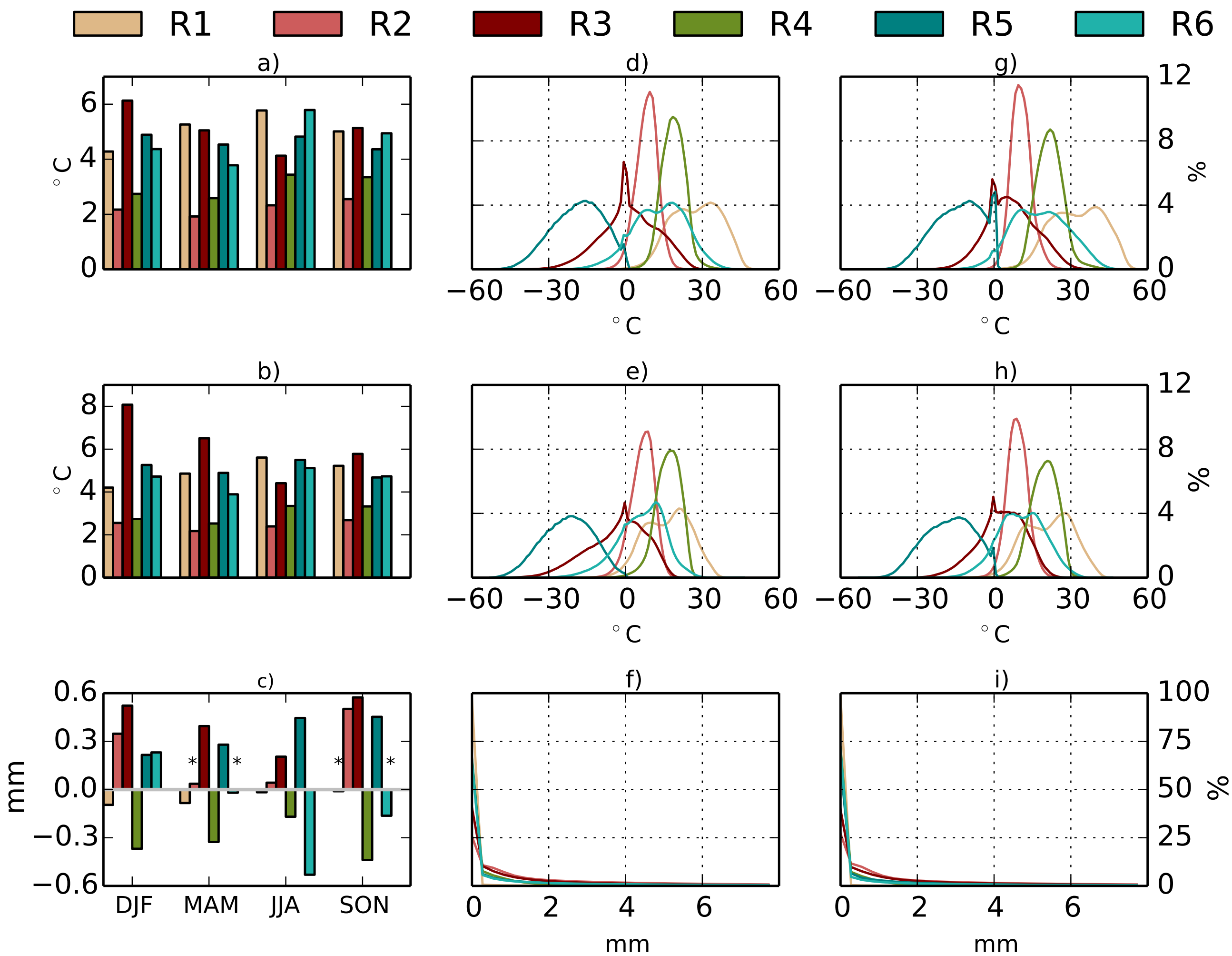

Figure 3: The seasonal average of the daily climatology difference between the recent-past (1986 – 2005) and long-term future (2081 – 2100) for a) maximum temperature, b) minimum temperature and c) daily total precipitation, for each of the six regions obtained using the multivariate K-Means clustering analysis is on the left column. Asterisks mark where the mean seasonal climatology difference is not statistically significant at the 95% confidence level. The middle and right column represent the Probability Distribution Functions of d) maximum temperature, e) minimum temperature and f) precipitation for the 1986 – 2005 and 2081 – 2100 periods respectively. These regions are color-coherent with the regions in Figure 2.b).