Segmentation of WorldClim dataset reveals new insight into spatial variability of global climate

P. Netzel, T. F. Stepinski

Space Informatics Lab, Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131, USA
Telephone: +1 513 .556.3583
Fax: +1 513.556.3370
Email: pawel@netzel.pl , stepintz@uc.edu

Abstract

We applied a segmentation algorithm to a high resolution global grid of climatic variables (WorldClim dataset) to delineate Earth's land surface into spatial climate units (SCUs) having levels of climate inhomogenity tightly controlled by algorithm's merging threshold parameter. The result is an accurate global map of spatial variability of climate. Interestingly, this variability varies greatly with geographical location. Using resultant segmentation we show that some climatic zones in the widely-used Köppen–Geiger classification (KGC) are climatically homogeneous while other are not, underscoring a qualitative character of the KGC. We attribute this result to a non-linear relationship between changes in vegetation types (on which the KGC is based) and changes in values of climatic variables. We also demonstrate the utility of climate segmentation for mapping specific climate types using the island of Great Britain as an example.

Keywords: global climate; segmentation; regional climates; Köppen climate classification

1. Introduction

The only widely-used delineation of land surface into different climatic zones is the Köppen–Geiger classification (KGC). However, the KGC can only be considered as the first-order, broad-scale qualitative overview of the spatial variability of climate. The KGC classifies climates using vegetation zones as proxies for climate; it expresses observed boundaries between different vegetation zones in terms of climatic variables. Thus the resolution of the KGC can only be as high as the resolution of the vegetations zones. Availability of high resolution world-wide grids of climatic variables makes possible to classify global climates directly without resorting to proxies such as vegetation types. Data clustering techniques were used ((Zscheischler et al. 2012; Zhang and Yan 2014; Netzel and Stepinski, 2016) to divide global climate data grid cells into clusters which are identified with different climatic zones. However, this approach did not yield much insight beyond and above of what we already knew from the KGC. This is because in the aforementioned studies the selected numbers of clusters were guided by the number of zones in the KGC. As a result, clustering-based classifications of climates are also broad-scale and, in general, similar to the KGC. Also, in clustering, the levels of clusters's inhomogeneities are not controlled, so different clusters (climatic zones) have different levels of inhomogeneity.

To get more insight into spatial variability of global climate we performed a segmentation of climatic grid. Analogous to the technique employed in image analysis, climate segmentation partitions climatic grid into multiple segments (spatial climate units or SCUs). The result is a set of SCUs that collectively cover the entire land surface. A level of dissimilarity between grid cells (local climates) in the SCU is controlled by segmentation algorithm's merging threshold parameter and is low. Adjacent SCUs have noticeably different climates. Segmentation is thus well-suited to reveal spatial variability of global climate; where climate changes on the small length-scale SCUs are small, but where climate changes on the large length-scale they are large. Juxtaposition of the KGC and segmentation maps shows a different degree of climate inhomogeneity in different KGC zones. In zones with the high degree of inhomogeneity vegetation types remain the same despite significant changes in climatic variables. In addition to revealing spatial variability of global climate, segmentation also identifies different climates in a region totally assigned to a single KGC zone. We also demonstrate an ability of segmentation to yield more specific maps of climate types using the island of Great Britain as an example.

2. Data and Methods

We use monthly sum of precipitation (*P*) and average temperature (*T*) from WorldClim (http://www.worldclim.org) 2.5 arc second (~5 km at the equator) global grid of climatic variables. This data are long term averages calculated from measurements taken between 1960 and 1990 in a world-wide network of climate stations and interpolating to the grid. Data has been normalized using procedure described in details by Netzel and Stepinski (2016).

A climate for a given grid cell, labelled *i,* is mathematically represented by a bivariate cyclic time series denoted by $C_i = \{(T_{\mathrm{i}}^1, P_{\mathrm{i}}^1), \ldots, (T_{\mathrm{i}}^{12}, P_{\mathrm{i}}^{12})\}$, where the time series progresses through 12 months. A time series representation of climate has advantage of taking into consideration month-to-month sequencing information. To measure dissimilarity, $D(C_{\mathrm{m}},C_{\mathrm{n}})$, between two climates we use a time-shift invariant version of the Euclidean distance introduced in Netzel and Stepinski (2016). This function is a computationally efficient approximation of the DTW distance (Berndt and Clifford 1994) and thus takes advantage of sequencing information in time series representing climates. Our dissimilarity function is normalized to yield values between 0 and 1. *D* facilitates a holistic comparison of two climates taking into account not only the values of T^{j} and P^{j} $j=\{1, \ldots, 12\}$ in the two locations but also their month-to-month progression over the year. *D* is designed to be small when observers at the two locations experience similar progression and character of seasons. Given the large size of the grid we use a modification of the fast scanning segmentation algorithm (Ding et al., 2009). The algorithm has a single parameter – a merging threshold which limits a level of allowable climate inhomogeneity in each segment.

3. Results

Running segmentation algorithm with the merging threshold set to 0.2 we obtained 20,488 segments (SCUs) ranging in size from 5 km^2 to 10,995,775 km^2; the mean area of a SCU is 6687 km^2. Many small segments are islands or are located in the mountainous areas. The average segment's internal dissimilarity is 0.11 (standard deviation is 0.06) and the average segment's external dissimilarity (mean dissimilarity between a focus segment and its neighbouring segments) is 0.37 (standard deviation is 0.1). This indicates a good quality of segmentation as an average segment is almost four times more cohesive than its neighbourhood.

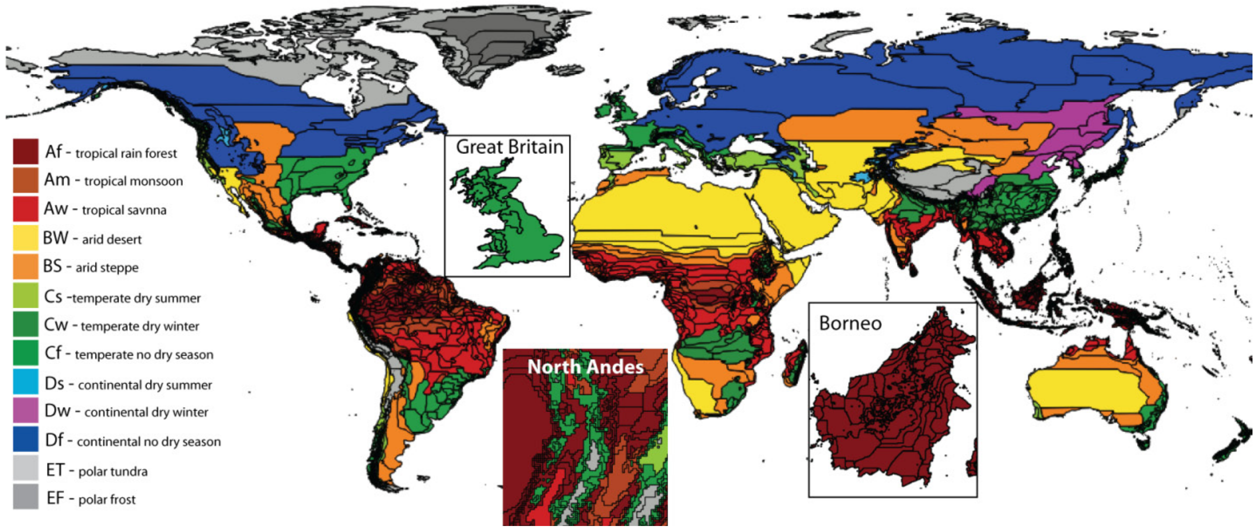

Figure 1. Boundaries of spatial climate units (SCUs) are shown by black lines. SOCs are classified using KGC; their classes are shown in colors according to the legend.

Threshold 0.2 means that dissimilarity between cells in a segment is smaller than 0.2. Because our dissimilarity measure is holistic, the meaning of the value of 0.2 can best be appreciated by giving an example: *D*(London, city) < 0.2 for Oxford, Cambridge, York. For those familiar with climates in these cities this example gives an idea of restriction on variability of climates within a segment. Others can explore similarity/dissimilarity of climates worldwide using our online application ClimateEx (http://sil.uc.edu/webapps/climateex/).

Fig.1 shows the global segmentation of climatic grid. SOCs are also classified according to KGC (using their centroids). Spatial variability of climate is indicated by density of SOCs, which is the highest in the mountainous areas (Himalaya, Andes, etc) and in the tropical regions. It is the lowest at the high northern latitudes and in the deserts. A mismatch between SOCs and KGC is most pronounced for zones Af, Am, and Aw and also present for the Cf. Zones BW, BS, and Df show the least amount of mismatch with SOCs. Thus, high climate variability in tropical zones (due to the wide range of precipitation) is not captured by the KGC vegetation-based rules. Desert and continental zones have less variation in climatic variables and fit more closely to SOCs.

Fig. 2 shows boundaries of SCUs in Great Britain, an island that is classified by the KGC as belonging in its entirety to the Cf zone. The background is the map of *D*(cell, London), a dissimilarity between a climate at a given location and the climate in London. Such map indicates climate variability at the cell level. Clearly, the boundaries of SCUs enclose locations with similar values of *D*(cell, London). For the nine largest SCUs in Great Britain we show climatograms of their centroids. They show variety of climates, with the discriminant being the amount of precipitation. The nine SCUs can be further classified on the basis of similarities of their climatograms; we choose to classify them into three classes as shown in the inset to Fig.2. The first climate class includes a single SCU located along the eastern part of the island; compared to the rest of the island it is characterized by lower precipitation. The second class include three SCUs, two in Scotland and one in Wales; it is characterized by the largest amount of precipitation. The third class, which includes five SCUs, represents a transition between the first two.

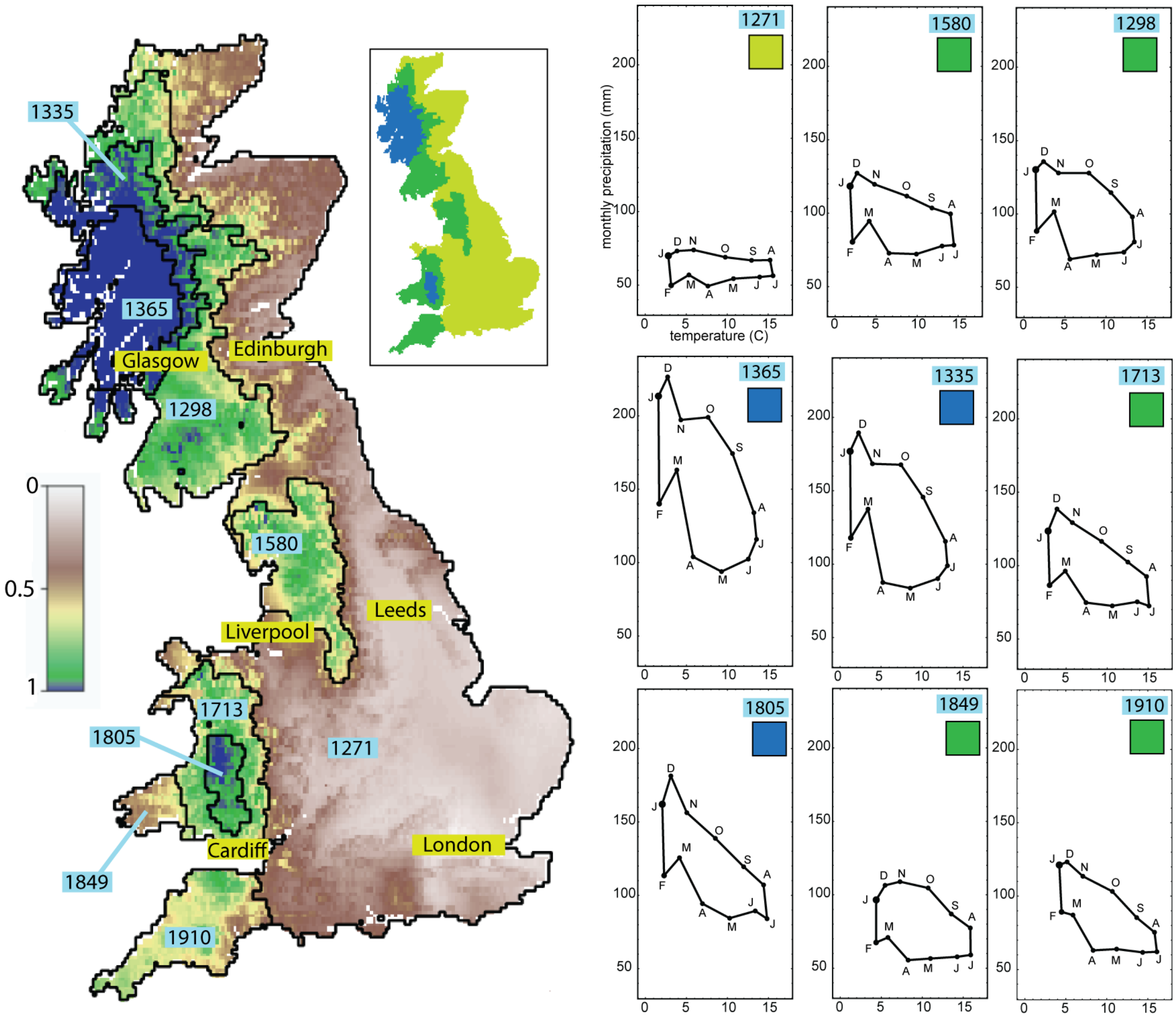

Figure 2. (Left) Boundaries of SCUs in Great Britain overlaid on the map of climate dissimilarities from the climate in London; inset shows a classification of SCUs into three climate types. Numbers of nine largest SCOs are given in blue. (Right) Climatograms for the nine largest SCOs.

4. Conclusions

One could expect that spatial variability of global climate is a problem that has been already sufficiently addressed, but it hasn't probably because of overconfidence in the KGC. Availability of global, high resolution climatic grids did not change the situation because of the focus on "reproducing" the KGC using data clustering methods. Our thesis is that grid segmentation rather than clustering is the most fruitful data analysis method to get new insight into spatial variability of global climate. The key to this advantage is its tight control over inhomogeneity of segments, a property that clustering lacks. Thus, spatial distribution of climate variability is revealed by just displaying a map of SCUs (Fig.1). Using climate segmentation we were able to understand limitations of the KGC. KGC has been relying on the assumption that vegetation is a good proxy for climate. Our results show that this may be true but only up to the point. A mismatch between our segmentation and KGC in troopics can be explained by findings (Schuur, 2003) that the value of net primary productivity (NPP) starts to decline at high precipitation (> 2400 mm/year) in tropical ecosystems thus decoupling the variation in climate from variation in vegetation and undermining the principle of the KGC. Does it matter? Conceptually, it certainly does. In practise, the KGC is most often utilized in the fields of agriculture, ecology, or forestry, all concerned with vegetation zones rather than climate zones per se. Segmentation of global climate bring attention to a difference between climatic and vegetation zones which is too often blurred.

A segmentation of global climate is a showcase of the methodology, but we also have shown, using the Great Britain as an example, that segmentation offers a principled and quantitative means for delineation of areas with different climates on a scale of a single country. On such scale segmentation may be used to delineate individual mesoclimates if weekly or daily climatic data are available at sufficiently high spatial resolution.

6. Acknowledgements

This work was supported by the University of Cincinnati Space Exploration Institute and by the grant NNX15AJ47G from the National Aeronautics and Space Administration.

7. References

Berndt, DJ, Clifford, J, 1994, Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD workshop*, vol. 10(16): 359-370.

Ding, JJ, Kuo, CJ, Hong, WC, 2009, An efficient image segmentation technique by fast scanning and adaptive merging. In *proceedings of CVGIP 2009 Conference*.

Netzel, P, Stepinski, TF, 2016, On using a clustering approach for global climate classification. *J. Climate* 29(9), 3387–3401.

Schuur, EAG, 2003, Productivity and global climate revisited: the sensitivity of tropical forest growth to precipitation. *Ecology*, 84(5), 1165-1170.

Zhang, X, Yan, X, 2014, Spatiotemporal change in geographical distribution of global climate types in the context of climate warming. *Climate Dynamics*, 43(3–4), 595–605.

Zscheischler, J, Mahecha, MD, Harmeling S, 2012 Climate classifications: the value of unsupervised clustering. *Procedia Computer Science*, 9: 897-906.