Constrained clustering of the precipitation regime in Greece

Eftychia Rousi*[1], Christina Anagnostopoulou†[1],
Angelos Mimis‡[2] and Marianthi Stamou§[2]

[1]Department of Meteorology and Climatology, Aristotle University of Thessaloniki, Greece
[2]Department of Economic and Regional Development, Panteion University of Athens, Greece

November 4, 2014

Summary

The aim of this paper is an objective clustering of the precipitation regime in Greece. The data consists of winter daily precipitation values obtained from a Regional Climate Model, the RACMO2/KNMI, for the period 1971-2000. The constrained clustering method is implemented by using three different linkages, single, complete and average, and for three different cluster numbers, 10, 20 and 30. Average and complete linkage both performed well, with the latter proving to be more detailed and its spatial resolution presents many similarities to the original data. The 20 and 30 clusters are clearly more representative than the 10 cluster results.

KEYWORDS: Constrained Clustering, Precipitation, Greece, Regional Climate Model.

1. Introduction

The problem of defining the various climate zones has a wide range of uses (Iyigun et al., 2013). These include the redefinition of climate zones and rainfall regimes as a result of ongoing climate changes while at the same time examining the reasons that lead to those changes. Also these have a direct effect to hydrology and flora. So the regional water management as well the farming strategies are affected.

In this context, the famous classification system of Köppen – Geiger has emerged, which was originally published by Köppen in 1918. This system provides a set of rules applied to variables derived from long term values for temperature and precipitation. In these, with several rules at hand, various locations are classified into climate types (Cannon, 2012). This rule based approach has been adopted and extended by various researchers, as for example by Thornthwaite who by following manual classifications projected the various locations into climate regions which exhibit climate homogeneity.

With the widespread use of personal computers a different approach has emerged. In this, climate classification is performed by clustering algorithms based on the assumption that areas with similar values of variables characterizing climate, such as temperature or precipitation, can be classified in the same climate type. In this way, the climate types are directly defined by the data. In this methodology, usually a two step approach is adopted. Firstly, a principal component analysis (PCA), followed by clustering analysis (CA) (Fovell and Fovell, 1993; Cannon, 2012).

Those approaches treat the spatial problem of climate zones in an aspatial way, meaning that an area

* erousi@geo.auth.gr
† chanag@geo.auth.gr
‡ mimis@panteion.gr
§ marianthi.stamou@panteion.gr

is included into a class regardless of its location. As a result, the existence of small patches of different classes within regions of a specific climate type that can be seen in Fovell and Fovell (1993) and an attempt to treat this was proposed by Pawitanand Huang (2003).

In our approach, we are extending the preliminary study of Pawitan and Huang (2003) by applying a robust constrained hierarchical algorithm and by examining the validity of the results in the wider area of Greece. Further we test the methodology in a regular grid of points and most importantly, as Fovell and Fovell (1993) encourages, data are not limited to the main land but climate characteristics in the Mediterranean Sea are also considered.

2. Materials and Methods

The data used in this study are provided by the Regional Climate Model KNMI-RACMO2 of the Royal Netherlands Meteorological Institute (van Meijgaard et al., 2008). The model has a spatial resolution of 25km x 25km and it is driven by the General Circulation Model ECHAM5/MPI. A broader area around Greece has been chosen, composed by 1064 grid points. The data consist of winter daily precipitation values over Greece for the 30 year period 1971-2000 (Figure 1).

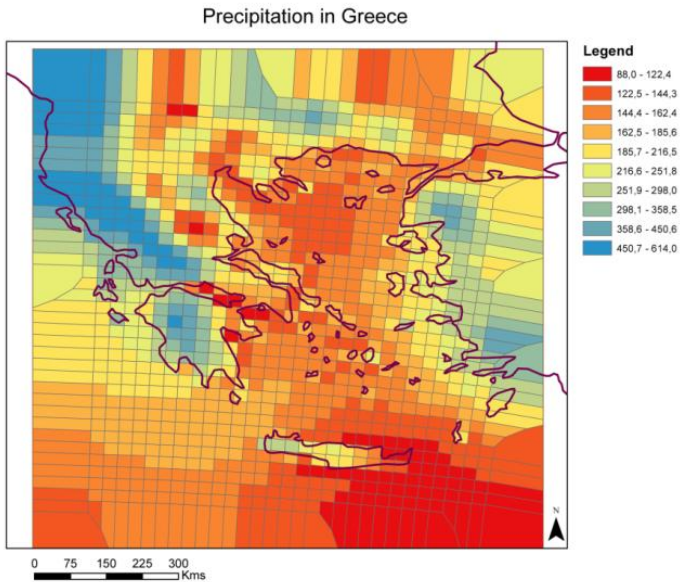

Figure 1. Choropleth map of winter daily precipitation values (in mm).

In our case a 'flavor' of agglomerative hierarchical clustering is used (Murtagh 1985). In this method, at the beginning, each area is considered as a region and at each step (iteration) the closest regions, in terms of a given metric, are merged till only one region is left. Different techniques can be devised by following different strategies to define the new distance between the newly merged and the other regions. So, by merging region R_i and R_i , the newly defined region $R_i \cup R_j$ will have distance in respect with the other regions R_k given by the Lance and Williams formula (1967):

$$\mathrm{d}(\mathrm{R_i} \cup \mathrm{R_j}, \mathrm{R_k}) = \mathrm{a_i} \cdot \mathrm{d}(\mathrm{R_i}, \mathrm{R_k}) + \mathrm{a_j} \cdot \mathrm{d}(\mathrm{R_j}, \mathrm{R_k}) + \mathrm{b} \cdot \mathrm{d}(\mathrm{R_i}, \mathrm{R_j}) + \quad (1)$$

$$+\mathrm{c} \cdot \left|\mathrm{d}(\mathrm{R_i}, \mathrm{R_k}) - \mathrm{d}(\mathrm{R_j}, \mathrm{R_k})\right|$$

where $\mathrm{d}(\mathrm{R_i}, \mathrm{R_j})$ is the distance between regions R_i and R_j in terms of the characteristics of each region which are precipitation values in our case. Also a_i, a_j , b and c are parameters whose values depend on the method (Gordon 1996). For example for a_i=a_j =1/2, b=0 and c=-1/2 you have single linkage.

Having described the hierarchical clustering techniques, we have to incorporate explicitly the spatial contiguity constraint. For that purpose a "Sorted Dictionary" structure has been adopted where we only keep the pair of regions that are contiguous and are sorted by their relative distance. Following this approach, at each step of the algorithm, we know the pair of regions having the minimum distance while satisfying the contiguity constraint. After removing from the structure the pair with minimum distance, a new region is created by merging them and the structure is updated accordingly (by using Equation 1).

In our implementation, the single, average and complete linkages are used by adopting the Lance and Williams formula. It has been developed in Python 2.7 has been imported as a script in ArcGIS.

3. Results and Discussion

The maps of the hierarchical constrained clustering method on the data are presented in Figures 2, 3 and 4, where winter daily precipitation values in Greece are illustrated for 10, 20 and 30 clusters. The average precipitation (in mm) of all grid points grouped in each cluster is shown in the legend of each figure. The daily precipitation simulated values are point data and in order to use them in the analysis, a Dirichlet tessellation was applied to define the neighbourhood structure.

Figure 2 presents the maps for 10, 20 and 30 clusters based on single linkage constrained clustering. It is obvious that this kind of linkage does not provide representative results for the precipitation regime in Greece.

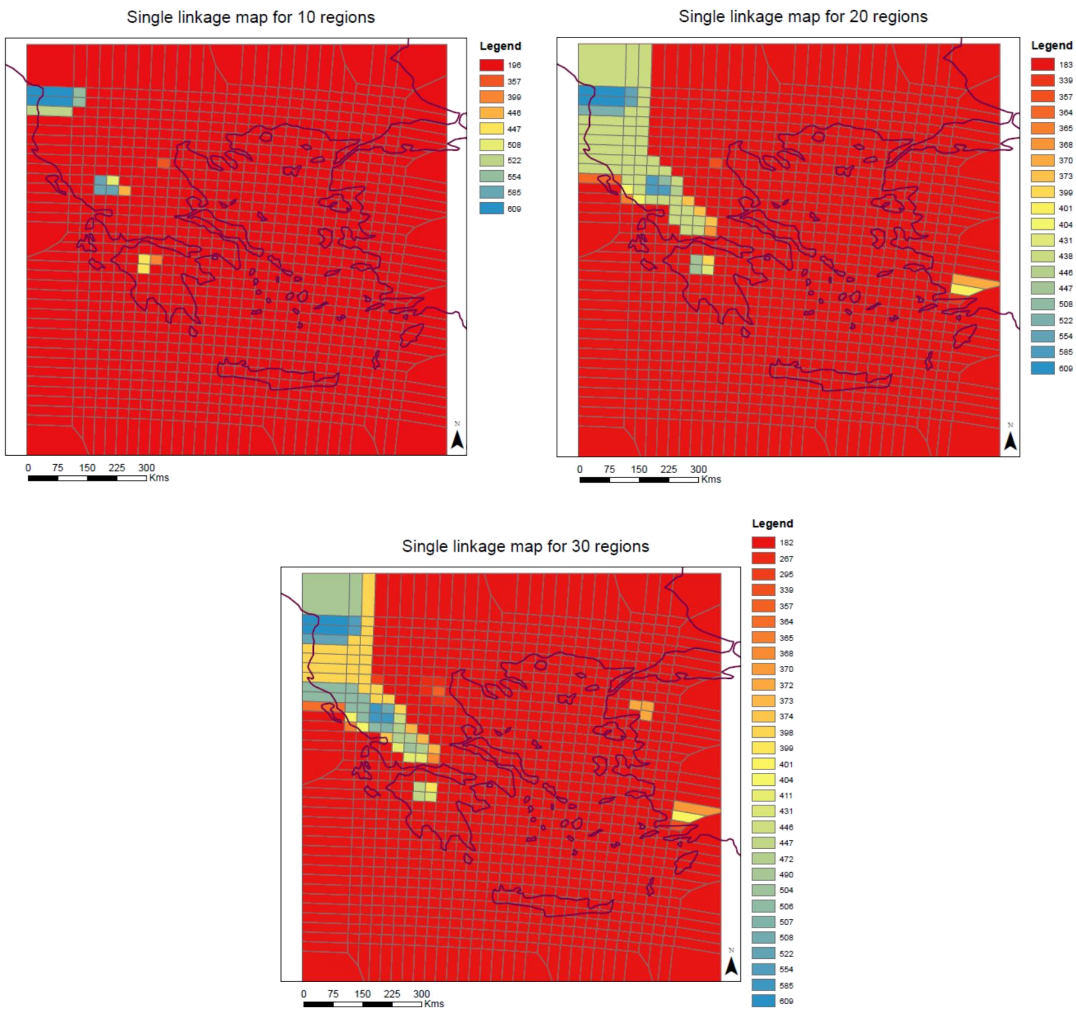

Figure 2. Single linkage constrained clustering.

Results of the average linkage for 10, 20 and 30 clusters are presented in the three maps of Figure 3. This linkage provides a better classification of the precipitation in Greece, especially for 20 and 30 clusters. In these corresponding two maps, the methodology captures not only the high precipitation amounts in the windward areas of Pindus mountains, but it also distinguishes the Olympus peak, the eastern Macedonia-western Thrace region and the increase of precipitation in the eastern Aegean and Asia Minor (Metaxas and Kallos, 1980).

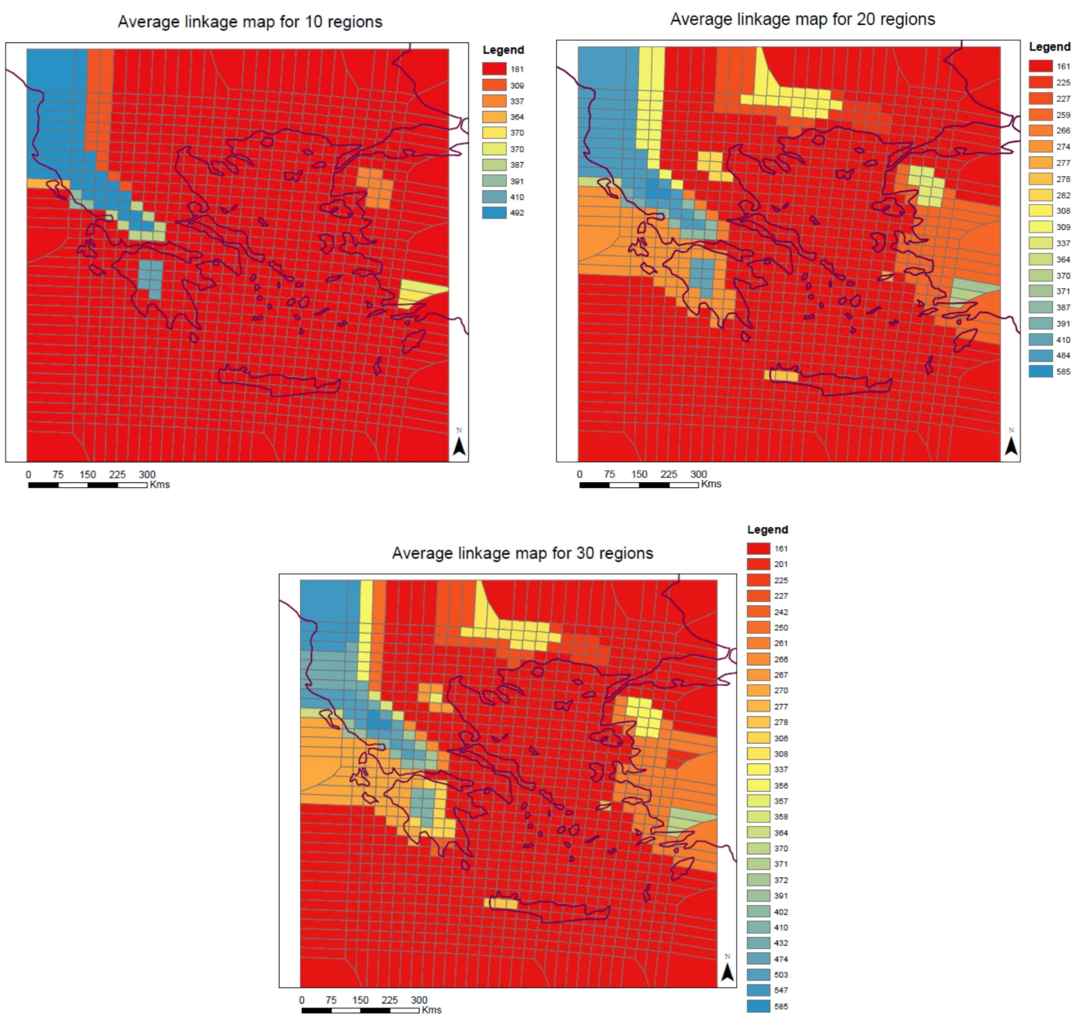

Figure 3. Average linkage constrained clustering.

Finally, the complete linkage was tested and the results are shown in the maps of Figure 4. This method seems to provide greater resemblance to the spatial distribution of the precipitation in Greece (Hatzianastassiou et al., 2008). In this case, even the 10 cluster solution gives a good image, although, 20 and 30 clusters are clearly more representative.

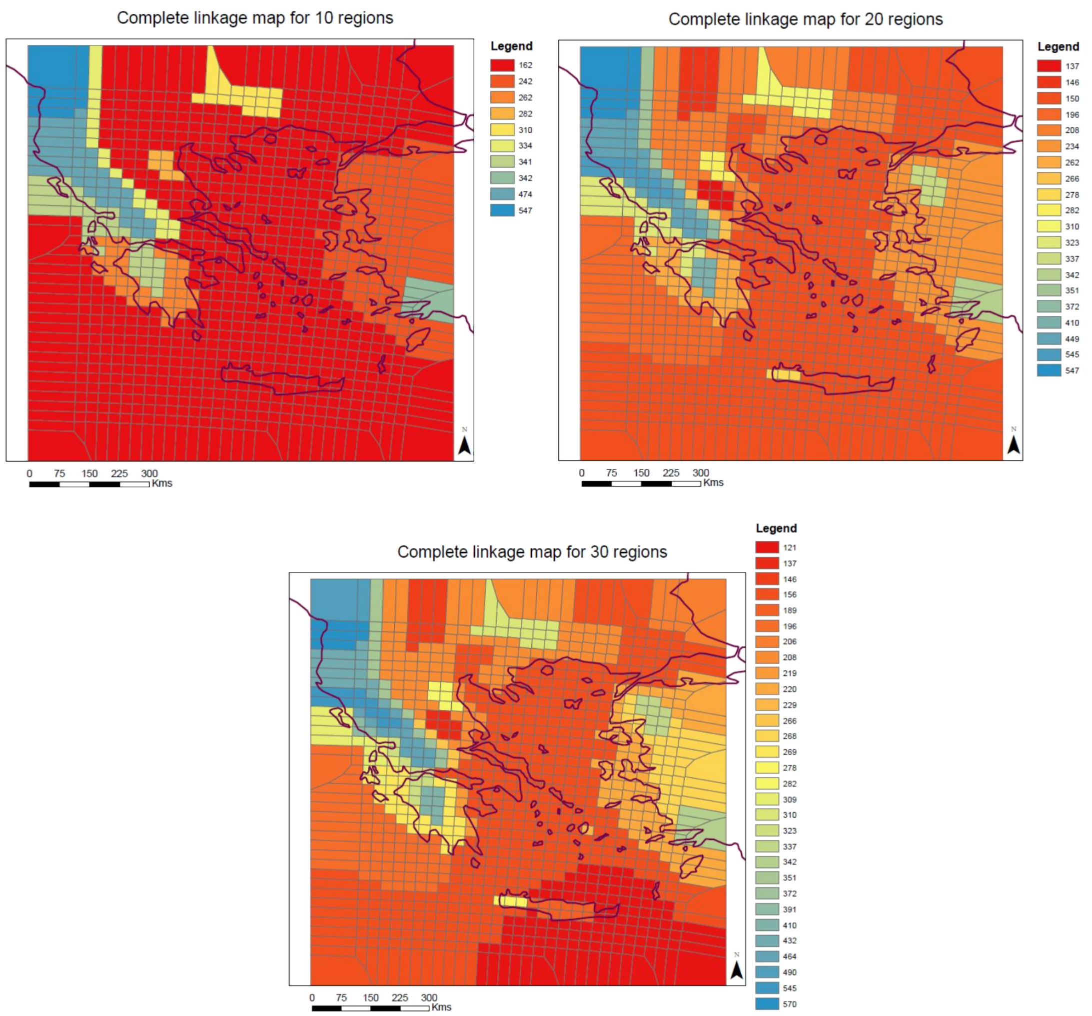

Figure 4. Complete linkage constrained clustering.

4. Conclusions

A hierarchical constrained clustering was performed in order to classify winter precipitation in Greece. Of the three different linkage methods tested, the single linkage was the one that performed poorly. Average and complete linkage both performed well, with the latter proving to be more detailed and its spatial resolution presents many similarities to the original data. Regarding the number of clusters, the 10 cluster solution is clearly not suitable for the precipitation data since it does not capture any of the data variability. On the contrary, 20 clusters could be considered good while 30 clusters do not add any value to the classification results. Future work could test the hierarchical constrained clustering on a combination of meteorological parameters that affect Greek climate.

5. Biography

Efi Rousi is a geographer with a Phd in Climatology. Her research interests include climate change, climate modelling, dynamic and synoptic climatology.

Christina Anagnostopoulou is an Assistant Professor of Climatology at the Department of Meteorology and Climatology, Aristotle University of Thessaloniki, Thessaloniki, Greece. Her research interests include synoptic and dynamic climatology, extreme weather events, methods of climate analysis, climate change.

Angelos Mimis is an assistant professor of spatial analysis in Panteion University of Athens, Greece. His interests include GIS, spatial analysis, computational geometry and optimization.

Marianthi Stamou is a PhD candidate in the Department of Economic and Regional Development, Panteion University of Athens, Greece. Her research interests include spatial econometric models and GIS.

References

Cannon A J (2012). Köppen versus the computer: comparing Köppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity. *Hydrology and Earth Systems Science*, 16, 217–229.

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

Gordon AD (1996). A survey of constrained clustering. *Computational Statistics and Data Analysis*, 21, 17-29.

Hatzianastassiou N Katsoulis B Pnevmatikos J and Antakis V (2008). Spatial and temporal variation of precipitation in Greece and surrounding regions based on global precipitation climatology project data. *Journal of Climate*, 21, 1349-1370.

Iyigun C Turkes M Batmaz I Yozgatligil C Purutcuoglu V Koc E K and Ozturk M Z (2013). Clustering current climate regions of Turkey by using a multivariate statistical model. *Theoretical and Applied Climatology*, 114(1-2), 95-106.

Lance G N and Williams W T (1967). A general theory of classificatory sorting strategies, I. Hierarchical Systems. *The Computer Journal*, 9, 373-380.

Metaxas D A Kallos G (1980). High rainfall amounts over western Greek mainland during December and January. In Proceedings of the 1st Hellenic–British Climatological Meeting, Hellenic Meteorological Society, Athens, 5–11 September, 125–137.

Murtagh F (1985). A Survey of Algorithms for Contiguity-constrained Clustering and Related Problems. *The Computer Journal*, 28(1), 82-88.

Pawitan Y and Huang J (2003). Constrained clustering of irregularly sampled spatial data. *Journal of Statistical Computation and Simulation*, 73(12), 853-865.

van Meijgaard E van Ulft L van den Berg W Bosveld F van den Hurk B Lenderink G and Siebesma A (2008).*The KNMI regional atmospheric climate model RACMO version 2.1*. Tech. Rep. TR-302, Royal Netherlands Meteorological Institute.