Long-Term Variability of Daily North Atlantic–European Pressure Patterns since 1850 Classified by Simulated Annealing Clustering

A. PHILIPP,* P. M. DELLA-MARTA,[+] J. JACOBEIT,* D. R. FEREDAY,[#] P. D. JONES,[@] A. MOBERG,[&] AND H. WANNER[+]

*Institute of Geography, University of Augsburg, Augsburg, Germany
[+]Institute of Geography, University of Bern, Bern, Switzerland
[#]Hadley Centre for Climate Prediction and Research, Met Office, Exeter, United Kingdom
[@]Climatic Research Unit, University of East Anglia, Norwich, United Kingdom
[&]Department of Meteorology, Stockholm University, Stockholm, Sweden

ABSTRACT

Reconstructed daily mean sea level pressure patterns of the North Atlantic–European region are classified for the period 1850 to 2003 to explore long-term changes of the atmospheric circulation and its impact on long-term temperature variability in the central European region. Commonly used k-means clustering algorithms resulted in classifications of low quality because of methodological deficiencies leading to local optima by chance for complex datasets. In contrast, a newly implemented clustering scheme combining the concepts of simulated annealing and diversified randomization (SANDRA) is able to reduce substantially the influence of chance in the cluster assignment, leading to partitions that are noticeably nearer to the global optimum and more stable. The differences between conventional cluster analysis and the SANDRA scheme are significant for subsequent analyses of single clusters—in particular, for trend analysis. Conventional indices used to determine the appropriate number of clusters failed to provide clear guidance, indicating that no distinct separation between clusters of circulation types exists in the dataset. Therefore, the number of clusters is determined by an external indicator, the so-called dominance criteria for t-mode principal component analysis. Nevertheless, the resulting partitions are stable for certain numbers of clusters and provide meaningful and reproducible clusters. The resulting types of pressure patterns reveal pronounced long-term variability and various significant trends of the time series of seasonal cluster frequency. Tentative estimations of central European temperature changes based solely on seasonal cluster frequencies can explain between 33.9% (summer) and 59.0% (winter) of temperature variance on the seasonal time scale. However, the signs of long-term changes in temperature are correctly reproduced even on multidecadal–centennial time scales. Moreover, linear warming trends are reproduced, implying from one-third up to one-half of the observed temperature increase between 1851/52 and 2003 (except for summer, but with significant trends for spring and autumn), indicating that changes in daily circulation patterns contribute to the observed overall long-term warming in the central European region.

1. Introduction

There is increasing evidence that the recent temperature increase in the European region (Jones and Moberg 2003; Moberg et al. 2006) is attributable to anthropogenic influences (e.g., Stott et al. 2004). The basic physical mechanisms responsible for this temperature increase are largely understood, however, less is known about the manifestation of global warming at the regional scale. As discussed in Folland and Karl (2001), the observed recent trend in surface air temperature in Europe may be induced to a high degree by changes in circulation dynamics of the North Atlantic–European sector, while the atmospheric circulation itself is possibly affected by anthropogenic forcing resulting in long-term changes of frequencies of distinct large-scale circulation states (Palmer 1999; Corti et al. 1999). However, doubts on a deterministic view of circulation changes have been raised by Stephenson et al. (2000) indicating that long-term changes of circulation-

type frequencies may only be the result of stochastic processes without any deterministic influences. A different point of view includes a separation between a so-called background mode of the North Atlantic–European temperature field, which is mostly responsible for the observed warming, and an independent so-called dynamic temperature mode, which is mostly driven by the North Atlantic Oscillation (NAO) but does not show an increasing trend itself at the present and thus may act just as a modulation of the recent temperature trend (Slonosky and Yiou 2001). To provide a solid observational database for the important discussion about reasons for present and past climatic changes, it is necessary to extend the knowledge about circulation changes as far back into the past as possible. In this context the relatively long time series of historical instrumental records in the North Atlantic–European region are of especially high value for reconstructing former climate variability.

While reconstructions of monthly or seasonally resolved sea level pressure (SLP) grids for the North Atlantic–European sector are advancing further back in time (e.g., Luterbacher et al. 2000, 2002; Jones et al. 1997, 1999), daily resolved gridded datasets have been limited to the starting year of 1881 up to now (Ansell et al. 2006). Within the European and North Atlantic Daily to Multidecadal Climate Variability project (EMULATE) a new reconstruction of daily mean SLP patterns based on early instrumental data has been realized by Ansell et al. (2006), offering to extend the view on daily resolved circulation variability back to the year 1850. Circulation data of daily resolution introduce a new quality into long-term synoptic climatology since they allow a distinctly better allocation between particular states of the atmospheric circulation and associated phenomena of the surface climate than data with a coarser resolution.

Long-term changes of the atmospheric circulation in the North Atlantic–European sector have been subject to various studies before, using long monthly and seasonally resolved station records (e.g., Slonosky et al. 2000), pressure field reconstructions (e.g., Luterbacher et al. 2000, 2002; Jacobeit et al. 2003), or classification-based time series (e.g., Bárdossy and Caspary 1990). Most of them use a conceptual categorization between pressure patterns of enhanced zonal flow and meridional circulation patterns throughout the North Atlantic–European sector.

Zonal westerly flow over the North Atlantic Ocean has been found to be significantly intensified as a result of an increase of strong positive NAO anomalies in the winter season from around 1980 to the mid-1990s (Jones et al. 1999). This has been interpreted by Hurrell

and van Loon (1997) as an outstanding feature for the period 1864 to 1995 and was referred to as the main reason for the recent warming of European surface air temperatures in winter.

Advancing farther back into time, a second phase of strengthened zonal airflow was established at the beginning of the twentieth century when positive NAO phases showed an increased persistence especially between 1903 and 1925 (Rogers et al. 2000). This is also apparent from frequency changes within the weather-type classification by Hess and Brezowsky (1977, hereinafter HB77) (Bárdossy and Caspary 1990). In contrast to Hurrell and van Loon (1997), Slonosky and Yiou (2001) found that this NAO trend in the early twentieth century was at least as strong as the 1980–95 NAO trend by using newly homogenized station data, which put the strength of the most recent trend into perspective.

The two more recent phases of strengthened zonal flow types over the North Atlantic have been accompanied by a decrease of meridional patterns over Europe. A phase of moderately enhanced positive NAO at the very beginning of the period covered by this study, in the 1860s, was reported to show also a strengthened meridional configuration over Europe (Luterbacher et al. 1999), with correspondingly high frequencies of meridional circulation patterns (Jacobeit et al. 2003). Using seasonally resolved data, Luterbacher et al. (1999) report a decreasing strength of linkage between a reconstructed NAO index and a meridional Eurasian (EU) circulation index (representing the pressure difference between western and eastern Europe) during the period 1840 to 1860 for winter, indicating a decoupling between North Atlantic and continental European airflow in the middle of the nineteenth century. Also Slonosky et al. (2000) confirm by EOF analysis of monthly station data that zonal flow and meridional flow over Europe have both been relatively strong in the period 1822 to 1870. However, Jacobeit et al. (2003), using monthly resolved pressure field reconstructions, found a distinct within-type change of the Russian high pattern type in January, suggesting that there might be no real decoupling but a spatial shift of the involved pressure centers during this period.

Monthly or seasonally resolved data allow a relatively wide range of interpretations, because similar monthly mean pressure maps may be set up by different intramonthly or intraseasonal variations of single pressure configurations as represented by daily maps. Therefore, the daily pressure patterns used in this study should shed substantially more light on general circulation changes. Furthermore, indices describing the

zonal airflow over the North Atlantic (like the NAO index) show an important influence on European climate. However, it seems that they are largely insufficient to give a complete picture of variations in pressure configuration affecting European climate, which is needed, however, to assess the reasons for past and present climate changes.

Appropriate methods for analyzing circulation changes may be grouped into three main types: 1) analyses of time coefficients achieved by eigenvector methods, like principal component analysis (PCA; Barnston and Livezey 1987; Jacobeit et al. 2003), 2) circulation indices calculated as differences between stations or grid points describing large-scale pressure gradients [e.g., NAO indices (Jones et al. 1997; Hurrell 1995; Luterbacher et al. 2002), European zonal indices (Jacobeit et al. 2001; Slonosky et al. 2000), or other pressure differences, for example, between London and Paris (Slonosky et al. 2000)], and 3) time series based on pressure pattern classification (see below).

This study focuses on the classification-based approach since it offers a relatively simple and direct way for interpretation (not always possible for eigenvector products) and provides a mostly complete representation of pressure configurations (in contrast to pressure gradient indices).

Often a general distinction is made between manual and automated classifications (Yarnal 1993; Yarnal et al. 2001), but modern approaches also include hybrids of both methods. Prominent examples of "subjective" or "manual" classifications are the catalogs of central European Grosswetterlagen (Hess and Brezowsky 1977) and of the Lamb weather types (Lamb 1972) based on the expert knowledge of meteorologists. "Automated" or "objective" approaches utilize statistical methods and computer technology and comprise two major groups, one based on correlation methods (e.g., Lund 1963; Schmutz and Wanner 1998) and the other based on different versions of cluster analysis (CA), e.g., hierarchical (Kalkstein et al. 1987; Chen and Harr 1993), nonhierarchical (Michelangeli et al. 1995), both combined (Jacobeit et al. 1998), or nonhierarchical based on simulated annealing (Bárdossy et al. 2002). The hybrid category (e.g., Sheridan 2002) generally includes a manual selection of conditions defining types or classes, such as key days, used as seeds for the different circulation types and various automated approaches for assigning individual patterns to these preselected types. Among the three groups of classifications—nonautomated (manual), partly automated (hybrid), and fully automated ones—the present study uses the latter type.

For the parameters used for classification a distinc-tion has to be drawn between weather type and circulation pattern classification. The first type includes variables of both circulation dynamics (e.g., pressure patterns) and meteorological surface conditions (e.g., temperature and humidity); see, for example, Goodess and Jones (2002) and Haylock and Goodess (2004). The second type, which is the one applied in this study, is based on just one parameter of atmospheric circulation (here SLP) and its classification into leading circulation patterns is completely independent from meteorological surface conditions, which may be analyzed afterward. According to Yarnal (1993) this may be called a "circulation-to-environment" type of study (e.g., Stehlík and Bárdossy 2003; Romero et al. 1999; Cannon et al. 2002; Linderson 2001). It implies a broader within-type range for meteorological surface conditions at particular locations but allows more general applications in synoptic climatology.

To avoid confusion, a distinction has to be drawn between the present study and approaches for regime analysis (see the overview of Stephenson et al. 2004). A practical definition of climate regimes refers to preferred quasi-stationary states of the large-scale circulation system where baroclinic disturbances acting on time scales shorter than 10 days are considered only as modulating the regimes (Michelangeli et al. 1995). Typical numbers of such climate regimes are relatively small, between 1 and 4 (e.g., Stephenson et al. 2004; Michelangeli et al. 1995; Corti et al. 1999). The present study differs fundamentally from climate regime studies in that it does not intend to identify just persistent regimes but aims to classify all different states of the circulation system on a daily scale. Persistent regimes are especially able to explain distinct types of surface climate (e.g., hot or cold spells), whereas short-term phenomena (e.g., frontal temperature changes or heavy precipitation events) are better reflected by patterns derived from a daily classification.

About the question of whether some methods are more suitable for classification than others, Stehlík and Bárdossy (2003) suggest that manual techniques could be influenced by methodological changes and are probably not suited for investigations of long-term changes in circulation. Jones et al. (1993), however, in comparing the subjective Lamb weather types with an objective method based on empirical formulas applied to gridpoint SLP data, conclude that there are no major differences between these two approaches. Huth (1996) compares a number of different circulation classifications and concludes that PCA offers the most robust results, but that the choice of the method depends on the aim of the classification. For CA, Kalkstein et al. (1987) conclude that hierarchical CA using the average-

linkage method is suitable for synoptic classification, while Chen and Harr (1993) prefer Ward's method for hierarchical CA of circulation patterns as do Gong and Richman (1995) for classifying precipitation data. However, the latter show that nonhierarchical k-means CA (also favored by Michelangeli et al. 1995) outperforms hierarchical CA in general. Additionally Gong and Richman (1995) recommend the use of fuzzy clustering—that is, objects may be member of more than one cluster—for datasets with no inherent separation into nonoverlapping clusters of objects. In this study, however, distinct classification of patterns provides a more concise interpretation, which is favored here for investigations on long-term variations in circulation and climate. Therefore in the present study the focus will be on nonhierarchical k-means CA.

The paper starts with an overview of the data sources in section 2 and then addresses the classification methods used in this study in section 3, while reasons for the choice of the simulated annealing clustering method are discussed in section 4. The decision for certain numbers of classes is addressed in section 5. Last, characteristics of the resulting classes and aspects of their long-term variability are presented in section 6, as well as an attempt to estimate its role for long-term temperature changes in central Europe in section 7. Methodological and climatological conclusions are drawn in section 8.

2. Data

The daily mean SLP dataset used for classification has been reconstructed by Ansell et al. (2006) within the European EMULATE project using reduced-space optimal interpolation (RSOI). The dataset spans the period between 1850 and 2003. A variety of different sources of data have been scrutinized and corrected for inhomogeneities in order to create the dataset, including 86 land and island station series, marine data from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS), and daily gridded products from other sources. The spatial grid of reconstruction data points provides a resolution of $5° \times 5°$ and spans between $25°$ and $70°N$ and between $70°W$ and $50°E$, resulting in a total of 250 grid points. For further details see Ansell et al. (2006).

To counteract the artificial northward increase in variance caused by meridian convergence, the gridded SLP data are weighted by the square root of the cosine of latitude for each grid point: $W_a = [\cos(\text{latitude})]^{1/2}$.

The reconstruction skill varies spatially between the central region [more than 90% explained variance of the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) data] and the periphery (a minimum of about 10% explained variance of reanalysis data in the northern African region). The varying reconstruction skill is well documented by an accompanying RSOI error field (Ansell et al. 2006), which is further used to account for the varying confidence in the data by the application of an error weight W_e defined as

$$W_e = (E/s)^{-1}, \qquad (1)$$

where E is the long-term mean RSOI error and s is the sample standard deviation for the corresponding grid point. By this way, grid points with lower reliability will have lower influence onto the classification of the pressure patterns.

Figures 1a–c show the average winter [December–February (DJF)] RSOI error field, the average standard deviation of DJF SLP, and the resulting W_e given by Eq. (1), which is representative for all seasons. Generally the RSOI error is greater in the north of the domain, as is the SLP standard deviation. The W_e field shows that the largest weights are centered on the region extending from the Azores to western Europe where they peak and remain on a high level toward central and eastern Europe with high confidence in pressure reconstructions for these regions.

In addition to the four 3-month seasons [DJF, March–May (MAM), June–August (JJA), September–November (SON)], 12 overlapping 2-month seasons [January–February (JF), February–March (FM), etc.] have been analyzed. This paper, however, will concentrate on the usual 3-month seasons, whose sample sizes vary between 13 897 and 14 168 daily SLP fields.

For characterizing the resulting pressure patterns, daily surface temperatures from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996; Kistler et al. 2001) are used for the period from 1948 to 2003, converted to anomalies from the corresponding long-term monthly mean.

Additionally monthly mean temperature reconstructions from Jones and Moberg (2003) for the period from 1851 to 2003 are used for proving the skill of the classification results to explain variability in surface climate. For this purpose a central European index was calculated as the spatial mean temperature anomaly (referring to the long-term mean for 1851/52 to 2003) for the area $45°–55°N$ and $5°–20°E$ where no missing values occur within the period. The monthly index values have been finally aggregated to seasonal means.

3. Methods

The main principle of classifying a sample of objects (in this study daily SLP patterns) is to generate clusters
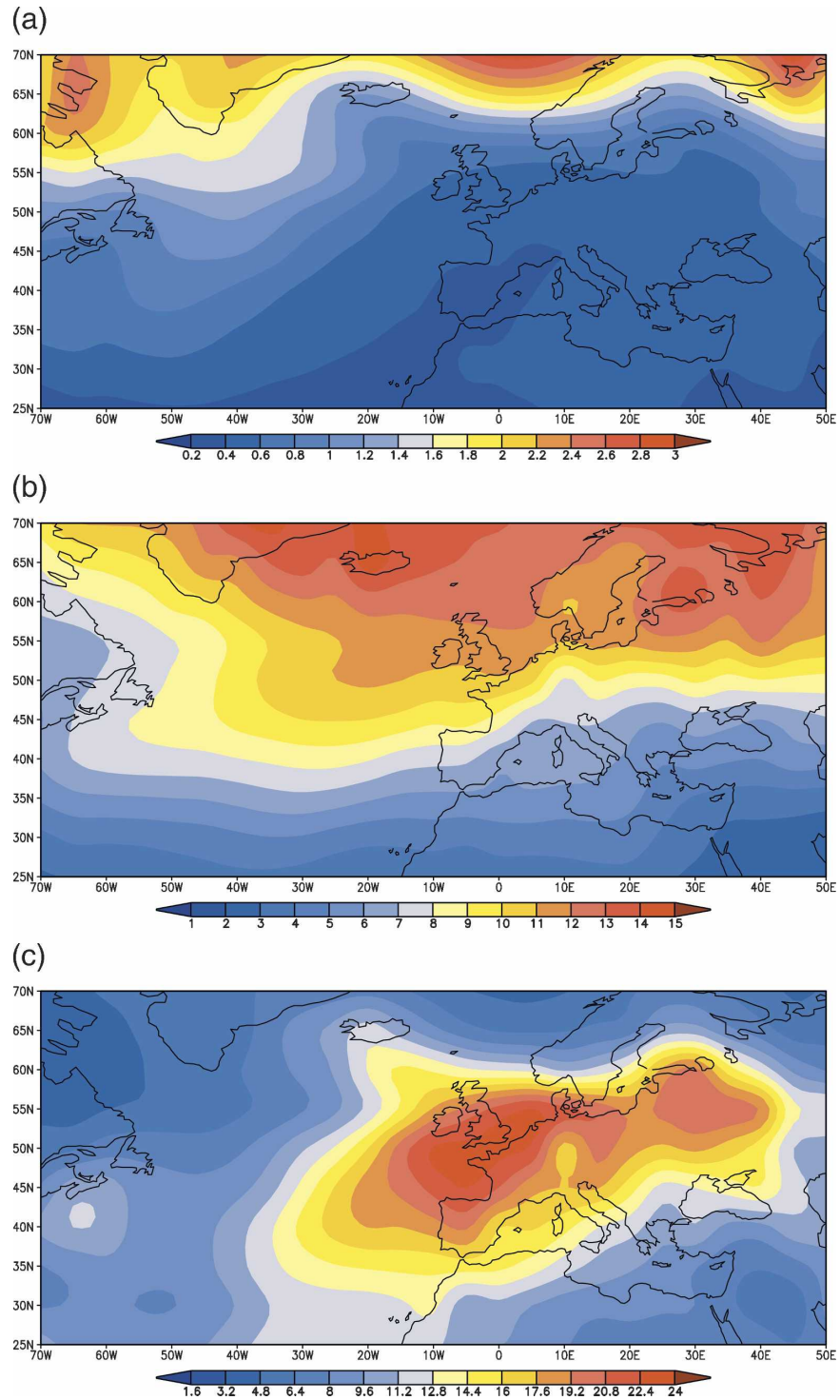
FIG. 1. (a) Winter (DJF) daily SLP mean optimum-interpolation error (hPa), (b) daily standard deviation (hPa), and (c) weights applied to the raw data before clustering [nondimensional units, calculated using Eq. (1)]. Colored contour shading corresponds to the legend at the bottom of each panel.

in the multidimensional space spanned by the parameters (grid points in this case), each comprising objects of highest similarity. The degree of dissimilarity within the resulting clusters of a certain partition can be measured by the so-called within cluster sum of squares of deviations (WSS):

$$\text{WSS} = \sum_{j=1}^{k} \sum_{i \in C_j} D(X_i, \overline{X}_j)^2, \qquad (2)$$

where k is the number of clusters C, i is the object number, $\overline{X}$ is the centroid of each cluster, and D is the Euclidean distance between the objects and its cluster centroids:

$$D(X_i, \overline{X}_j) = \left[\sum_{l=1}^{m} (X_{il} - \overline{X}_{jl})^2 \right]^{1/2}, \qquad (3)$$

where m is the number of parameters (grid points in this case) describing the objects. The use of Euclidean distances is preferred because they are able to reflect the absolute gradients in the daily pressure patterns, which is not the case, for example, for correlation-based measures for dissimilarity (Gong and Richman 1995). In contrast to WSS the so-called explained cluster variance (ECV) allows comparison of the quality of a cluster solution between different datasets:

$$\text{ECV} = 1 - \frac{\text{WSS}}{\text{TSS}}, \qquad (4)$$

where TSS is the "total cluster sum of squares" calculated as the sum of the Euclidean distances between all objects and the overall centroid ($k = 1$). Thus, ECV can vary only between 0 and 1 and should be maximized by the classification methods.

A crucial point for classifying datasets with large sample sizes and objects of transient character is the existence of an unknown number of various solutions that cannot be improved further by rearranging objects from one cluster to another (convergence of optimization), though the solution may be more or less far away from the best solution possible. The former solutions of less quality are called local optima, while the best solution is called the global optimum. The only way to determine the global optimum is to check all possible combinations of objects, which is impossible for large datasets. This also implies that there is no way to prove any optimum to be the global optimum or how far any local optimum is away from the global optimum. Some methods are not able to prevent solutions of relatively low quality (e.g., conventional k-means CA, see below); however, there are strategies based on multistart techniques and simulated annealing algorithms (see below) that are able to get over local minima and approximate the global optimum.

a. Conventional k-means CA

A common way of applying CA is to create seed partitions as first guesses and to use them for further optimization with k-means CA. Three methods to create seed partitions from the SLP dataset are used in this study: 1) s-mode rotated principal component analysis (RPCA) of the SLP data by assigning each daily pattern to the principal component (PC) whose loading pattern shows the highest correlation onto it, 2) t-mode RPCA (see Jacobeit et al. 2003) by assigning each pattern to the PC with the highest loading onto it, and 3) hierarchical CA utilizing Ward's method (e.g., Gong and Richman 1995).

Starting from such a seed partition the objects are checked sequentially and are reassigned from one cluster to another if that reduces within-cluster variance. After each reassignment the cluster centroids have to be updated for subsequent tests on reassignments. The iterative optimization process ends if no further reassignment can increase the ECV, that is, convergence to a local optimum has been reached. This algorithm is referred to as "straightforward k-means algorithm," henceforth addressed as SFA.

SFA may produce different results for different starting partitions (e.g., Michelangeli et al. 1995). This is because the starting partition determines the sequencing of reassignments of objects since each object is compared with the cluster centroids resulting from the previous step. Thus the starting partition steers the optimization process into a certain direction and into a certain local optimum. However, the quality of the results is unpredictable from the quality of the starting partition (cf. section 4).

Another limitation of conventional k-means CA is the dependence on the ordering of checks and reassignments. For example, when using the optimized algorithm (henceforth called HWA) published by Hartigan and Wong (1979), in most cases a different local optimum is reached as compared with SFA since HWA alters the sequence of checks and reassignments. Thus the probability to step into the next best local optimum is sometimes reduced. Hence the ordering of objects—which is just a matter of chance—together with the actual starting partition causes conventional k-means CA to step into a certain local optimum without any possibility to determine a priori, which one will be addressed. In particular, conventional k-means CA offers no strategy to approximate the global optimum, thus, largely reducing the confidence on the result.

b. Multistart technique and diversified randomization

To reduce the probability of converging to a local optimum of very low quality, k-means CA may be run multiple times by using different randomly created starting partitions, the best, in terms of ECV being selected. This scheme is called the "multistart" technique. Since the ordering of objects plays a similar role as the seed partition, another variant, called the "diversified randomization" technique, is used acting in the same manner: CA is run several times on randomized starting partitions but, in addition, randomizes the ordering of objects and cluster numbers throughout the iterative process of checking and reassigning. The combination of these randomization techniques increases the probability of hitting the global optimum. However, since the resulting optima of each single run are still intrinsically dependent on chance, and the k-means optimization process itself does not include a target-oriented strategy to avoid local minima, there is no evidence that the solution is in fact close to the global optimum.

c. Clustering by simulated annealing

An advanced approach explicitly designed to approximate the global optimum is the simulated annealing technique, for example, used to solve combinatorial optimization problems like the "traveling salesman problem" (Aarts and Korst 1989) and applied to CA, for example, in biotechnology for genome clustering (Lukashin and Fuchs 2001). Applications within climatology are still rare (e.g., Bárdossy et al. 2002; Hannachi and Legras 1995) and its potential for classification of meteorological fields has not been evaluated systematically up to now. Clustering by simulated annealing simulates the process of tempering by reheating in which particles of a hot melt correspond to the objects in CA. It differs from common CA mainly because it is not bound to the concept of irreversible paths throughout the optimization process; that is, it does not converge to a local optimum that cannot be left anymore but allows an object to leave its cluster at any stage (with a specific probability), even if this step at the first instance leads to an increase of the cost function (WSS in this case). While the main principle of different annealing algorithms is always the same, its implementation was redesigned as described below.

Starting from a random partition, each object is checked at each iteration for being in the appropriate cluster and is reassigned if not (addressed as "right" reassignment), but it could also be reassigned to another cluster (called "wrong" reassignment) if the acceptance probability P is higher than a number between 0.0 and 1.0 achieved by a random number generator, where

$$P = \exp\left(\frac{D_{\text{old}} - D_{\text{new}}}{T}\right) \quad (5)$$

and D_{new} is the Euclidean distance between the object and a potential new cluster, D_{old} is the Euclidean distance between the object and its present cluster, and T is the so-called temperature control parameter (Aarts and Korst 1989), which is initially a big number but is slowly decreasing during the optimization process according to the so-called cooling factor C, which should be close below 1.0:

$$T_{i+1} = CT_i, \quad (6)$$

where T_{i+1} is the temperature following T_i and i is the temperature state with a duration defined as the time needed to check all objects once (i.e., one iteration in our implementation). The initial value for T is determined empirically in order to ensure that 99% of the n objects are moved during the first iteration, allowing for free movement of objects between the clusters but also reducing the idle time of decreasing T without effect. For each single iteration, the ordering of objects is randomized. Within each iteration, each object is checked to be moved to any of the other clusters whose ordering is randomized as well for each check. This approach ensures that all possibilities are checked out and allows for a quick and simple stopping criterion. After each iteration, T is reduced by multiplication with C. At the very end of the process T is low; thus no single wrong reassignment is possible anymore. The process is stopped when no right reassignment is possible (convergence has been reached) and no wrong reassignment has appeared for a complete iteration. For C being relatively close to 1 (e.g., $C = 0.999\,999\,0$), the procedure needs a runtime of an unacceptable number of days for the datasets used in this study and there was still a slight dependence on the series of random numbers. The most effective way to achieve high confidence that the solution is near the global optimum was found to be a combination of simulated annealing and the diversified randomization technique (henceforth called simulated annealing and diversified randomization, or SANDRA). This procedure performs simulated annealing CA 1000 times—the first reference run with $C = 0.999\,90$ and the following 999 runs with $C = 0.90$—and finally selects the best one. Experimental comparisons have shown that this scheme at least performs as well as single runs with very slow cooling rates but decreases runtime significantly.

TABLE 1. Clustering quality as a function of the number of clusters and the classification method for the winter season (DJF) expressed in terms of explained cluster variance, ECV*100 [Eq. (4)]. Rows show results by the number of clusters k from 2 to 14, and the columns show the methods, where (from left to right) Smod is s-mode RPCA, Tmod is t-mode RPCA, Ward is hierarchical CA using Ward's method, SFA is straightforward k-means CA using the results of the first three methods for starting partitions, HWA is Hartigan and Wong k-means CA algorithm using the results of the first three methods for starting partitions, and SANDRA is CA by simulated annealing and diversified randomization. Here $++$ denotes the best result in each line, $+$ denotes the second best, $--$ denotes the worst, and $-$ denotes the second worst result among the converging methods (excluding the first three methods). Note that small differences have large effects for similarity between cluster partitions (see text).

k	Smod	Tmod	Ward	SFA$_{smod}$	SFA$_{tmod}$	SFA$_{Ward}$	HWA$_{smod}$	HWA$_{tmod}$	HWA$_{Ward}$	SANDRA
2	7.7852	11.2966	9.6001	20.3533++	20.3533++	20.3533++	20.3533++	20.3533++	20.3533++	20.3533++
3	17.1204	22.1652	21.2351	30.5214−−	30.5215++	30.5215++	30.5214−−	30.5215++	30.5214−−	30.5215++
4	21.4049	24.1212	25.2365	35.5763−	35.5753−−	35.6181++	35.5765	35.6084+	35.6181++	35.6181++
5	23.4293	25.4664	27.8730	39.2823−	39.1275−−	39.3307+	39.2823−	39.1275−−	39.3307+	39.3362++
6	25.7191	25.9036	29.7329	42.4295−−	42.4301+	42.4295−−	42.4301+	42.4301+	42.4295−−	42.4309++
7	25.8725	26.2340	31.6719	44.7430	44.7111−	44.7604+	44.7518	44.6469−−	44.7601	44.7605++
8	28.0494	26.2586	33.1934	46.5524	46.7104++	46.6823	46.5384−−	46.7096+	46.5503−	46.7104++
9	28.6244	26.3103	34.5733	48.2637−	48.3156	48.2649	48.2680	48.3191+	48.2050−−	48.3193++
10	30.5541	26.3262	36.8952	49.6135−	49.5868−−	49.6170	49.6190	49.6228+	49.6182	49.6265++
11	30.1236	26.3358	37.8739	50.6342	50.7141+	50.6096−−	50.6341−	50.6546	50.6344	50.7161++
12	32.2345	26.3790	38.6134	51.6494	51.6098−−	51.6643	51.6414	51.6103−	51.6916+	51.7189++
13	31.9204	26.4168	39.4559	52.5275	52.5270−	52.5854+	52.5753	52.5269−−	52.5794	52.6569++
14	32.3580	26.4628	40.0205	53.1989−−	53.2670−	53.4946	53.5035+	53.3478	53.4916	53.5308++

4. Evaluation of classification methods

As an example Table 1 shows the ECV [Eq. (4)] for the winter season and for cluster numbers from 2 to 14, including the different classification techniques used in this study. The first three methods used to create seed partitions differ considerably from those converging to a (local) optimum. Obviously a better starting partition does not necessarily lead to a better convergence for the optimizing methods, supporting the argument that they converged to a certain optimum just by chance. The differences in ECV between the optimization methods increase with the number of clusters (from no difference for 2 clusters to the greatest differences for 14 clusters). This reflects the increasing number of possible partitions with an increasing number of clusters allowing for more local optima to be established. In comparison with the other methods the SANDRA technique reaches the optimum nearest to the global optimum (i.e., its ECV is highest) for all numbers of clusters. This is due to the largely reduced probability for SANDRA to converge to a local optimum of low quality.

Even if it is favorable to achieve clusters with the lowest possible within-cluster variability, the question remains whether an enhanced optimization of ECV is actually important concerning the composition of the clusters, the shape of their centroids, and their time series of cluster frequencies. In other words, Does an enhanced optimization of ECV help in finding nearly identical partitions (regardless of the starting partition and ordering of objects)? This question seems even more important as the differences in quality (expressed by ECV in Table 1) for a certain number of clusters at the first glance seem small and negligible. To find an answer, the interrelation between quality and similarity of two partitions has been examined as follows.

By varying the cooling rate for the SANDRA clustering method between 0.0 and 0.90, a set of 1000 solutions was produced comprising 615 different local optima of vastly varying quality that have been compared pairwise. The similarity between two partitions a and b of a pair $S_{a,b}$ is defined to be the count of pairs P of objects that are in the same cluster in partition a as well as in partition b plus the pairs that are in different clusters in partition a as well as in partition b divided by the overall number of pairs; thus $S_{a,b}$ may range between 0 and 1 [also referred to as the "Rand index" (Rand 1971)]:

$$S_{a,b} \frac{\sum P_{a=b} + \sum P_{a \neq b}}{n}. \tag{7}$$

The joint quality of a pair of partitions ($Q_{a,b}$) may be defined as

$$Q_{a,b} = (\mathrm{ECV}_a \mathrm{ECV}_b)^{1/2}. \tag{8}$$

To examine the interrelation between similarity $S_{a,b}$ and quality $Q_{a,b}$ of pairs of partitions, the corresponding values are presented as scatterplots in Fig. 2, with each point representing a pair of partitions. Figure 2a shows all of the possible combinations between two partitions. Even though there is a strong scattering of points, this scattering is not evenly distributed but
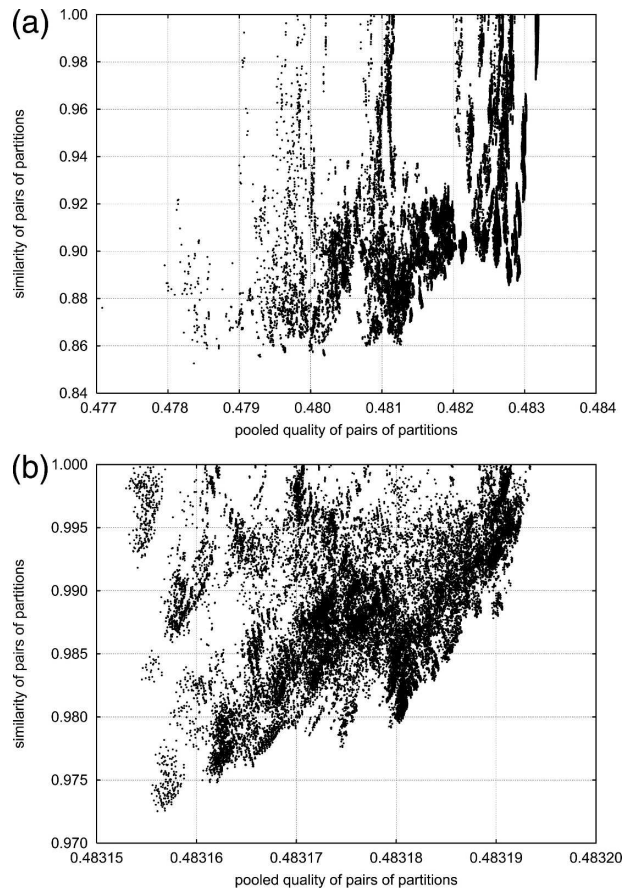
FIG. 2. Similarity of pairs of partitions (y axis) depending on their pooled clustering quality [x axis; Eq. (8)] for the winter season (DJF). (a) The complete sample of partition pairs; (b) enlarged section of the upper right agglomeration in (a). Note that even small improvements in quality lead to a strong reduction in potential dissimilarity between two partitions.

shows strong irregularities and distinct structures in the form of clusters of points and more or less continuous linear structures. These are interpreted as preferred pathways (linear structures) into central local optima with high attracting properties (high-density clusters of points) for the optimization process. Figure 2b shows the enlarged area of the very upper right corner of Fig. 2a, including the cluster of points with partitions of highest quality and similarity.

The most important feature is that the scattering of points only covers the upper and left areas while the area in the lower right corner is left totally void. This indicates that high similarity between partitions is possible for all levels of quality, but low similarity is shown only for partitions of low quality. In other words, the probability of low similarity decreases strongly with increasing quality.

The explanation for this phenomenon is that for lev-

els of low quality many largely dissimilar partitions of local optima are possible; however, for higher levels of quality the possibility for different partitions to be dissimilar decreases as they become more and more similar to the partition of the global optimum. If, finally, two partitions are nearly identical to the partition of the global optimum, then they must also be nearly identical to each other. This means, the nearer any partition is to the global optimum in terms of ECV, the more similar to the partition of the global optimum it must be.

The sensibility of this rule is pointed out by the scaling of the axes in Fig. 2b, where the cluster of points shows a strong tendency for a reduced possibility of being dissimilar ($S_{a,b}$ ranging between 0.9725 and 1.0) with just a slightly increased quality ($Q_{a,b}$ ranging between 0.4831525 and 0.483194). This means even if the quality of partitions produced by any method of CA differs just slightly from the global optimum, the unsteadiness of the method grows rapidly. Therefore, the steadiness of any method in finding methodologically stable and reproducible clusters depends strongly on its ability to approximate the global optimum.

Using the SANDRA method cannot guarantee to find the global optimum; it is not even possible to estimate how far away from the global optimum the resulting partition is, because the global optimum remains unknown. However, it seems evident that it is necessary to derive highly optimized partitions in order to exclude the influence of chance on the created classification as much as possible.

The enhanced reproducibility of partitions achieved using the SANDRA method relative to those of conventional CA methods can be demonstrated by a subsampling experiment. The dataset was split up into two halves, one comprising the days with an odd sequence number and the other half comprising days with an even sequence number. After applying CA to each subsample separately, two full-sample partitions have been reconstructed by assigning the days not used for CA to the clusters. The resulting two partitions, each based on another subsample, are compared by calculating the Rand index [Eq. (7)]. This has been done for all 3-month seasons using conventional clustering as well as the SANDRA method. Table 2 shows the results as a measure of reproducibility. The similarity is higher and relatively constant for the SANDRA method in all seasons with a minimum of 0.9452 in spring, while conventional CA reaches a minimum of 0.8311 in summer, indicating that conventional CA can be more strongly affected by subsampling. The relatively low similarity in summer for conventional CA may suggest that no sufficiently stable clusters exist in the dataset. However, in comparing with the result for the SANDRA method it

TABLE 2. Similarity between partitions based on odd and even subsamples for conventional and SANDRA CA. The table shows the similarity between partitions measured by the Rand index after assigning all days to the clusters based on subsamples of days selected by odd and even sequence numbers. HWA denotes the conventional method using Ward's hierarchical clusters as seed partitions for optimization with nonhierarchical cluster analysis, while SANDRA denotes the results of advanced optimization.

Season	HWA	SANDRA
DJF	0.8894	0.9489
MAM	0.9055	0.9452
JJA	0.8311	0.9536
SON	0.9601	0.9701

is evident that the reason is just the methodological unsteadiness of conventional CA, which produces a relatively low reproducibility.

The differences between conventional and SANDRA clustering are further evident by the possible bandwidth of the shape of the resulting patterns and even more from the temporal characteristics of the pattern frequency time series. As mentioned above, this difference grows with the number of clusters and the complexity of the dataset, which is highest for the spring season.

Figure 3 shows the centroid pattern and time series plots for the seasonal cluster frequency of cluster 1 (out of $k = 11$) in spring, comparing results derived by SFA k-means of Ward seeds with those derived by the SANDRA scheme. Both methods find that the blocking high over the British Isles is the most important pattern in spring. Because of different assignments of days to the cluster, however, the two methods result in a different shape of the cluster centroid and the variability in time of seasonal cluster frequency is extremely different as they have only 32% of their variance in common (i.e., an r value of 0.57).

While the SANDRA cluster reveals a significant trend in blocking situations, the SFA k-means cluster is blended with other pattern types that hide the signal of a long-term trend. However, it also happens that conventional CA also shows clusters featuring significant trends, whereas equivalent clusters achieved with the SANDRA scheme do not. This demonstrates the danger of using a method achieving its results largely influenced by chance, especially concerning sensitive analyses like trend statistics. Therefore, in this study the SANDRA technique is used to determine highly optimized and methodologically stable clusters.

5. Estimation of appropriate numbers of clusters

A systematic examination of methods for estimating the appropriate number of clusters is given by Milligan

and Cooper (1985). They are designed to find a clear natural clustering in a given dataset and examine the different solutions achieved for a range of cluster numbers and choose the best according to some criteria. Milligan and Cooper (1985) show that the so-called pseudo-F indicator (Calinski and Harabasz 1974) performs very well in identifying the number of clusters in an artificial dataset with inherent clearly separated clusters. It measures the variation between objects of different clusters related to the within-cluster variance; that is, large values indicate solutions where the clusters are well separated.

A somewhat different method is given by Kaufmann and Rousseeuw (1990), who define a silhouette index with maxima indicating that the cluster centroids are located in high-density clouds of objects.

A third indicator used in this study is the so-called overlap ratio based on a method derived by Gerstengarbe and Werner (1997) that examines the overlap of clouds of objects between two clusters, with a ratio of 0(1) for completely separated (overlapping) clusters.

The resulting values for the three indicators plus the ECV index [Eq. (4)] are presented in Fig. 4. Both the ECV and the scaled pseudo-F index are continuously increasing in each season and do not offer any hint to the number of clusters, such as a pronounced elbow in the line. In contrast to the former indices, the silhouette index and the overlap ratio show some irregularities. However, the silhouette index does not show any primary maximum after $k = 2$. Only a secondary maximum is shown for $k = 7$ in autumn. The overlap ratio shows a minimum for $k = 3$ and $k = 4$ in winter, no clear minimum for $k < 14$ in spring, a minimum for $k = 4$ and $k = 13$ in summer, and minima for $k = 4$ and $k = 11$ in autumn. Additionally, it does not fall below 0.5, which means that more than one-half of all gridpoint values of all days overlap into other clusters (to be expected since there are no groups of days with SLP values altogether at all grid points higher or lower than in the other groups). Therefore, any further test on significance of separation as applied in Gerstengarbe and Werner (1997) is superfluous in this context.

Since the principal meaning of all these indices is the same, they should show more or less clear hints for the same number of clusters if there was a clearly preferred separation into clusters. Obviously, this is not the case for this dataset. Even calculating the first difference series of the indices (not shown) does not help, since contradictory results are obtained.

Michelangeli et al. (1995) suggest another method for estimation of the numbers of clusters by a so-called classifiability index and a reproducibility index based on an approach by Chen and Harr (1993) for EOF
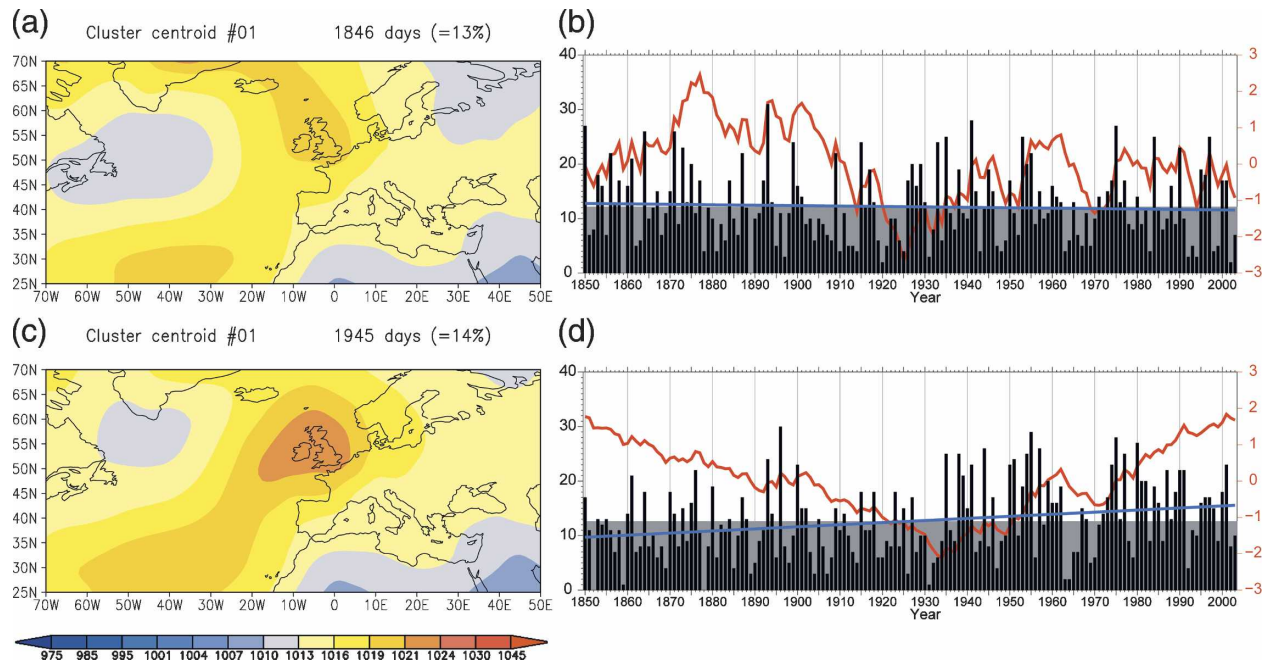
FIG. 3. Comparison of spring cluster 1 derived by conventional *k*-means CA initiated by seeds obtained from hierarchical CA (a),(b) using Ward's method and (c),(d) derived by SANDRA CA. The SLP cluster centroids are shown in (a) (conventional *k*-means CA; 13% of the observations) and (c) (SANDRA CA; 14% of the observations). (b),(d) The seasonal cluster frequency (black bars; left *y* axis), cumulative anomalies of seasonal cluster frequencies (red line; right *y* axis; see section 1), long-term mean frequency (gray shading; left *y* axis), and linear trend line (blue; left *y* axis) using (b) conventional *k*-means CA and (d) the SANDRA scheme. The correlation coefficient of seasonal cluster frequencies is 0.56.

analysis. Since this approach has limitations caused by the usage of conventional *k*-means CA, we applied a modified scheme of the reproducibility index (see the appendix) to our data; however, we did not find any clear suggestion for the number of clusters. Altogether we conclude that there is no reliable determination of an optimal number of clusters, which is supported by the different indices.

Therefore, an external indicator was used allowing a rough estimate of the number of pressure patterns that are needed to represent the majority of the variance within the set of pressure fields. For this objective a *t*-mode PCA of the dataset was carried out using the correlation matrix of the pressure fields. The PCs were varimax rotated for numbers of 2 up to 20 PCs. The so-called dominance test (Jacobeit 1993) gives an estimate of the number of PCs that represents the highest amount of explained variance; at the same time all extracted PCs were still realistic manifestations of the atmospheric circulation and not purely artificial constructs by linear combinations of the input data. To ensure this criterion, each PC of each rotation result was tested to show up with a loading dominating at least one input variable and fulfilling the following three conditions.

1) The loading must be greater than one standard deviation above all other loadings.
2) The loading must be the uniquely leading one for this object, which ensures that the PC is the only one explaining a high amount of variance of this variable instead of linear combinations of two or more PCs. This condition is assumed to be fulfilled if the loading is more than one standard deviation higher than for the second highest PC on this variable.
3) For the case of a *t*-mode PCA the loading must represent a significant correlation between the field of scores and the original pressure field taking into consideration the spatial autocorrelation in both fields. For testing the significance of a spatial correlation the degrees of freedom have been reduced using the procedure of Moran's *I* (Dutilleul 1993). A significance level of 0.01 was chosen in order to allow only loadings indicating a high degree of similarity to pass the test.

The application of this extended dominance test for RPCA may be related to the decision on the number of clusters since it gives information on how many clearly different (i.e., orthogonal in case of RPCA) and important patterns exist in the dataset as analogues to cluster
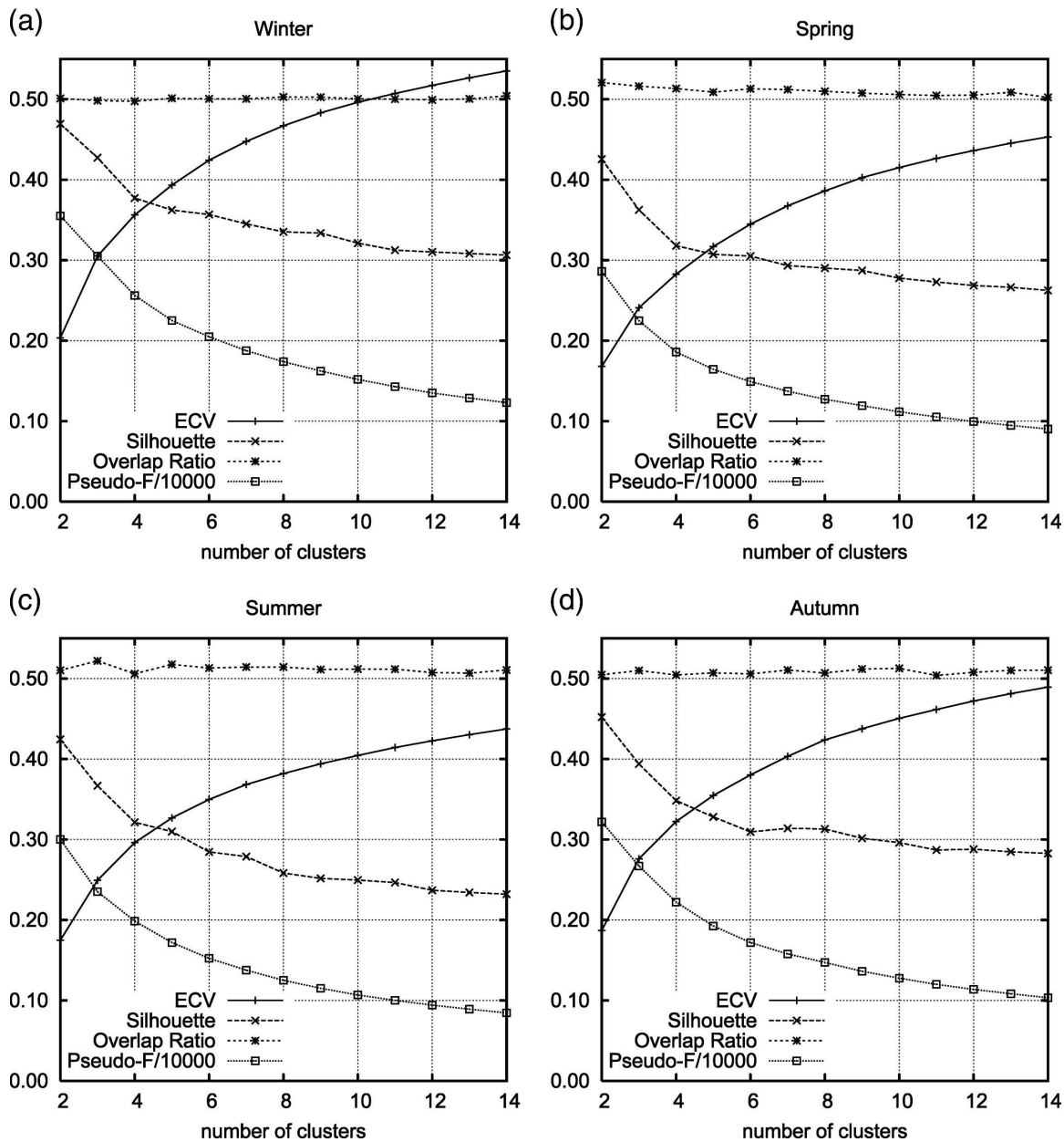
FIG. 4. Various statistics for estimating the appropriate number of clusters for (a) winter (DJF), (b) spring (MAM), (c) summer (JJA), and (d) autumn (SON). Each plot shows explained cluster variance (plus signs), silhouette index (times signs), cluster overlap ratio (asterisks), and pseudo-F scaled as a fraction of 10 000 (open squares).

centroids. The clustering process itself just determines the shape of these patterns by leaving the constraint of orthogonality but enhancing the similarity between the objects in each cluster. In this sense CA serves as a kind of empirical oblique rotation of patterns representing vectors in the multidimensional space targeted on optimizing ECV. Since there is a discrepancy in orthogonality between the two methods, this test cannot precisely indicate the appropriate number of clusters, but

in lack of any other reliable hint it suggests a suitable minimum number of patterns needed to adequately reflect the variability of pressure configurations in the dataset.

The resulting numbers of clusters are given in Table 3 for the 2- and 3-month seasons and may confirm the usefulness of this method since they reflect the seasonal cycle by relatively high numbers in the transitional seasons (characterized by high variability of pressure con-

TABLE 3. The number of clusters for each 2- and 3-month season assessed using the dominance criteria.

Season	k
DJ	7
JF	7
FM	9
MA	11
AM	10
MJ	8
JJ	5
JA	6
AS	8
SO	10
ON	7
ND	7
DJF	9
MAM	11
JJA	6
SON	8

figurations), high to intermediate numbers for the winter, and lowest numbers for the summer seasons (reduced variability in the large-scale pressure field).

6. Classification results

Figures 5–8 give a comprehensive characterization of the resulting cluster sets for the four 3-month seasons by showing the cluster centroid patterns, the seasonal cluster frequencies (SCFs), and the associated 1948–2003 mean temperature anomalies for each cluster. The latter were calculated as composites of departures of each daily mean within each cluster from the long-term mean using NCEP–NCAR reanalysis data (Kalnay et al. 1996; Kistler et al. 2001).

In addition to the SCF time series, the normalized series of SCF cumulative anomalies are shown. It is the sum of anomalies (SCF of each season minus the long-term mean frequency) accumulated over time, starting from the first year until the reference year. An advantage of cumulative anomalies relative to low-pass-filtered time series is that they point out periods of predominant negative or positive anomalies more clearly: in case of predominant negative anomalies the curve is decreasing while it is increasing in phases of predominant positive anomalies.

The SCF time series have been tested for significant trends (95% confidence level) over the whole study period using the Mann–Kendall trend test as well as the parametric trend-to-noise ratio test for linear trends. The latter is simply the ratio between the linear trend value and the standard deviation of the time series, which indicates a linear trend at the 5% significance level by values above 1.96 if normal distribution of the

sample may be assumed (Rapp 2000). This test is known to be very rigorous since any linear trend is easily disturbed by temporal interruptions of the trend.

The trend-to-noise ratio indicates that none of the SCF series show an uninterrupted linear trend over the whole period, whereas the Mann–Kendall trend test indicates for a number of SCF series that increases within the time series are significantly more than decreases or vice versa. Results are given in detail in Table 4, while the most prominent features are highlighted below.

In winter (Fig. 5) cluster 4, a westerly pattern with a strong cyclone in the northeastern Atlantic advecting warm maritime air onto the European continent, shows a significant positive trend (according to the Mann–Kendall test) while cluster 6 significantly decreases in SCF. It resembles the so-called *Winkelwestlage* of the HB77 classification with a westerly flow over the Atlantic Ocean but a sharp turn to southerly flow over the continent caused by a strong continental high pressure cell. The latter causes low winter temperatures over the eastern part of Europe.

Other clusters partly show remarkable long-term variations but no overall trend. The most prominent case is cluster 1, which reflects a NAO-like pattern and causes warm winter conditions over central and northern Europe. This pattern shows the well-known steep increase in the last two decades but also high frequencies in the first two decades 1850–70. Remarkably, the beginning of the twentieth century is not even a positive phase (as in Hurrell and van Loon 1997). The main reason for this might be that cluster 1 does not represent very much variability of the NAO (correlation of 0.62 between SCF and a seasonal NAO index derived from the same dataset) but more variability of pressure gradients over Europe (because of the optimum-interpolation error weighting of the data used for clustering).

Other remarkable changes concern cluster 7 (a cutoff low in the Gulf of Genoa often arising with upper cold troughs) and cluster 8 (high pressure bridge over Europe and relatively low pressure over the Mediterranean). Both patterns are associated with cool winter conditions over Europe and show a phase of relatively high SCF between 1930 and 1970, a period when the warming trend of the twentieth century in Europe was interrupted. In summary the winter long-term changes of daily pressure patterns largely support a warming throughout the period since 1850 as well as its interruption in the mid-twentieth century.

Out of 11 clusters in spring (Fig. 6) four clusters show significant long-term trends. Cluster 1, a blocking high over Great Britain with northerly cold-air advection
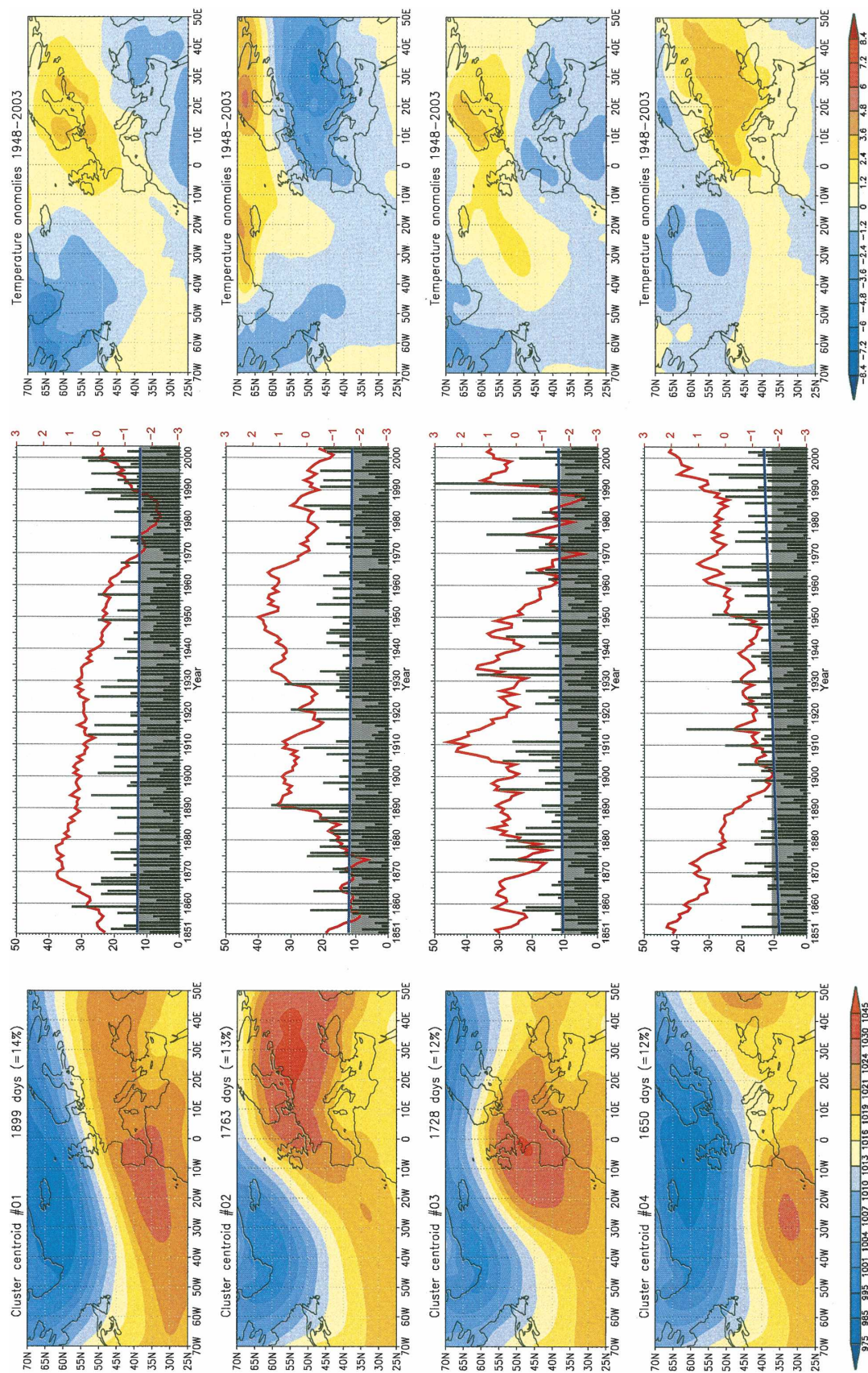
FIG. 5. Winter-season (DJF) clusters generated using the SANDRA scheme: (left) SLP centroid patterns (hPa). (center) Seasonal cluster frequency (black bars; left y axis), their normalized cumulative anomalies (red line; right y axis; see text), long-term mean frequency (gray shaded; left y axis), and linear trend line (blue line; left y axis). (right) Associated mean temperature anomaly field for 1948–2003 (K). Clusters are ordered by the highest to lowest number of days in each cluster.
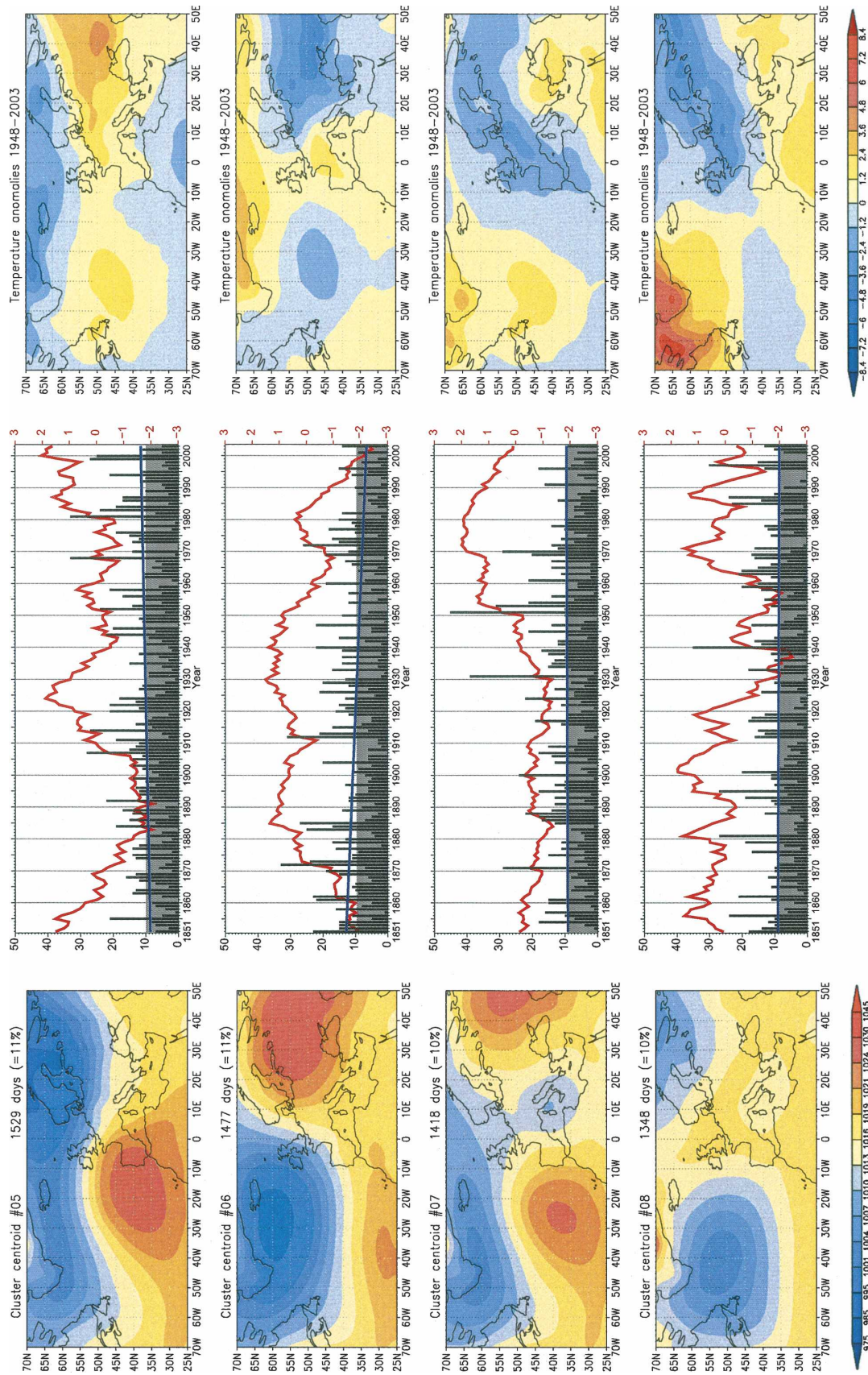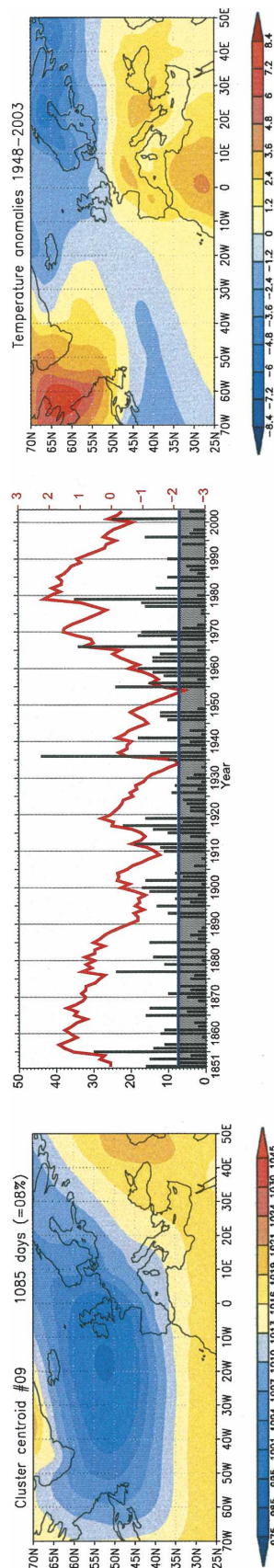
FIG. 5. (*Continued*)

FIG. 5. (Continued)

for central Europe, is strongly increasing as well as cluster 3, a strong Azores high and weak Icelandic low with cold conditions for southwestern Europe but close to average ones for central Europe. In contrast to this, the SCF of cluster 5, showing a low pressure cell over the western Atlantic Ocean and the eastern Baltic region and cold conditions for most of Europe, decreased strongly. Also cluster 8 decreased significantly but to a lesser extent. This cluster represents a strong cyclone over Great Britain and a warm sector in central and eastern Europe while cold sectors cover the northern and western borders of Europe. In summary SCF changes in spring do not imply clear impacts for European temperature anomalies: the cool clusters 1 and 5 show trends of opposite direction as well as the indifferent or warm clusters 3 and 8.

Three of the six summer clusters (Fig. 7) show significant trends: cluster 2, a retreated Azores high and cyclonic influence over Scandinavia leading to cold northerly airflow toward Europe, has increased. Cluster 5, an extended high over western Europe with warm/cold conditions over western/eastern Europe, and cluster 6, a low pressure system over the northwest of the British Isles, with the warm sector over central Europe and cold conditions in the westernmost part, have decreased. Besides these significant trends, cluster 1, an Azores high extension over central Europe with associated warm conditions, shows distinct decadal variability (high-frequency periods until 1880, from 1930 to circa 1948, in the 1960s and early 1970s, and again in the most recent years; see Fig. 7).

In autumn (Fig. 8), cluster 1 shows a significant positive trend. This cluster represents slightly warmer conditions than are normal over eastern Europe that are due to a retreated southeastern European high. In contrast to this, cluster 4, including a strengthened and northward extended high with strong cooling throughout continental Europe, shows a significant strong downward trend, as does cluster 7, a weaker but northwesterly extended high with cool conditions over eastern Europe.

7. Application for diagnostics of European temperature change

The mean surface temperature anomaly fields for each cluster in the period 1948 to 2003 (right column in Figs. 5–8) may further be used as calibration for estimating past temperature anomalies. This allows an additional comparison between SANDRA and conventional clustering and addresses the general ability of the SLP pattern classifications to explain the link between atmospheric circulation and surface air temperature anomalies, including their long-term variability. This is
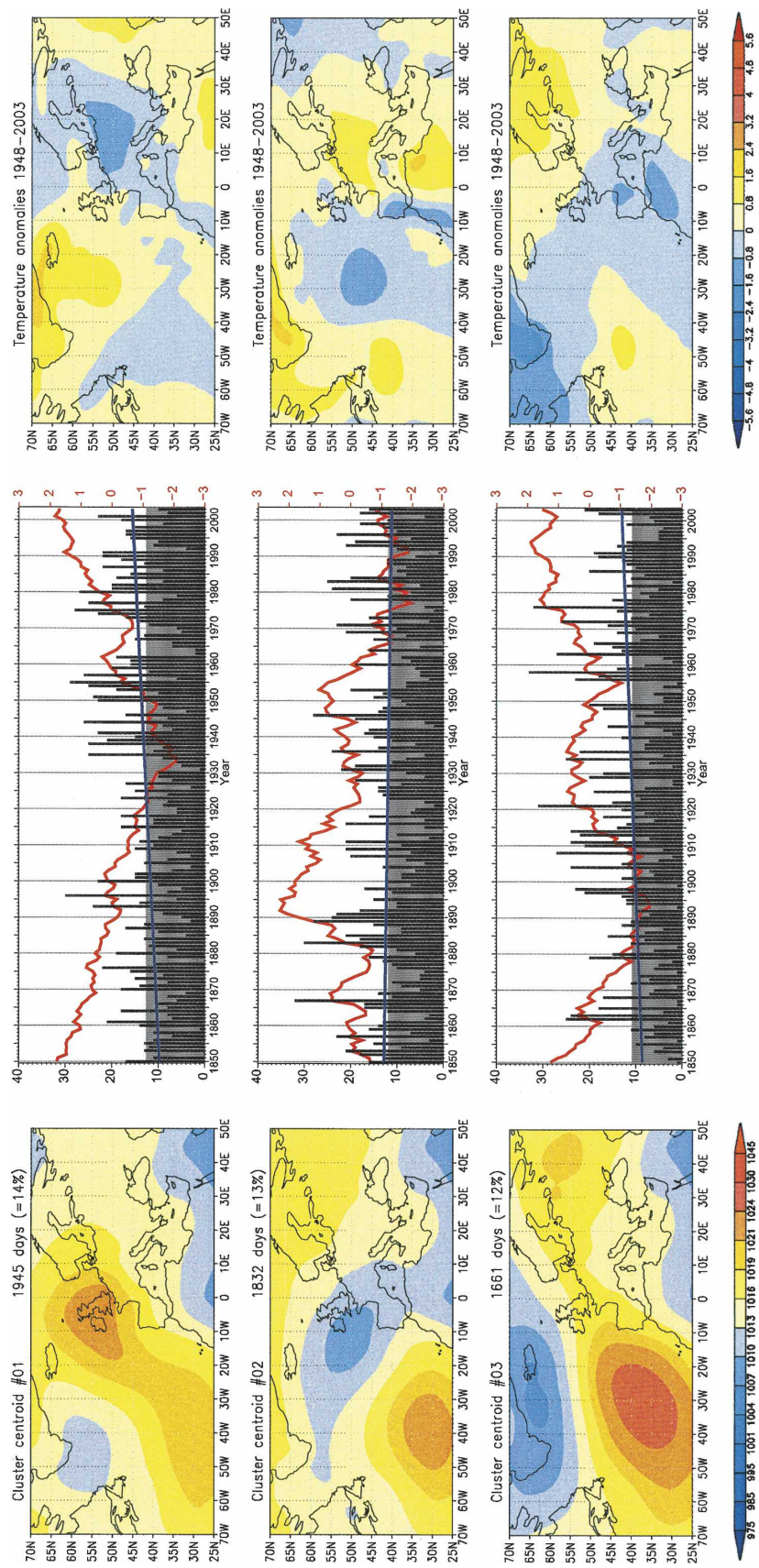
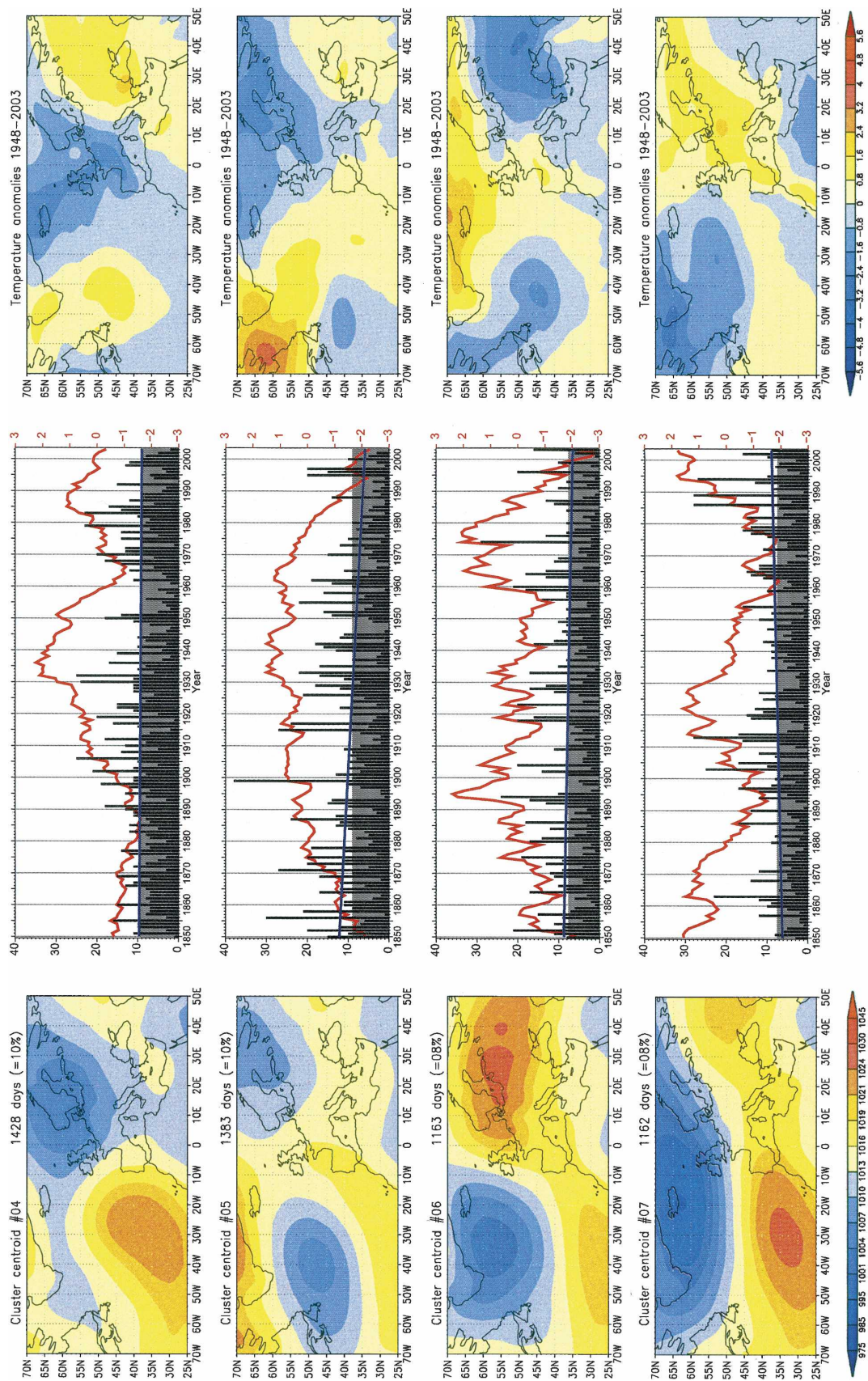FIG. 6. As in Fig. 5, but for spring (MAM).
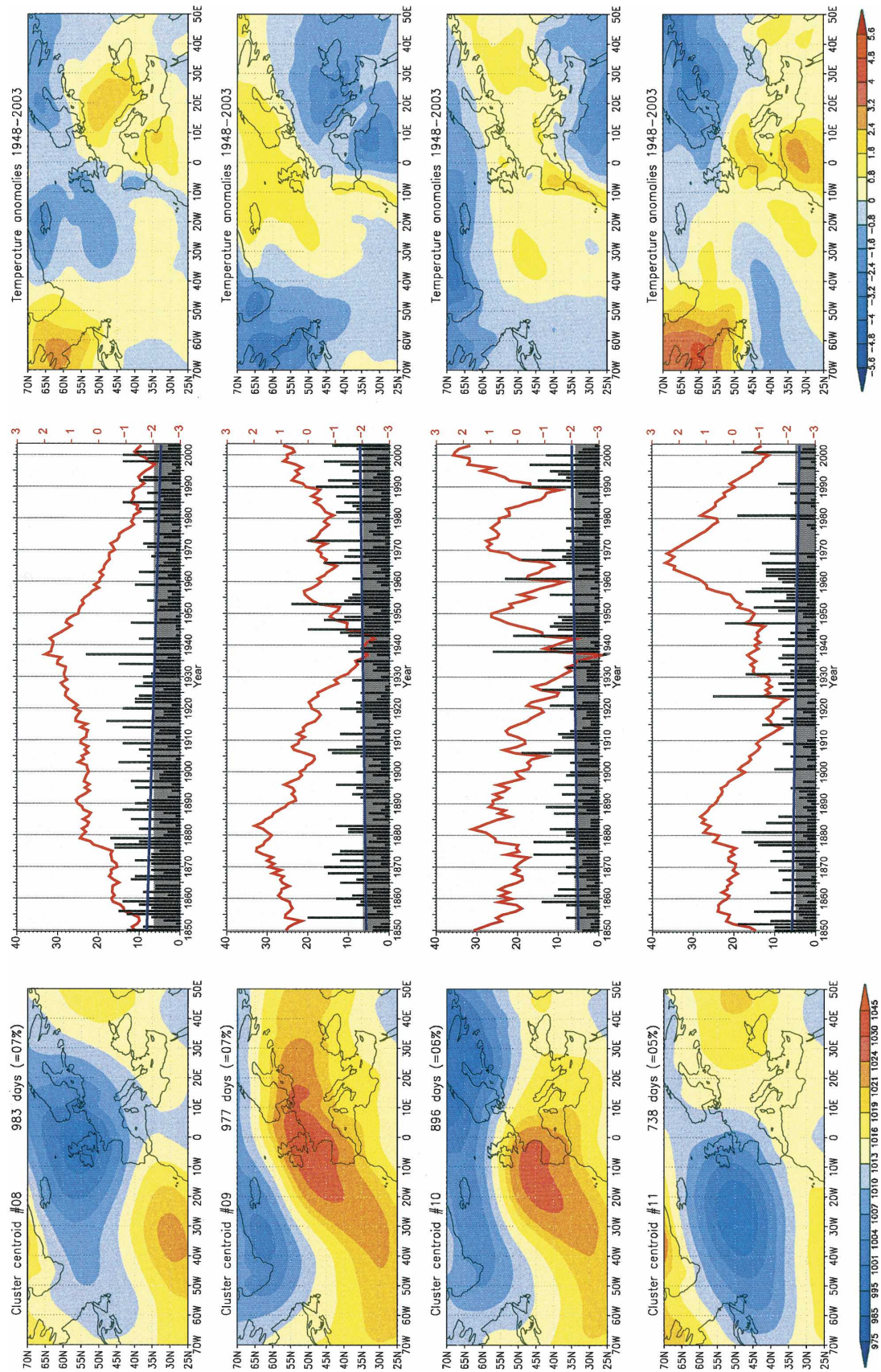
FIG. 6. (Continued)

FIG. 6. (*Continued*)

realized by a simple composite model and a multiple regression analysis.

The composite model estimation for the seasonal temperature anomalies is derived by building the sum of the cluster-specific temperature anomalies for each day in the season according to the cluster assignment of that day and dividing the sum by the number of days in the season for each year. In this way seasonal temperature anomalies are reproduced back to the year 1851/52 for the central European area from 45° to 55°N and from 5° to 20°E. These reconstructions are compared with the central European temperature index based on the data of Jones and Moberg (2003) (called "observed" hereinafter).

For comparison, temperature reconstructions are included based on the SANDRA classification and on conventional k-means CA (initiated by Ward seeds for winter, by t-mode PCA seeds for spring, by s-mode PCA seeds for summer, and by k-mode PCA seeds for autumn). The latter classifications do not represent the worst local minima but represent typical solutions obtained by conventional k-means clustering. Additionally multiple least squares regression models have been set up using the SCF time series of the SANDRA CA and the conventional CA results mentioned before, as predictors within the whole period.

Table 5 shows the amount of temperature variance explained by the estimation models, as well as indicators for trends in the observed and estimated temperature time series. As expected the model's qualities are generally low except for the winter season with more than 50% of explained variance. The explained variances of the models based on the SANDRA clustering and those based on conventional k-means CA show a tendency for higher performance of some few percent for the SANDRA-based models. In winter and autumn, all of the estimations explain largely the same variance (minimum of 80.8% shared variance among the estimates; not shown in Table 5), whereas in spring and summer the differences between the composite estimates and the regression estimates increases (minimum of 31.9% shared variance between SANDRA regression estimates and conventional k-means CA composites for spring and minimum of 29.4% shared variance between SANDRA CA composites and conventional k-means CA regression estimates for summer).

Figure 9 shows the observed and predicted temperature time series, including their normalized cumulative anomalies representing periods of generally equal signs for anomalies by increase or decrease of the curve. An outstanding feature is the high degree of correspondence in the cumulative anomaly plots for winter suggesting a close link between long-term variability of
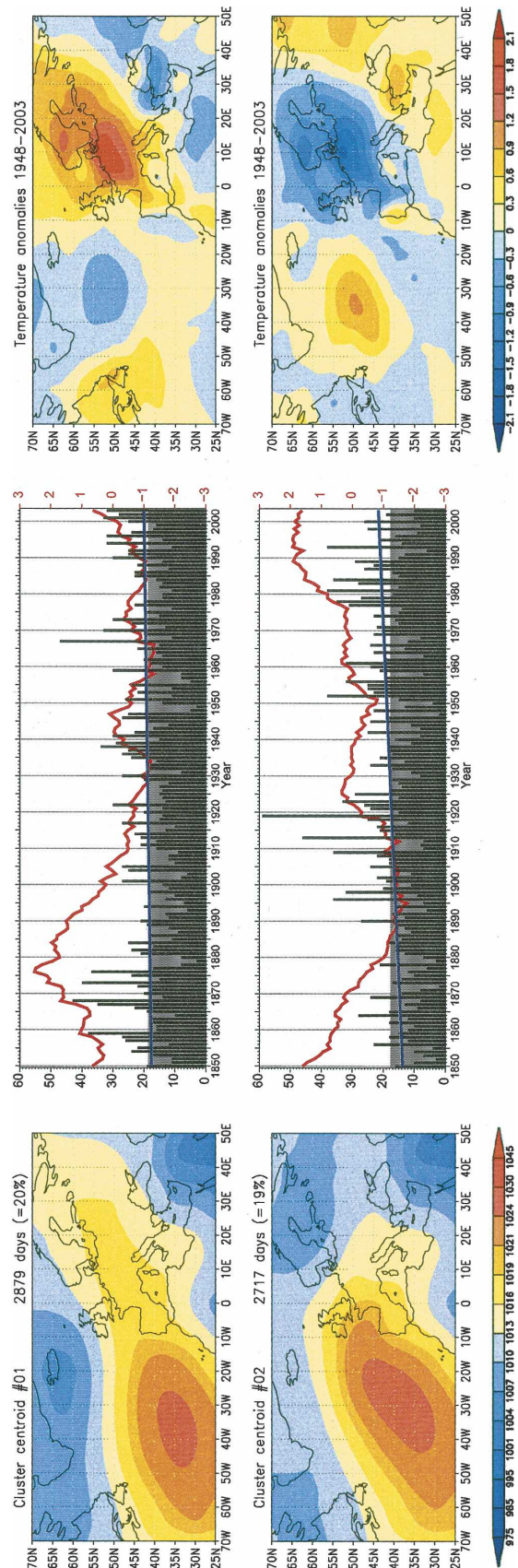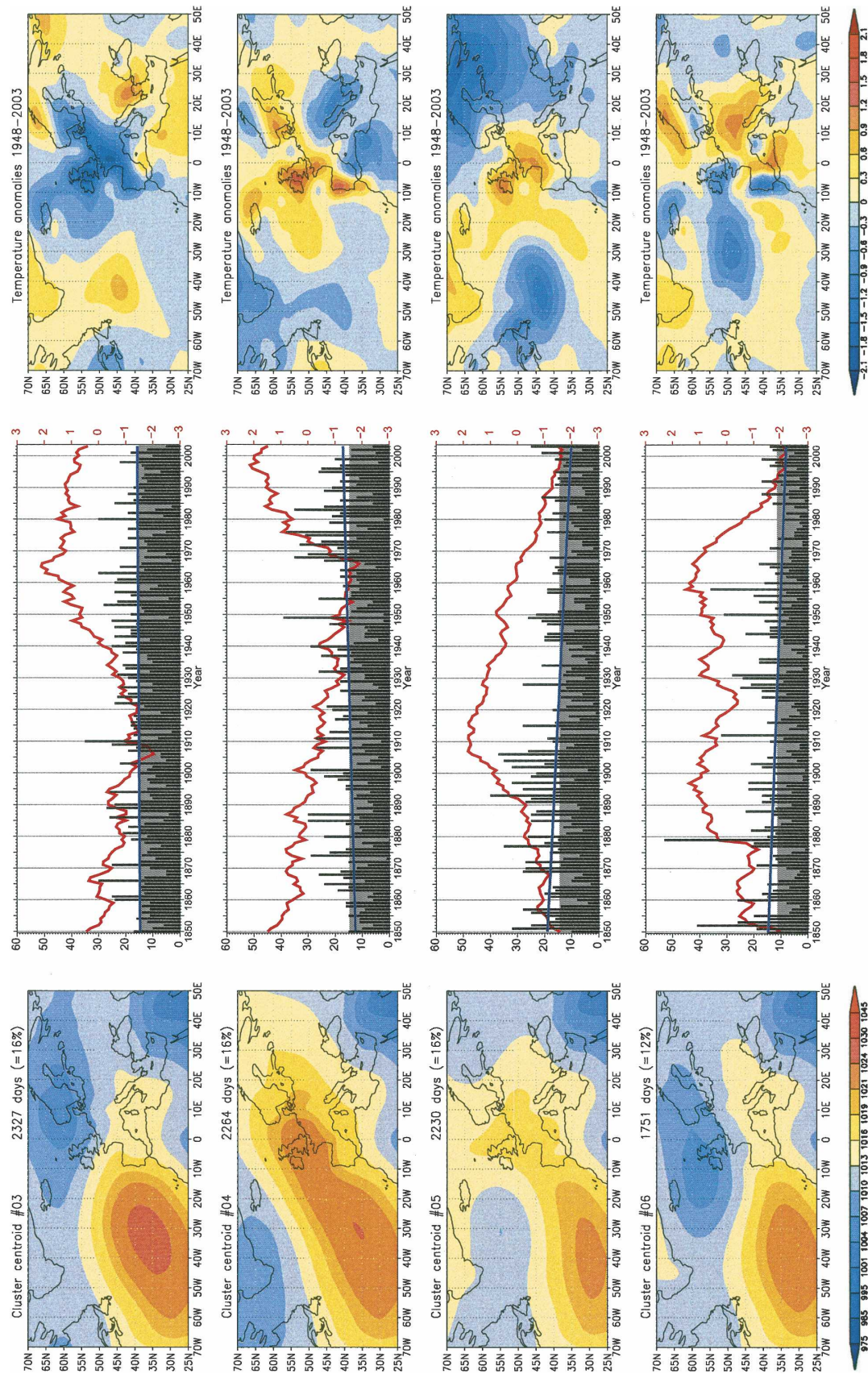

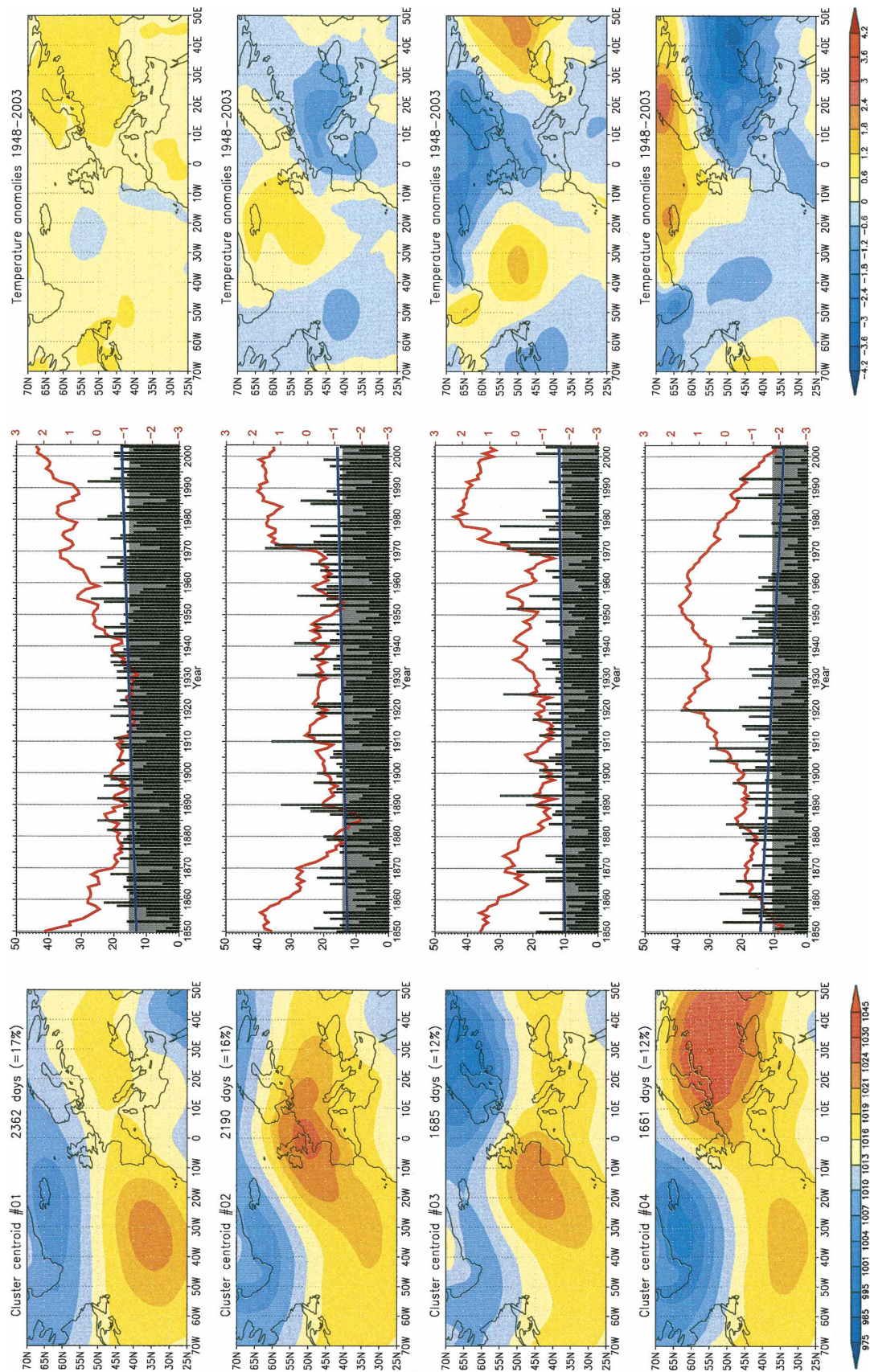FIG. 7. As in Fig. 5, but for summer (JJA).

FIG. 7. (*Continued*)
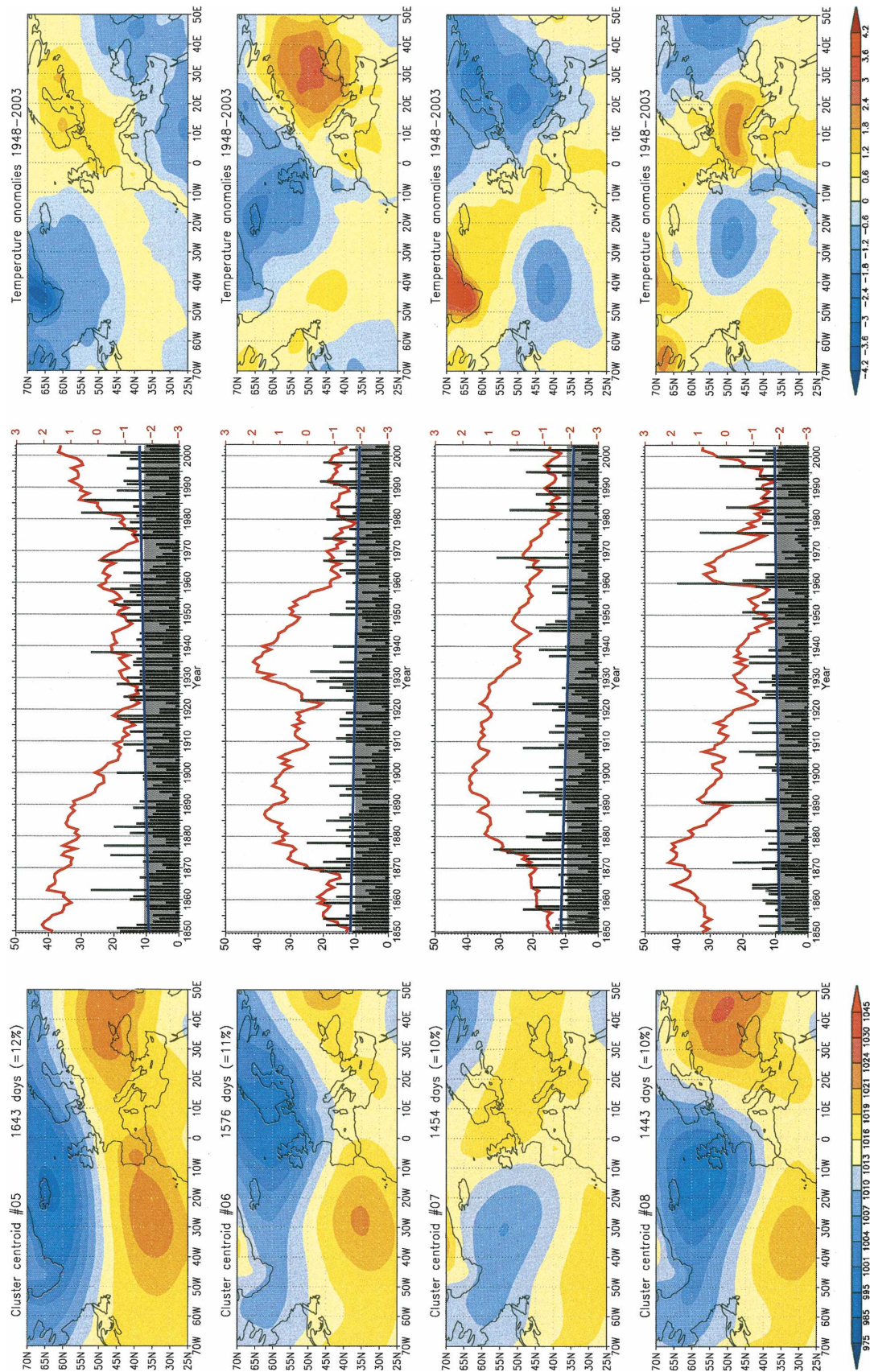
FIG. 8. As in Fig. 5, but for autumn (SON).

FIG. 8. (*Continued*)

TABLE 4. Seasonal cluster frequency trends and their significance for clusters obtained using the SANDRA clustering scheme for all 3-month seasons within the period 1850–2003. Trend/noise is the trend-to-noise ratio. The P is the Mann–Kendall trend-test probability, and the Z is the trend-test score. Significant values for Mann–Kendall P (95% significance level) are highlighted in boldface.

Cluster	Linear trend	Trend/ noise	Mann– Kendall P	Mann– Kendall Z
		DJF		
1	−0.7670	0.0993	0.683	−0.409
2	−1.0502	0.1441	0.772	−0.290
3	1.4306	0.1569	0.907	0.117
4	4.6947	0.6468	**0.045**	2.002
5	2.8867	0.4123	0.430	0.789
6	−6.3652	0.9013	**0.001**	−3.311
7	0.3692	0.0495	0.986	0.017
8	−0.6101	0.0862	0.408	−0.828
9	−0.5943	0.0746	0.271	−1.101
		MAM		
1	5.9020	0.8730	**0.005**	2.823
2	−1.7535	0.2539	0.392	−0.856
3	4.4041	0.5912	**0.015**	2.443
4	−0.7350	0.1371	0.442	−0.769
5	−6.2730	0.9052	**0.000**	−3.496
6	−2.1823	0.3674	0.092	−1.686
7	2.6662	0.4468	0.153	1.428
8	−3.3884	0.7255	**0.007**	−2.714
9	1.6168	0.3210	0.283	1.073
10	1.6366	0.3136	0.448	0.759
11	−1.8935	0.3504	**0.035**	−2.103
		JJA		
1	2.3917	0.2931	0.152	1.433
2	7.7749	0.8605	**0.001**	3.468
3	0.9559	0.1471	0.421	0.805
4	4.4898	0.5493	0.083	1.734
5	−8.8655	1.0540	**0.001**	−3.446
6	−6.7468	0.8168	**0.006**	−2.761
		SON		
1	4.5552	0.8651	**0.002**	3.037
2	3.1380	0.4269	0.136	1.492
3	1.8266	0.2871	0.300	1.037
4	−7.2244	0.9988	**0.000**	−3.876
5	2.8452	0.4597	0.103	1.629
6	−2.7036	0.4596	0.119	−1.559
7	−3.9780	0.5678	**0.007**	−2.681
8	1.5411	0.2234	0.945	−0.069

temperature and circulation pattern types at least in terms of the sign of anomalies.

Further evidence is achieved by comparing the SCF series of individual pattern types with the temperature time series in Fig. 9. Except for the years 1865 to 1870, when clusters 1, 4, and 5 had a small SCF peak phase, mainly cool winter conditions are found for the second half of the nineteenth century accompanied with relatively low frequencies for the warm clusters 4, 5, and 9. Cluster 1 had a mainly positive phase up to 1870 but

was declining afterward. On the other hand, the cold clusters 2 and 6 had a high SCF phase until 1890. The warming in the first 20 years of the twentieth century seems to be induced by a high phase of cluster 5 and partly of cluster 4, whereas the cool cluster 8 had a decline between 1900 and 1910 and the cool cluster 2 had a decline between 1910 and 1920. Notably, the warm winter phase at the beginning of the twentieth century is reconstructed by all models without a distinct high phase of the westerly NAO-like pattern 1. High SCF of cluster 5 (westerly flow over the continent) and low SCF of cluster 2 (continental high) seem to be the main reasons.

In spring, the composite models fail to approximate temperature variations except for the signal of recent warming, which is reproduced (second row in Fig. 9, red lines). Even for this event that is better reproduced by the multiple regression analysis, however, it is difficult to identify responsible clusters, since there is a complex mixture of increasing and declining SCFs for cold and warm clusters.

Also in summer the explained variance is generally low (see Table 5). Nevertheless, major aspects of long-term variability are reproduced (Fig. 9, third row) like the relatively warm period up to 1880 (though overestimated). The same is true for the relatively cool phase from 1880 to the mid-1920s, the following warm phase, and the recent enhanced warming. The early warm period seems to be induced by cluster 1 with its Azores high extending on the European continent, whereas cluster 2 (northern cool air influx due to northward extension of the Azores high) plays a major role for the subsequent cooling around 1900. The following warm–cool–warm sequence is traced again by the SCF series of cluster 1 supported by complementary frequency changes of the cold clusters 2 and 3.

In autumn, high correspondence between reproduced and observed long-term temperature variations is found again, even for the composite model results. The main feature in the cumulative anomalies—the breakpoint in 1922 between cold conditions before and warm conditions afterward—is reproduced by all models precisely. The comparison between the curves of normalized cumulative temperature anomalies in Fig. 9 (bottom row) and the cumulative anomalies of SCF in Fig. 8 reveals a high degree of correspondence with warm clusters 1, 5, and 8 and cool clusters 4 and 7: correspondingly they show turning points, an onset of a high phase (cool cluster 7), or an offset of a low phase (warm cluster 8) around 1922, respectively.

In all seasons, positive trends are indicated for the observed time series, ranging between 0.46 K in sum-

TABLE 5. Comparison of statistical models for estimating observed seasonal temperature variability based on circulation pattern classification for the 3-month seasons (DJF, MAM, JJA, SON) for 1851/52–2003. Comparisons are made between classifications based on SANDRA CA and conventional k-means CA using a primitive composite model and a multiple least squares regression model, both including seasonal cluster frequency of all clusters as predictors. Classifications used for comparison are conventional k-means CA initiated by Ward seeds for winter, by t-mode PCA seeds for spring, by s-mode PCA seeds for summer, and by t-mode PCA seeds for autumn. Here, "explained var." denotes the fraction (%) of explained variance of observed seasonal temperature, "linear trend" denotes the linear trend of the observed/estimated time series (K), "trend/noise" is the trend-to-noise ratio (absolute linear trend divided by standard deviation), and "Mann–Kendall P" denotes the probability for making an error in assuming a trend according to the Mann–Kendall trend test. Significant values of Mann–Kendall P (95% significance level) are highlighted in boldface.

		Composite models		Regression models	
	Observed	SANDRA CA	Conventional CA	SANDRA CA	Conventional CA
			DJF		
Explained var.	100.00	55.59	50.49	59.04	58.43
Linear trend	1.39	0.26	0.24	0.57	0.63
Trend/noise	0.84	0.46	0.44	0.45	0.50
Mann–Kendall P	**0.00**	0.17	0.16	0.13	0.07
			MAM		
Explained var.	100.00	16.03	12.93	35.88	33.11
Linear trend	1.23	−0.01	−0.03	0.42	0.34
Trend/noise	1.25	0.04	0.15	0.71	0.61
Mann–Kendall P	**0.00**	0.71	0.67	**0.01**	**0.03**
			JJA		
Explained var.	100.00	13.11	12.18	33.92	31.89
Linear trend	0.46	−0.10	−0.11	−0.17	−0.12
Trend/noise	0.59	0.48	0.59	0.37	0.26
Mann–Kendall P	0.06	0.06	0.05	0.18	0.44
			SON		
Explained var.	100.00	31.16	30.72	36.51	33.99
Linear trend	0.95	0.18	0.16	0.48	0.46
Trend/noise	1.07	0.65	0.58	0.90	0.89
Mann–Kendall P	**0.00**	**0.03**	**0.05**	**0.00**	**0.00**

mer and 1.39 K in winter ("linear trend" in Table 5). Except for summer, these trends are significant according to the Mann–Kendall trend test (95% confidence level), whereas the trend-to-noise ratio indicates no significant linear overall trends (values greater than 1.96; all samples are normal distributed). Generally the estimation models show weaker positive trends, especially the composite models. The estimations for winter temperatures all include a positive trend, with maxima of 0.57 and 0.63 K for the regression estimates (observed 1.39 K), the latter being significant at the 90% confidence level only though. In spring, both composite estimates show nearly zero trends, whereas the two regression-based time series indicate significant positive trends (according to the Mann–Kendall trend test) of 0.42 and 0.34 K (observed 1.23 K). In summer the observation shows an insignificant trend of 0.46, whereas all estimate models wrongly show weak negative trends (but even significant for the composite model using conventional CA). The positive trend in autumn (0.95 K observed) is indicated by all estimation models significantly (Mann–Kendall), with maxima for both regression estimates of 0.48 and 0.46 K (i.e., roughly one-half of the observed trend). Concerning the comparison

between SANDRA CA and conventional CA in depicting the observed trends, no general preference is found.

8. Conclusions and discussion

Analysis of synoptic and dynamic climatology usually involves the use of large and complex datasets of circulation patterns. A number of studies indicate that there might be preferred numbers of particular circulation states (e.g., Michelangeli et al. 1995; Corti et al. 1999). However, we cannot find a clear separation of clusters inherent in the dataset used in this study. This is consistent with some recent publications and there is increasing evidence that there is no significant natural separation between clusters of circulation patterns (Stephenson et al. 2004; Christiansen 2007). As a probable result of a high number of smooth transitions in the shape between different individual patterns, any indices designed to discover an appropriate number of clusters fail with respect to this decisive issue in classification techniques. This shortcoming has been compensated in the present study by a rough external estimator suggesting how many t-mode principal components do fit into the dataset.
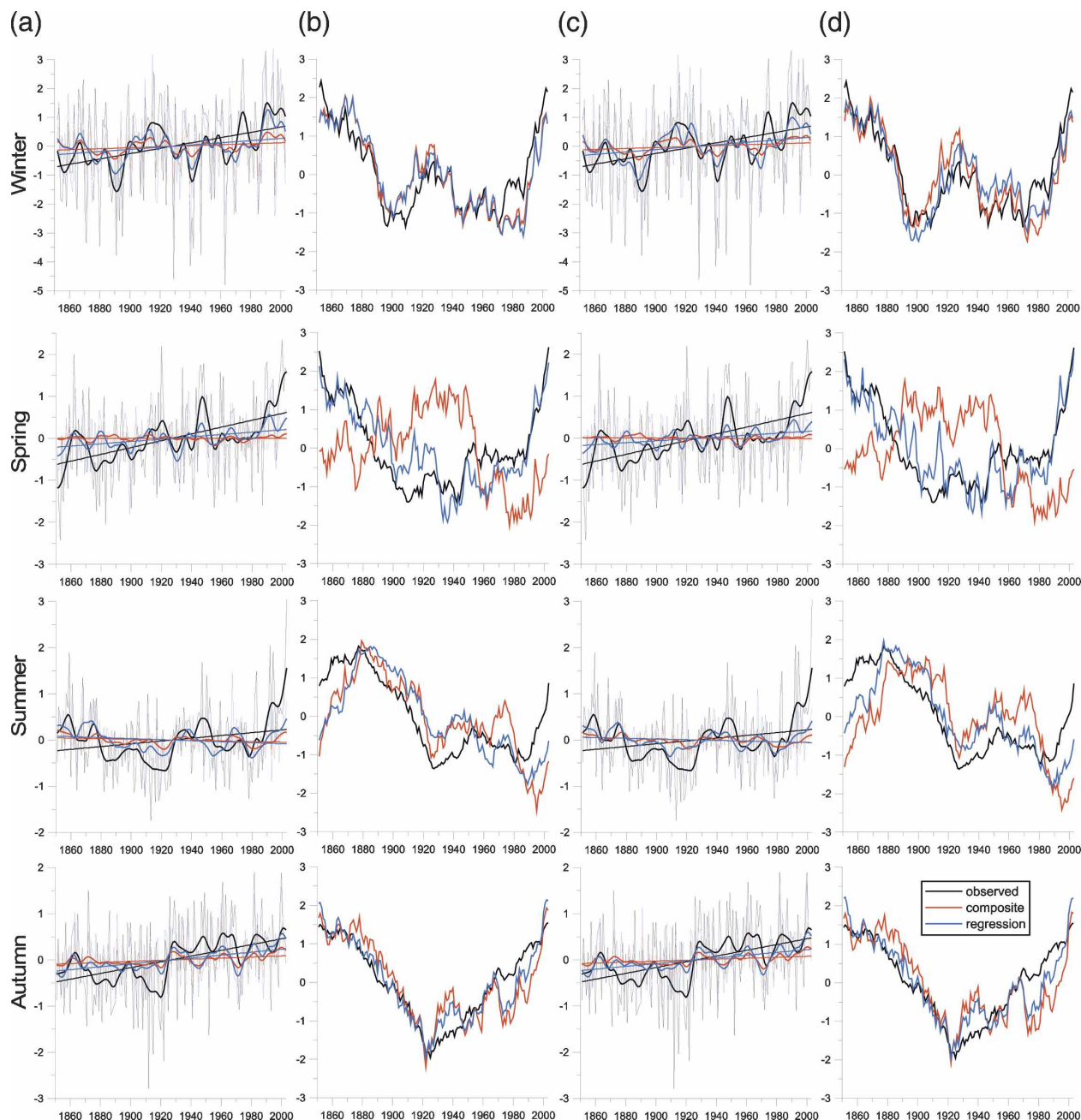
FIG. 9. Seasonal temperature anomalies for central Europe (45°–55°N, 5°–20°E) as observed (black), estimated by classification-based composite models (red), and estimated by multiple least squares regression models using seasonal frequencies of circulation pattern clusters as predictors (blue). (a) anomalies (K; thin lines), low-pass-filtered anomalies (thick curves), and linear trend (thick lines) for observed temperatures and those estimated using conventional k-means cluster analysis. (b) Normalized cumulative anomalies for observed temperatures and those estimated using conventional k-means CA. (c) As in (a), but for estimations using SANDRA CA. (d) As in (b), but for estimations using SANDRA CA. All anomalies are calculated according to the 1851/52–2003 long-term mean for each season.

Another and even more important consequence of the complexity of this dataset is the failure of conventional k-means cluster analysis in finding a stable result of high quality. Conventional k-means CA converges to an undetermined local optimum by chance and should not be used for datasets offering various local optima, since the differences between local optima can be significant for trend analysis of pattern frequencies.
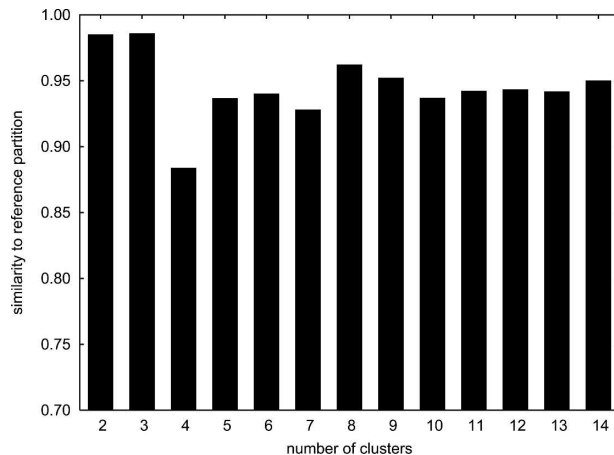
FIG. 10. Reproducibility of the winter-season (DJF) clustering for number of clusters between 2 and 14. The results are based on the average similarity between 100 half-sample clustering results and full-sample clustering results using the SANDRA clustering scheme; see text for more details.

However, even if there seems to be no clear separation of clusters in the dataset and the complexity of the dataset leads to the above mentioned problems for conventional CA, there are a number of hints that the presented clusters are sufficiently stable and reproducible to speak of meaningful patterns.

- In a subsampling experiment using separately odd and even days there is a high similarity (Rand index >0.94) in all seasons.
- Another subsampling experiment used to examine reproducibility by selecting subsets of one-half of the sample size by random results in a Rand index not below 0.93 in winter for numbers of clusters between 2 and 14 except for the number of four clusters (Fig. 10, discussed in more detail in the appendix).
- Patterns of cluster centroids based on three subperiods (1850–1900, 1901–51, and 1952–2003, not shown) enable us to clearly identify the patterns of cluster centroids based on the whole period in all cases for winter and summer and in most of the cases for spring and autumn, while only some clusters are merged or split as a result of the long-term changes of the circulation within the whole 154-yr period.
- If comparing patterns for increasing numbers of clusters [not shown, available at http://www.cru.uea.ac.uk/projects/emulate (atmospheric circulation patterns)] almost all patterns of the centroids can be recognized except those that are split up in order to build a new additional cluster.

This means there is some structure in the dataset, but it is not clear enough to define a certain unique number of clusters, and it is so complex that conventional CA can be confused by local optima. The latter unfavorable effect has been substantially reduced by a new combination of simulated annealing clustering and multistart techniques (implemented in a common procedure called SANDRA) that achieves results of highest clustering quality and reproducibility in comparison with all other methods for CA.

Using the SANDRA procedure, circulation types have been determined for the period 1850–2003 represented by their centroid patterns, seasonal frequency time series, and associated temperature anomaly fields (the latter based on the reanalysis period 1948–2003).

The derived cluster frequency time series show remarkable long-term variability on the decadal to multidecadal time scales and some significant overall trends are indicated in all seasons. In winter, an NAO-like pattern (Fig. 5, cluster 1) shows high frequencies between 1850 and 1870 and again since 1985, but no overall trend is observed. However, there is a significant increase of a westerly pattern mainly steered by cyclonic activity north of the British Isles (cluster 4). Both of these zonal patterns are characterized by generally warm conditions in Europe. A significant negative trend is observed for a meridional pattern (cluster 6) representing strong Russian highs with cold conditions especially over eastern Europe but also a pronounced North Atlantic low (known as *Winkelwestlage* from HB77). Its decrease can explain a strong decrease of cold extreme events in southeastern Europe, whereas the mean temperature and the positive winter extremes increased less in this region (Moberg et al. 2006). Moreover, the high frequencies of this cluster between 1851 and 1885 with its strong North Atlantic low in conjunction with the strong Russian high may explain the decoupling of the usual complementary link between meridional and zonal pressure configurations for this period as described by Luterbacher et al. (1999) and Slonosky et al. (2000). In spring, the most pronounced trends are shown for increasing blocking high situations (Fig. 6, cluster 1) and decreasing patterns with cyclones over eastern Scandinavia (cluster 5) leading to cold-air advection for Europe. The summer season has seen dramatic long-term changes. While three out of six types, mainly differing in the direction and intensity of the Azores high extension onto Europe, show significant trends, cluster 1 (Fig. 7), characterized by high pressure cells centered directly over central Europe, does not show a significant overall trend but rather shows distinct spells of above- and below-normal frequencies: until 1875 there has been the most pronounced accumulations of these patterns followed by the 1930s and finally by a spell since around 1980. All of

these phases have been unusually warm in central Europe. These results agree with the findings of Della-Marta et al. (2006, manuscript submitted to *Climate Dyn.*) who show that summer heat wave events have a distinct multidecadal oscillation. None of the summer patterns shows an exceptional frequency for the extremely warm summer 2003, but cluster 1 and cluster 5 high pressure and warm conditions over western Europe both show above-normal frequencies. As in winter, there is a significant decline of strengthened and/or westward extended Russian high patterns in autumn (Fig. 8, clusters 4 and 7) but an increase of situations with southerly shifted Russian highs (cluster 1). Beside these overall trends interdecadal variability seems less pronounced than in other seasons.

The physically meaningful accordance between temperature and the circulation types allows the setup of descriptive diagnostic models for estimating the seasonal mean temperature anomalies for central Europe. The first attempt is a primitive composite of characteristic temperature anomalies for the clusters weighted by their seasonal cluster frequency. The second approach is a multiple least squares regression predicting the seasonal mean temperature anomalies using seasonal cluster frequencies as predictors. Both estimation methods are implemented using the SANDRA clustering results, on the one hand, and a conventional *k*-means clustering result, on the other hand, for comparison. Overall, there is a tendency for better performance from the models based on SANDRA clustering results, but the differences are small relative to the conventional *k*-means clustering at least with respect to qualitative conclusions drawn from this diagnostic approach.

In contrast to the trend analysis of single clusters, the clustering quality has less impact on classification-based applications if the whole set of clusters is used because of their integrating nature. In case of temperature estimations, for example, the quality of the separation between two cold/warm clusters is less relevant if they are both relatively cold/warm. Therefore, the total estimation differs only gradually between two cluster sets of different quality, even though the SANDRA clustering results show advantages concerning the estimation skill. Therefore, it is concluded that classification based analyses of circulation dynamics integrating the whole of the patterns (like downscaling applications) are less affected by minor clustering quality, while examination of the dynamics of single patterns is vulnerable to the unsteadiness of conventional CA.

Besides the comparison of clustering methods, the results of this diagnostic modeling approach allow conclusions concerning the link between daily circulation patterns and temperature variations in central Europe.

In particular the question whether changes of circulation patterns are contributing to regional temperature change not only on the intra- and interdecadal but also on the multidecadal and centennial time scales is addressed.

The classifications and thus the temperature reconstructions are not adapted for temperature estimation because the clustering procedure has been run independently from any temperature information just on SLP data. Nevertheless, the models show that a notable amount of temperature variation in central Europe is dependent on changes in cluster frequencies. This is not only true for short-term variability but also for long-term changes. The composite model approach is able to simulate the sign of temperature changes in winter and partly in summer and autumn but fails in spring, whereas the multiple regression approach produces a good fit concerning long-term changes for all seasons: the accumulation of relatively cold winters, springs, and autumns since 1850 until 1900 and 1920 is reproduced as well as an (relatively overestimated) initial warm phase until 1880 and a cold spell until 1930 in summer. The most significant discrepancies are recognized in the course of the second half of the twentieth century for all seasons, whereas the recent warming since 1985 can be simulated correctly for all seasons. Except for summer for which the observed warming trend is weak and insignificant, the circulation-based regression estimates indicate a long-term warming of roughly one-third up to one-half of the observed temperature increase, including significant trends for spring and autumn.

This means that changes in frequency of daily circulation patterns have clearly contributed to the long-term overall warming of central Europe in the period between 1850 and 2003. However, an important role of the NAO for this overall trend as suggested, for example, by Hurrell and van Loon (1997) is not indicated, even though an NAO-like pattern contributes to the enhanced warming since 1985. Instead, an increasing westerly pattern representing strong cyclonic activity north of the British Isles and a decrease of meridional patterns are consistent with the overall warming. However, the driving force of changes in daily circulation patterns itself remains unclear. It may be a random process (Stephenson et al. 2000), a reaction to global warming (Palmer 1999; Corti et al. 1999), or a mixture of different factors.

However, it has to be emphasized that the contribution of circulation changes to warming in central Europe as indicated in the present study just refers to relative changes of temperature and not to their absolute extent. To provide a precise estimation of the absolute contribution of changes in circulation patterns to

the observed warming, further advanced statistical models have to be used (including an advanced calibration, e.g., by taking into account the effect of persistence of patterns or operating on a subseasonal time scale) that are beyond the scope of this study. Moreover it has to be emphasized that the findings refer only to regional temperature change. Since changes in circulation patterns mainly cause just a redistribution of given temperature contents in the atmosphere, their effects should be balanced out by opposite temperature changes in other regions (if secondary effects like altered exchange rates of energy between ocean and atmosphere or altering of the degree and character of the cloud cover may be ignored). Therefore, these results may not be used to falsify the concept of an overall long-term, circulation-independent warming mode on the centennial time scale. Such an independent warming mode should be visible in a time-variant link between cluster frequencies and temperature, known as so-called within-type changes as presented, for example, by Beck et al. (2006, manuscript submitted to *Int. J. Climatol.*) for monthly data. Therefore, further studies are necessary to narrow down the observed fraction of warming contributed by circulation changes. The enhanced daily resolution of SLP data in conjunction with the presented advanced classification technique and a more sophisticated calibration of temperatures to the circulation types is expected to offer a more precise look in the future.

APPENDIX

Limitations of the Classifiability and Reproducibility Index for Cluster Analysis

The so-called classifiability index and reproducibility index are used by Michelangeli et al. (1995) to estimate the appropriate number of clusters based on an approach by Chen and Harr (1993) for EOF analysis. The classifiability index is used to test the dependence of several SFA k-means results on their randomly initiated seed partitions. If the results of 50 different SFA k-means runs with varying seed partitions are more similar among themselves for a certain number of clusters than for other numbers it is concluded that this is a

hint to the appropriate number. The intention is to find numbers of clusters where the separation of objects into clusters is very clear and the optimization mostly finds this separation regardless of the starting partition.

This approach, however, assumes that there is a clear separation inherent in the dataset. If this is not the case, as it is for circulation patterns, the result of this test mainly depends on two factors: first, on chance, since conventional k-means CA can be trapped by any local optima (see section 4 and Fig. 2), and, second, on the numbers of clusters itself, since the number and scattering of possible local optima naturally grow with the number of clusters.

The unreliability of conventional k-means CA is also significant regarding the reproducibility index: 100 random subsamples of one-half of the data were used to examine their similarity to a reference partition that was chosen to be the one most similar to all others. According to section 4 it seems questionable why it should be preferable to achieve by chance a number of partitions being relatively similar to a reference partition, which itself is also indirectly determined by chance and might be a local optimum of minor quality. Because of the random nature of SFA k-means results another set of 100 partitions might show a very different scattering. An important factor in this test is again the number of clusters itself, which determines the number of possible local optima. Thus it would be surprising if this test would ever show higher numbers of clusters for any complex dataset.

It is argued that an unsteady method (SFA k means) is used to test the steadiness of clusters in the dataset. Therefore a clear distinction between the unsteadiness of the method and the unsteadiness of clusters arising from the character of the dataset seems improbable when using this method. This is not the case for EOF analysis as presented by Chen and Harr (1993) since EOFs are not affected by such inherent methodological instabilities as is conventional k-means CA.

In contrast to the classifiability index, the reproducibility index can be modified to allow meaningful statements on the cluster structure of the dataset in conjunction with simulated annealing CA since there is some confidence on the robustness of this method. As in Chen and Harr (1993) and Michelangeli et al. (1995) 100 subsamples of half of the dataset are drawn by random and subjected to the SANDRA scheme. Afterward the objects in the other half that were not used for optimization are assigned to their nearest cluster.

The whole-sample SANDRA result is used as the reference partition and is compared with the 100 partitions achieved by the subsample runs. The average similarity [see Eq. (7)] of the 100 subsample cluster

solutions to the whole sample reference partition indicates whether this number of clusters is more reproducible with different datasets than others. By using the SANDRA results it is possible to virtually exclude the unsteadiness of the method for being responsible for differences between the reference partition and the partitions based on the subsamples. Thus, this approach reflects the robustness of the clustering in relation to the character of the dataset rather than that of the method. The result for the winter season is shown as an example in Fig. 10.

In Fig. 10 the reproducibility index generally does not decrease with increasing numbers of clusters as it does in Michelangeli et al. (1995) (except for their cluster solution of $k = 4$ in the Atlantic sector). Because Michelangeli et al. (1995) use correlation coefficients between the cluster centroids for determination of similarity, differences may occur partly by the use of different measures for similarity. Nevertheless, decreasing index values support the former statement that this method applied with conventional k-means clustering mainly reflects the growing number of local optima with increasing numbers of clusters, which are counteracted by using the more stable SANDRA scheme. Surprisingly the modified reproducibility has its minimum for four clusters, which was the number of regimes selected in the North Atlantic–European region by Michelangeli et al. (1995) for 700-hPa geopotential height data of the season that goes from mid-November to the end of March. However, obviously Fig. 10 does not show any clear result for a specific number of clusters. Suggestions for two or three clusters or (as indicated by a second maximum) for eight or nine clusters do not include a convincing single peak as would be expected if there was a separation into distinct clusters.

REFERENCES

Aarts, J., and E. H. Korst, 1989: *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing.* John Wiley and Sons, 272 pp.

Ansell, T. J., and Coauthors, 2006: Daily mean sea level pressure reconstructions for the European–North Atlantic region for the period 1850–2003. *J. Climate,* **19,** 2717–2742.

Bárdossy, A., and H. Caspary, 1990: Detection of climate change in Europe by analyzing European atmospheric circulation patterns from 1881 to 1989. *Theor. Appl. Climatol.,* **42,** 155–167.

——, J. Stehlík, and C. Hans-Joachim, 2002: Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Climate Res.,* **23,** 11–22.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.,* **115,** 1083–1126.

Beck, C., J. Jacobeit, and P. Jones, 2006: 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.

Calinski, T., and J. Harabasz, 1974: A dendrite method for cluster analysis. *Commun. Stat.,* **3,** 1–2.

Cannon, A. J., P. H. Whitfield, and E. R. Lord, 2002: Synoptic map-pattern classification using recursive partitioning and principal component analysis. *Mon. Wea. Rev.,* **130,** 1187–1206.

Chen, J., and P. Harr, 1993: Interpretation of extended empirical orthogonal function (EEOF) analysis. *Mon. Wea. Rev.,* **121,** 2631–2636.

Christiansen, B., 2007: Atmospheric circulation regimes: Can cluster analysis provide the number? *J. Climate,* **20,** 2229–2250.

Corti, S., F. Molteni, and T. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature,* **398,** 799–802.

Della-Marta, P. M., J. Luterbacher, H. von Weissenfluh, E. Xoplaki, M. Brunet, and H. Wanner, 2007: Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Climate Dyn.,* **29,** 251–275.

Dutilleul, P., 1993: Modifying the *t*-test for assessing the correlation between two spatial processes. *Biometrics,* **49,** 305–314.

Folland, C. K., and T. R. Karl, 2001: Observed climate variability and change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 99–181.

Gerstengarbe, F., and P. Werner, 1997: A method to estimate the statistical confidence of cluster separation. *Theor. Appl. Climatol.,* **57,** 103–110.

Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate,* **8,** 897–931.

Goodess, C., and P. Jones, 2002: Links between circulation and changes in the characteristics of Iberian rainfall. *Int. J. Climatol.,* **22,** 1593–1615.

Hannachi, A., and B. Legras, 1995: Simulated annealing and weather regimes classification. *Tellus,* **47A,** 955–973.

Hartigan, J. A., and M. Wong, 1979: Algorithm AS 136: A *K*-means clustering algorithm. *Appl. Stat.,* **28,** 100–108.

Haylock, M., and C. Goodess, 2004: Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. *Int. J. Climatol.,* **24,** 759–776.

Hess, P., and H. Brezowsky, 1977: Katalog der Großwetterlagen Europas 1881–1976, 3e verbesserte und ergänzte Auflage. Berichte des Deutschen Wetterdienstes 113, Offenbach, Germany, 70 pp.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillations: Regional temperatures and precipitation. *Science,* **269,** 676–679.

——, and H. van Loon, 1997: Decadal variations in climate associated with the North Atlantic Oscillation. *Climatic Change,* **36,** 301–326.

Huth, R., 1996: An intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.,* **8,** 893–922.

Jacobeit, J., 1993: Regionale Unterschiede im atmosphärischen Zirkulationsgeschehen bei globalen Klimaveränderungen (Regional differences of the atmospheric circulation under conditions of global climate change). *Die Erde,* **124,** 63–77.

——, C. Beck, and A. Philipp, 1998: Annual to decadal variability in climate in Europe—Objectives and results of the German

contribution to the European climate research project AD-VICE. Würzburger Geographische Manuskripte 43, 163 pp.

——, P. Jönsson, L. Bärring, C. Beck, and M. Ekström, 2001: Zonal indices for Europe 1780–1995 and running correlations with temperature. *Climatic Change,* **48,** 219–241.

——, H. Wanner, J. Luterbacher, C. Beck, A. Philipp, and K. Sturm, 2003: Atmospheric circulation variability in the North-Atlantic-European area since the mid-seventeenth century. *Climate Dyn.,* **20,** 341–352.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate,* **16,** 206–223.

——, M. Hulme, and K. R. Briffa, 1993: A comparison of Lamb circulation types with an objective classification scheme. *Int. J. Climatol.,* **13,** 655–663.

——, T. Jonsson, and D. Wheeler, 1997: Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int. J. Climatol.,* **17,** 1433–1450.

——, and Coauthors, 1999: Monthly mean pressure reconstructions for Europe for the 1780–1995 period. *Int. J. Climatol.,* **19,** 347–364.

Kalkstein, L. S., G. Tan, and J. A. Skindlov, 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Appl. Meteor.,* **26,** 717–730.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kaufmann, L., and P. Rousseeuw, 1990: *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley and Sons, 342 pp.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247–267.

Lamb, H., 1972: *British Isles Weather Types and a Register of Daily Sequence of Circulation Patterns, 1861–1971. Geophysical Memoirs,* Vol. 116, Her Majesty's Stationery Office, 85 pp.

Linderson, M., 2001: Objective classification of atmospheric circulation over southern Scandinavia. *Int. J. Climatol.,* **21,** 155–169.

Lukashin, A., and R. Fuchs, 2001: Analysis of temporal gene expression profiles: Clustering by simulated annealing and determining the optimal number of clusters. *Bioinformatics,* **17,** 405–414.

Lund, I. A., 1963: Map-pattern classification by statistical methods. *J. Appl. Meteor.,* **2,** 56–65.

Luterbacher, J., C. Schmutz, D. Gyalistras, E. Xoplaki, and H. Wanner, 1999: Reconstructions of monthly NAO and EU indices back to AD 1675. *Geophys. Res. Lett.,* **26,** 2745–2748.

——, and Coauthors, 2000: Monthly mean pressure reconstruction for the late maunder minimum period (AD 1675–1715). *Int. J. Climatol.,* **20,** 1049–1066.

——, and Coauthors, 2002: Extending North Atlantic Oscillation reconstructions back to 1500. *Atmos. Sci. Lett.,* **2,** 114–124.

Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.,* **52,** 1237–1256.

Milligan, G., and M. Cooper, 1985: An examination of procedures for determining the number of clusters in a data set. *Psychometrika,* **50,** 159–179.

Moberg, A., and Coauthors, 2006: Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *J. Geophys. Res.,* **111,** D22106, doi:10.1029/2006JD007103.

Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. *J. Climate,* **12,** 575–591.

Rand, W. M., 1971: Objective criteria for the evaluation of clustering methods. *J. Amer. Stat. Assoc.,* **66,** 846–850.

Rapp, J., 2000: Eine erweiterte Definition des Begriffes "Trend" in der Klimadiagnose (An extended definition of trends in climate diagnostics). Klimastatusbericht 1999 des Deutschen Wetterdienstes, 107–110.

Rogers, J. C., S.-H. Wang, and D. H. Bromwich, 2000: On the role of the NAO in the recent northeastern Atlantic arctic warming. *Geophys. Res. Lett.,* **31,** L02202, doi:10.1029/2003GL018857.

Romero, R., G. Sumner, C. Ramis, and A. Genovs, 1999: A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *Int. J. Climatol.,* **19,** 765–785.

Schmutz, C., and H. Wanner, 1998: Low frequency variability of atmospheric circulation over Europe between 1785 and 1994. *Erdkunde,* **52,** 81–94.

Sheridan, S., 2002: The redevelopment of a weather-type classification scheme for North America. *Int. J. Climatol.,* **22,** 51–68.

Slonosky, V. C., and P. Yiou, 2001: The North Atlantic Oscillation and its relationship with near surface temperature. *Geophys. Res. Lett.,* **28,** 807–810.

——, P. D. Jones, and T. D. Davies, 2000: Variability of the surface atmospheric circulation over Europe, 1774–1995. *Int. J. Climatol.,* **20,** 1875–1897.

Stehlík, J., and A. Bárdossy, 2003: Statistical comparison of European circulation patterns and development of a continental scale classification. *Theor. Appl. Climatol.,* **76,** 31–46.

Stephenson, D. B., V. Pavan, and R. Bojariu, 2000: Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.,* **20,** 1–18.

——, A. Hannachi, and A. O'Neill, 2004: On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc.,* **130,** 583–605.

Stott, P., D. Stone, and M. Allen, 2004: Human contribution to the European heat wave of 2003. *Nature,* **432,** 610–614.

Yarnal, B., 1993: *Synoptic Climatology in Environmental Analysis: A Primer.* Belhaven Press, 256 pp.

——, A. Comrie, B. Frakes, and D. Brown, 2001: Developments and prospects in synoptic climatology. *Int. J. Climatol.,* **21,** 1923–1950.