Physics and Chemistry of the Earth 35 (2010) 341–351

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

A two-tier atmospheric circulation classification scheme for the European–North Atlantic region

Galina S. Guentchev [a,*], Julie A. Winkler [b,1]

[a] UCAR CLIVAR PACE, 1D-513 David Skaggs Research Center, 325 Broadway, Boulder, CO, United States
[b] Department of Geography, Michigan State University, East Lansing, MI, United States

ARTICLE INFO

Article history:
Received 21 March 2009
Received in revised form 23 November 2009
Accepted 29 December 2009
Available online 14 January 2010

Keywords:
Circulation classification
Europe
Sea-level pressure
Cluster analysis
Temporal changes
Temperature extreme events

ABSTRACT

A two-tier classification of large-scale atmospheric circulation was developed for the European–North-Atlantic domain. The classification was constructed using a combination of principal components and k-means cluster analysis applied to reanalysis fields of mean sea-level pressure for 1951–2004. Separate classifications were developed for the winter, spring, summer, and fall seasons. For each season, the two classification tiers were identified independently, such that the definition of one tier does not depend on the other tier having already been defined. The first tier of the classification is comprised of *supertype* patterns. These broad-scale circulation classes are useful for generalized analyses such as investigations of the temporal trends in circulation frequency and persistence. The second, more detailed tier consists of circulation *types* and is useful for numerous applied research questions regarding the relationships between large-scale circulation and local and regional climate. Three to five *supertypes* and up to 19 circulation *types* were identified for each season. An intuitive nomenclature scheme based on the physical entities (i.e., anomaly centers) which dominate the specific patterns was used to label each of the *supertypes* and *types*. Two example applications illustrate the potential usefulness of a two-tier classification. In the first application, the temporal variability of the *supertypes* was evaluated. In general, the frequency and persistence of *supertypes* dominated by anticyclonic circulation increased during the study period, whereas the *supertypes* dominated by cyclonic features decreased in frequency and persistence. The usefulness of the derived circulation *types* was exemplified by an analysis of the circulation associated with heat waves and cold spells reported at several cities in Bulgaria. These extreme temperature events were found to occur with a small number of circulation *types*, a finding that can be helpful in understanding past variability and projecting future changes in the occurrence of extreme weather and climate events.

1. Introduction

Atmospheric circulation catalogs have historically been an important tool for investigating the role of large-scale atmospheric circulation on local and regional climate. These classifications collapse circulation patterns observed over a long period of record into a manageable number of categories or types. Differences in the frequency and character of the circulation types can then be associated with seasonal and interannual variations in local climate parameters such as temperature and precipitation or can be used to better understand spatial variations in climate conditions. Circulation classifications vary from those designed to be potentially applicable for a wide range of uses to those developed for specific regions and purposes (Huth et al., 2008). Two well-known examples of the former are the Hess–Brezowsky catalog (Baur et al., 1944; Hess and Brezowsky, 1969) which summarizes surface airflow over central Europe and the Lamb airflow types which are defined by airflow direction over the British Isles (Lamb, 1972). On the other hand, classifications have been developed to investigate very specific research questions such as the occurrence of temperature anomalies in the Greek Islands (Maheras et al., 2000; Kassomenos et al., 2003), heavy snowfall days in Andorra (Esteban et al., 2005), and heavy precipitation events in the Alps (Plaut et al., 2001).

The nature of the classification varies with the purpose for which it was developed. In particular, "circulation" can be defined in a number of different ways. Many of the early catalogs, such as the Lamb weather types, focused on airflow direction over a specified region. On the other hand, more recent classifications attempt to capture differences in the configuration of sea-level pressure fields (e.g., Plaut and Simonnet, 2001; Esteban et al., 2006; Philipp et al., 2007) or fields of upper-level (e.g., 500 hPa) geopotential height (e.g., Huth, 2000; Maheras et al., 2004; Casado et al., 2008; Anagnostopoulou et al., 2008). For some classifications, the

* Corresponding author. Fax: +1 303 497 7013.
E-mail addresses: Galina.Guentchev@noaa.gov (G.S. Guentchev), winkler@msu.edu (J.A. Winkler).
[1] Fax: +1 517 432 1671.

1474-7065/$ - see front matter
doi:10.1016/j.pce.2009.12.011

pressure or height fields are constrained to relatively small geographic regions such as the alpine area of Europe (Schüepp, 1979), whereas other classifications reflect continental-scale circulation (e.g., Philipp et al., 2007; James, 2007). More complex classification schemes incorporate circulation at multiple levels of the atmosphere (e.g., Maheras et al., 2000; Post et al., 2002; James, 2007) or combine location-specific surface air mass conditions with airflow characteristics (e.g., Kassomenos et al., 2003).

Another feature that varies greatly among the different circulation classifications is the number of categories or types. A recent inventory by Huth et al. (2008) revealed that the number of circulation types for classifications available for Europe ranged from 4 to 40 and was relatively uniformly distributed within that range. This finding is not unexpected as the multiple purposes for which circulation classifications are employed often require different levels of detail. For example, fewer categories are useful when the emphasis is on long-term trends in broadly-defined circulation characteristics such as the relative frequency or persistence of cyclonic and anticyclonic airflow (e.g., Kyselý and Domonkos, 2006; Kyselý and Huth, 2006). In contrast, a larger number of categories is useful when investigating day-to-day weather variability or the relationship between circulation and the occurrence of extreme weather events (e.g., Kyselý, 2002). However, the varying number of circulation types makes it difficult to compare circulation classifications, even when the classifications are based on similar circulation parameters and were developed for the same geographic area. Another complication occurs when using the same classification for multiple purposes. In many instances, both the temporal changes in broad-scale circulation and the influence of circulation on local climate conditions are of interest. In this case, either two circulation classifications need to be used, or alternatively, detailed circulation types need to be aggregated into a smaller number of "basic" categories (James, 2007) or a classification with a small number of types needs to be subdivided into a larger number of categories. Whereas the former approach is infrequently used, there are numerous instances of modifying existing classifications. Most often, detailed circulation types are subjectively grouped into broader categories (see, for example, Galambosi et al., 1996; Huth, 2001; Kyselý, 2002; Kyselý and Huth, 2006; James, 2007), introducing considerable subjectivity into what often was initially an objective (i.e., computer assisted) classification.

Here we recognize that a circulation classification is likely to be used for multiple purposes and develop a two-tier circulation classification for the European–North Atlantic region. The two, objectively- and independently-defined solutions represent different levels of circulation complexity. The coarse solution, which we refer to as *supertypes*, is applicable for more synthetic analyses, such as the long-term trends in broad-scale circulation. The fine solution is a detailed classification consisting of numerous circulation *types* and is applicable when investigating the relationships between circulation and local climates. The usefulness of the two tiers is illustrated by employing the first tier of the classification scheme to estimate the temporal changes in the frequency and persistence of large-scale circulation over the European region, while the second tier is used to identify the circulation patterns associated with summertime heat waves and wintertime cold spells reported at four observing stations in Bulgaria. Bulgaria was selected as the example country, as the role of circulation on local temperature extremes has not been as extensively studied as for other parts of Europe.

2. Data, study area, and methods

2.1. Development of the circulation classification

The circulation classification was developed from daily mean sea-level pressure (SLP) fields for January 1, 1951–December 31, 2004 from the NCEP/NCAR Reanalysis (Kalnay et al., 1996; Kistler et al., 2001). The reanalysis fields have a 2.5° latitude × 2.5° longitude spatial resolution. The study area, represented by 777 grid points (21 latitudinal rows by 37 longitudinal columns), is enclosed between 40°W and 50°E, and 20°N and 70°N. This broad geographic region was used so that the classification would be widely applicable for different regions across Europe.

Circulation *supertypes* and *types* were identified independently for each of four seasons, defined as winter (December–February), spring (March–May), summer (June–August), and fall (September–November). This seasonal development is similar to that recently used by Philipp et al. (2007), but differs from a number of other circulation classifications which were developed either for the entire year (e.g., Hess and Brezowsky, 1969; Lamb, 1972; Esteban et al., 2006) or for "cool" and "warm" seasons (e.g., Kassomenos et al., 2003; James, 2007). The finer seasonal definitions were used since the European–North Atlantic region, like other midlatitude regions, is characterized by seasonally diverse atmospheric circulation. Computer-assisted classification procedures are likely to emphasize circulation patterns that occur during those times of the year when variability is large and miss some of the nuances in the circulation during the period of the year (usually summer) when variability is small.

Similar to numerous previous classifications (e.g., Corte-Real et al., 1999; Plaut and Simonnet, 2001; Kassomenos et al., 2003; Esteban et al., 2006), a combination of principal component analysis (PCA) and cluster analysis was used to develop the classification. An S-mode PCA, where the rows (cases) were the days in each season and the columns (variables) were the grid points, was used to remove collinearity among the classification variables (i.e., grid points). The S-mode PCA was performed on a correlation similarity matrix, calculated from time standardized SLP grid point values. A Varimax orthogonal rotation was applied to facilitate interpretation. Using an iterative approach only rotated components with at least one loading >0.5 were retained for the cluster analysis. The number of retained components varied by season, i.e.: 11 PCs in winter (86.9% explained variance), 14 PCs in spring (87.9% explained variance), 16 PCs in summer (85.3% explained variance), and nine components in fall (80.8% explained variance).

The component scores were then subjected to a cluster analysis. A combination of hierarchical and non-hierarchical procedures was employed. The hierarchical procedures (average distance and Ward's minimum variance methods) were used to suggest the number of seeds or clusters using Euclidean distance as the measure of similarity between the observations. The cluster analysis was applied to non-standardized rather than standardized component scores following Johnson (1998), who points out that standardized scores do not realistically represent the distances between observations. Five stopping criteria were used to estimate the number of candidate clusters: (a) Cubic Clustering Criterion (Ray, 1982), (b) Hotelling Pseudo-t^2 statistic (Johnson, 1998), (c) the combined Pseudo-*F* and Pseudo-t^2 statistics (Fovell and Fovell, 1993), (d) the proportion of the explained variance (Kalkstein et al., 1987; Davis and Kalkestein, 1990), and (e) the distance between the clusters merged at each step (Wilks, 1995; Fovell and Fovell, 1993). Most of these rules recommend finding local peaks in the plots of the criteria and considering the cluster steps corresponding to or before these peaks as the appropriate numbers of clusters. Given that many of the plots feature multiple peaks at different levels of detail, the stopping rules yielded multiple candidate numbers of clusters per criterion for each hierarchical clustering method. Each of the candidate numbers of clusters identified by the five stopping criteria for the hierarchical procedures served as an estimate of the number of clusters in a non-hierarchical, k-means clustering procedure applied using the SAS FASTCLUS software. A randomization option was used that allowed for the selection of