Classifications of Winter Euro-Atlantic Circulation Patterns: An Intercomparison of Five Atmospheric Reanalyses

JAN STRYHAL

Department of Physical Geography and Geoecology, Faculty of Science, Charles University, Prague, Czech Republic

RADAN HUTH

Department of Physical Geography and Geoecology, Faculty of Science, Charles University, and Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

(Manuscript received 1 February 2017, in final form 19 June 2017)

ABSTRACT

Atmospheric reanalyses have been widely used to study large-scale atmospheric circulation and its links to local weather and to validate climate models. Only little effort has so far been made to compare reanalyses over the Euro-Atlantic domain, with the exception of a few studies analyzing North Atlantic cyclones. In particular, studies utilizing automated classifications of circulation patterns—one of the most popular methods in synoptic climatology—have paid little or no attention to the issue of reanalysis evaluation. Here, five reanalyses [ERA-40; NCEP-1; JRA-55; Twentieth Century Reanalysis, version 2 (20CRv2); and ECMWF twentieth-century reanalysis (ERA-20C)] are compared as to the frequency of occurrence of circulation types (CTs) over eight European domains in winters 1961–2000. Eight different classifications are used in parallel with the intention to eliminate possible artifacts of individual classification methods. This also helps document how substantial effect a choice of method can have if one quantifies differences between reanalyses. In general, ERA-40, NCEP-1, and JRA-55 exhibit a fairly small portion of days (under 8%) classified to different CTs if pairs of reanalyses are compared, with two exceptions: over Iceland, NCEP-1 shows disproportionately high frequencies of CTs with cyclones shifted south- and eastward; over the eastern Mediterranean region, ERA-40 and NCEP-1 disagree on classification of about 22% of days. The 20CRv2 is significantly different from other reanalyses over all domains and has a clearly suppressed frequency of zonal CTs. Finally, validation of 32 CMIP5 models over the eastern Mediterranean region reveals that using different reanalyses can considerably alter errors in the CT frequency of models and their rank.

1. Introduction

Atmospheric reanalyses represent a widely used tool in the research of climate. Reanalyses have evolved over the last two decades into what the community now accepts as a quasi-realistic representation of the evolution of the atmosphere spanning—depending on the dataset—from several years to more than a century. The reanalyses have been considered confident particularly in midlatitudes of the Northern Hemisphere, for which an abundance of observations were assimilated into reanalyses.

Recently, an increasing number of papers have been dealing with various aspects of large-scale circulation over Europe and the North Atlantic, such as its long-term variability [see Hertig et al. (2015) and references therein], recent trends (Kučerová et al. 2016), and effects of both on (inter alia) temperature and precipitation variables (Beck et al. 2007; Beranová and Huth 2008; Casado et al. 2010; Küttel et al. 2011; Plavcová and Kyselý 2012; Cahynová and Huth 2016). Not only reanalyses but also global climate model (GCM) and regional climate model (RCM) output have been scrutinized for various regions: Europe and the North Atlantic (e.g., Rust et al. 2010; Pastor and Casado 2012; Perez et al. 2014), Asia (Finnis et al. 2009b), North America (McKendry et al. 2006; Finnis et al. 2009a), Australia (Gibson et al. 2016), and polar regions (Cassano et al. 2006; Lynch et al. 2006). In this kind of study, reanalyses have been the most prominent source representing the reality. However, little to no attention has usually been paid to their evaluation

Corresponding author: Jan Stryhal, jan.stryhal@natur.cuni.cz

DOI: 10.1175/JCLI-D-17-0059.1

TABLE 1. List of atmospheric reanalyses used in the study. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

Reanalysis	Institution	Resolution of data (lon × lat)	Reference
ERA-40	European Centre for Medium-Range Weather Forecasts	2° × 2°	Uppala et al. (2005)
NCEP-1	National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR)	2.5° × 2.5°	Kalnay et al. (1996)
JRA-55	Japan Meteorological Agency	1.25° × 1.25°	Kobayashi et al. (2015)
20CRv2	NOAA/Earth System Research Laboratory, University of Colorado CIRES Climate Diagnostics Center	2° × 2°	Compo et al. (2011)
ERA-20C	European Centre for Medium-Range Weather Forecasts	2° × 2°	Poli et al. (2016)

(against nonassimilated observations or other reanalyses) and discussion on how their selection may have affected the results.

Moreover, only few studies aimed to intercompare reanalyses. Additionally, these studies primarily focused on regions with large observational uncertainty, such as the Arctic and the Antarctic (e.g., Bracegirdle and Marshall 2012; Lindsay et al. 2014; Nygård et al. 2016), midlatitudes of the Southern Hemisphere (Bromwich and Fogt 2004), and the tropics (Stickler and Brönnimann 2011; Kumar et al. 2013). Global studies indicate that the agreement among reanalyses is closest over regions with densest data coverage, such as Europe and North America (Greatbatch and Rong 2006; Wang et al. 2006), and claim that reanalyses agree fairly well on the wave activity in the synoptic and low frequencies in the northern extratropics after 1980 (Dell'Aquila et al. 2016). As far as atmospheric circulation over the Euro-Atlantic domain is concerned, only a handful of studies compared selected circulation features—mostly cyclones and storm tracks—in two or more reanalyses. Trigo (2006) analyzed North Atlantic cyclones in ERA-40 and NCEP-1 (for the explanation of abbreviations and more information on the reanalyses, see Table 1) and found discrepancies that were primarily attributed to different spatial resolution of assimilation models. Hanson et al. (2004) found out that there were only weak correlations between time series of cyclone frequency—in particular for low-intensity cyclones—derived from NCEP-1 and a slightly extended variant of ERA-15. Kouroutzoglou et al. (2011) corroborated the crucial effect of grid resolution in their analysis of Mediterranean cyclones in ERA-40 data. For more information on cyclone representation by reanalyses and GCMs, readers are also referred to Ulbrich et al. (2009) and Wang et al. (2016). Comparisons of reanalyses in the Euro-Atlantic domain were also conducted for the North Atlantic Oscillation, which was shown to differ only negligibly between NCEP-1 and ERA-40 (Greatbatch and Rong 2006). Unlike cyclones, no study has so far utilized circulation classifications (classifications of atmospheric circulation patterns) toward evaluating reanalyses.

Classifications represent a different approach to analyze atmospheric circulation. This tool has been widely used in synoptic climatology to describe the highly variable circulation—usually expressed by daily or monthly mean sea level pressure (SLP) or geopotential height (GPH) patterns—by a relatively low number of circulation types (CTs). Both the definition of CTs and the attribution of patterns to the CTs can be achieved by various statistical methods; for their review, see Huth et al. (2008). Only a few studies have, nevertheless, used more than one reanalysis to define CTs or compare the circulation statistics (Rust et al. 2010; Belleflamme et al. 2013; Perez et al. 2014; Gibson et al. 2016); moreover, the primary goal of these studies was validation of GCMs and not intercomparison of reanalyses. Other studies have arbitrarily used either ERA-40 or NCEP-1 as quasi observations.

For the North Atlantic domain, Perez et al. (2014) defined 100 CTs by *k*-means clustering of principal components based on 3-day averaged SLP anomalies in NCEP-1 data and, subsequently, projected these CTs onto ERA-40 and Twentieth Century Reanalysis, version 2 (20CRv2), data and a set of GCMs. Although the authors claimed the distribution of the frequency of occurrence across the CTs to be ''similar'' for the three reanalyses, quantitative indices showed nonnegligible differences: for annual data, root-mean-square errors of the relative frequency of CTs in ERA-40 and 20CRv2 relative to NCEP-1 were 0.16% and 0.26%, respectively, while over 20% of GCMs scored under 0.5%, with the minimum value being 0.33%. For winter, the respective values were 0.34% and 0.39%, while the best GCMs scored slightly over 0.5%. These results suggest that when evaluating CTs in GCM output, we need to better assess the observation uncertainty, especially if future generations of GCMs produce increasingly more reliable circulation climatologies. For a similar domain, Rust et al. (2010) showed that the differences between ERA-40 and NCEP-1 in shapes of CT centroid patterns and CT frequencies depended on the imposed cluster shape, being a result of both the definition of not well-separated clusters and suboptimal number of CTs. The former issue relates to the fact that the atmospheric circulation is rather a continuum than a set of well distinguishable states (Philipp et al. 2016). Consequently, results of synoptic-climatological studies tend to depend on the selection of classification criteria including the classification method itself. Therefore, the need for a parallel usage of more than one method has been stressed several times (e.g., Huth et al. 2008) in order that reliable results are obtained and artifacts of individual methods not overinterpreted.

In the study, multiple circulation classifications are used to define CTs in five global reanalyses over the Euro-Atlantic domain and its seven subdomains. Although CTs in reanalyses undoubtedly differ throughout the whole year, all analyses are conducted only for winter (DJF). Winter has been an extensively studied season because the links between large-scale circulation (including CT frequencies) and local-scale climatic elements are strongest in winter over various Euro-Atlantic regions (see, e.g., Beck et al. 2007; Pasini and Langone 2012; Plavcová and Kyselý 2013; Broderick and Fealy 2015; Cahynová and Huth 2016). Consequently, studying the uncertainty of winter reference circulation data is of utmost importance, as errors in the data and faulty assumptions regarding the data could negatively affect the results of many studies.

The paper is organized as follows: The datasets and methods are described in section 2. In section 3, the main results are presented and discussed. The paper seeks answers to the following questions: 1) Do different reanalyses have notably different CT frequencies over any of the tested domains? 2) Do the eventual differences depend on the classification method used to define the CTs? 3) Can notably different results be obtained in a GCM validation if different reanalyses are used as a benchmark? The main conclusions are presented in section 4.

2. Data and methods

a. Reanalysis data

Five global reanalyses are used in the study (Table 1). The selection was influenced by the choice of the time period (DJF 1960/61–1999/2000), which in turn respects the time scale typically used in synoptic-climatological studies. Although new generations of reanalyses, such as ERA-Interim, were shown to better represent reality, inter alia owing to the inclusion of satellite data, their shorter span considerably limits trend analyses. Therefore, the original ERA-40 and NCEP–NCAR reanalysis

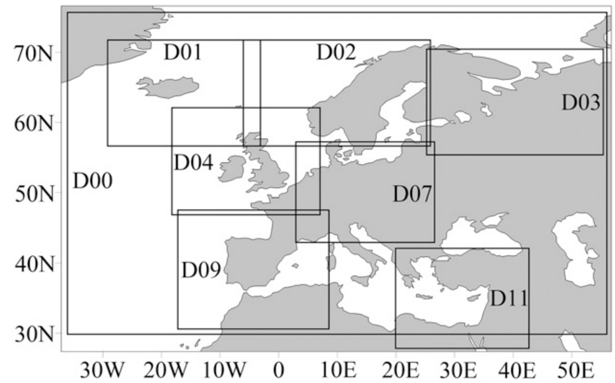


FIG. 1. Location of spatial domains over which classifications were calculated and reanalyses compared.

(NCEP-1), as well as the relatively newer JRA-55, are and will be widely used. Moreover, the recent reanalyses, 20CRv2 and ECMWF twentieth-century reanalysis (ERA-20C), are becoming popular in the research of long-term climate variability; they need to be assessed against more "traditional" reanalyses as they assimilate only a few surface variables. This attribute has been proven useful because the reanalyses are not constrained by inhomogeneities in the data type assimilated (upper-air and satellite observations). Nevertheless, the data are still constrained by changes in the density of observations, notably over oceans and farther in the past (Gibson et al. 2016).

The reanalyzed daily mean SLP patterns were interpolated by bicubic splines onto the longitude–latitude grid of $3° \times 2°$ over the Euro-Atlantic domain (D00) and $1° \times 1°$ over its seven subdomains (Fig. 1): Iceland (D01), western Scandinavia (D02), northeastern Europe (D03), British Isles (D04), central Europe (D07), western Mediterranean (D09), and eastern Mediterranean (D11). Note that these domains are a subset of domains defined within the European Cooperation in Science and Technology Action 733 (COST733). To aid comparison with other studies, the spatial extent and codes of the domains and also the spatial resolution of the classified patterns and the number of CTs follow conventions introduced in the action. For more information on the action, refer to editorials of special issues by Huth et al. (2010) and Tveito and Huth (2016).

b. Circulation classifications

The research method used in the study is that of classifications of circulation patterns. In this case, the classified patterns are the reanalyzed gridded daily mean SLP maps. The general goal of a circulation classification is to substitute a wide variety of patterns with a few CTs (forming a so-called catalog of CTs), which would simplify the complexity of atmospheric circulation and

TABLE 2. List of classification methods used in the study.

Method abbreviation	Method name	No. of CTs	Reference
GWT	Grosswettertypes	10	Beck et al. (2007)
JCT1	Jenkinson–Collison	10	Jones et al. (1993)
JCT2			
LND	Lund	9	Lund (1963)
PCT	T-mode PCA obliquely rotated	9	Huth (1993)
CKM	k-means by dissimilar seeds	9	Enke and Spekat (1997)
SAN	Simulated annealing and diversified randomization (SANDRA)	9	Philipp et al. (2007)
KMD	k-medoids	9	Kaufman and Rousseeuw (1990)

thus facilitate its analysis. There are many approaches how CTs can be defined and the patterns classified to these CTs (so-called classification methods), each resulting in a unique classification result. For a detailed review of classification methods, readers can refer to Huth et al. (2008) and Philipp et al. (2016).

To obtain reliable results, a total of eight methods were used in the study (Table 2). The selected methods are among the most widespread in recent literature, and all main automated classification approaches are represented by at least one method. Moreover, all these methods are included in the COST733 classification software version 1.2, which is freely available online (http://cost733.geo.uni-augsburg.de/cost733wiki). A brief description of the methods follows.

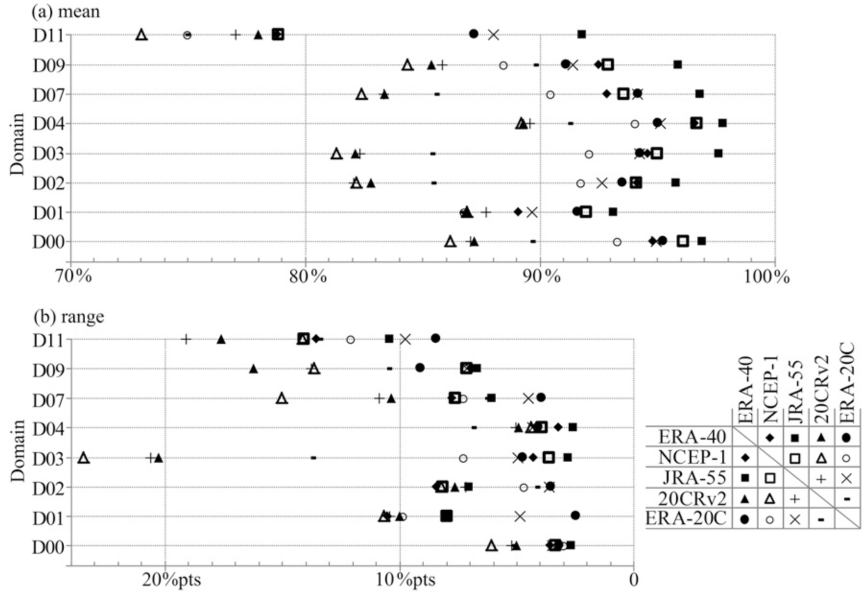
GWT, JCT1, and JCT2 represent hybrid (also threshold based) methods, which subjectively predefine CTs and, subsequently, automatically assign patterns using threshold values of certain indices, such as vorticity and direction of airflow. In our case, the catalogs consist of eight directional types—one for advection from each directional octant [further referred to as west (W), northwest (NW), north (N), northeast (NE), east (E), southeast (SE), south (S), and southwest (SW)]—one cyclonic CT (C), and one anticyclonic CT (A). The remaining methods define CTs automatically as part of the classification process; thus, the resulting classification mirrors only the ability of the respective algorithm to divide the data cloud into clusters. Consequently, these methods are somewhat more objective, although multiple more or less subjective choices (such as defining the number of CTs) still have to be made prior to the classification. LND is the oldest automated method and one of the so-called leader-algorithm-based classifications. It finds key (leader) patterns that well represent (i.e., highly correlate with) relatively large groups of individual patterns. PCT is a method based on principal component (PC) analysis with the input data matrix in a T mode [i.e., grid points correspond to columns of the data matrix and time realizations (days) to its rows], followed by the direct oblimin rotation of PCs. The scores of the rotated PCs represent spatial structure (maps) of CTs, and their loadings are used to assign the patterns to classes. For more information on rotation of PCs and modes of PCA, refer to Richman (1986) and Compagnucci and Richman (2008). CKM, SAN, and KMD are algorithms of nonhierarchical cluster analysis, also called optimization methods, since they incorporate steps that help find a solution closer to the globally optimal partitioning (that with minimum within-type variance) for the number of CTs selected in advance.

It is clear that not all results can be shown in the paper and that a different approach is required to analyze results than to simply interpret individual classifications. Instead, the approach we propose uses simultaneously all 75 CTs (eight catalogs of 9 or 10 CTs each), regardless the classification. These CTs are plotted and analyzed together, and differences in their frequency of occurrence in each pair of reanalyses are quantified in a far more robust way than if based on a single classification. Furthermore, this approach makes it possible to group CTs with similar centroid patterns and elicit whether the differences between reanalyses are common to various classifications (i.e., systematic) or rather accidental. To avoid the necessity to project CT centroids from one reanalysis to another (in other words, to assign daily patterns from one dataset to CTs that were sooner defined on another dataset), the classifications are applied on all five reanalyses together (i.e., on 18 000 daily patterns). In section 3f, nevertheless, a projection will be utilized for classification of outputs of a GCM ensemble in order to illustrate the effect of the choice of the reanalysis on GCM validation. This approach is highly beneficial since projection is often several orders of magnitude faster than running classifications on a large number of datasets.

3. Results and discussion

a. Intercomparison of reanalysis datasets

A straightforward way to assess the congruency of classifications in different datasets is to compute the



(c) number of significantly different classifications

	ERA-40	NCEP-1	JRA-55	20CRv2	ERA-20C
ERA-40		1 4 0 0	0 3 0 0	8 4 7 4	0 0 0 0
NCEP-1	0 1 0 8		0 1 0 0	8 5 8 4	1 5 0 3
JRA-55	0 0 0 0	0 1 0 8		8 2 8 4	0 4 0 0
20CRv2	4 7 6 8	6 6 8 8	4 7 6 8		8 4 6 4
ERA-20C	0 0 4 3	0 4 8 8	0 0 5 4	3 8 3 8	

D00 D01 D02 D03 (above diagonal)
D04 D07 D09 D11 (below diagonal)

FIG. 2. Differences of classifications in reanalyses: (a) percent of days classified with the same CT; each value is an average of eight classifications for the respective domain. (b) As in (a), but for the range of classifications (in percentage points). (c) Number of classifications with significantly different CT frequencies based on the chi-square test and the 5% level. Cases with collective significance of 0.01 are in boldface.

relative frequency of days that are classified with different CTs by a pair of reanalyses. A more common approach is to compare CT frequencies; however, even a near-perfect fit of the frequencies does not rule out the possibility that a certain amount of days is classified differently if the classes are not unambiguously constrained, which is rarely the case in the atmosphere. One would expect that differences in assignment would be relatively rare events related to circulation patterns far from CT centroids, at least over regions with enough integrated observations. The percentages for all pairs of reanalyses and domains are shown in Fig. 2a; the results are composites of all classifications. There is indeed a good agreement among ERA-40, NCEP-1, and JRA-55 (less than 8% of days classified differently) except for D01 and D11. Over D01, ERA-40 and NCEP-1 disagree on about 11% of days. Over D11, ERA-40 and JRA-55 are alike; however, NCEP-1 differs from both ERA-40 and JRA-55 in about 22% of days. The two twentieth-century reanalyses show different behavior. ERA-20C seems to be very consistent with ERA-40 and JRA-55, even over D01 and D11. On the other hand, 20CRv2 leads to considerably differing classifications compared to all four remaining reanalyses. This suggests that ERA-20C might be closer to reality and, therefore, more appropriate to be used to classify CTs should the recent-climate reanalyses have too limited a time span. However, it is not possible to test the validity of this hypothesis for the period prior to 1957 using the traditional reanalyses such as ERA-40. Moreover, it has to be remembered that an accord among reanalyses does not necessarily mean that the reanalyses are not biased from reality.

While Fig. 2a provides a robust estimate of the accordance between reanalyses by showing the average of eight classifications, Fig. 2b illustrates to what extent the accordances can differ between individual classifications. The values show the range (in percentage points) between the classification with the highest accordance and the classification with the lowest accordance. For example, the range of accordances between NCEP-1 and 20CRv2 over D03 is about 23 percentage points with the minimum of 67% in KMD and the maximum of 90% in LND. Such a wide range indicates that there is some kind of difference between the two datasets that only cluster analysis methods detect. In general, the wider the

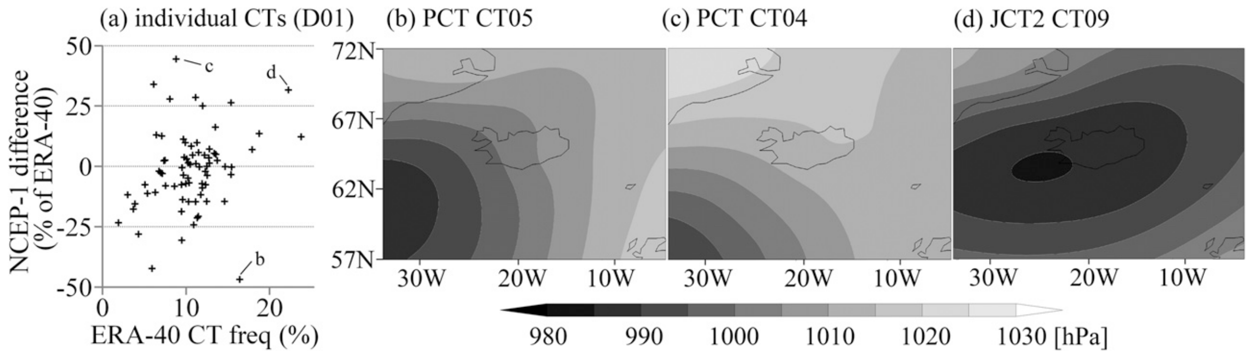


FIG. 3. Comparison of ERA-40 and NCEP-1 over D01: (a) relative frequency of CTs in ERA-40 and respective NCEP-1 anomalies (in percent of ERA-40); (b)–(d) centroids of selected CTs. The three values denoted by "b," "c," and "d" in (a) refer to the respective CTs depicted in (b)–(d). Note that in (a) all CTs are plotted together regardless the classification.

range is, the stronger the need for more classifications, for example, to assess the significance of differences between two reanalyses. The differences between classifications will be analyzed in detail in section 3e.

Figure 2c shows how many classifications (out of eight) have statistically different frequencies of CTs. The significance is tested separately for each domain and each pair of reanalyses using the chi-square test at the 5% level. More than 40% of these tests (640 tests in total) detect a significant difference. ERA-40, NCEP-1, and JRA-55 usually do not significantly differ, except for D11 and D01. On the other hand, significant differences are much more numerous when 20CRv2 and D11 are involved in comparisons (75% and 79%, respectively). There are several cases in which the frequencies significantly differ only in one classification (e.g., JCT2 over D07 between ERA-40 and NCEP-1). However, to declare circulation in two reanalyses significantly different, one positive chi-square test out of eight is not enough; see, for example, the multiplicity problem for independent tests by Wilks (2006). With eight tests at the 5% level, declaring collective significance at the 10% (1%) level requires at least two (three) local tests to be positive.

The following sections further analyze the most significant differences between the reanalyses. First, in section 3b, the somewhat striking difference between reanalyses over D01 is further analyzed and discussed. We focus on ERA-40 and NCEP-1 since these two reanalyses are most widespread in studies of North Atlantic and since D01 is the only domain where CT frequencies of these two reanalyses significantly differ (except D11, which will be shown in a separate section). Second, in sections 3c and 3d, respectively, results obtained for 20CRv2 and D11 are analyzed. Third, section 3e focuses on differences between classification methods, and, finally, section 3f illustrates how the choice of reanalysis can influence results of GCM validation.

b. ERA-40 and NCEP-1 over Iceland

The difference in CT frequency in NCEP-1 relative to ERA-40 is, on average, about 12% over D01; however, some CTs—including several rather frequent ones—deviate considerably more (Fig. 3a). Among all classifications, PCT detects the most profound differences. The most frequent CT in ERA-40—PCT circulation type 5 (CT05), depicted in Fig. 3b—has in NCEP-1 about half the frequency compared to ERA-40. When patterns assigned to PCT CT05 are assigned differently in NCEP-1, the latter favors either CT04 with the low shifted southward and a ridge over and east of Greenland (Fig. 3c) or one of two CTs that place the cyclone farther eastward, closer to Iceland (not shown). The latter deviation concurs with an increased frequency of CTs with the cyclone close to the center of the domain in NCEP-1 (see, e.g., Fig. 3d).

The described differences between ERA-40 and NCEP-1 over Iceland seem to concur with Trigo (2006), who found considerable differences in the ability of these reanalyses to capture the frequency and spatial distribution of cyclogenesis and the location where cyclones reach minimum SLP during winter. While the maxima were localized southwest of Iceland in NCEP-1, they extended over the whole Denmark Strait and along the eastern coast of Greenland in ERA-40—compare Figs. 1 and 2 in Trigo (2006). The differences were attributed primarily to the coarser horizontal resolution of the NCEP-1 integration model. Although circulation pattern classifications should theoretically be less affected by reanalysis model resolution than cyclone-tracking algorithms, the results suggest that in some cases CT frequencies do differ over D01. It is, however, debatable what the effect of resolution in this case is

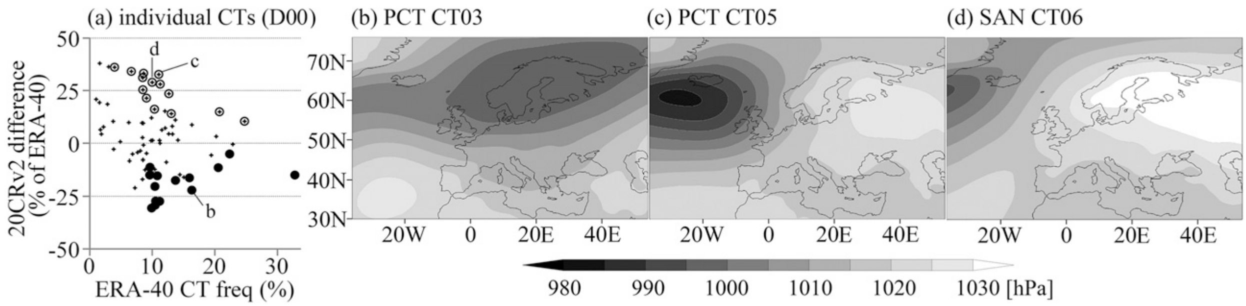


FIG. 4. As in Fig. 3, but comparing ERA-40 with 20CRv2 over the Euro-Atlantic domain (D00). In (a), filled circles highlight zonal CTs with W and NW advection, and open circles highlight CTs with high pressure over the continent.

since the finer-resolution JRA-55 produces results similar to NCEP-1; moreover, a coarser-resolution version of ERA-40 was used in the present study, which should obliterate the benefits of the finer resolution if there are any. To conclude, the differences between ERA-40 and NCEP-1 CTs over D01 seem to reflect real biases rather than artifacts of the classification methodology.

c. 20CRv2

Figure 2 indicates that classifications in 20CRv2 deviate from the remaining datasets considerably more than any other reanalysis and that this behavior is spatially consistent. Excluding 20CRv2, reanalyses classify differently, averaged over all pairs of reanalyses and all domains, about 7.7% of days. On the other hand, 20CRv2 classifies differently, on average, about 15.5% of days; that is, twice as many. The ratio grows larger to 2.4 if ERA-20C is excluded, owing to the overall good agreement between ERA-40, NCEP-1, and JRA-55. These results document that using classifications based (only) on the 20CRv2 ensemble mean is problematic even for regions with relatively dense observation networks such as central Europe. It seems, therefore, advisable to utilize the whole set of ensemble runs (a unique feature of the reanalysis is that it also includes a 56-member ensemble) instead of using only the ensemble mean in synoptic-climatological studies restricted to 20CRv2 data, in order to account for the observation uncertainty.

Consequent to its anomalous behavior is the question whether 20CRv2 is biased in favor of CTs with certain properties (e.g., strong vorticity or direction of flow). To answer this question, 20CRv2 is compared in detail with ERA-40. Note that substituting ERA-40 with JRA-55 or NCEP-1 would lead to very similar results. Over D00, the difference in the CT frequency in ERA-40 and 20CRv2 is, on average, 14%; individual CTs are plotted in Fig. 4a. The highlighted CTs show that there is a tendency in 20CRv2 data—independent of the classification method—toward a lower frequency of W and NW CTs, such as PCT CT03 (Fig. 4b), and a higher frequency of CTs with high SLP over the continent, especially over northeastern Europe, such as PCT CT05 (Fig. 4c) and SAN CT06 (Fig. 4d). Note that the codes used here to describe the CTs (e.g., W for western advection) are analogous to codes of 10 CTs defined by hybrid methods (see section 2b). The attribution of CTs to the 10 groups is based on the shape of CT centroids and is a result of authors' expert judgement guided by pattern correlations of individual CTs with the CTs defined by hybrid methods.

Results for D02 and D03 (Fig. 5) corroborate those for D00. The 20CRv2 clearly underestimates the frequency of CTs with cyclones along the western and northern coast of Scandinavia, over the White Sea, and Karelia. On the contrary, anticyclonic CTs are more frequent over both domains in 20CRv2. Higher SLP over northeastern Europe increases the frequency of southerly advection over Scandinavia, as well as advection from the whole eastern (southeastern) quadrant over D07 (D09). Zonal advection from over the North Atlantic is suppressed over D02, D07, and D09 in 20CRv2. All these differences are, on average, at the rate of approximately 15%–25% of respective ERA-40 values and are apparent in all classifications, although individual values vary depending on the method. Groups of CTs with consistent differences in frequency in 20CRv2 and ERA-40 are highlighted in Fig. 5 for D02, D03, D07, and D09 by filled and open circles.

d. Eastern Mediterranean

Among the evaluated domains, reanalyses differ the most over D11. Only ERA-40, JRA-55, and ERA-20C have relatively similar CT statistics. The best agreement is between ERA-40 and JRA-55: the mean absolute difference of CT frequency in JRA-55 relative to ERA-40 is less than 5% (see Fig. 6a), about 8% of days differ in their classification (Fig. 2a), and significance testing failed to return a positive test in any classification (Fig. 2c). NCEP-1 disagrees with other reanalyses on

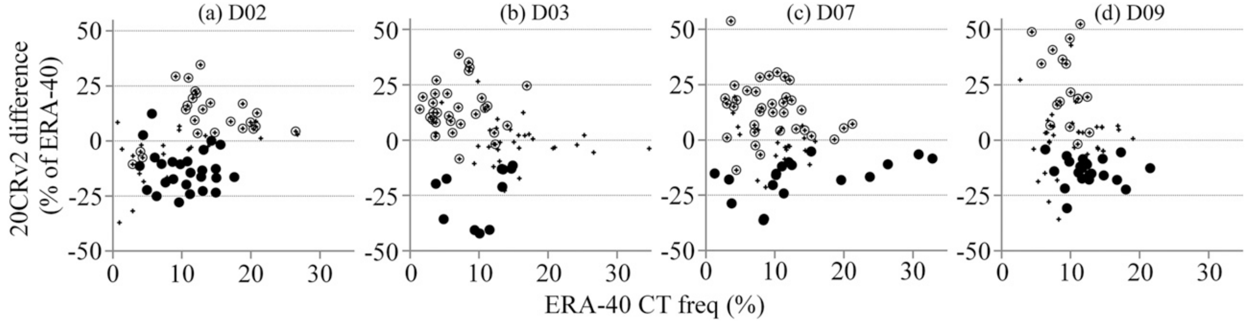


FIG. 5. Comparison of ERA-40 and 20CRv2 over selected subdomains (a) D02, filled circles: CTs with N and NW advection and C CTs (N+NW+C; see text for further explanation) and open circles: S+SW+A; (b) D03, filled circles: C and open circles: A+N+NE+E; (c) D07, filled circles: W+C and open circles: A+NE+E+SE+S; and (d) D09, filled circles: W+NW+N and open circles: S+SE+E.

classification of more than 20% of daily patterns. Figure 6 compares CT frequencies in selected pairs of reanalyses. There is a clear difference in the frequency of CTs with advection from the SE quadrant (low pressure over the central Mediterranean region) and from N and NE quadrants (low pressure to the east of the domain): the former (latter) direction is less (more) frequent in NCEP-1 and especially in 20CRv2 data, compared to ERA-40, ERA-20C, and JRA-55. The averages of relative differences in CT frequency for pairs of reanalyses shown in Figs. 6b–d are in turn about 17%, 22%, and 29%.

The overall worse correspondence of reanalyses over D11 relative to other domains is somewhat expectable owing to generally weaker horizontal pressure gradients. Even minor (in absolute terms) differences in SLP patterns can be expected to have a relatively profound impact on the classification of the patterns. Several conclusions can be drawn from these results. First, circulation classifications can be seen as a powerful tool in investigating and quantifying differences between reanalysis datasets. Second, bearing in mind these differences, one should be careful when using classifications for synoptic-climatological studies of both the real climate and its model simulations.

e. Differences between classification methods

The differences between reanalyses shown in previous sections are to a large extent present in all classifications. There are, however, several cases with an interesting variability of results if these differences are quantified and significance tested. Three pairs of reanalyses are selected as examples: 20CRv2 versus ERA-40, NCEP-1 versus ERA-40, and ERA-20C versus JRA-55. Note that CTs defined by methodologically similar methods are grouped together since they produce catalogs that (in this case) tend to behave similarly.

The results in Fig. 7a show a spatial pattern similar to those in Fig. 2, with a better agreement between 20CRv2 and ERA-40 over D01 and D04 and worse over D11. However, the values obtained by different groups of methods—in particular by hybrid (GWT+JCTs) and cluster analysis (SAN+KMD+CKM) methods—considerably vary, especially over D01, D03, D09, and D11. Figures 7b and 7c illustrate how the frequency of individual CTs differs in the two datasets over D01 and D03,

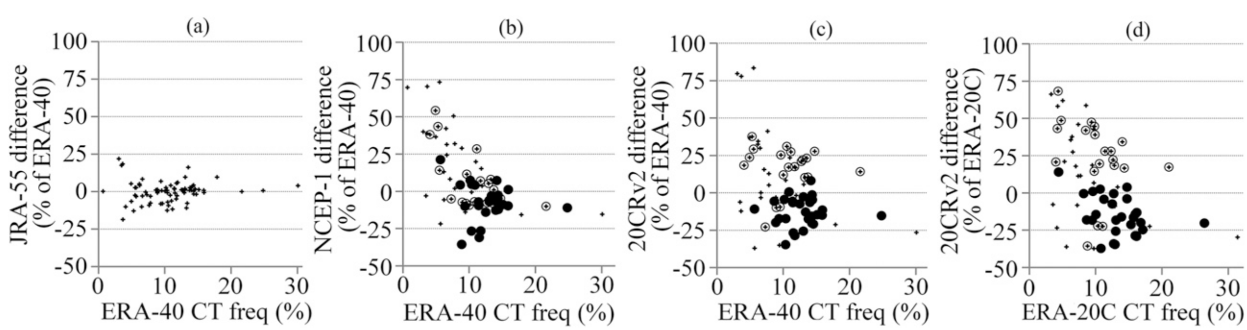


FIG. 6. (a)–(d) Comparison of CT frequency in reanalyses over D11, whereEach panel shows a different pair of reanalyses. In (b)–(d), filled circles highlight SW+S+SE CTs, and open circles highlight N+NE CTs. One and three outlying CTs (with frequency under 5% and overestimation of 105%–190%) are not shown in (c) and (d), respectively.

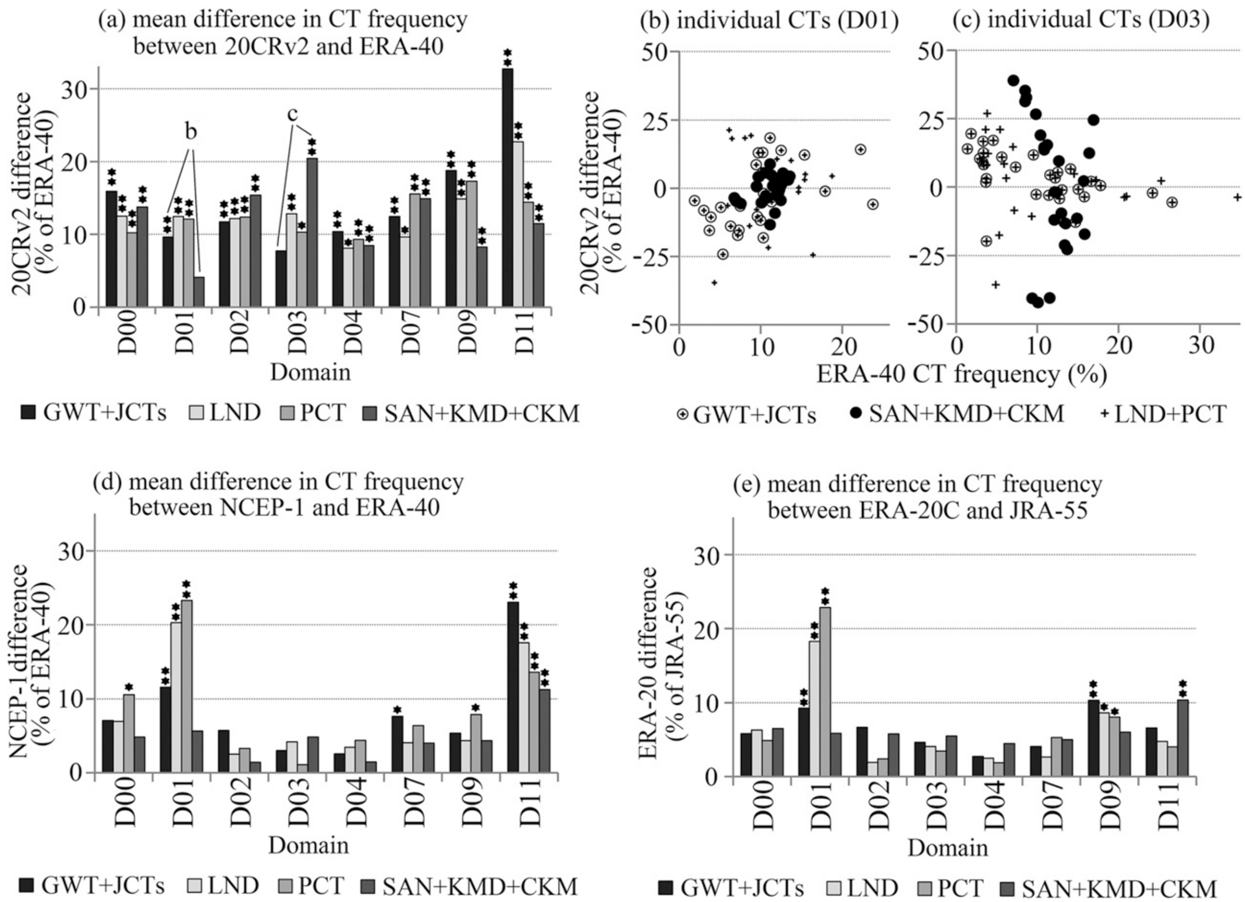


FIG. 7. The effect of classification methods on the difference of the CT frequency in reanalyses. (a) Spatial variability of the dependence of the mean absolute difference of the CT frequency in 20CRv2 and ERA-40 on classification methods. Note that methodologically similar methods were grouped together: hybrid (GWT+JCTs), LND, PCT, and cluster analysis (SAN+KMD+CKM). One and two asterisks indicate classifications with significantly different CT frequencies at 10% and 1% levels, respectively. (b) Differences in the relative frequency of individual CTs for D01 between 20CRv2 and ERA-40; filled (open) circles highlight CTs by cluster analysis (hybrid) methods. (c) As in (b), but for D03. (d) As in (a), but for NCEP-1 and ERA-40. (e) As in (a), but for ERA-20 and JRA-55.

respectively. Over D01, the cluster analysis methods define CTs that have fairly similar frequencies in both reanalyses—note the mean difference in Fig. 7a being only a fraction of the values obtained by the remaining methods and CT frequencies being significantly indifferent based on chi-square test. Over D03, cluster analyses yield far greater discrepancies between the datasets compared to other methods, while hybrid methods lead to indifferent CT frequencies. By repeating the same analysis for all remaining reanalysis pairs and domains (two more examples are shown in Figs. 7d,e), it becomes evident that no method gives systematically either too small or too large a difference between two reanalyses. The results agree with and enhance what Rust et al. (2010) showed for ERA-40 and NCEP-1, that is, that the difference in CT frequencies in two datasets depends on the shape of clusters of daily fields within the phase space imposed by the classification method. Here, the intercomparison of multiple methods shows that this dependence is spatially inconsistent and reflects the spatially varying ability of classification methods to separate the rather continuous data space into classes. Consequently, whether a certain method is or is not able to recognize existing differences is highly unpredictable. In reality, the clusters are functions of several factors, some being independent of the classification method (such as character of circulation), others chosen by the researcher or directly imposed by the method (e.g., number of CTs, method used to define CTs, and measure of similarity used to assign patterns). Undoubtedly, the measure of similarity (here, the Euclidean distance for clustering methods versus pattern correlation for LND and PCT versus various flow indices for hybrid methods) strongly influences which differences will be captured and which ignored. For instance, while spatial correlation cannot distinguish between patterns that have the same structure of isobars but differ in mean SLP, the Euclidean distance makes it possible to find differences in the mean SLP but may fail to detect differences in structure. A more in-depth analysis into which factors cause the particular differences between methods shown here would be beyond the scope of the paper.

To conclude, it is evident that using a single classification to compare two datasets and to analyze the spatial patterns of their mutual relationship (i.e., to say where the differences between reanalyses are smaller and where larger) cannot provide reliable results. Relying on one classification will likely cause two studies that utilize the same data but different classification methods to arrive at different or even contradictory conclusions.

f. *Case study: Validation of GCM output*

There is no doubt that the presented differences between reanalyses will to some extent influence the results of any analysis and that the extent will depend on the selection of (classification) methods, domains, and likely also on the research objective. To illustrate this issue, the following analysis tests the influence of the choice of reanalysis on validation of GCM winter circulation over D11. This analysis is one part of a broader ongoing research on the applicability of circulation classifications to validation of historical climate runs and interpretation of future climate runs by GCMs.

An ensemble of historical runs by 32 CMIP5 GCMs (Table 3) was accessed online (http://cmip-pcmdi.llnl.gov). See Taylor et al. (2012) for more information on the CMIP5 experiment. The simulated winter 1961–2000 daily mean SLP patterns were interpolated onto the same grid as the reanalyses. Subsequently, each catalog of CTs defined from reanalyses was projected onto the model data, resulting in eight classifications for each model. Finally, for each of the classifications, the relative frequencies of CTs and their errors with respect to the relative frequencies of the same CTs in each reanalysis were computed. The model errors are further evaluated in the same manner as were the differences between reanalyses in previous sections of the text; that is, errors of all 75 CTs are analyzed together. Medians of absolute values of the errors are used as a basis for model rankings, and since five reanalyses are used to compute the errors, five different rankings are created.

In Fig. 8, five median absolute errors and five rankings are shown for each model. There are several models that rank among the best or worst regardless the reanalysis (e.g., EC-EARTH, CMCC-CM, and MIROC4h on one hand, and BCC_CSM1.1, GFDL-ESM2M, and MIROC-ESM on the other). However, about one-third of the models display a high variability of the median errors and, consequently, also the rankings; note, for example, HadGEM-CC, MRI-CGCM3, and MRI-ESM1 for which the rankings differ extremely even between ERA-40 and NCEP-1. In the analysis, medians were used rather than averages to limit the effect of outliers; CTs occurring rarely in reanalyses can be vastly overestimated by some models. Basing the model rankings on mean errors or on fewer classifications can lead to considerably different results for some models (not shown). Therefore, one ought to be cautious when evaluating circulation in GCM output as relatively minor changes in the experiment setup—such as replacing one classification method or one reanalysis for another—can potentially lead to diverging results.

4. Conclusions

The main goal of the paper was to compare daily SLP patterns produced by five global reanalyses for the Euro-Atlantic region. We aimed at the winter season since during winter the links between the large-scale circulation and climatological elements are strongest over the domain, and, therefore, synoptic-climatological studies have preferably focused on this season. So far, studies have compared reanalyses over regions with large observation uncertainty, since it has been presumed that differences between reanalyses are negligible, and so not worth looking at, over regions with abundant observations. The present study suggests that both the differences between reanalyses and the effect of the choice of reanalyses on results may have been underestimated in synoptic climatology.

The article aimed to address three questions: 1) Does the CT frequency differ between reanalyses over Europe and the North Atlantic? 2) Do the differences between reanalyses depend on the classification method? 3) Does the choice of reference reanalysis influence results of GCM validation? The questions are answered in the following three paragraphs.

ERA-40, NCEP-1, and JRA-55 agree on classification of most days relatively well (less than 8% of days are classified with different classes if pairs of reanalyses are compared) except for Iceland (D01) and the eastern Mediterranean (D11). Over D01, NCEP-1 differs from ERA-40 in the frequency of cyclonic CTs; cyclones seem to be displaced southward and eastward in NCEP-1 relative to ERA-40. Over D11, NCEP-1 differs from both ERA-40 and JRA-55 in the classification of about 22% of days: relative frequency of CTs with advection

TABLE 3. List of GCMs used in the study.

Model name	Institution abbreviation	Ensemble member	Modeling center or group
BCC_CSM1.1	BCC	r1i1p1	Beijing Climate Center, China Meteorological Administration
CanESM2	CCCma	r1i1p1	Canadian Centre for Climate Modelling and Analysis
CCSM4	NCAR	r1i2p1	National Center for Atmospheric Research
CESM1(CAM5)	NSF–DOE–NCAR	r1i1p1	Community Earth System Model contributors
CMCC-CESM	CMCC	r1i1p1	Centro Euro-Mediterraneo per I Cambiamenti Climatici
CMCC-CM	CMCC	r1i1p1	
CMCC-CMS	CMCC	r1i1p1	
CNRM-CM5	CNRM–CERFACS	r1i1p1	Centre National de Recherches Météorologiques–Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique
CSIRO-Mk3L-1.2	CSIRO–QCCCE	r1i2p1	Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence
EC-EARTH	EC-EARTH	r1i1p1	EC-EARTH consortium
FGOALS-g2	LASG–CESS	r1i1p1	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, and Center for Earth System Science, Tsinghua University
GFDL CM3	NOAA/GFDL	r1i1p1	NOAA/Geophysical Fluid Dynamics Laboratory
GFDL-ESM2G	NOAA/GFDL	r1i1p1	
GFDL-ESM2M	NOAA/GFDL	r1i1p1	
HadGEM2-AO	NIMR/KMA	r1i1p1	National Institute of Meteorological Research/Korea Meteorological Administration
HadCM3	MOHC	r1i1p1	Met Office Hadley Centre
HadGEM2-CC	MOHC	r1i1p1	
HadGEM2-ES	MOHC	r5i1p1	
INM-CM4.0	INM	r1i1p1	Institute of Numerical Mathematics
IPSL-CM5A-LR	IPSL	r6i1p1	L'Institut Pierre-Simon Laplace
IPSL-CM5A-MR	IPSL	r3i1p1	
IPSL-CM5B-LR	IPSL	r1i1p1	
MIROC4h	MIROC	r1i1p1	Atmosphere and Ocean Research Institute (The University of Tokyo),
MIROC5	MIROC	r1i1p1	National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
MIROC-ESM	MIROC	r1i1p1	Japan Agency for Marine-Earth Science and Technology, Atmosphere
MIROC-ESM-CHEM	MIROC	r1i1p1	and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies
MPI-ESM-LR	MPI-M	r1i1p1	Max Planck Institute for Meteorology
MPI-ESM-MR	MPI-M	r1i1p1	
MPI-ESM-P	MPI-M	r1i1p1	
MRI-CGCM3	MRI	r1i1p1	Meteorological Research Institute
MRI-ESM1	MRI	r1i1p1	
NorESM1-M	NCC	r1i1p1	Norwegian Climate Centre

from the north and northeast tends to be higher in NCEP-1, while advection from the southern quadrant tends to be less frequent. Classifications in ERA-20C are quite consistent with those in ERA-40 even over D01 and D11. The 20CRv2, on the other hand, leads to considerably—and often significantly—different classifications; relative to the four remaining reanalyses and averaged over all eight domains, it classifies differently over 15% of days. Furthermore, over the Euro-Atlantic domain (D00), it is biased in favor of CTs with high SLP over the continent, whereas the frequency of CTs with zonal advection is underestimated. These biases were further shown to correspond with differences in the frequency of CTs defined for individual geographical domains. This case demonstrates that recently produced reanalyses that stretch farther and farther into the past should be used with utmost caution.

Eight classification methods were used in the study. This choice makes it possible to select and describe only those differences between reanalyses that are present in multiple classifications, therefore being very likely related to real features and not statistical artifacts of particular methods. It is evident that profound differences between two datasets are detected in multiple classifications. Nevertheless, if the differences between the datasets are quantified, one can get considerably diverging results if one uses different classification methods, which can lead to even completely contradictory interpretations in some cases. Therefore, one should avoid relying not only on one reanalysis but also

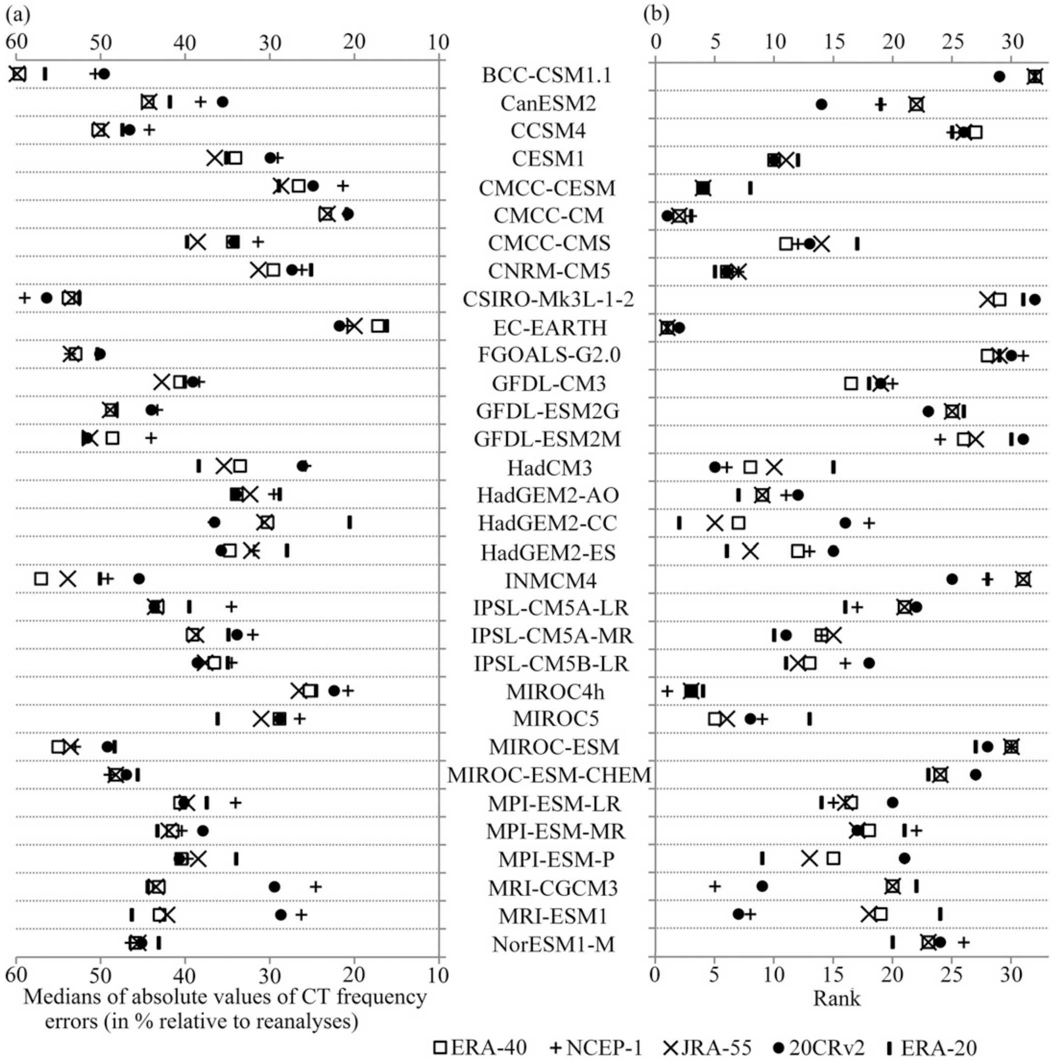


FIG. 8. Validation of 32 CMIP5 GCMs according to their ability to simulate the frequency of winter CTs over D11. (a) The symbols show the dependence of the median error in CT frequency of the respective model on the reference reanalyses. Each median is computed from absolute values of errors in the frequency of 75 CTs, the errors being expressed in percent of the CT frequency in the respective reanalysis. (b) As in (a), but for rank based on the median errors.

on one classification. Using several statistically similar methods, such as several algorithms of cluster analysis, does not seem to have a tangible effect, since the catalogs of CTs produced by similar methods are very similar. Contrariwise, using a relatively small set of distinct classifications based on different families of algorithms is much more advisable as it can identify various kinds of differences between datasets. Our results corroborate conclusions of previous studies that relying on one classification (method) in synoptic-climatological studies is dangerous. We suggest that it is used with utmost caution and is reserved only for special situations. In the context of the present study, such situations might be, for example, theoretical studies such as an in-depth analysis of causes of differences between datasets (which would help us interpret these differences correctly in the future) and how factors such as the definition of the spatial domain and the number of CTs influence these differences.

Utilizing output of historical runs of an ensemble of 32 CMIP5 GCMs, it was illustrated that the choice of different reanalyses can have a profound effect in GCM validation over D11 in winter. The rank of several models heavily depends on the benchmark reanalysis; in some cases even changing ERA-40 for NCEP-1 can lead to shifts in rankings by as many as 10–15 positions out of

32 (e.g., HadGEM2-CC, MRI-CGCM3, and MRI-ESM1). For other European domains, the effect can be expected to be less substantial than for D11; however, for most regions around the globe it will likely be much stronger. To conclude, the uncertainty intrinsic to atmospheric reanalyses must not be neglected, not only for those parts of Earth for which we have a minimum of in situ measurements but even for regions with an abundance of observations. All presented results suggest that 20CRv2 should not be used as a reference dataset. However, to say which of the remaining reanalyses is closest to reality is not possible without a direct comparison with independent observations. If such a comparison is not available, using an ensemble of reanalysis datasets should become a norm. This recommendation seems not only relevant for CT-based studies but also needs to be addressed by the downscaling community. The choice of reanalysis data in downscaling was shown to be significant at lower latitudes and in regions with sparse observation networks (e.g., Brands et al. 2012; Manzanas et al. 2015). In light of the results presented here, it should be reassessed also over midlatitude regions over which perfect agreement of reanalyses has been taken for granted.

Acknowledgments. The work was funded by the Grant Agency of Charles University, Project 188214. We acknowledge the following organizations for providing their reanalysis datasets: NOAA/OAR/ESRL PSD, Boulder, Colorado, for the NCEP–NCAR reanalysis and Twentieth Century Reanalysis, version 2; ECMWF for ERA-40 and ERA-20C; and JMA for JRA-55. We thank all climate modeling groups for their GCM simulations and the PCMDI for enabling access to the data. Thanks are also due to all developers of the COST733 software. Parts of the study were carried out during stays of the first author at the Institute of Geography, University of Augsburg, Germany. The authors would like to express their deep gratitude to the Institute, and namely to Dr. Andreas Philipp for his friendly attitude and generous help.

REFERENCES

Beck, C., J. Jacobeit, and P. D. Jones, 2007: Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central Europe since 1780. *Int. J. Climatol.*, **27**, 473–491, doi:10.1002/joc.1410.

Belleflamme, A., X. Fettweis, C. Lang, and M. Erpicum, 2013: Current and future atmospheric circulation at 500 hPa over Greenland simulated by the CMIP3 and CMIP5 global models. *Climate Dyn.*, **41**, 2061–2080, doi:10.1007/s00382-012-1538-2.

Beranová, R., and R. Huth, 2008: Time variations of the effects of circulation variability modes on European temperature and precipitation in winter. *Int. J. Climatol.*, **28**, 139–158, doi:10.1002/joc.1516.

Bracegirdle, T. J., and G. J. Marshall, 2012: The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. *J. Climate*, **25**, 7138–7146, doi:10.1175/JCLI-D-11-00685.1.

Brands, S., J. M. Gutiérrez, S. Herrera, and A. S. Cofiño, 2012: On the use of reanalysis data for downscaling. *J. Climate*, **25**, 2517–2526, doi:10.1175/JCLI-D-11-00251.1.

Broderick, C., and R. Fealy, 2015: An analysis of the synoptic and climatological applicability of circulation type classifications for Ireland. *Int. J. Climatol.*, **35**, 481–505, doi:10.1002/joc.3996.

Bromwich, D. H., and R. L. Fogt, 2004: Strong trends in the skill of the ERA-40 and NCEP–NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001. *J. Climate*, **17**, 4603–4619, doi:10.1175/3241.1.

Cahynová, M., and R. Huth, 2016: Atmospheric circulation influence on climatic trends in Europe: An analysis of circulation type classifications from the COST733 catalogue. *Int. J. Climatol.*, **36**, 2743–2760, doi:10.1002/joc.4003.

Casado, M. J., M. A. Pastor, and F. J. Doblas-Reyes, 2010: Links between circulation types and precipitation in Spain. *Phys. Chem. Earth*, **35**, 437–447, doi:10.1016/j.pce.2009.12.007.

Cassano, J. J., P. Uotila, and A. Lynch, 2006: Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries, part 1: Arctic. *Int. J. Climatol.*, **26**, 1027–1049, doi:10.1002/joc.1306.

Compagnucci, R. H., and M. B. Richman, 2008: Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int. J. Climatol.*, **28**, 703–726, doi:10.1002/joc.1574.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.

Dell'Aquila, A., and Coauthors, 2016: Benchmarking Northern Hemisphere midlatitude atmospheric synoptic variability in centennial reanalysis and numerical simulations. *Geophys. Res. Lett.*, **43**, 5442–5449, doi:10.1002/2016GL068829.

Enke, W., and A. Spekat, 1997: Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Res.*, **8**, 195–207, doi:10.3354/cr008195.

Finnis, J., J. Cassano, M. Holland, and P. Uotila, 2009a: Synoptically forced hydroclimatology of major Arctic watersheds in general circulation models; Part 1: The Mackenzie River Basin. *Int. J. Climatol.*, **29**, 1226–1243, doi:10.1002/joc.1753.

——, ——, ——, and ——, 2009b: Synoptically forced hydroclimatology of major Arctic watersheds in general circulation models; Part 2: Eurasian watersheds. *Int. J. Climatol.*, **29**, 1244–1261, doi:10.1002/joc.1769.

Gibson, P. B., P. Uotila, S. E. Perkins-Kirkpatrick, L. V. Alexander, and A. J. Pitman, 2016: Evaluating synoptic systems in the CMIP5 climate models over the Australian region. *Climate Dyn.*, **47**, 2235–2251, doi:10.1007/s00382-015-2961-y.

Greatbatch, R. J., and P.-P. Rong, 2006: Discrepancies between different Northern Hemisphere summer atmospheric data products. *J. Climate*, **19**, 1261–1273, doi:10.1175/JCLI3643.1.

Hanson, C. E., J. P. Palutikof, and T. D. Davies, 2004: Objective cyclone climatologies of the North Atlantic—A comparison between the ECMWF and NCEP reanalyses. *Climate Dyn.*, **22**, 757–769, doi:10.1007/s00382-004-0415-z.

Hertig, E., C. Beck, H. Wanner, and J. Jacobeit, 2015: A review of non-stationarities in climate variability of the last century with focus on the North Atlantic–European sector. *Earth-Sci. Rev.*, **147**, 1–17, doi:10.1016/j.earscirev.2015.04.009.

Huth, R., 1993: An example of using obliquely rotated principal components to detect circulation types over Europe. *Meteor. Z.*, **2**, 285–293.

——, C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O. E. Tveito, 2008: Classifications of atmospheric circulation patterns. *Ann. N. Y. Acad. Sci.*, **1146**, 105–152, doi:10.1196/annals.1446.019.

——, ——, and O. E. Tveito, 2010: Classifications of atmospheric circulation patterns—Theory and applications—Preface. *Phys. Chem. Earth*, **35**, 307–308, doi:10.1016/j.pce.2010.06.005.

Jones, P. D., M. Hulme, and K. R. Briffa, 1993: A comparison of Lamb circulation types with an objective classification scheme. *Int. J. Climatol.*, **13**, 655–663, doi:10.1002/joc.3370130606.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–470, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kaufman, L., and P. J. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, 368 pp.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:10.2151/jmsj.2015-001.

Kouroutzoglou, J., H. Flocas, I. Simmonds, K. Keay, and M. Hatzaki, 2011: Assessing characteristics of Mediterranean explosive cyclones for different data resolution. *Theor. Appl. Climatol.*, **105**, 263–275, doi:10.1007/s00704-010-0390-8.

Kučerová, M., C. Beck, A. Philipp, and R. Huth, 2016: Trends in frequency and persistence of atmospheric circulation types over Europe derived from a multitude of classifications. *Int. J. Climatol.*, **37**, 2502–2521, doi:10.1002/joc.4861.

Kumar, A., L. Zhang, and W. Wang, 2013: Sea surface temperature–precipitation relationship in different reanalyses. *Mon. Wea. Rev.*, **141**, 1118–1123, doi:10.1175/MWR-D-12-00214.1.

Küttel, M., J. Luterbacher, and H. Wanner, 2011: Multidecadal changes in winter circulation-climate relationship in Europe: Frequency variations, within-type modifications, and long-term trends. *Climate Dyn.*, **36**, 957–972, doi:10.1007/s00382-009-0737-y.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, doi:10.1175/JCLI-D-13-00014.1.

Lund, I. A., 1963: Map-pattern classification by statistical methods. *J. Appl. Meteor.*, **2**, 56–65, doi:10.1175/1520-0450(1963)002<0056:MPCBSM>2.0.CO;2.

Lynch, A., P. Uotila, and J. J. Cassano, 2006: Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries, part 2: Antarctic. *Int. J. Climatol.*, **26**, 1181–1199, doi:10.1002/joc.1305.

Manzanas, R., S. Brands, D. San-Martín, A. Lucero, C. Limbo, and J. M. Gutiérrez, 2015: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines. *J. Climate*, **28**, 4171–4184, doi:10.1175/JCLI-D-14-00331.1.

McKendry, I. G., K. Stahl, and R. D. Moore, 2006: Synoptic sea-level pressure patterns generated by a general circulation model: Comparison with types derived from NCEP/NCAR reanalysis and implications for downscaling. *Int. J. Climatol.*, **26**, 1727–1736, doi:10.1002/joc.1337.

Nygård, T., and Coauthors, 2016: Validation of eight atmospheric reanalyses in the Antarctic Peninsula region. *Quart. J. Roy. Meteor. Soc.*, **142**, 684–692, doi:10.1002/qj.2691.

Pasini, A., and R. Langone, 2012: Influence of circulation patterns on temperature behavior at the regional scale: A case study investigated via neural network modeling. *J. Climate*, **25**, 2123–2128, doi:10.1175/JCLI-D-11-00551.1.

Pastor, M. A., and M. J. Casado, 2012: Use of circulation types classifications to evaluate AR4 climate models over the Euro-Atlantic region. *Climate Dyn.*, **39**, 2059–2077, doi:10.1007/s00382-012-1449-2.

Perez, J., M. Menendez, F. J. Mendez, and I. J. Losada, 2014: Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Climate Dyn.*, **43**, 2663–2680, doi:10.1007/s00382-014-2078-8.

Philipp, A., P. M. Della-Marta, J. Jacobeit, D. R. Fereday, P. D. Jones, A. Moberg, and H. Wanner, 2007: Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *J. Climate*, **20**, 4065–4095, doi:10.1175/JCLI4175.1.

——, C. Beck, R. Huth, and J. Jacobeit, 2016: Development and comparison of circulation type classifications using the COST 733 dataset and software. *Int. J. Climatol.*, **36**, 2673–2691, doi:10.1002/joc.3920.

Plavcová, E., and J. Kyselý, 2012: Atmospheric circulation in regional climate models over Central Europe: Links to surface air temperature and the influence of driving data. *Climate Dyn.*, **39**, 1681–1695, doi:10.1007/s00382-011-1278-8.

——, and ——, 2013: Projected evolution of circulation types and their temperatures over Central Europe in climate models. *Theor. Appl. Climatol.*, **114**, 625–634, doi:10.1007/s00704-013-0874-4.

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, doi:10.1175/JCLI-D-15-0556.1.

Richman, M. B., 1986: Rotation of principal components. *J. Climatol.*, **6**, 293–335, doi:10.1002/joc.3370060305.

Rust, H., M. Vrac, M. Lengaigne, and B. Sultan, 2010: Quantifying differences in circulation patterns based on probabilistic models: IPCC AR4 multimodel comparison for the North Atlantic. *J. Climate*, **23**, 6573–6589, doi:10.1175/2010JCLI3432.1.

Stickler, A., and S. Brönnimann, 2011: Significant bias of the NCEP/NCAR and twentieth-century reanalyses relative to pilot balloon observations over the West African monsoon region (1940–1957). *Quart. J. Roy. Meteor. Soc.*, **137**, 1400–1416, doi:10.1002/qj.854.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Trigo, I. F., 2006: Climatology and interannual variability of storm-tracks in the Euro-Atlantic sector: A comparison between ERA-40 and NCEP/NCAR reanalyses. *Climate Dyn.*, **26**, 127–143, doi:10.1007/s00382-005-0065-9.

Tveito, O. E., and R. Huth, 2016: Circulation-type classifications in Europe: Results of the COST 733 Action. *Int. J. Climatol.*, **36**, 2671–2672, doi:10.1002/joc.4768.

Ulbrich, U., G. C. Leckebusch, and J. G. Pinto, 2009: Extra-tropical cyclones in the present and future climate: A review. *Theor. Appl. Climatol.*, **96**, 117–131, doi:10.1007/s00704-008-0083-8.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

Wang, X. L., V. R. Swail, and F. W. Zwiers, 2006: Climatology and changes of extratropical cyclone activity: Comparison of ERA-40 with NCEP–NCAR reanalysis for 1958–2001. *J. Climate*, **19**, 3145–3166, doi:10.1175/JCLI3781.1.

——, Y. Feng, R. Chan, and V. Issac, 2016: Inter-comparison of extra-tropical cyclone activity in nine reanalysis datasets. *Atmos. Res.*, **181**, 133–153, doi:10.1016/j.atmosres.2016.06.010.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 648 pp.