On Using a Clustering Approach for Global Climate Classification

PAWEL NETZEL AND TOMASZ STEPINSKI

Space Informatics Laboratory, Department of Geography, University of Cincinnati, Cincinnati, Ohio

(Manuscript received 6 September 2015, in final form 18 December 2015)

ABSTRACT

Classifying the land surface into climate types provides means of diagnosing relations between Earth’s physical and biological systems and the climate. Global climate classifications are also used to visualize climate change. Clustering climate datasets provides a natural approach to climate classification, but the rule-based Köppen–Geiger classification (KGC) is the one most widely used. Here, a comprehensive approach to the clustering-based classification of climates is presented. Local climate is defined as a multivariate time series of mean monthly climatic variables and the authors propose to use dynamic time warping (DTW) as a measure of dissimilarity between local climates. Also discussed are the choice of climatic variables, the importance of their proper normalization, and the advantage of using distance-based clustering algorithms. Using the WorldClim global climate dataset and different combinations of clustering parameters, 32 different clustering-based classifications are calculated. These classifications are compared between themselves and to the KGC using the information-theoretic V measure. It is found 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. Two such classifications are compared in detail between each other and to the KGC. About half of 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, it is demonstrated 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; a 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), and distributions of invasive species (Werier and Naczi 2012) and virus vectors (Bruger 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 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 reflecting 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, worldwide 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 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 LTMM of temperature and precipitation. This system lacks the notion of similarity between LCs (see section 3e 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 is in the data without any prior assumptions about their expected character and/or location. A clustering approach is feasible because of 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 unlikeness 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 PCA to reduce the dimensionality of vectors representing LCs.

More recently, with the increased availability of climate data from worldwide 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. 2013). 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. (2013) 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 studies of clustering-based classifications of climates 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 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 reflect 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

<table>
<thead>
<tr>
<th>Table 1. List of acronyms used in the paper.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
</tr>
<tr>
<td>CT</td>
</tr>
<tr>
<td>KGC</td>
</tr>
<tr>
<td>LTMM</td>
</tr>
<tr>
<td>PCA</td>
</tr>
<tr>
<td>DTW</td>
</tr>
<tr>
<td>EUC</td>
</tr>
<tr>
<td>HC</td>
</tr>
<tr>
<td>PAM</td>
</tr>
</tbody>
</table>
process (the choice of variables, the choice of normalization procedure, the choice of dissimilarity function, and the choice of clustering algorithm) have the greatest influence 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 KGCs 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 2004; Hijmans et al. 2005). WorldClim data are given on a 30-arc-s grid and has the spatial extent of 90°N–60°S over all longitudes. The grid cells contain mean-monthly climatic variables interpolated from a meteorological time series measured from a worldwide network of meteorological stations between 1950 and 2000. We use mean monthly values of the following variables: temperature $T$, which is a measure of average climatic thermal conditions; precipitation $P$, which is a measure of climate humidity; and temperature range $D = T_{\text{max}} - T_{\text{min}}$, which is a measure of variability of the climatic thermal condition. The terms $T_{\text{min}}$ and $T_{\text{max}}$ 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 data from the Mollweide grid to a grid with a resolution of 75 km × 75 km per grid cell resulting in a 213 × 482 grid of which 23 979 cells represent land surface and the rest represent water. We use such a relatively coarse grid because of 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. 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 overinfluenced by the variable with the largest range. We use two different normalization transformations. The first is the standard normalization $X_i \leftarrow [X_i - \min(X_i)]/\max(X_i - \min(X_i)) $, where the $X_i$ terms are climatic variables. We refer to this type of normalization as global. However, we note that the precipitation variable, with a range of 0–1550 mm, has a distribution that 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 influence of precipitation on the overall dissimilarity between LCs would be artificially diminished. Thus, we introduced a second normalization procedure referred to as modified. This procedure transforms variables $T$ and $D$ according to the normalization formula as given above, but the variable $R$ is transformed as follows:

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

Thus, the top 1% of precipitation values, those in a range of 350–1550 mm, are all transformed to the value of 1, while the remaining 99% of precipitation values, those in a range between 0 and 350 mm, are transformed to values between 0 and 1. This has the effect of restoring the influence 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, an LC at location $i$ is given as $LC_i = \{M_i^1, \ldots, M_i^{12}\}$, where the time series progresses through 12 months $M_i = (T_i^1, R_i^1, D_i^1)$, for $i = \{1, \ldots, 12\}$, 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 $d^{\text{DTW}}_{\text{bst}}(LC_1, LC_2)$ and 2) a time-shift invariant version of the Euclidean distance $d^{\text{EUC}}_{\text{bst}}(LC_1, LC_2)$. To calculate $d^{\text{EUC}}_{\text{bst}}$ the standard Euclidean distance $\sum_{i=1}^{12} d^2(M_i^1, M_i^2)^{1/2}$ between $LC_1$ and $LC_2$ is calculated 12 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 $d^{\text{EUC}}_{\text{bst}}(LC_1, LC_2)$ is the minimum of the 12 calculated values. This ensures that
the dissimilarity between two LCs is independent from time shift in their seasons.

Our preferred dissimilarity function is 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 nonlinear alignments between two time series to accommodate sequences that are similar but locally out of phase (Rabiner and Juang 1993; Nafiz 2005). Figure 1 illustrates this difference; two local climates (for Dallas and Los Angeles) are shown as bivariate \((T \text{ and } R)\) monthly series. Standard Euclidean distance aligns the two series month to month, as shown by dashed lines in Fig. 1a1. This corresponds to taking a diagonal (possibly suboptimal) path in the matrix of distances between two series, as shown in 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 minimum value of the overall distance) path through the matrix of distances (Fig. 1b2). Dashed lines in Fig. 1b1 show the alignment between the two 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 nonlinear alignment also uses the same minimization over the cyclic shifts of months as described above for the \(d^{EUC}_{\text{EUC}}\).

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 HC with Ward linkage (Ward 1963), and the second is the PAM algorithm (Kaufman and Rousseau 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). Figure 2a illustrates the principle of \(V\) measure using the specific example of a classification of KGC with 5 CTs (KG5) and classification obtained using variables \(T, R, \text{ and } D\) with modified normalization, DTW dissimilarity function, and PAM clustering algorithm using 5 clusters (\(TRDi\) DTW PAM5; see the next section for an explanation of the classification naming convention). The KG5 partitions the world into five climate classes; spatial extents of these classes are shown by nongray areas in the left column of Fig. 2a. The \(TRDi\) DTW PAM5 (DTW5 for short) partitions the world into five climate types; spatial extents of these types are shown by nongray 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 its constituent types measures a level of class homogeneity with respect to types. Homogeneous classes of KG5 (like A) have small values of entropy, and inhomogeneous classes of KG5 (like D) have large values of entropy. The ratio of class entropy to the entropy of the
entire DTW5 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 KG5 classes with respect to DTW5 types. Reversing the roles of classes and types (right column in Fig. 2a), we calculate a completeness of KG5 classes with respect to DTW5 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 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:

1. the number of climatic variables used to describe an LC, either two variables ($T$ and $R$) or three variables ($T$, $R$, and $D$);
2. the method of variable normalization, either global $g$ or modified $l$;
3. the choice of a dissimilarity function, either DTW or the EUC;
4. the choice of a clustering algorithm, either HC or PAM; and

---

**Fig. 2.** (a) Illustrating the concept of $V$ measure. (left) Spatial extents of KG5 climate classes and how they intersect TRD/DTW PAM5 climate types. (right) Spatial extents of TRD/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.
5) 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.

Figure 2b shows the results of the comparisons between 34 climate classifications [the KGC for 5 and 13 CTs (KG5 and KG13, respectively) 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 classifications of greater similarity, 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 KGC 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, and D with global normalization, DTW dissimilarity function, and PAM clustering algorithm using 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 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 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 D.

For classifications derived using the 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 KGC

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, depending on the number of CTs), and KG5 or KG13, consists of comparing their spatial delineations of land surface and a comparison of their medoids. The 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 do not assign names to CTs obtained using clustering algorithms; we simply refer to them as DTW5–1 through DTW5–5 and EUC5–1 through 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 $\rightarrow$ EUC5–5, DTW5–1 $\rightarrow$ EUC5–1, DTW5–4 $\rightarrow$ EUC5–3 and EUC5–4, and DTW5–2 $\rightarrow$ EUC5–2.

The CT DTW5–3 has no equivalent in the EUC5 classification. EUC5 has two CTs that can be described as tropical but does not 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 classifications can be matched as follows: DTW5–4 $\rightarrow$ A, DTW5–5 $\rightarrow$ B, DTW5–2 $\rightarrow$ C, DTW5–3 $\rightarrow$ D, and DTW5–1 $\rightarrow$ E. However, the exemplar for DTW5–2

**FIG. 3.** Comparison of three different classifications—TRDI DTW PAM5, TRDI EUC PAM5, and KG5—with five climate types each. (left) Geographical extents of climate types and (right) climate curves of average $T$ (°C) vs average monthly $R$ (mm) for each climate type. The white-to-black color gradient in the climate curves indicates the monthly range of temperatures (°C) from small to large.
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 13 CTs—the number of CTs at the second level of the KGC. At that level KGC CTs are tropical rain forest (Af), 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 through DTW13–13 and EUC13–1 through EUC13–13, respectively.

Figure 4 shows the maps and exemplars of CTs for the three classifications. The exemplars are color coded to correspond to map legends, but they do not 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 CT exemplars as shown in Fig. 4. As indicated by the map legends, 12 of 13 CTs can be matched to each other. The only unmatched CTs are the DTW13–12, which is an extremely dry and hot desert climate that 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, reflecting 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 $\rightarrow$ Af, DTW13–13 $\rightarrow$ BW, DTW13–2 $\rightarrow$ ET, DTW13–2 $\rightarrow$ DF, DTW13–5 $\rightarrow$ Dw, and DTW13–11 $\rightarrow$ 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 that covers the central and eastern United States, portions of Argentina and Uruguay, and the southeastern coast of Australia and the DTW13–6 that 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 a possibility and came to the 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.

Figure 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, and 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. Five 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.

Figure 6 shows maps of climate inhomogeneity for each CT in the KGC. Exemplars’ locations are shown by a black dot, and 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 the central and eastern United States, portions of Argentina and Uruguay, and the southeastern coast.
of Australia, another that covers southeastern China, and another that covers northern 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 DB. The KGC is characterized by \( DB = 3.42 \).

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 more homogeneous CTs is unsurprising since 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. A \( DB = 1.92 \) for the DTW13 further indicates that, on 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 \( D \).

Figure 8 shows 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 coincides with the region labeled 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 coast 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 previously.

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. Because of 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 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.

- 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 advantages, which could be more pronounced if we used daily mean instead of monthly mean data.

- 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 influence of precipitation on similarity between LCs and results in a classification based predominantly on temperature. This can be observed in Fig. 1 of Zhang and Yan (2014), 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.

- 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 do not consider KGC 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 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 KG13 beyond what has been already 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 KG13. Whether DTW13 is better than KG13 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 Grant NNX15AJ47G from the National Aeronautics and Space Administration.

REFERENCES


Ganti, V., R. Ramakrishnan, J. Gehrke, A. Powell, and J. French, 1999: Clustering large datasets in arbitrary metric spaces. Proc. 15th Int. Conf. on Data Engineering, Sydney, NSW, Australia, Institute of Electrical and Electronics Engineers, 502–511.


