JOURNAL OF CLIMATE

On using a clustering approach for global climate classi cation

PAWEL NETZEL AND TOMASZ STEPINSKI*

Space Informatics Lab, Department of Geography, University of Cincinnati, Cincinnati, OH, USA

ABSTRACT

Classifying the entire land surface into different climate types provides a convenient means of diagnosing the existence of spatial relations between Earth s various physical and biological systems and the climate. Global climate classifications are also used to visualize climate change. Clustering of climate datasets seems like a natural approach to climate classification but instead it is the Köppen-Geiger classification (*KGC*) that is the most widely used. Here, we present a comprehensive approach to the clustering-based classification of climates. Going beyond previous research, we define local climate as a multivariate time series of mean-monthly climatic variables and propose to use Dynamic Time Warping (*DTW*) as a measure of dissimilarity between local climates. We also discuss the choice of climatic variables, the importance of their proper normalization, and point out the advantage of using distance-based clustering algorithms. Using the WorldClim global climate dataset and different combinations of clustering parameters we calculate 32 different clustering-based classifications. These classifications are compared between themselves and to the *KGC* using the information-theoretic V-measure. We find that the best classifications are obtained using three climate variables (temperature, precipitation, and temperature range), a data normalization that takes into account the skewed distribution of precipitation values, and the Partitioning Around Medoids clustering algorithm. We compare in detail two such classifications both to each other and to the *KGC*. About half the climate types found by clustering can be matched to the familiar *KGC* classes but the rest differ in their climatic character and spatial distribution. Finally, we demonstrate that clustering-based classification results in climate types that are internally more homogeneous and externally more distinct than climate types in the *KGC*.

1. Introduction

Global climate classification schemes aim to identify distinct climate types and map their geographical extents. By discretizing a multitude of local climates (LCs) into a manageable number of climate types (CTs, list of all acronyms is given in Table 1.) classification simplifies the spatial variability of climates into a form that is more meaningful and easier to analyze. Thus, climate classification provides intuitive and valuable insight into the relationships between climate and Earth s physical and biological systems, such as erosion (Peel et al. 2001), soils (Rohli et al. 2015), the biota (Baker et al. 2010; Garcia et al. 2014), as well as distributions of invasive species (Werier and Naczi 2014) and virus vectors (Brugger and Rubel 2013). Climate classification is also used to provide visualization of global climate datasets (Fraedrich et al. 2001; Diaz and Eischeid 2007; Zhang and Yan 2014; Chen and Chen 2013; Spinoni et al. 2015) in order to illustrate climate change in terms of shifting geographical boundaries of major climate types. Similarly, it is used to visualize future spatial distributions of climate types (Beck et al. 2005; Gallardo et al. 2013; Hanf et al. 2012; Mahlstein et al. 2013) as predicted by climate models. Some studies (Rubel and Kottek 2010; Feng et al. 2014) applied climate classification to a combination of historical data and model predictions to illustrate climate shifts over the longer time periods. Finally, climate classification was used to interpret the results of models designed to simulate paleoclimates (Guetter and Kutzbach 1990).

From a methodological point of view, widely used global climate classifications (Köppen 1936; Thornthwaite 1948; Trewartha and Horn 1980) are heuristic schemes re ecting environmental and geographical knowledge accumulated over decades of research. In particular, the Köppen-Geiger classification (*KGC*) scheme (Köppen 1936) has become a de facto standard for global climate classification especially as its modern implementations (Kottek et al. 2006; Peel et al. 2007; Spinoni et al. 2015) allow for convenient mapping of CTs from climatic data collected from an extensive, world-wide network of weather stations.

Despite its popularity the *KGC* has a number of shortcomings, the chief among them being the core methodology itself. The *KGC* is based on the assumption that

*Corresponding author address: Space Informatics Lab, Department of Geography, University of Cincinnati, Cincinnati, OH 45221, USA

E-mail: stepintz@uc.edu

TABLE 1. **List of acronyms used in the paper**

LC	local climate
CT	climate type
KGC	Köppen-Geiger classification
LTMM	long-term monthly-means
PCA	principal components analysis
DTW	dynamic time warping dissimilarity function
EUC	Euclidean dissimilarity function
HC	hierarchical clustering
PAM	partitioning around medoids

delineation of CTs can be guided by the extents of different plant regions (Thornthwaite 1943) by expressing their boundaries in terms of temperature and precipitation. It is classification by a hierarchy of predicate statements (Spinoni et al. 2015) that assigns a class to an LC on the basis of the values of long-term monthly-means (LTMM) of temperature and precipitation. This system lacks the notion of similarity between LCs (see section 3c for elaboration) making it impossible to assess natively the uniformity of climates within a given CT. Similarly, it lacks the notion of similarity between CTs beyond organizing them into a hierarchy. In addition, the CTs are permanently set by the *KGC*. This is a potential issue when using climate classification to visualize global climate change. Applying the *KGC* to the results of climate models to map future climatic zones does not account for the possibility of an emergence of new CTs.

In this paper we investigate a clustering approach to the problem of global classification of climates. A clustering process groups LCs into clusters based on their mutual similarities using an automatic (unsupervised) algorithm. We associate these clusters with CTs. CTs are "discovered" by an algorithm on the basis of what s in the data without any prior assumptions about their expected character and/or location. A clustering approach is feasible due to the availability of extensive, global LTMM climatic datasets (Hijmans et al. 2005; Harris et al. 2014). Each dataset record is used to describe an LC at the location of a station, or at a grid cell if the data is gridded. The degree of dislikeness between a given pair of LCs is measured by a dissimilarity function. LC representation and the dissimilarity function are also used to assess the climatic uniformity of any CTs regardless of whether they originated from clustering or were delineated by the *KGC*.

Using clustering to delineate climatic zones has been proposed previously, albeit mostly in a regional rather than global context (Stooksbury and Michaels (1991); DeGaetano (1996); Bunkers et al. (1996); Fovell and Fovell (1993); Unal et al. (2003). All these early studies used very similar techniques – representing LC as a vector of LTMM of climatic variables, using the Euclidean distance as the dissimilarity function between LCs, and applying the hierarchical clustering algorithm to obtain the set of CTs. Most studies also used the Principal Component Analysis (PCA) to reduce the dimensionality of vectors representing LCs.

More recently, with the increased availability of climate data from world-wide networks of stations, a clustering methodology has been applied to global classification of climates (Zscheischler et al. 2012; Zhang and Yan 2014; Metzger et al. 2012). The clustering techniques used in these studies follow the methods applied to regional classifications. LCs are represented by vectors, with Zhang and Yan (2014) using LTMM for temperature and precipitation, Zscheischler et al. (2012) using LTMM of three remotely sensed indices including two vegetation indices in addition to LTMM for temperature and precipitation, and Metzger et al. (2012) using a vector of 42 bioclimatic variables. All three studies use the Euclidean distance and various versions of *K*-means clustering to obtain CTs.

All previous clustering-based classifications of climates studies used generic, off-the-shelf techniques without taking into account the specificity of climate data. In this paper we revisit this problem by introducing climate-data-specific modifications to the clustering procedure. Because LC is an intra-annual pattern of weather conditions at a given location we propose to represent it as a cyclic time series of local climatic variables rather than as a feature vector of these variables. A time series representation takes into consideration month-to-month sequencing information which the vector representation lacks; it corresponds more closely to the human perception of climate. To take advantage of sequencing information we also propose to use the Dynamic Time Warping (*DTW*) distance (Berndt and Clifford 1994) rather than the Euclidean distance as a dissimilarity function. To account for the cyclic nature of climate we use a version of *DTW* designed for cycling time series. We submit that LCs, as represented by time series, should not be averaged because the mean may not re ect correctly properties of the set of LCs from which it was derived. Consequently, our analysis is performed in a distance space (Ganti et al. 1999) rather than in a more common feature space; the only allowable operation on a pair of LCs is the calculation of their dissimilarity value.

Using our enhancements to the clustering method we calculate and compare 32 different climate classifications obtained using different clustering protocols. The goal is to determine which elements of the clustering process (the choice of variables, the choice of normalization procedure, the choice of dissimilarity function, and the choice of clustering algorithm) have the greatest in uence on the result and to select the preferred classifications. Then, the two preferred classifications (one using the *DTW* and another using the Euclidean distance as the dissimilarity function) are compared to each other and to the *KGC* from the perspective of how they partition the land surface and the character of their CTs. Finally, the *DTW*-based and *KG*

classifications are examined for spatial and climatic inhomogeneities of their CTs.

2. Data and Methods

a. Data, variables, and normalization

We use the WorldClim global climate dataset (Hijmans et al. 2005). WorldClim data is given on a 30 arc. sec. grid and has the spatial extent of (180W, 180E) - (90N, 60S). The grid cells contain mean-monthly climatic variables interpolated from a meteorological time series measured from a world-wide network of meteorological stations between 1950 and 2000. We use mean-monthly values of the following variables: temperature, denoted by T, which is a measure of average thermal conditions; precipitation, denoted by R, which is a measure of climate humidity; and temperature range, $D = T_{max} - T_{min}$, which is a measure of thermal condition variability. T_{min} and T_{max} are mean-monthly values of minimum and maximum temperature, respectively.

We reprojected the WorldClim grid to the Mollweide projection which is a near-equal area global projection (Usery and Seong 2001). We require an equal area projection for evaluating similarity between different classifications (Cannon 2012). We then resampled the Mollweide grid to the resolution of 75 km $\times$ 75 km per cell resulting in a 213 $\times$ 482 grid of which 23,979 cells represent land surface and the rest represent water. We use such a relatively coarse grid due to computational considerations. Because we work in the distance space we need to generate and work with dissimilarity matrices having a size equal to the square of the number of cells. However, as our goal in this paper is to compare different clustering-based classifications rather than to produce the most accurate map of climate types, this resolution is sufficient.

Climatic variables have different meanings and different ranges of values. In order to contribute equally to the value of dissimilarity between two LCs they need to be scaled to have identical ranges (using normalization, as performed by Zhang and Yan (2014)) or, at least, similar ranges (using z-score standardization, as performed by Zscheischler et al. (2012)), otherwise the value of dissimilarity would be over-in uenced by the variable with the largest range. We use two different normalization transformations. The first is the standard normalization $X_i \leftarrow (X_i - \text{Min } X_i)/(\text{Max } X_i - \text{Min } X_i)$ where X_i are climatic variables. We refer to this type of normalization as "global" or (g). However, we note that precipitation variable, with a range of 0mm to 1550mm, has a distribution which is highly skewed toward large values. This means that an overwhelming number of normalized values for precipitation will be very small. As a result the in uence of precipitation on the overall dissimilarity between LCs would be artificially diminished. Thus, we introduced a second normalization procedure referred to as "modified" or (l). This procedure transforms variables T and D according to the normalization formula as given above, but the variable R is transformed as follows:

$$R \leftarrow \begin{cases} \frac{R}{350} & \text{if } R \leq 350 \\ 1 & \text{if } R > 350 \end{cases} \quad (1)$$

Thus, the top 1% of precipitation values, those in a range of 350mm to 1550mm, are all transformed to the value of 1 while the 99% of precipitation values, those in a range between 0mm and 350mm, are transformed to values between 0 and 1. This has the effect of restoring the in uence of precipitation on the overall dissimilarity between LCs while preserving 99% of precipitation data unmodified.

b. Climate representation, dissimilarity functions, and clustering methods

We represent local climate as a trivariate (bivariate if only T and R are used) cyclic time series. Thus, a LC at location i is given as $\text{LC}_i = \{M_i^1, \ldots, M_i^{12}\}$, where the time series progresses through 12 months $M_i = (T_i^1, R_i^1, D_i^1)$, $i = \{1, \ldots, 12\}$, from January to December. We use multivariate time series representation of an LC to account for the interactions and comovements between the three (or two) climate variables.

We utilize two different dissimilarity functions appropriate for multivariate time series; (1) time-shift invariant versions of the Dynamic Time Warping $d_{tsi}^{DTW}(\text{LC}_1, \text{LC}_2)$ and (2) time-shift invariant version of the Euclidean distance $d_{tsi}^{EUC}(\text{LC}_1, \text{LC}_2)$. To calculate d_{tsi}^{EUC} the standard Euclidean distance, $\sqrt{\sum_{k=1}^{12} d^2(M_1^k, M_2^k)}$, between LC_1 and LC_2 is calculated twelve times. During this calculation the time series representing LC_1 is kept unchanged whereas the time series representing LC_2 undergoes a cyclic shift in the sequence of months. The $d_{tsi}^{EUC}(\text{LC}_1, \text{LC}_2)$ is the minimum of the twelve calculated values. This ensures that the dissimilarity between two LCs is independent from time shift in their seasons.

Our preferred dissimilarity function is Dynamic Time Warping (*DTW*) (Berndt and Clifford 1994). *DTW* is widely used for calculating dissimilarity between two time series. The difference between *DTW* and Euclidean distance is that *DTW* allows non-linear alignments between two time series to accommodate sequences that are similar, but locally out of phase (Rabiner and Juand 1993; Nafiz 2005)). Fig. 1 illustrates this difference; two local climates (for Dallas and Los Angeles) are shown as bivariate (T and R) monthly series. Standard Euclidean distance aligns the two series month-to-the-same-month as shown by dashed lines in Fig. 1A1. This corresponds to taking a diagonal (possibly sub-optimal) path in the matrix of distances between two series as shown in an Fig. 1A2. The matrix of distances consists of Euclidean distances between each possible pair of months. *DTW* calculates distance between two series using an optimal (resulting in the

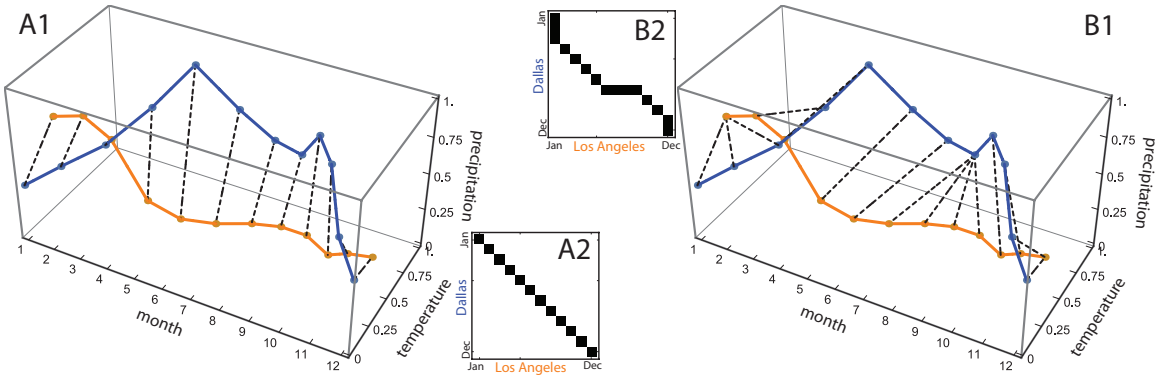

FIG. 1. Comparison between *DTW* and Euclidean distances using bivariate (T and R) time series for Los Angeles (orange) and Dallas (blue). (A1) Dashed lines show an alignment of pairs used by Euclidean distance. (A2) A sub-optimal path through the distance matrix corresponding to Euclidean distance. (B1) Dashed lines show an alignment of pairs used by *DTW*. (B2) An optimal path through the distance matrix corresponding to *DTW*.

minimum value of the overall distance) path through the matrix of distances (Fig. 1B2). Dashed lines in Fig. 1B1 shows the alignment between the two time series resulting from the optimal path. In our calculations we use a cyclic version of *DTW*, similar to the one described in Nafiz (2005), which in addition to non-linear alignment also uses the same minimization over the cyclic shifts of months as described above for the d_{tsi}^{EUC}.

As we work in the distance space the first step for any clustering-based classification is the calculation of dissimilarity between all pairs of LCs resulting in a 23,979 $\times$ 23,979 dissimilarity matrix. A separate dissimilarity matrix needs to be calculated for all combinations of the choice of normalization, number of variables, and the choice of dissimilarity measure. To obtain a classification we use two popular clustering algorithms that take a dissimilarity matrix as their only input. The first is a hierarchical clustering (*HC*) with Ward linkage (Ward 1963) and the second is the Partitioning Around Medoids (*PAM*) algorithm (Kaufman and Rousseeuw 1987). We use implementations of these algorithms in the R software environment.

c. *Classification evaluation methods*

We use two different evaluations of climate classifications, one involves evaluating the degree to which two classifications result in similar spatial partitioning, and the other evaluates the clustering quality of a single classification.

First, we want to quantify the degree to which two different classifications partition the world into similar climatic zones. For this purpose we use an information theoretic index called the V-measure (Rosenberg and Hirschberg 2007). Fig. 2A illustrates the principle of V-measure using the specific example of *KG5* and *TRDl DTW PAM5* classifications (see the next section for explanation of classifications naming convention). The *KG5* (*KG* for short) partitions the world into five climate classes; spatial extents of these classes are shown by non-gray areas in the left column of Fig. 2A. The *TRDl DTW PAM5* (*DTW* for short) partitions the world into five climate types; spatial extents of these types are shown by non-gray areas in the right column of Fig. 2A. We observe that an area of each class intersects multiple types as indicated by type-specific colors. For each class the entropy of a histogram of it constituent types measures a level of class homogeneity with respect to types. Homogeneous classes (like *A*) have small values of entropy and inhomogeneous classes (like *D*) have large values of entropy. The ratio of class entropy to the entropy of the entire *DTW* partition of the world indicates how much more uniform the distribution of types in a given class is with respect to the entire world. The area-weighted average of such ratios is a measure (the smaller the better) of homogeneity of *KG* classes with respect to *DTW* types. Reversing the roles of classes and types (right column in Fig. 2A) we calculate a completeness of *KG* classes with respect to *DTW* types. The V-measure is the harmonic mean of homogeneity and completeness. Note that V-measure is symmetric with respect to the partitions.

Second, we want to evaluate the quality of a single classification. Here, we assume that a classification is of good quality if LCs within each CT are all highly similar and CTs are highly dissimilar from each other. Thus, we need to treat all classifications as clusterings (even the *KGC*) and perform an internal evaluation of these clusterings using the Davies–Bouldin index *DB* (Davies and Bouldin 1979). The smaller the value of *DB* the better the quality of the classification. Note that clustering quality is only one of many criteria we use to evaluate classifications, however it is the only one which is quantitative. As the value

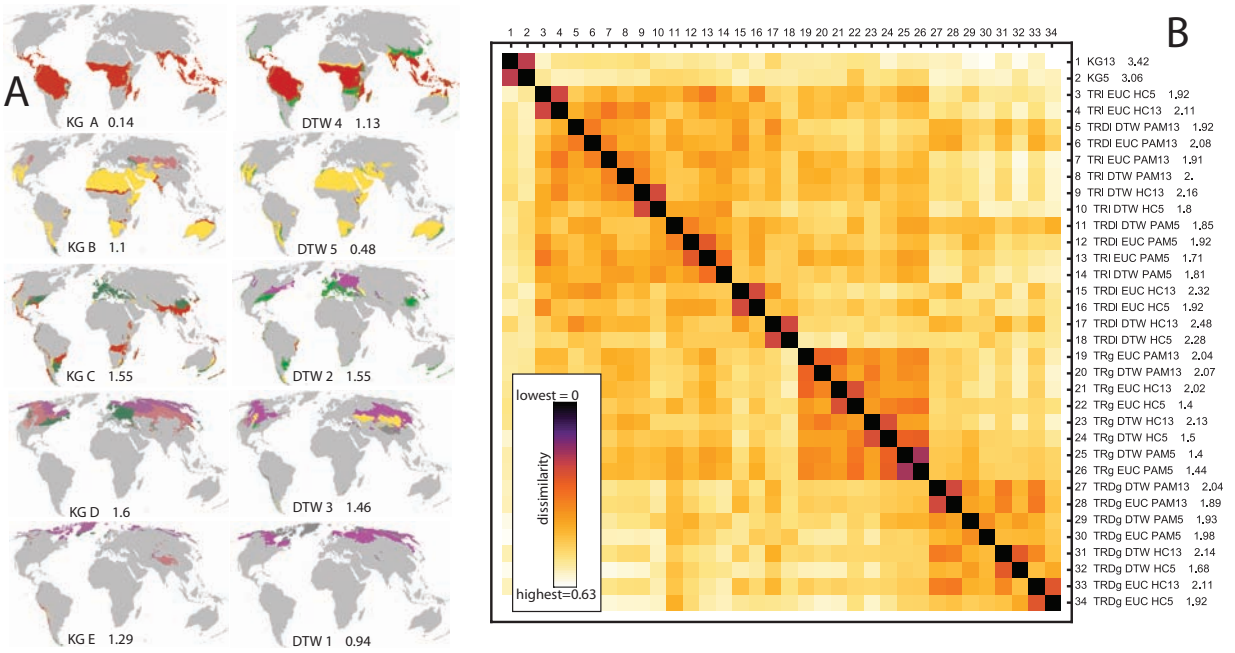

FIG. 2. (A) Illustrating the concept of V-measure. The right column shows spatial extents of *KG5* climate classes and how they intersect *TRDI DTW PAM5* climate types. The left column shows spatial extents of *TRDI DTW PAM5* climate types and how they intersect *KG5* climate classes. The value of entropy is given for each class and type. See Fig. 3 for legends linking colors to specific climatic classes and types. (B) Heat map illustrating the V-measure-based comparison between 34 different climate classifications. The black-to-white color gradient indicates dissimilarities between pairs of classifications from small to large. Note the significant degree of dissimilarity between the clustering-based classification and the *KGC*.

of clustering quality is dependent on the definition of the dissimilarity function, it is possible, in principle, to get a bad classification with high clustering quality if an inappropriate dissimilarity function is used.

3. Results

The results are grouped into three parts: (1) comparison of 32 different clustering-based classifications stemming from different choices of free parameters in the clustering process, (2) detailed comparison of the two preferred classifications between themselves and the *KGC*, and (3) examination of climate inhomogeneities within CTs.

a. Comparison of clustering-based classifications

There are five different parameters in the clustering procedure.

- The number of climatic variables used to describe an LC, either two variables (T and R) or three variables (T, R, and D).
- The method of variable normalization, either global (g) or modified (l).
- The choice of a dissimilarity function, either *DTW* or the Euclidean distance.
- The choice of a clustering algorithm, either *HC* or *PAM*.
- The number of climate types (clusters), either 5 or 13 to match the number of classes in the first two levels of the *KGC*.

Determining the optimal number of clusters directly from the data is possible but was not attempted because the focus was on comparison to the *KGC*. Altogether there are 32 combinations of these parameters resulting in 32 possible protocols for clustering procedure and leading to 32 different classifications.

Fig. 2B shows the results of the comparisons between 34 climate classifications (the *KGC* for 5 and 13 CTs are also included) using the values of the V-measure as the basis for comparison. The graph in Fig. 2B is a heat map (Wilkinson and Friendly 2009) – a graphical representation of the V-measure-based dissimilarity matrix. Darker colors indicate more similar classifications with black corresponding to a dissimilarity equal to 0 (identical partitionings) and the lightest color corresponding to the largest dissimilarity equal to 0.63 (between *KG5* and *TRDg EUC HC5*). Notice that the largest dissimilarity is still significantly smaller than the absolute upper limit of 1, thus all classifications have some level of spatial correspondence

to each other, but some more so than the others. Classifications are numbered from 1 to 34, the first two being the *KGCs* with 13 and 5 CTs, respectively. The remaining classifications are labeled to indicate the choice of free parameters used, for example, *TRDg DTW PAM13* indicates classification obtained using variables T, R, D, global normalization, DTW dissimilarity function, PAM clustering algorithm, and 13 clusters.

From examining Fig. 2B our first observation is that none of our clustering-based classifications delineate CTs in close spatial correspondence with the *KGC*. The second observation is that for any given choice of variables, normalization, and dissimilarity function, the classifications with 5 and 13 CTs are similar if obtained using hierarchical clustering. This is an expected result as hierarchical clusterings (and the *KGC*) subdivide a more broadly-defined CT into constituent, more narrowly-defined CTs resulting in a high spatial correspondence between partitionings. The third observation is that when using the global (g) normalization the classification depends mostly on the number of variables, with lesser dependence on the choice of dissimilarity function and clustering algorithm. This is because global normalization reduces the contribution of precipitation (R) to an overall value of dissimilarity between two LCs. Thus, a two-variable classification is predominantly a temperature classification while a three-variable classification delineates land surface differently as it also depends on temperature range (D).

For classifications derived using the modified (l) normalization there is no clear pattern of similarities (apart from the coupling of corresponding classifications delineated using hierarchical clustering). This indicates that variable normalization is the most important parameter of the clustering protocol; using improper normalization (like the g normalization) will lead to improper classifications no matter what the other parameters are. The heat map in Fig. 2B does not offer further insights on the relative importance of the remaining parameters. To proceed we visually assessed the 16 classifications calculated using the l normalization procedure for their relative merits. In our judgment the classifications using three variables are preferred over classifications using only two variables and those obtained using the *PAM* clustering algorithm are preferred over those obtained using the *HC* algorithm. Consequently we selected *TRDl DTW PAM* and *TRDl EUC PAM* classifications for further, more detailed comparison.

b. Comparison between TRDl DTW PAM, TRDl EUC PAM, and KG classifications

A more detailed comparison between the two selected clustering-based classifications, *TRDl DTW PAM* (hereafter referred to as *DTW5* or *DTW13* depending on the number of CTs), *TRDl EUC PAM* (hereafter referred to as *EUC5* or *EUC13*), and *KG5/13*, consists of comparing their spatial delineations of land surface and a comparison of their medoids. Medoid of a CT is its constituent LC that has the smallest average dissimilarity to all other LCs in this CT; it corresponds to a centroid in coordinate space. We use medoids as exemplars of CTs and compare different CTs by comparing their medoids.

The left panels in Fig. 3 show maps of the three classifications, assuming five CTs. This number has been chosen because there are five major types of climate in the *KGC* – tropical (A), arid (B), temperate (C), continental (D), and polar (E) – and we want to examine their relation to CTs delineated by our algorithms. We don t assign names to CTs obtained using clustering algorithms, we simply refer to them as *DTW5-1* to *DTW5-5* and *EUC5-1* to *EUC5-5*, respectively. The right panels in Fig. 3 show the exemplars of corresponding CTs. An exemplar (or any other LC) is visualized by a climate curve – a parametric curve in the (T,R,D) space with time being the parameter. Only projections of 3D climate curves onto the (T,R) plane are shown, but the changes in the value of D are encoded by the colors along a climate curve. Dots indicate months, with January singled out by the larger black dot and February singled out by the larger gray dot. A climate s character can be inferred from the location and the shape of the climate curve.

The V-measure-based dissimilarity between *DTW5* and *EUC5* is only 0.34, but not all CTs can be matched between the two classifications, and those that can be matched have different spatial extents. Roughly, the CTs in the two classifications can be matched as follows: *DTW5-5*→*EUC5-5*, *DTW5-1*→*EUC5-1*, *DTW5-4*→*EUC5-3* and *EUC5-4*, *DTW5-2*→*EUC5-2*. The CT *DTW5-3* has no equivalent in the *EUC5* classification. *EUC5* has two CTs that can be described as tropical but doesn t have separate CTs that distinguish between continental and polar climates.

The V-measure-based dissimilarity between *DTW5* and *KG5* is 0.5, but visually they appear to match better than *DTW5* and *EUC5*. Roughly, the CTs in the two classification can be matched as follows: *DTW5-4*→A, *DTW5-5*→B, *DTW5-2*→C, *DTW5-3*→D, and *DTW5-1*→E. However, the exemplar for *DTW5-2* has a very different character than the exemplar for C, which displays a tropical-like shape albeit at the lower range of temperatures and with the smaller amplitude of precipitation. The V-measure-based dissimilarity between *EUC5* and *KG5* is 0.51, the biggest difference between them being the extents and climatic profiles of temperate CTs and the existence of two tropical CTs in the *EUC5*. In our opinion, the *DTW5* offers the most reasonable division of land surface into only five CTs.

Next, we compare these three classifications assuming thirteen CTs – the number of CTs at the second level of the *KGC*. At that level *KGC* CTs are: tropical rainforest *Af*,

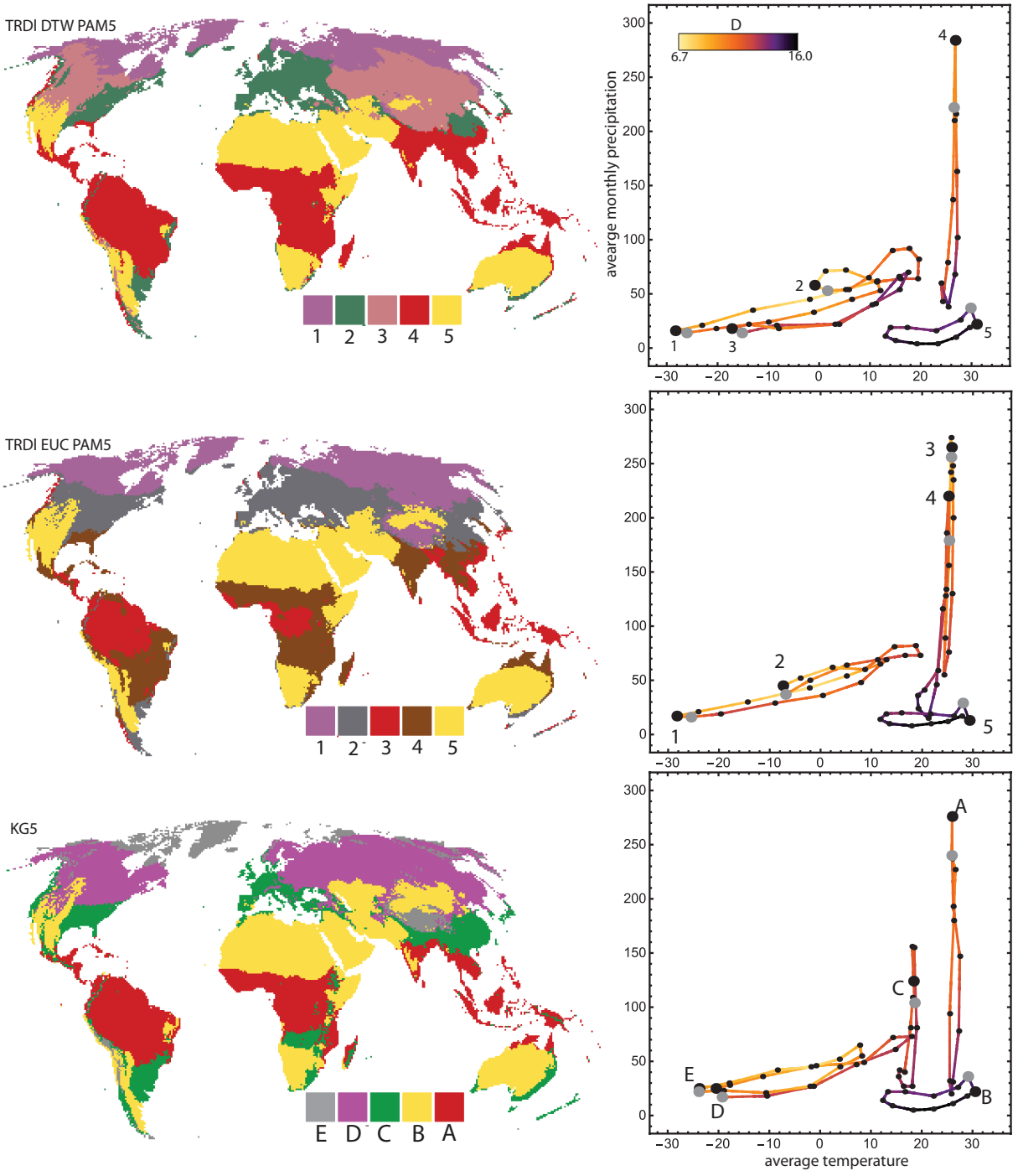

FIG. 3. Comparison of three different classifications, *TRDI DTW PAM5*, *TRDI EUC PAM5*, and *KG5*, with five climate types each. Panels on the left show geographical extents of climate types and panels on the right show climate curves for each climate type. The white-to-black color gradient indicates the monthly range of temperatures from small to large.

tropical monsoon (*Am*), tropical savanna (*Aw*), arid desert (*BW*), arid steppe (*BS*), temperate dry summer (*Cs*), temperate dry winter (*Cw*), temperate without dry season (*Cf*), continental dry summer (*Ds*), continental dry winter (*Dw*), continental without dry season (*Df*), polar tundra (*ET*), and polar frost (*EF*). We refer to clustering-based CTs as *DTW13-1* to *DTW13-13* and *EUC13-1* to *EUC13-13*, respectively.

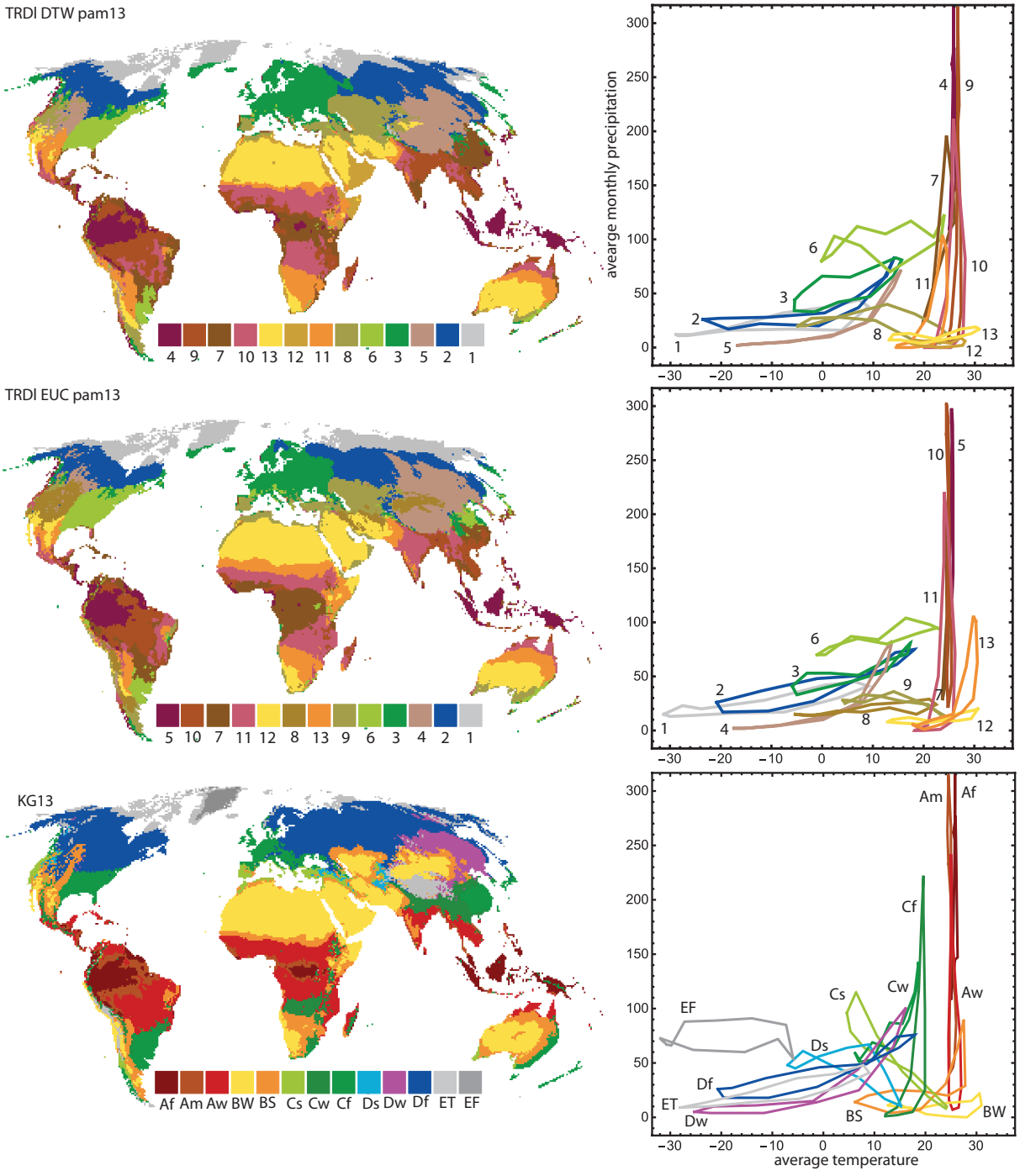

FIG. 4. Comparison of three different classifications, *TRDI DTW PAM13*, *TRDI EUC PAM13*, and *KG13*, with thirteen climate types each. Panels on the left show geographical extents of climate types and panels on the right show climate curves for each climate type.

Fig. 4 shows the maps and exemplars of CTs for the three classifications. The exemplars are color-coded to correspond to map legends but they don t show the variation of D. The V-measure-based dissimilarity between *DTW13* and *EUC13* is 0.33 and their overall similarity is confirmed by examining the maps and CTs exemplars as shown in Fig. 4. Twelve of thirteen CTs can be matched to each other as indicated by the map legends. The only unmatched CTs are the *DTW13-12* which is an extremely dry and hot desert climate which has no equivalent in the

EUC13 classification and the *EUC13-8* which is an arid climate at moderate temperatures and has no equivalent in the *DTW13* classification. In addition, the boundaries of corresponding CTs are shifted relative to each other.

The *V*-measure-based dissimilarity between *DTW13* and *KG13* is 0.53, re ecting differences that can be observed on the maps and by comparing their respective CTs. We can closely match exemplars for the following six pairs of CTs: *DTW13-4*→*Af*, *DTW13-13*→*BW*, *DTW13-2*→*ET*, *DTW13-2*→*DF*, *DTW13-5*→*Dw*, and *DTW13-11*→*BS*. Note, however, that despite close matches in exemplars, some of these paired CTs have markedly different spatial extents.

The remaining CTs cannot be closely matched. The geographical extent of the *Cf* is partially covered by two distinct *DTW13* climates, the *DTW13-6*, which covers central and eastern U.S., portions of Argentina and Uruguay, and the southeastern coast of Australia, and the *DTW13-6*, which covers Europe. The combined extent of *Am*, *Aw*, and *Cw* coincides with the combined extent of *DTW13-9*, *DTW13-7* , and *DTW13-10*, but there is no good one-to-one matching between individual CTs. The *Ds* and the *DTW13-8* have somewhat similar exemplars but very different geographical ranges. Finally, *Cs* and *EF* have no equivalents in the *DTW13* classification, and the *DTW13-12* has no equivalent in the *KG13* classification.

c. Climate inhomogeneities within CTs

Most applications of climate classifications implicitly assume that climate within a single CT is relatively uniform. However, this assumption cannot be verified within the scope of the *KGC* as it lacks a native notion of climate similarity. In a modern implementation of *KGC*, which has a form of a decision tree (Spinoni et al. 2015), the 11 derived climatic variables (Peel et al. 2007; Cannon 2012) used to steer an LC through the tree could be thought of as a vector description of the LC. Could the Euclidean distance between such vectors define a viable measure of climate similarity native to *KGC*? We have tested such possibility and came to a conclusion that it does not offer a good measure of similarity. This is because the 11 variables were designed to be used for predicate statements and not for assessment of similarity.

Our approach is built from the ground up on the notion of climate similarity so we are in a position to investigate the homogeneity of various CTs. We apply *DTW* similarity to investigate climate homogeneity within a CT even if the CTs are delineated using the *KGC*. The homogeneity assessments are visualized in two different ways. First, for each CT we compare the climate curve of its exemplar with climate curves of a representative sample of fifty LC randomly selected from this CT. Representative sample means that LCs are randomly drawn from a distribution representing the spread of dissimilarities between LCs and the exemplar. This illustrates the range of different climates that are grouped into a single CT. Second, we map the geography of climate inhomogeneity for each CT. Such a map shows the location of an exemplar; other locations belonging to a given CT are color-coded according to their dissimilarity from the exemplar.

Fig. 5 shows the climate inhomogeneities of CTs in the *KGC*. Panels show one or two CTs; two CTs are shown when possible to decrease the size of the figure. Black climate curves pertain to exemplars, other climate curves pertain to the representative sample of LCs. Each CT is labeled by its name and the color it was depicted by in Fig. 4. The two numbers following the CT s name are the maximum dissimilarity to the exemplar, and the 90th percentile of dissimilarities to the exemplar. This last number is a good indicator of the spread of LCs within the CT. The following CTs, *Af*, *Am*, *BW*, *Dw*, and *Ds*, are characterized by a relatively small spread of LCs so they are relatively homogeneous, as could be confirmed visually by observing that the climate curves of LC are close to the climate curve of the exemplar. On the other hand, CTs such as *ET*, *Cf*, *Cw*, and *BS* are highly inhomogeneous.

Panels in Fig. 6 show maps of climate inhomogeneity for each CT in the *KGC*. Exemplars locations are shown by a black dot, locations progressively more dissimilar to the exemplar are shown in progressively darker colors. The most striking geographical inhomogeneities are observed for *ET*, *Cf*, *Cw*, and *BS*. The *KGC* assigns Tibet to the *ET* climate type, but clearly it has a climate markedly different from that of the polar regions which constitute the rest of the region identified as *ET*. The *Cf* climate type appears to group two or maybe even three distinct climates, one that covers central and eastern U.S., portions of Argentina and Uruguay, and the southeastern coast of Australia, another which covers southeast China, and another that covers norther Europe. The *Cw* climate appears to lack any coherent form or multiple forms and it is questionable whether it should be considered as a distinct climate type. Finally, the *BS* climate type appears to be geographically too spread to be climatically homogeneous.

We can also quantify the quality of the entire *KGC* from a clustering perspective by using the *DB* index which is equal to *DB*=3.42.

We performed the same analysis for the *DTW13* classification. The results are shown in Figs. 7 and 8 which are the *DTW13* equivalents of Figs. 5 and 6. Comparing the results in Fig. 5 with the results in Fig. 7 we observe that the *DTW13* classification leads to a higher degree of CT homogeneity than the *KGC*. The range of the 90th percentile of dissimilarities to the exemplar is $(0.4, 0.75)$ for the *DTW13* classification, with a mean of 0.57 and standard deviation of 0.09. For the *KGC* the range is $(0.57, 1.09)$, with a mean of 0.75 and standard deviation of 0.15. The fact that the *DTW13* classification yields

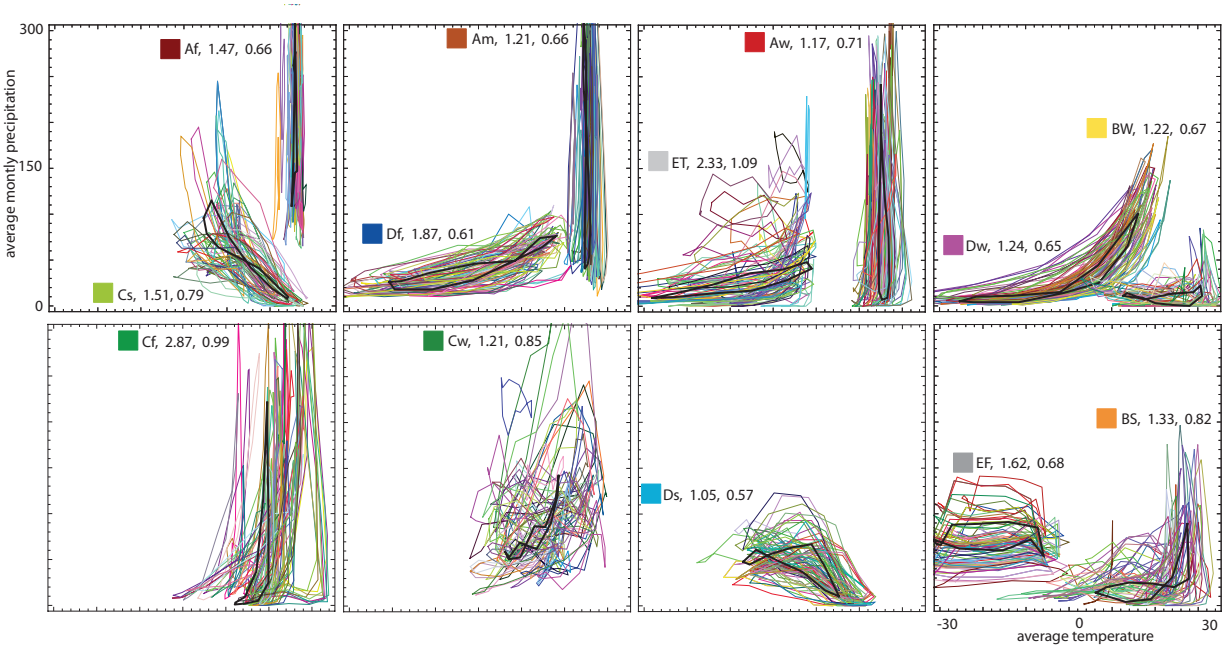

FIG. 5. Inhomogeneity of local climates within each of thirteen *KG* climate types. For each climate type the black curve corresponds to the exemplar climate and colored curves correspond to the representative sample of 50 local climates within this climate type. Climate types are labeled with the names and colors from Fig. 4. The values of the maximum dissimilarity to the exemplar, and the 90th percentile of dissimilarities to the exemplar are also shown.

more homogeneous CTs is not a surprise as the goal of a clustering algorithm, such as *PAM*, is to maximize the homogeneity within individual CTs and the disparity between exemplars of different CTs. The *DB*=1.92 for the *DTW13* further indicates that, an average, *DTW13* CTs are more uniform and more distinct from each other than the CTs in the *KGC*.

Examining Fig. 7 we observe that all CTs in the *DTW13* classification are relatively homogeneous. The largest inhomogeneity value (0.75) is assigned to the *DTW13-4*, even so the climate curves of the representative sample of LCs appear to cluster closely around the exemplar. This is because only the projections of climate curves on the (T, R) plane are shown. In these two variables the *DTW13-4* is more homogeneous than when all three variables (T, R, D) are used. The reverse situation is observed for *DTW13-11*, where a small value of inhomogeneity (0.53) is given, but the figure shows a moderate spread of the representative sample. This is because the *DTW13-11* is very homogeneous in the values of D.

Panels in Fig. 8 show maps of climate inhomogeneity for each CT in the *DTW13* classification. Overall, the *DTW13* CTs are characterized by higher degree of geographical homogeneity than the *KGC* CTs. The *DTW13-1* climate shows that the middle of Greenland has climate significantly different from the rest of this CT. Recall that the *DTW13* did not yield an equivalent of *EF* climate in the *KGC*. The region indicated as not fitting the rest of the *DTW13-1* coincide with the region labeled as *EF* by the *KGC*. We conclude that the absence of an *EF* equivalent in the *DTW13* classification is due to our restricting the number of CTs to 13. Within this limit the *PAM* algorithm did not separate this region as an individual CT, but if one more climate type would be allowed, this region would become a new CT corresponding closely to the *EF*. The *DTW13-3* climate, which covers most of Europe but extends to Iceland and the southeastern cost of Greenland, is relatively homogeneous with the exception of the coast of Greenland, which has a climate markedly different from the exemplar. This region would join the new CT in the *DTW14* classification as discussed above.

4. Conclusions

The *KGC* is the most widely used climate classification system and has been so for over 100 years (Peel et al. 2007). This does not mean that research toward a more complete understanding of the spatial distribution of climates across terrestrial land surface should cease. Due to its exploratory character clustering offers a different point of view on how the world s climates can be grouped into CTs. Our aim was to critically examine various elements of the clustering process to arrive at a protocol that results in the most acceptable clustering-based climate classification. What made this task difficult was the lack of “ground

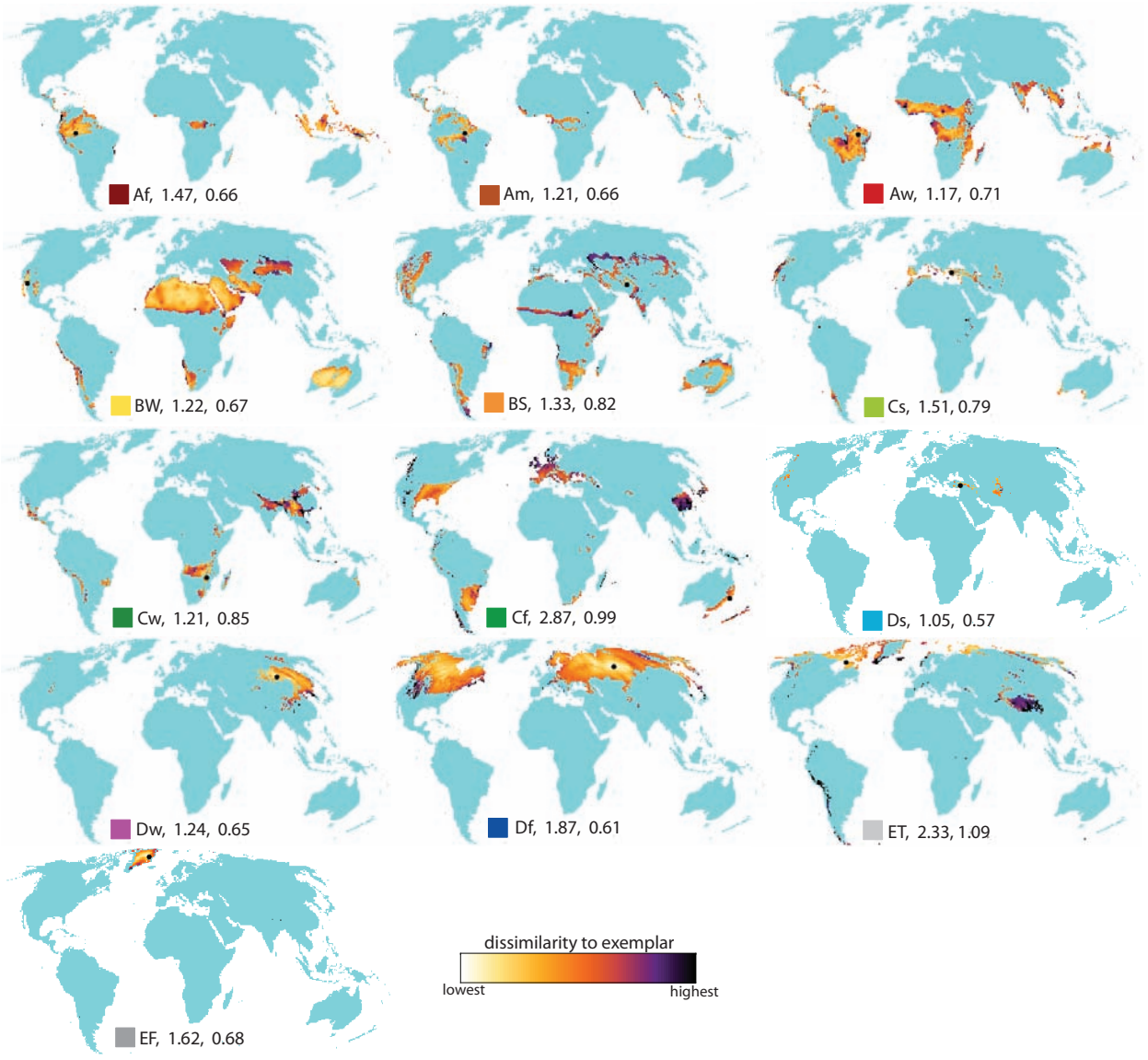

FIG. 6. Geographical depiction of inhomogeneities of local climates within each of thirteen *KG* climate types. For each climate a black dot indicates the location of an exemplar and the white-to-black color gradient indicates locations having climates with small to large dissimilarity to the exemplar. Climate types are labeled with the names and colors used in Fig. 4. The values of the maximum dissimilarity to the exemplar, and the 90th percentile of dissimilarities to the exemplar are also shown.

truth" to measure against. Certainly, the *KGC* cannot be considered ground truth because our goal is not to reproduce it but rather to arrive at a useful alternative for the grouping of climates. After examining a large number of possible clustering protocols (of which only 32 are documented in this paper) we arrived at the following conclusions.

(1) A good mathematical representation of local climate and an appropriate choice of dissimilarity function matters. Defining LC as a cyclic time series and using a dissimilarity function that takes this definition into account results in an automatic adjustment for the seasons, taking into account not only the location of the LC (northern or southern hemisphere) but also local physical conditions. This results in a better quality of classification. The use of *DTW* instead of the Euclidean distance had a smaller impact than we expected. This is because the major advantage of *DTW* – its ability to adjust the dissimilarity of two LCs for season-related time shifts has already been accounted for by using time-shift invariant dissimilarity functions. Even so, *DTW* still offers some small advan-

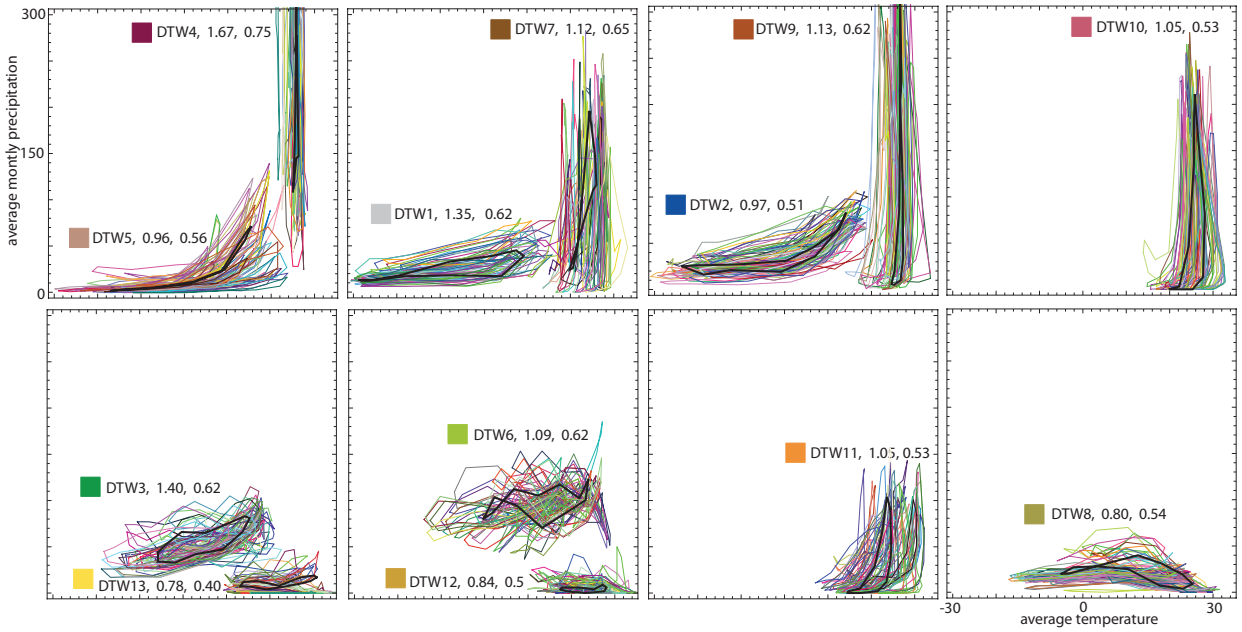

FIG. 7. Inhomogeneity of local climates within each of thirteen *DTW13* climate types. For each climate type the black curve corresponds to the exemplar climate and colored curves correspond to the representative sample of 50 local climates within this climate type. Climate types are labeled with the colors from Fig. 4. The values of the maximum dissimilarity to the exemplar, and the 90th percentile of dissimilarities to the exemplar are also shown.

tages which could be more pronounced if we used daily-means instead of monthly-means data.

(2) Proper normalization of variables is important. Using a standard normalization (as in Zhang and Yan (2014)) or standardization (as in Zscheischler et al. (2012)) of data effectively reduces the in uence of precipitation on similarity between LCs and results in a classification based predominantly on temperature. This can be observed in Fig. 1 in the (Zhang and Yan 2014) paper where climate types have markedly longitudinal character consistent with overreliance on temperature. To avoid this problem the skewed distribution of monthly-mean precipitation values toward the large values needs to be taken into consideration when normalizing the variables. Adding the monthly-mean amplitude of temperature (D) as the third climatic variable provides additional information to the clustering process and changes the classification. Whether or not to utilize D depends on one s concept of what constitutes climate. We think that D is pertinent to the perception of climate, but whether it should carry the same weight as T and P remains an open question.

(3) Finally, we have found that using the *PAM* algorithm gives better results than using the *HC* algorithm. This is in agreement with the earlier findings (Gerstengarbe et al. 1999) that the K-means algorithm (which is different from the *PAM* algorithm we used, but is based on a similar principle) should be preferred over the *HC* algorithm in the context of climate classification. However, when using *PAM* the resultant classification would not form a hierarchy of climates. We have also tested a divisive hierarchical clustering algorithm *DIANA* (Kaufman and Rousseeuw 2009). *DIANA*, like *PAM*, produces clusters by dividing all LCs rather than agglomerating them like in the *HC*. Thus, it may produce clusterings comparable in quality to those yielded by *PAM* while also preserving a hierarchical structure. We found that *DIANA* yields good classification into 13 CTs when using *DTW* dissimilarity function but not for other combinations of number of clusters/dissimilarity function. Moreover, *DIANA* is an order of magnitude more computationally expensive than *PAM* and thus not practical for clustering a large grid of LCs. Clustering with *PAM*, *HC*, or *DIANA* requires storing a distance matrix in a computer memory. This is plausible given the 75 km resolution cells used in our present calculations. Distance space-based clustering of the higher resolution grid would require using an online distance space-based clustering algorithm such as *BUBBLE* (Ganti et al. 1999), which dynamically clusters an incoming stream of data points (LCs) without storing a distance matrix.

As there is no strict agreed upon, specific, definition of climate (we don t consider *KG* definitions to be definitive) there are no criteria to determine which climate classification is better than the other. Moreover, regardless of its definition, climate changes continuously across the land surface which means that boundaries between different

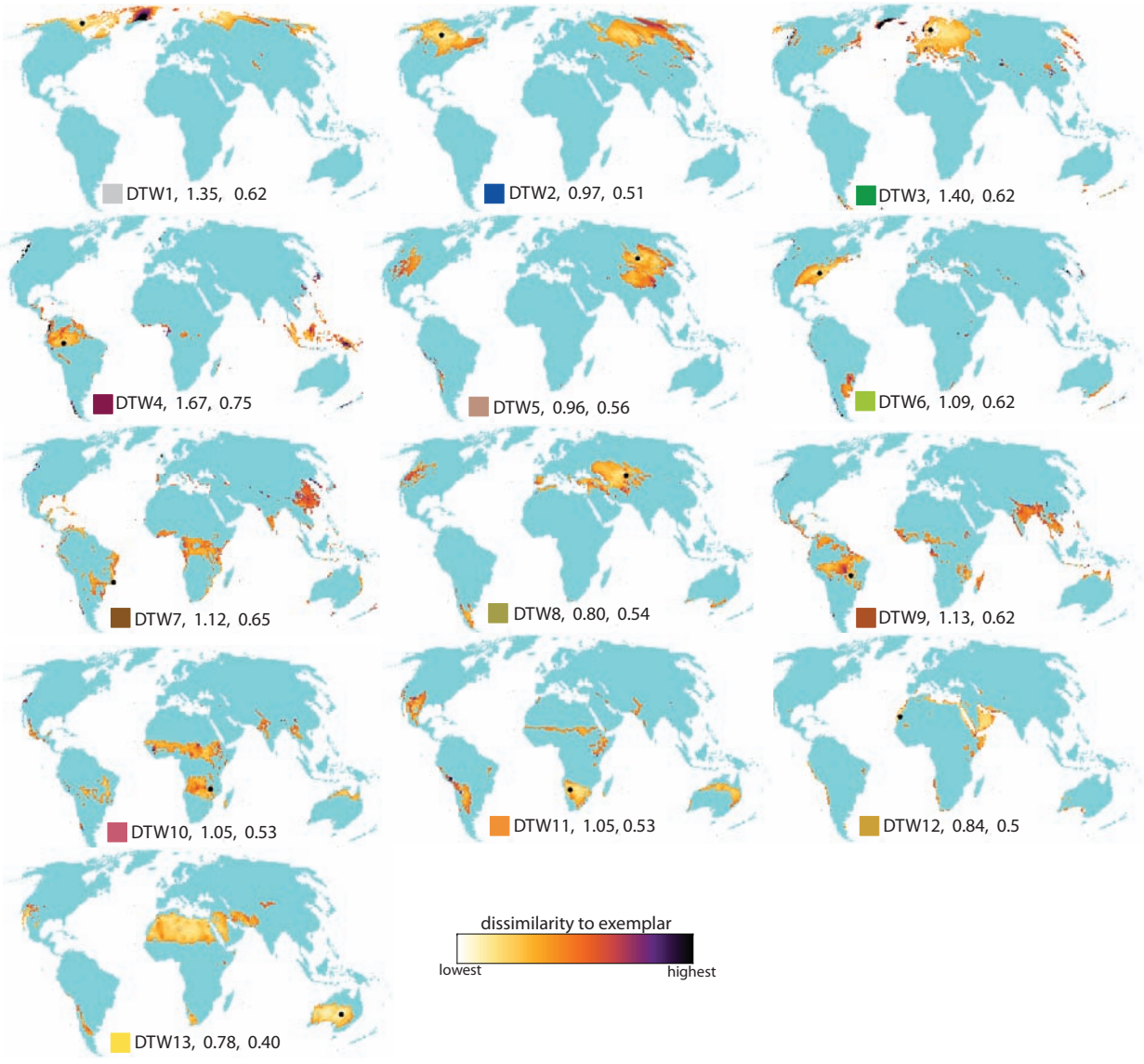

FIG. 8. Geographical depiction of inhomogeneities of local climates within each of thirteen *DTW13* climate types. For each climate a black dot indicates the location of an exemplar and the white-to-black color gradient indicates locations having climates with small to large dissimilarity to the exemplar. Climate types are labeled with the colors from Fig. 4. The values of the maximum dissimilarity to the exemplar, and the 90th percentile of dissimilarities to the exemplar are also shown.

climate types may easily shift when using different classification methods even if the definition of climate remains the same. In short, we cannot expect to find one classification that delineates CTs optimally from all possible points of view. In this context our investigation resulted in establishing a framework for exploring different climatic partitioning.

The *KGC* appeals to many because it is very familiar, and the names and meanings of its individual CTs have achieved a status of textbook knowledge. In addition, the *KGC* scheme has the appealing quality of being presentable in the form of a decision tree (Spinoni et al. 2015). On the other hand, as we demonstrated in this paper, the clustering approach has its own advantages including a capability to delineate custom classifications and the ability to assess the uniformity of climates within single CTs as well as diversity between different CTs.

Discussing the relative merits of the *DTW13* classification (the one we singled out from the set of 32 investigated classifications) versus the *KGC13* above what has been discussed in section 3 is beyond the scope of this paper. *DTW13* uses a different definition of climate, and,

within this definition, it formally outperforms the *KGC13*. Whether *DTW13* is better than *KGC13* depends on acceptance or rejection of this definition. One application where clustering-based classification should be used instead of *KGC* is the visualization of climate change based on predictions from global climate models. This is because there is no reason to believe that the effect of climate change will be limited to shifting boundaries of present-day CTs. Demonstrating differences between *KGC*-based and classification-based mapping of future climates is beyond the scope of this paper but will be a topic of future research.

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

References

Baker, B., H. Diaz, W. Hargrove, and F. Hoffman, 2010: Use of the KöppenTrewartha climate classification to evaluate climatic refugia in statistically derived ecoregions for the Peoples Republic of China. *Climatic Change*, **98**, 113–131.

Beck, C., J. Grieser, M. Kottek, F. Rubel, and B. Rudolf, 2005: Characterizing global climate change by means of Köppen Climate Classification. *Klimastatusbericht*, **51**, 139–149.

Berndt, D. J., and J. Clifford, 1994: Using Dynamic Time Warping to Find Patterns in Time Series. *KDD Workshop*, **10(16)**, 359–370.

Brugger, K., and F. Rubel, 2013: Characterizing the species composition of European Culicoides vectors by means of the Köppen-Geiger climate classification. *Parasit Vectors*, **6(1)**, 333.

Bunkers, M. J., J. R. Miller, and A. T. DeGaetano, 1996: Definition of climate regions in the Northern Plains using an objective cluster modification technique. *Journal of Climate* , **9(1)**, 130–146.

Cannon, A. J., 2012: Hydrology and Earth System Sciences. *Köppen versus the computer: comparing Kppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity*, **16(1)**, 217–229.

Chen, D., and H. W. Chen, 2013: Using the Köppen classification to quantify climate variation and change: An example for 19012010. *Environmental Development*, **6**, 69–79.

Davies, D. L., and D. W. Bouldin, 1979: A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1(2)**, 224–227.

DeGaetano, A. T., 1996: Delineation of mesoscale climate zones in the northeastern United States using a novel approach to cluster analysis. *Journal of Climate*, **9(8)**, 1765–1782.

Diaz, H. F., and J. K. Eischeid, 2007: Disappearing "alpine tundra" Köppen climatic type in the western United States. *Geophysical Research Letters*, **34**, L18 707.

Feng, S., Q. Hu, W. Huang, C.-H. Ho, R. Li, and Z. Tang, 2014: Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. *Global and Planetary Change*, **112**, 41–52.

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

Fraedrich, K., F.-W. Gerstengarbe, and P. C. Werner, 2001: Climate shifts during the last century. *Climatic Change*, **50(4)**, 405–417.

Gallardo, C., V. Gil, E. Hagel, C. Tejeda, and M. de Castro, 2013: Assessment of climate change in Europe from an ensemble of regional climate models by the use of KöppenTrewartha classification. *International Journal of Climatology*, **33(9)**, 2157–2166.

Ganti, V., R. Ramakrishnan, J. Gehrke, A. Powell, and J. French, 1999: Clustering large datasets in arbitrary metric spaces. *15th International Conference on Data Engineering*, IEEE, 502–511.

Garcia, R. A., M. Cabeza, C. Rahbek, and M. B. Araújo, 2014: Multiple dimensions of climate change and their implications for biodiversity. *Science*, **344**, 1247 579.

Gerstengarbe, F.-. W., P. C. Werner, and K. Fraedrich, 1999: Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution. *Theoretical and Applied Climatology*, **64**, 143–150.

Guetter, P. J., and J. E. Kutzbach, 1990: A modified Köppen classification applied to model simulations of glacial and interglacial climates. *Climatic Change*, **16(2)**, 193–215.

Hanf, F., J. Körper, T. Spangehl, and U. Cubasch, 2012: Shifts of climate zones in multi-model climate change experiments using the Köppen climate classification. *Meteorologische Zeitschrift* , **21(2)**, 111–123.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observationsthe CRU TS3. 10 Dataset. *International Journal of Climatology*, **34(3)**, 623–642.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G.Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25(15)**, 1965–1978.

Kaufman, L., and P. Rousseeuw, 1987: Statistical Data Analysis Based on the L1 Norm and Related Methods. *Clustering by means of medoids*, Y. Dodge, Ed., North-Holland, 405–416.

Kaufman, L., and P. J. Rousseeuw, 2009: *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Köppen, W., 1936: Das geographische System der Klimate. *Handbuch der Klimatologie*, W. Köppen, and R. Geiger, Eds., Gebrder Borntraeger, Berlin, 1–44.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, **15(3)**, 259–263.

Mahlstein, I., J. S. Daniel, and S. Solomon, 2013: Pace of shifts in climate regions increases with global temperature. *Nature Climate Change*, **3(8)**, 739–743.

Metzger, M. J., R. G. H. Bunce, R. H. G. Jongman, R. Sayre, A. Trabucco, and R. Zomer, 2012: A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, **22(5)**, 630–638.

Nafiz, A., 2005: Cyclic Sequence Comparison Using Dynamic Warping. *Lecture Notes in Computer Science, vol. 3568*, Springer, 328–335.

Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.,*, **11**, 1633–1644.

Peel, M. C., T. A. McMahon, B. L. Finlayson, and F. R. G. Watson, 2001: Identification and explanation of continental differences in the variability of annual runoff. *Journal of Hydrology*, **250**, 224–240.

Rabiner, L., and B. Juand, 1993: *Fundamentals of speach recognition*. Printice-Hall International Inc., 507 pp.

Rohli, R. V., T. A. Joyner, S. J. Reynolds, and T. J. Ballinger, 2015: Overlap of global KöppenGeiger climates, biomes, and soil orders. *Physical Geography*, **36(2)**, 158–175.

Rosenberg, A., and J. Hirschberg, 2007: V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 410–420.

Rubel, F., and M. Kottek, 2010: Observed and projected climate shifts 19012100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, **19(2)**, 135–141.

Spinoni, J., J. Vogt, G. Naumann, H. Carrao, and P. Barbosa, 2015: Towards identifying areas at climatological risk of desertification using the KöppenGeiger classification and FAO aridity index. *International Journal of Climatology*, **35(9)**, 2210–2222.

Stooksbury, D. E., and P. J. Michaels, 1991: Cluster analysis of southeastern US climate stations. *Theoretical and Applied Climatology*, **3–4**, 143–150.

Thornthwaite, C. W., 1943: Problems in the Classification of Climates. *Geographical Review*, **33(2)**, 233–255.

Thornthwaite, C. W., 1948: An approach toward a rational classification of climate. *Geographical review*, **38(1)**, 55–94.

Trewartha, G. T., and L. H. Horn, 1980: *Introduction to climate*. McGraw Hill, New York, NY.

Unal, Y., T. Kindap, and M. Karaca, 2003: Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology*, **23(9)**, 1045–1055.

Usery, E. L., and J. Seong, 2001: All Equal-Area Map Projections Are Created Equal, But Some Are More Equal Than Others. *Cartography and Geographic Information Science*, **28(3)**, 183–193.

Ward, J. H., 1963: Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, **58(**, 236–244.

Werier, D. A., and R. F. C. Naczi, 2014: Carex secalina (Cyperaceae), an introduced sedge new to North America. *Rhodora*, **114(960)**, 349–365.

Wilkinson, L., and M. Friendly, 2009: The history of the cluster heat map. *The American Statistician*, **63(2)**, 179–184.

Zhang, X., and X. Yan, 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., M. D. Mahecha, and S. Harmeling, 2012: Climate classifications: the value of unsupervised clustering. *Procedia Computer Science*, **9**, 897–906.