# Quantifying Differences in Circulation Patterns Based on Probabilistic Models: IPCC AR4 Multimodel Comparison for the North Atlantic*

Henning W. Rust and Mathieu Vrac

*LSCE-IPSL, CNRS/CEA/UVSQ, Gif-sur-Yvette, France*

Matthieu Lengaigne and Benjamin Sultan

*LOCEAN, Université Pierre et Marie Curie, Paris, France*

ABSTRACT

The comparison of circulation patterns (CPs) obtained from reanalysis data to those from general circulation model (GCM) simulations is a frequent task for model validation, downscaling of GCM simulations, or other climate change–related studies. Here, the authors suggest a set of measures to quantify the differences between CPs. A combination of clustering using Gaussian mixture models with a set of related difference measures allows for taking cluster size and shape information into account and thus provides more information than the Euclidean distances between cluster centroids. The characteristics of the various distance measures are illustrated with a simple simulated example. Subsequently, a five-component Gaussian mixture to define circulation patterns for the North Atlantic region from reanalysis data and GCM simulations is used. CPs are obtained independently for the NCEP–NCAR reanalysis and the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40), as well as for twentieth-century simulations from 14 GCMs of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) database. After discussing the difference of CPs based on spherical and nonspherical clusters for the reanalysis datasets, the authors give a detailed evaluation of the cluster configuration for two GCMs relative to NCEP–NCAR. Finally, as an illustration, the capability of reproducing the NCEP–NCAR probability density function (pdf) defining the Greenland anticyclone CP is evaluated for all 14 GCMs, considering that the size and shape of the underlying pdfs complement the commonly used Euclidean distance of CPs' mean values.

## 1. Introduction

Certain states of the atmospheric circulation are thought to be more prominent than others, they recur more frequently and/or persist longer (e.g., Michelangeli et al. 1995). These states are usually defined on variables such as sea level pressure (SLP), geopotential heights, and/or wind fields (e.g., Simonnet and Plaut 2001; Moron et al. 2008). A finite sets of these states, called circulation patterns (CPs) or circulation regimes (Stephenson et al. 2004; Philipp et al. 2007), have become attractive for a discrete description of the complex atmospheric system. In particular, in synoptic meteorology, CPs (sometimes called weather types) can be used in forecasting as they can be related to local weather (e.g., Yarnal 1993); or they can be defined on a combination of large-scale circulation indices and meteorological surface conditions. For a discrimination of the terms weather types and CPs we refer to Philipp et al. (2007).

Early classifications of large-scale atmospheric circulation are the Lamb weather types for the British Isles (Lamb 1972), the European Großwetterlagen (Hess and Brezowsky 1977), or the Schüepp classification (Schüepp 1978) for Switzerland. These classifications have been obtained in a subjective manner by examining and manually classifying the synoptic situation. Since then, other, so called objective, weather-typing schemes have been developed, based on automated clustering of circulation

---

*Corresponding author address:* Henning W. Rust, Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL, CNRS/CEA/UVSQ), Orme des Merisiers, Bât. 701, F-91191 Gif-sur-Yvette, France.
E-mail: hrust@lsce.ipsl.fr

variables (e.g., Jones et al. 1993; Bárdossy 1994; Plaut and Simonnet 2001, hereafter referred to as PS01; Jacobeit et al. 2003). These methods consider two states (or instances of observations on a grid) of the atmospheric circulation as similar if they are close according to a certain metric. Days with similar circulation are assigned to the same CP. A CP can then be regarded as a set of elements confined in a volume in the multidimensional state space spanned by the gridded values. The assumption underlying the CP hypothesis is based on the existence of a finite and typically small number of such volumes in state space. In a probabilistic framework, these CPs are local maxima in the state space probability density function (pdf; e.g., Michelangeli et al. 1995; Stephenson et al. 2004).

A wide range of clustering methods is known to define CPs in an objective way. They can be subdivided into two main categories: the first one is based on probabilistic models describing the density of observations in a state space using pdfs, and the second is based on distances of pairs of observations used to partition the state space into multiple regions (clusters). Due to the absence of an explicitly defined model, the latter approaches are sometimes referred to as heuristic or model-free methods (Bock 1996). Some popular methods in this category are as follows: the iterative relocation method (or $k$ means, e.g., MacQueen 1967), based on minimizing the intracluster variance around $k$ centroids (or mean patterns) by exchanging cluster members, a powerful extension using simulated annealing has been suggested as well (e.g., Hannachi and Legras 1995; Philipp et al. 2007); hierarchical agglomerative clustering (HAC; Ward 1963; Casola and Wallace 2007), starting off from one cluster per observation, at each step, the two closest clusters are merged until a desired number of clusters is reached; and self-organized maps (SOMs; Kohonen 1998; Wehrens and Buydens 2007; Leloup et al. 2008), a clustering approach based on artificial neural networks (ANN) and providing additionally a two-dimensional topography relating the resulting cluster centroids. The latter can be very useful in special applications, for example, for the estimation of a probability density function (Brajard et al. 2008).

The other category, the probabilistic models (Bock 1996), provide a different approach to cluster analysis. A mixture of pdfs is used to represent the distribution of elements among the different clusters, for example, Gaussian pdfs in the Gaussian mixture model (GMM) approach (Fraley and Raftery 2002). It directly implements the idea that the probability density of atmospheric states is a multimodal pdf or can be approximated with a superposition of Gaussian pdfs (Branstator and Selten 2009). These models can be considered as a formalization and generalization of some heuristic methods as $k$ means and HAC. With a suitably chosen covariance structure

(Banfield and Raftery 1993), the cluster boundaries are not necessarily limited to spheres in multidimensional space, as is the case for $k$ means. They have been used for atmospheric circulation clustering in various previous works (Haines and Hannachi 1995; Hannachi 1997, 2007; Smyth et al. 1999). GMMs showed better CP results (i.e., producing more consistent CPs across various levels and being more sensitive to day-to-day variations in pattern frequencies) over the eastern United States in comparison to HAC (Vrac et al. 2007a) and also provided useful CPs for precipitation downscaling (Vrac et al. 2007b). The model-based and model-free approaches mentioned here, as well as many other clustering methods, are discussed, for example, in a very general manner by Duda et al. (2001) and with a focus on circulation patterns by Huth (1996). Applications other than atmospheric circulation clustering utilize data mining strategies; such as customer consumption behaviors, image analysis, and Internet usage.

Recently, weather types and CPs have become an important concept for climate change studies and related impact assessments, such as the description of the climatology of severe storms in Virginia (Davis et al. 1993), the investigation of the causes of extreme weather events in Europe (Yiou and Nogaj 2004), or the description of the North Atlantic Oscillation (e.g., Michelangeli et al. 1995). The climate change context also brings along the need for a quantification of differences in CPs; for example, for the validation of general circulation models (GCMs) (e.g., Huth 2000), the downscaling of GCM outputs (e.g., Conway and Jones 1998; Wilby et al. 1998; Fowler et al. 2007; Maraun et al. 2010), or the investigation of teleconnection patterns (e.g., Cassou 2008). The aim of this article is to propose and demonstrate the application of a novel set of measures to quantify differences in CPs. These measures are based on a probabilistic description of the clusters in the state space of atmospheric circulation, as provided by GMMs. This clustering method has the following characteristics: 1) the size and shape of the clusters are explicitly modeled and can directly be taken into account by the quantitative difference measures, 2) clusters are not limited to spheres in the state space but can take ellipsoidal shapes with various sizes and orientations, and 3) the estimation of the uncertainty of classification is straightforward. The proposed measures yield scalar values for CP differences and are thus particularly useful for studies comprising a large set of GCMs where a detailed individual comparison of CPs is not feasible anymore.

These difference measures are used in this study for two purposes: 1) to compare CPs based on spherical and nonspherical clusters, both obtained from reanalysis data, and 2) to compare CPs from GCM simulations to

reanalysis CPs. We base the discussion on five CPs for the North Atlantic region obtained from reanalysis SLP anomalies by PS01. Using PS01 as a reference has mainly practical reasons: 1) PS01 is one of the few references using actually SLP, most other works use geopotential heights, a quantity which is not available for all the GCM considered in this study, and 2) the goal of this paper is to present and exemplify a set of distance measures, and five CPs are convenient to work with. The discussion of an optimal number of CPs for the North Atlantic region is beyond the scope of this paper. To reproduce the results from PS01, we restrict the GMM to five spherical clusters. This restriction is then relaxed and nonspherical clusters will be obtained for the reanalysis data and the GCM simulations. The GMM's covariance structure defining the shape of the clusters is chosen by means of the Bayesian Information Criterion (BIC). However, for the reasons mentioned above, we keep the number of CPs equal to five.

The two sets of reanalysis data and the 14 GCMs' twentieth-century simulations used in this study are described in section 2. The clustering procedure using GMMs is subsequently presented in section 3. Spherical and nonspherical reanalysis CPs are presented and compared on the basis of population histograms in section 4. Subsequently, in section 5, the set of quantitative difference measures for pdfs is introduced and illustrated using a two-dimensional example. Section 6 repeats the comparison of reanalysis patterns now using the pdf-based measure instead of the population histograms. Furthermore, CPs are defined for the GCM simulations, and their configuration in state space is compared to the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis. The last part of this section focuses on the Greenland anticyclone (GA) CP and confronts all GCM simulations to this particular NCEP–NCAR CP.

## 2. Data

The North Atlantic is a well-studied region with respect to weather types or circulation regimes, (e.g., Vautard 1990; Michelangeli et al. 1995; PS01; Hewitson and Crane 2002; Jacobeit et al. 2003; Philipp et al. 2007; Cassou 2008; Casado et al. 2009) to name but a few. Many studies focus on geopotential heights, mostly at 500 hPa (e.g., Vautard 1990; Casado et al. 2009), and others study SLP (e.g., PS01; Hewitson and Crane 2002; Jacobeit et al. 2003; Philipp et al. 2007). Since daily data of geopotential heights are not available in all the GCMs considered, we choose SLP for this comparative study. We use two reanalysis products, the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF)

Re-Analysis project (ERA-40, available online at http://www.ecmwf.int; Uppala et al. 2005) and NCEP–NCAR reanalysis project (available online at http://www.cdc.noaa.gov/data/reanalysis; Kalnay et al. 1996). Additionally, we consider twentieth-century runs (20C3M) from 14 coupled ocean–atmosphere GCMs available via the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) Coupled Model Intercomparison Project phase three (CMIP3) database (Meehl et al. 2007), which offers references for the models at its Web site (available online at http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php). The model list is given in Table 1, including their short names used here and their atmospheric and oceanic resolutions. All datasets are interpolated [two-dimensional (2D) linear interpolation] onto the regular NCEP–NCAR grid with a zonal and meridional resolution of 2.5°. The analysis is carried out in the North Atlantic region defined as $(-60°, 40°E) \times (30°, 70°N)$, and thus comprising $41 \times 13 = 533$ grid points. We furthermore restrict the dataset to the winter months [November–March (NDJFM)] of the years 1975–2000. The SLP anomalies are obtained as differences to a spline-smoothed mean annual cycle for every grid point, calculated by averaging over the 26 yr used.

To avoid an overrepresentation of highly correlated dimensions, it is advantageous to reduce dimensionality using a principal component analysis (PCA; Davis and Kalkstein 1990; Huth 1996). PCA (Preisendorfer 1988; Jolliffe 2002; Hannachi et al. 2007) is applied to one concatenated dataset, comprising the two sets of reanalysis data and the 14 model simulations. This procedure ensures that we have one common basis of principal components (PCs) for all datasets in which the variability of all models is well represented. In the following, we use a projection of the data onto the first 10 PCs, retaining at least 85% of the total variability. The fractions for the individual models do vary around this value with L'Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4; 88%) at the upper end and the Commonwealth Scientific and Industrial Research Organisation Mark version 3.5 (CSIRO Mk3.5; 82%) at the lower end. We thus reduce the dimension of the problem from 533 highly correlated to 10 uncorrelated dimensions. Furthermore, being a linear combination of the gridded values, the transformed variables tend to be closer to a Gaussian distribution than the original values themselves (Stephenson et al. 2004).

## 3. Clustering with Gaussian mixture models

The underlying assumption for GMMs is that the elements of one cluster are distributed according to a

TABLE 1. Reanalyses and IPCC model abbreviations with their original names and their oceanic and atmospheric resolution. The common data period is 1975–2000.

| Abbreviation | Description | Atmospheric resolution | Oceanic resolution |
|---|---|---|---|
| Reanalyses | | | |
| ERA-40 | ERA-40 reanalysis | T159L60 | — |
| NCEP | NCEP–NCAR reanalysis | T62L28 | — |
| IPCC models | | | |
| cccma_cgcm3_1 | Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (CGCM3.1, T47) | T47L31 | $1.8° \times 1.8°$L29 |
| cnrm_cm3 | CNRM CM3.0 | T63L45 | $2° \times 2°$L31 |
| csiro_mk3_0 | CSIRO Mk3.0 | T63L18 | $1.875 \times 0.84°$L31 |
| csiro_mk3_5 | CSIRO Mk3.5 | T63L18 | $1.875 \times 0.84°$L31 |
| gfdl_cm2_0 | Geophysical Fluid Dynamics Laboratory Climate (GFDL) Model version 2.0 (CM2.0) | $2.5° \times 2°$L24 | $1° \times 1°$L50 |
| gfdl_cm2_1 | GFDL CM2.1 | $2.5° \times 2°$L24 | $1° \times 1°$L50 |
| ingv_echam4 | INGV ECHAM4 | T106L19 | $2° \times 2°$L31 |
| inmcm3_0 | INM-CM3.0 | $5° \times 4°$L21 | $2.5° \times 2°$L33 |
| ipsl_cm4 | IPSL CM4 | $2.5° \times 3.75°$L19 | $2° \times 2°$L31 |
| miroc3_2_hires | MIROC3.2(hires) | T106L56 | $0.28° \times 0.1875°$L47 |
| miroc3_2_medres | MIROC3.2(medres) | T42L20 | $1.4° \times 0.5°$L43 |
| miub_echo_g | MIUBECHOG | $3.75° \times 3.75°$L19 | $2.8° \times 2.8°$L20 |
| mpi_echam5 | MPI ECHAM5 | T63L31 | $1.5° \times 1.5°$L40 |
| mri_cgcm2_3_2a | MRI CGCM2.3.2a | T42L30 | $2.5° \times 2°$L23 |

multivariate normal distribution. The distribution of the complete set of observations, that is, elements of all clusters, can then be described using a weighted sum, or mixture, of multivariate Gaussian distributions (Pearson 1894):

$$p(\mathbf{x}) = \sum_{k=1}^{G} \tau_k f(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k), \qquad (1)$$

with $f(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)$ denoting a normal pdf with expectation $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\sigma}_k$. The coefficient $\tau_k$ gives the probability that an arbitrary observation comes from the component $k$ out of $G$ components. The model parameters $\theta = (\tau_k, \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)$ can be estimated from the data. Instead of prescribing the covariance matrices (Haines and Hannachi 1995; Hannachi 1997) or estimating all its free parameters (Smyth et al. 1999; Hannachi and O'Neil 2001; Hannachi 2007), parsimonious parameterizations of $\boldsymbol{\sigma}_k$ can be used to facilitate estimation. A flexible and easily interpretable set of those parameterizations includes spherical clusters of equal or different size ($\boldsymbol{\sigma}_k = a_k \mathbf{1}$), an ellipsoidal cluster with the principal axis parallel to the basis vectors ($\boldsymbol{\sigma}_k$ diagonal), or ellipsoidal clusters of equal or different size and orientation, see Table 2 (Banfield and Raftery 1993; Fraley and Raftery 2007). Given a set of observations $\mathbf{x}_i$, $i = 1, \ldots, N$, the parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\sigma}_k$, and $\tau_k$, as well as the class membership for the $\mathbf{x}_i$ can be estimated via a

maximum-likelihood technique using an expectation–maximization (EM) algorithm (Dempster et al. 1977). Observations $\mathbf{x}_i$ are associated to the mixture component $k$ with the largest associated probability

$$p_k(\mathbf{x}_i) = \tau_k f(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\sigma}_k) \bigg/ \sum_{j=1}^{G} \tau_j f(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\sigma}_j). \quad (2)$$

This allows a quantification of the uncertainty of classification of $\mathbf{x}_i$ via the sum of probabilities for being associated to the other clusters:

$$\overline{p}_k(\mathbf{x}_i) = \sum_{j \neq k} p_j(\mathbf{x}_i) = 1 - p_k(\mathbf{x}_i). \qquad (3)$$

If all components are well separated, the probability [Eq. (2)] for all elements and their associated clusters will be close to one and thus the uncertainty [Eq. (3)] close to zero. In less optimal situations with overlapping clusters, the probability of association [Eq. (2)] will be small but $p_j(\mathbf{x}_i) > 1/G$ will always hold. Thus, for the uncertainty in Eq. (3), it follows that $\overline{p}_j(\mathbf{x}_i) < (G - 1)/G$.

In general, mixture models are not limited to Gaussian distributions. In the univariate case, mixtures of arbitrary distributions can yield useful models, for example, Gamma and generalized Pareto (GP) distributions for precipitation modeling (Vrac and Naveau 2007) or Gaussian and GP distributions for finance data (Carreau

TABLE 2. Possible parameterizations of covariance matrices for the Gaussian mixture models (Banfield and Raftery 1993; Fraley and Raftery 2007).

| Abbreviation | Parameterization |
|---|---|
| EII | Spherical, equal volume |
| VII | Spherical, varying volume |
| EEI | Diagonal, equal volume, and shape |
| VEI | Diagonal, equal shape |
| EVI | Diagonal, equal volume |
| VVI | Diagonal, varying volume, and shape |
| EEE | Ellipsoidal, equal covariance matrix |
| EEV | Ellipsoidal, equal volume, and shape |
| VEV | Ellipsoidal, equal shape |
| VVV | Ellipsoidal, unconstrained |

and Bengio 2009). For clustering in a multidimensional space, parsimoniously parameterized forms of the pdfs are convenient, and thus in practical applications multivariate normal distributions are frequently preferred.

The number of components $G$ as well as the type of parameterization of the covariance matrices can be inferred using BIC (Schwarz 1978):

$$\text{BIC} = -2\,\log L(\theta|\mathbf{x}_1 \cdots \mathbf{x}_N) + m\,\log(N), \qquad (4)$$

with $L(\theta)$ being the likelihood of the Gaussian mixture model with parameters $\theta$ given the observations $\mathbf{x}_i$, $m$ being the number of parameters [depending on the number of components $G$ and on the complexity of the parameterization of $\boldsymbol{\sigma}_k$, see (Fraley and Raftery 2007)], and $N$ being the size of the dataset.

This offers a consistent and transparent way of determining the number of components, as no external measure has to be considered, as is the case, for example, in Philipp et al. (2007). As mentioned earlier, the goal of this study is to introduce and exemplify the pdf-based difference measures, and to not loose focus, we restrict ourselves to five clusters as in the SLP-based study of PS01. The BIC is used, however, to choose one of the 10 different parameterizations of the covariance matrix (Table 2). Other methods of selecting the number of components include, for example, cross validation (Smyth et al. 1999; Hannachi and O'Neil 2001). Recent results indicate, however, that the number of clusters is mostly a fragile quantity and depends not only on the clustering algorithm but also on the time span of the underlying data (Christiansen 2006).

## 4. North Atlantic circulation patterns

In a first step, we aim at reproducing the five CPs from PS01 defined with $k$ means on daily SLP anomalies in the North Atlantic region. We adopt their naming convention and denote CPs as Atlantic Ridge (AR), Blocking (BL), GA, Western Blocking (WBL), and Zonal (ZO). To be compatible with their $k$-means approach, we restrict the GMM to five spherical clusters. In a second step, we allow for nonspherical clusters and use the covariance structure that yields the smallest BIC value. Population histograms are a first attempt to discuss the resulting differences to the spherical clusters. In sections 6b and 6c, the NCEP–NCAR nonspherical CPs will serve as a frame of reference for evaluating CPs from GCM simulations.

### a. Reanalysis circulation patterns with spherical and nonspherical clusters

Mean values (centroid or composite patterns) of the five GMM spherical components obtained for NCEP–NCAR reanalysis are shown in the top row of Fig. 1. Color shading denotes SLP anomalies and absolute values are given as contour lines. The centroids compare well to the corresponding result of PS01, depicted in their Fig. 4. Visual association yields from left to right: AR, BL, GA, WBL, and ZO.

Using ERA-40, similar CPs can be obtained based on spherical clusters; the centroids are shown in the second row of Fig. 1. Three centroids can be visually well identified with AR (first column), BL (second column), and ZO (last column) from NCEP–NCAR or PS01 CPs. One of the two remaining patterns can be interpreted as GA with the cyclone shifted northward (third column). The centroid in the fourth column shows a strong anticyclone east of Greenland, similar to GA, but with the associated cyclone being too weak and slightly shifted westward. If to be interpreted within the frame set by NCEP–NCAR, it can be compared to WBL with the anticyclone shifted northward (as previously discussed for GA) and the associated cyclone westward. However, it remains as the pattern that is the most difficult to associate to any of the NCEP–NCAR centroids.

Next, we remove the constraint leading to spherical clusters and use the BIC to choose among the 10 parameterizations for the covariance matrices. For both, NCEP–NCAR and ERA-40, ellipsoidal clusters with equal size and orientation (EEE, in Table 2) yield the lowest BIC (NCEP–NCAR: 263 226, ERA40: 263 916) and a significant improvement on the spherical cluster variant (VII, in Table 2; NCEP–NCAR: 267 762, ERA40: 268 664). It is worth noting that the number of parameters for the five covariance matrices increases from 5 to 55 in this case. The resulting centroids are shown in Fig. 1 in the third and fourth row, respectively. For the EEE model type, the covariance matrices are identical for all clusters and describe an ellipsoid in 10-dimensional space
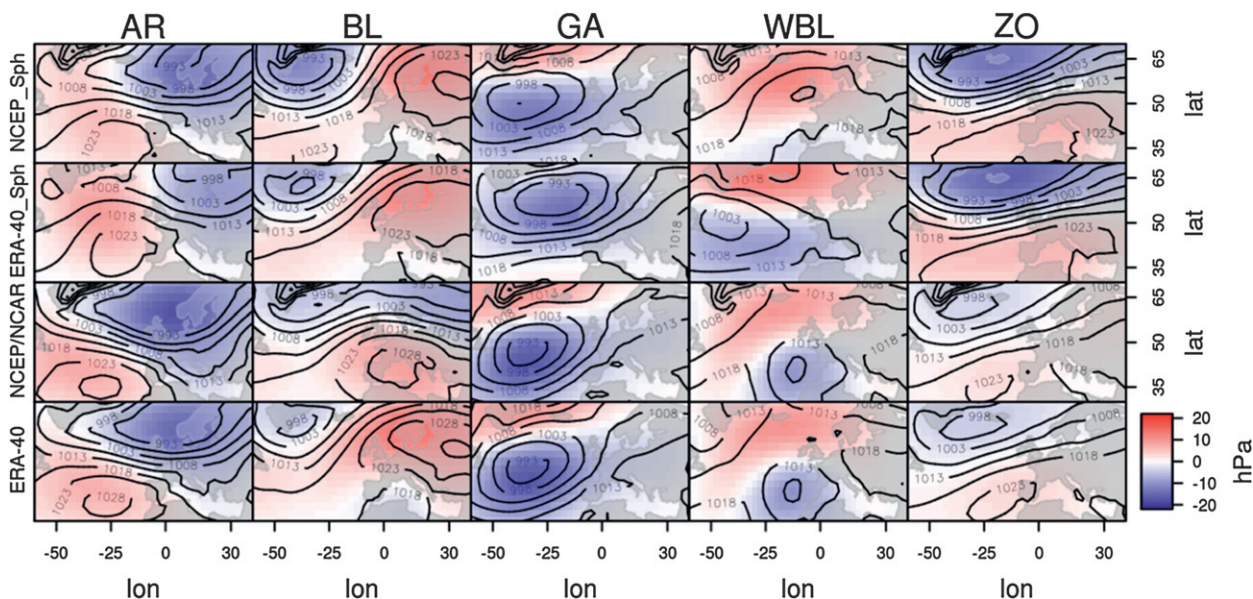
FIG. 1. Mean values (centroids) of CPs obtained from clustering NCEP–NCAR and ERA-40 SLP anomalies using (first and second row) spherical clusters and (third and fourth row) nonspherical clusters. The CPs are visually associated to the result from PS01 and thus named from left to right AR, BL, GA, WBL, and ZO. SLP anomalies are color coded and absolute values are given as contour lines, both in hPa.

(10 PCs). Its longest axis is about 8 times longer than its smallest. The axes are not parallel to the coordinate system (as for diagonal models) but are linear combinations of the 10 PCs. We thus refrain from a more detailed discussion of the covariance structure here.

The first centroids (Fig. 1, third and fourth row, first column) of NCEP–NCAR and ERA-40 show a pattern very similar to the AR with a large cyclonic structure extending farther west and south, pushing the anticyclone southward. While the second centroid (second column) for ERA-40 is almost identical to its spherical counterpart, the NCEP–NCAR centroid shows a quite different cyclonic structure extending westward, pushing the blocking anticyclone to the south, leading to a more ZO-like pattern. In the third column, the centroids resulting from nonspherical clusters are both similar to the corresponding NCEP–NCAR spherical GA. An interesting observation can be made in the fourth column: the patterns from the nonspherical clusters are different from their spherical counterparts in both cases. They do, however, resemble the PS01 WBL centroid more than the spherical NCEP–NCAR result does. The two are furthermore almost identical. The last column shows the patterns that are similar to ZO. For NCEP–NCAR, as well as for ERA-40, the contrast of the associated south–north SLP anomaly gradient is not as expressed as for the spherical patterns in the first two rows.

For spherical, as well as for nonspherical clusters, one could expect results being close for the two reanalysis datasets. There are, however, differences in resolution and parameterization of the reanalyses products' atmospheric models, as well as in their data assimilation schemes (e.g., Dell'Aquila et al. 2005). ERA-40 and NCEP–NCAR are found to differ the most where their observational data basis is sparse (Sterl 2004). The resulting disparities in state space are strong enough to yield different cluster mean patterns for the two datasets. The susceptibility of the clustering result to these state space disparities is supposedly increased by 1) clusters that are not well separated in space and 2) a suboptimal cluster number.

For nonspherical clusters, visual differences in the mean patterns are in general smaller and affect mainly BL. This could be an indication that those clusters are more robust against the differences in state space configuration.

It is interesting to observe that the population of the individual clusters, that is, their number of elements, is very different for spherical and nonspherical clusters. Figure 2 shows the population per CP for the two reanalysis datasets for spherical and nonspherical clusters. In the spherical case, the population is roughly evenly distributed among the clusters (Fig. 2, first and second panel). For nonspherical clusters, more than 60% of the total population (NCEP–NCAR: 70%, ERA-40: 61%) are associated to ZO. Two implications of this uneven distribution are as follows: 1) ZO contains more elements than in the spherical case and thus the centroid is calculated across a larger variety of elements leading to the less-pronounced cyclone/anticyclone contrast in Fig. 1 (third and fourth row, fifths column), and 2) as EEE
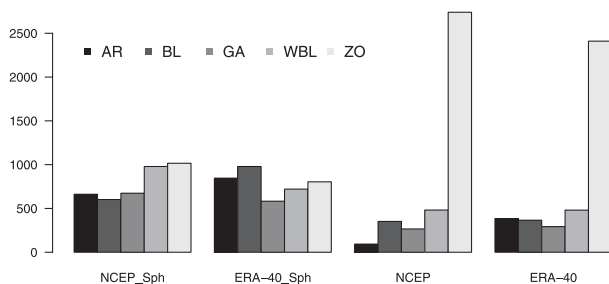
FIG. 2. (first and second grouping) Population of spherical and (third and fourth grouping) nonspherical patterns for the reanalysis datasets.

allows only for one covariance matrix common for all clusters the dominant ZO population is likely to determine its structure. This might lead to suboptimal model pdfs for low-populated clusters—for example, in the sense that state space regions with a considerable probability density according to the model pdf are basically empty because of a low population of that cluster in that region, which is a problem that we get back to in section 6a. It is evident from Fig. 2 that the nonspherical CPs are not constituted with the same elements as in the spherical case, they must be mixtures of multiple spherical CPs. Otherwise, such a difference in CP population would not be possible.

### b. Comparing spherical and nonspherical clusters

An intuitive and straightforward way to study the relationship between nonspherical and spherical clusters is a histogram with each nonspherical CP's population broken down into the contributions from the spherical clusters, as shown in Fig. 3. This representation allows for the evaluation of the composition of the nonspherical clusters in terms of contribution from the spherical clusters. The nonspherical AR consists mainly of the same elements also constituting AR for NCEP–NCAR; ERA-40 has also contributions from spherical GA and ZO. NCEP–NCAR's nonspherical BL has a strong component from spherical ZO, which is also visible in its centroid pattern (Fig. 1, third row, first column). Furthermore, spherical WBL and AR contribute more to nonspherical BL than the spherical BL. This is different for ERA-40. Here, the spherical BL makes up the dominant contribution to nonspherical BL, visible in very similar centroid patterns (Fig. 1). Nonspherical GA for NCEP–NCAR is basically constituted by its spherical counterpart. For ERA-40, this CP has a stronger contribution from spherical WBL that is consistent with the corresponding centroid patterns in Fig. 1. Nonspherical WBL is again dominated by its spherical counterpart in both cases with major contributions from spherical GA for NCEP–NCAR and
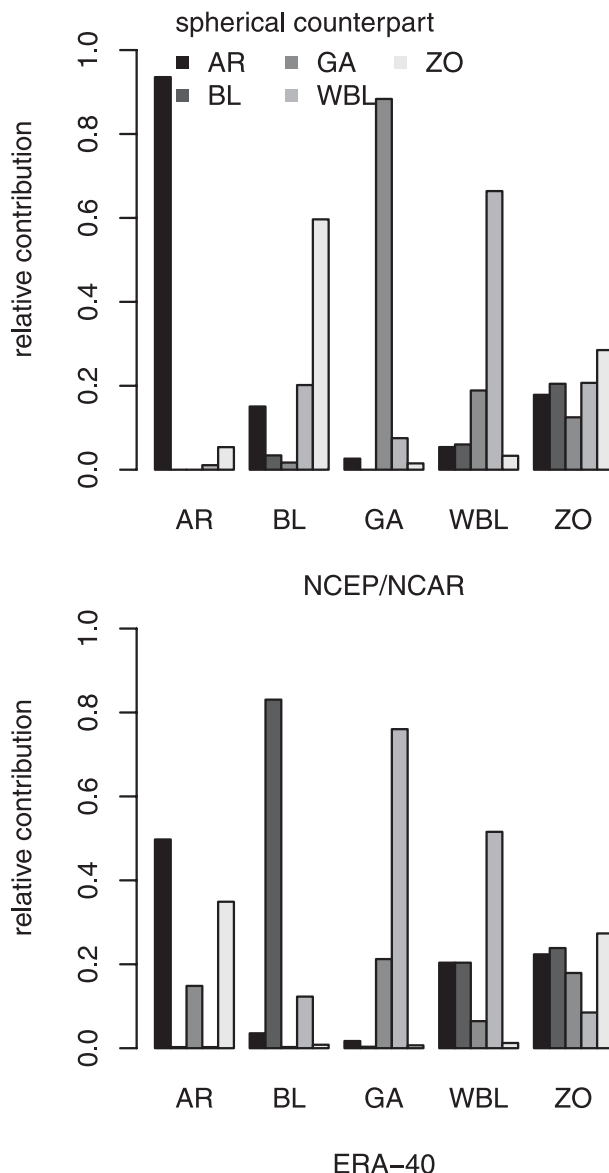


FIG. 3. Histogram for the nonspherical CPs showing their population by contributions from the spherical clusters for (top) NCEP–NCAR and (bottom) ERA-40. The contribution is measured relatively to the nonspherical CPs' total population.

spherical BL and AR for ERA-40. For ZO, we find for both reanalysis datasets a broad mixture of all five spherical CPs, this explains why the SLP anomaly contrast between the cyclonic and the anticyclonic structure is low.

In a more general context, this type of analysis can be regarded as a discrete estimation of overlap or similarity of the spherical and nonspherical clusters. If those two variants share a lot of elements, as, for example, the NCEP–NCAR nonspherical and spherical AR, their overlap in terms of pdf is large. On the other hand, the

NCEP–NCAR nonspherical BL shares almost no elements with its spherical counterpart and has thus, in terms of the associated pdfs, a small overlap or a large distance. In case of Gaussian mixture model clustering, where clusters are described with a pdf, these heuristic estimates of similarity and distance can be replaced with more sophisticated measures presented in the following section.

## 5. Quantifying differences in circulation patterns

So far, visual differences in the CPs' mean values have been discussed, as well as the differences quantified by means of counting elements of two classification approaches. In the following, we make use of a probabilistic description of CPs by means of pdfs. CPs from different datasets can now be compared using difference or similarity measures for pdfs accounting for more than the Euclidean distance between CPs' mean values. A set of those difference measures is introduced and exemplified with a simulated example. These measures are then used to quantify differences between CPs from NCEP–NCAR reanalysis and IPCC model simulations. The CPs are defined by clusters represented by the pdfs in the GMMs.

### a. Distance measures based on probabilistic models

Apart from a visual comparison of mean values (centroid patterns), a popular quantitative difference measure is the Euclidean distance in the multidimensional space

$$d_{\mathrm{Eucl}}(P, Q) = \sqrt{(\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q)^t (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q)}, \quad (5)$$

where $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_Q$ denote the mean values of CPs $P$ and $Q$. This measure can be used to quantify a difference, however, it includes only first-moment information, disregarding the higher moments describing size, shape, or orientation of the whole clusters (i.e., the CP). Strictly speaking, the Euclidean distance [Eq. (5)] does not define a distance for pdfs. This is the case for most of the measures introduced in the following. Nevertheless, we will use the terms distance and difference measure synonymously in the following to refer to a quantification of dissimilarities or differences between CPs or pdfs.

As a first generalization, we introduce the Mahalanobis distance (Mahalanobis 1936), designed for normal distributions with identical covariance matrices. It thus includes second-order information. The Mahalanobis distance is given by

$$d_{\mathrm{Maha}}(P, Q) = (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q)^t \boldsymbol{\sigma}_P^{-1} (\boldsymbol{\mu}_P - \boldsymbol{\mu}_Q), \quad (6)$$

with $\boldsymbol{\sigma}_P$ denoting the covariance matrix associated with the reference CP $P$. It reduces to the square of the Euclidean distance for $\boldsymbol{\sigma}_P = \mathbf{1}$, compare Eq. (5). For an ellipsoidal reference CP (i.e., the covariance matrix defines an ellipsoid), the Mahalanobis distance gives lower weights to distances along a major axis of the ellipsoid than to those along a minor axis. Figs. 4a and 4b exemplify this property. Equation (6) is symmetric and fulfils the criteria of a distance only for identical covariance matrices ($\boldsymbol{\sigma}_P = \boldsymbol{\sigma}_Q$). Otherwise it depends on the choice of the reference pdf. Equation (6) can be straightforwardly symmetrized as $d_{\mathrm{Maha,sym}}(P, Q) = d_{\mathrm{Maha}}(P, Q) + d_{\mathrm{Maha}}(Q, P)$.

Another popular approach for quantifying differences in pdfs is the Kullback–Leibler divergence [KL, also called information discrimination (Kullback 1987), relative entropy, or $I$ coefficient] defined as

$$d_{\mathrm{KL}}(P, Q) = I(P|Q) = \int_{\mathbb{R}} \log\left[\frac{q(x)}{p(x)}\right] q(x)\, dx. \quad (7)$$

It measures the difference between a reference pdf $p(x)$ (the truth) defining CP $P$ and another pdf $q(x)$ (the model) defining CP $Q$. The KL is a noncommutative that is a nonsymmetric measure and, as such, is not a distance in the mathematical sense. It is related to entropy measures from information theory, for example, mutual information and Shanon entropy (Kullback and Leibler 1951). For two normal distributions with identical covariance matrices, the KL reduces to $(1/2)d_{\mathrm{Maha}}$, compare Eq. (A1).

A symmetrized version of the KL, also called $J$ coefficient, can be formulated as

$$d_{\mathrm{J}}(P, Q) = J(P, Q) = I(P|Q) + I(Q|P). \quad (8)$$

If the pdfs defining CPs $P$ and $Q$ are identical, the KL and the $J$ coefficient yield zero. Both increase with diverging $P$ and $Q$. In some special cases, for example, two normal distributions with identical covariance matrix and appropriate symmetries as in Figs. 4b and 4d, we have $I(P|Q) = I(Q|P)$ and therefore $J(P, Q) = 2I(P, Q)$.

A conceptually different approach to quantify differences or similarities in pdfs is based on the Hellinger coefficient (Hellinger 1909)

$$d_{\mathrm{Hell}}^{(s)}(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\, dx, \quad (9)$$

with a parameter $0 < s < 1$. Choosing $s = \frac{1}{2}$ yields a symmetric measure with values between zero ($p$ and $q$ have disjoint supports) and one ($p$ and $q$ are identical). Roughly speaking, it can be thought of as a measure of overlap of the two distributions. The definition given in
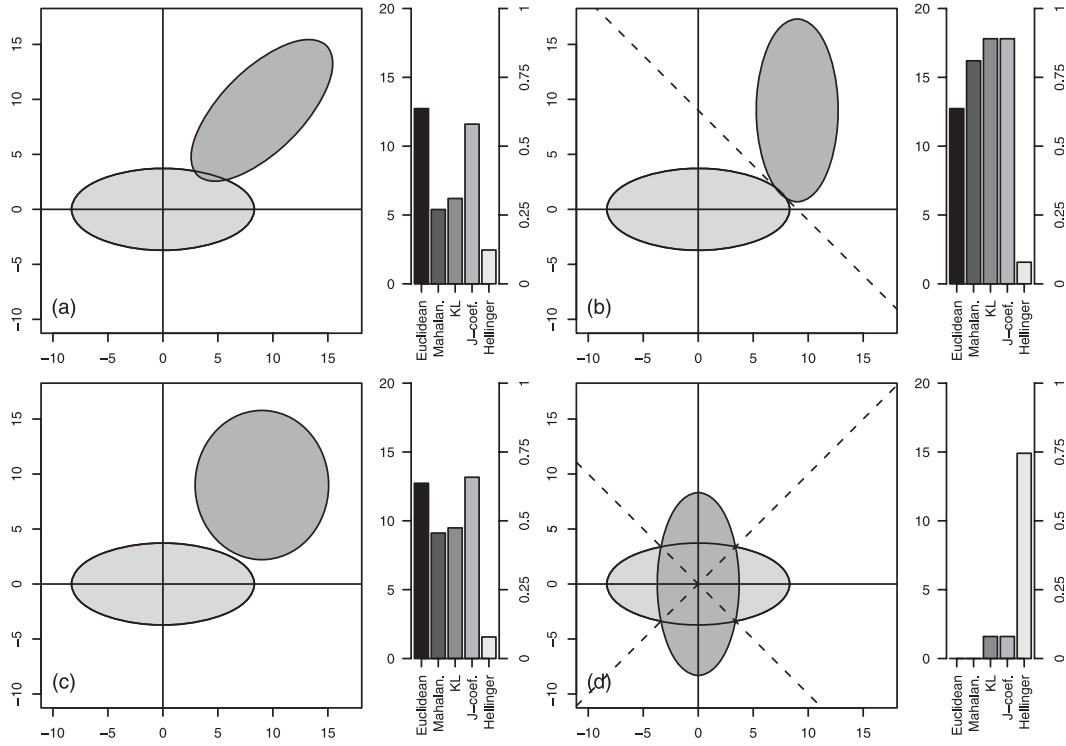
FIG. 4. Exemplifying the Euclidean and Mahalanobis distance, the Kullback–Leibler divergence, the $J$ coefficient, and the Hellinger similarity coefficients with two-dimensional normal pdfs. The values of the Mahalanobis distance and $J$ coefficient are divided by 2 for easier comparison with the Kullback–Leibler divergence. The reference pdf $p$ for the Mahalanobis distance and the KL divergence is the pdf with varying position and shape (dark gray), see Eq. (7). The bars on the right side of each panel show the values of the four measures. Recall that the Hellinger coefficient is a similarity measure and reacts contrariwise to changes in the pdfs. (a),(c) Two different asymmetric situations ($d_{KL} \neq d_J$) with pdf $p$ varying in shape. (b),(d) Two different symmetric situations ($d_{KL} = d_J$) with pdf $p$ changing the location. The dashed lines give the symmetry axes.

Eq. (9) is in fact a measure of similarity. In some cases, it is useful to use the complement, which defines a measure of difference consistent with the other measures.

Other than the Euclidean and Mahalanobis distances, the integral equations for the KL [Eq. (7)] and the Hellinger coefficient [Eq. (9)] are in general difficult to compute. In the case of multivariate normal distributions, closed form solutions exist (Bock and Diday 2000) and are given in the appendix. Table 3 gives a summary of the main features of these different measures.

### b. Exemplifying the distance measures

We exemplify the measures introduced above using two 2D Gaussian pdfs: one static pdf $q$, centered at the origin, and another pdf $p$ with varying location (thus a varying mean) and with different orientation and shape given by the covariance matrix $\boldsymbol{\sigma}_p$. Figure 4 shows the ellipses containing 90% of the mass of the distributions. The four panels depict four different situations to point out various characteristics of the difference measures introduced above. The bars at the side give the

corresponding values for the Euclidean and Mahalanobis distance, the KL divergence, and the $J$ and Hellinger coefficients. The Mahalanobis distance [Eq. (6)] is divided by two; thus, for identical covariance matrices, we obtain the same value as for KL. Similarly, we divide the $J$ coefficient by two; under certain conditions of symmetry, it thus equals the KL, compare Figs. 4b and 4d. Note that the comparison of the absolute values between various distance measures is in general not meaningful. However, keeping the above-mentioned relations into account, information can be gained from their intercomparison in particular situations.

In Figs. 4a–c, the centers of the two distributions do not change, but the orientation and shape of pdf $p$ (dark gray) does change. The Euclidean distance thus remains the same, unlike the Mahalanobis distance, which takes the shape of the covariance matrix into account. It increases from panel (a) to panel (b) because the line connecting the two centers does not run any more along the major axis of the covariance matrix of $p$ but along a direction of smaller spatial extend. In panel (b), the

TABLE 3. The various distance measures suggested for pdf-based clusters and their main characteristics.

| Measure | Notation | Range | Symmetric | Characteristics |
|---|---|---|---|---|
| Euclidean | $d_{\mathrm{Eucl}}$ | $[0, \infty)$ | Yes | Distance of means only |
| Mahalanobis | $d_{\mathrm{Maha}}$ | $[0, \infty)$ | No | Difference in means with metric depending on one covariance matrix |
| KL | $d_{\mathrm{KL}}$ | $[0, \infty)$ | No | Takes both covariance matrices into account, reduces to Mahalanobis for $\sigma_p = \sigma_q$ |
| $J$ coefficient | $d_J$ | $[0, \infty)$ | Yes | Symmetrized KL, $2d_J = d_{\mathrm{KL}}$ for certain symmetric conditions |
| Hellinger ($s = 0.5$) | $d_{\mathrm{Hell}}$ | $[0, 1]$ | Yes | Similarity measures, "overlap" |

Kullback–Leibler and $J$ coefficient (divided by two) are the same because of the symmetry along the dashed line. In panel (c), a change in the shape of pdf $p$ breaks this symmetry, now KL and $J$ coefficient yield different values. Panel (d) shows two concentric distributions and, thus, the value of the Euclidean and Mahalanobis distances are zero. The covariance matrices, however, have different orientation and, therefore, the KL and $J$ coefficient yield nonzero values, and the Hellinger coefficient is smaller than 1, indicating differences in the pdfs. Here, the values for KL and $J$ coefficient are equal—again, because of symmetry—and small compared to panel (a) and (b) because of the identical mean values. The Hellinger coefficient has strongly increased compared to the previous three panels, indicating that important regions of the pdfs overlap.

This example demonstrates that the Mahalanobis distance, KL, $J$ coefficient, and Hellinger coefficient yield information about the relative position, shape, and orientation of the pdfs that is complementary to the Euclidean distance. An R package called gaussDiff, allowing for the calculation of the different measures, has been developed for this study.[1] It is freely available on the Comprehensive R Archive Network (CRAN) Web page (available online at http://cran.r-project.org/).

## 6. Quantitative comparison of CPs in the North Atlantic region

We first get back to a comparison of spherical and nonspherical CPs from the two reanalysis datasets and augment the discussion in section 4b, which was solely based on population histograms. Now, we obtain CPs for the 14 GCM simulations using GMM with nonspherical clusters. For two selected GCMs, we evaluate the cluster configuration in state space in detail by comparison to NCEP–NCAR reanalysis. Finally, a difference to NCEP–NCAR GA is computed and compared for CPs from all 14 GCMs.

### a. Comparing spherical and nonspherical clusters based on pdfs

Instead of counting elements of spherical clusters contributing to nonspherical clusters, as in section 4b, we use the Hellinger measure [Eq. (9)] to assess cluster similarities. Figure 5 shows the similarity coefficients of the nonspherical clusters (abscissa) with their spherical counterparts (gray shadings). Figure 5 can be qualitatively compared with Fig. 3, which was based on evaluating counts of cluster memberships. Although in both cases the ordinate spans the range from zero to one, the absolute values of the two figures are not to be directly confronted. Their relative magnitudes give comparable information about the cluster configuration.

A Hellinger coefficient of 0.6 (Fig. 5, top) indicates that the pdf of NCEP–NCAR's nonspherical AR shares a large fraction of mass with its spherical counterpart, being in line with the large relative contribution of spherical AR to nonspherical AR shown in Fig. 3 (top). Hellinger coefficients around 0.2–0.3 can be observed for the similarity of the other spherical pdfs with nonspherical AR, while the bar plot in Fig. 3 shows almost no relative contribution of the other spherical CPs to nonspherical AR. There is thus a small overlap of the pdfs, but no or only a very small number of elements are present in these areas of overlap because of a small overall cluster population. The count-based measure does not therefore show a relative contribution. The situation is more difficult for BL. The count-based measure in Fig. 3 shows almost no contribution of spherical BL to its nonspherical counterpart, instead a large contribution of ZO. In fact, the mean pattern (centroid) in Fig. 1 (third row, second column) also exhibits strong zonal characteristics. On the other hand, the similarity coefficients in Fig. 5 (top) indicate that spherical BL and nonspherical BL share a large fraction of their mass. Two facts can help to explain this seemingly contradictory result: 1) the pdf of the spherical BL has a small variance (the smallest of all five CPs), and its mass is thus concentrated in a small region of the state space; and 2) as mentioned earlier, parts of the nonspherical cluster must
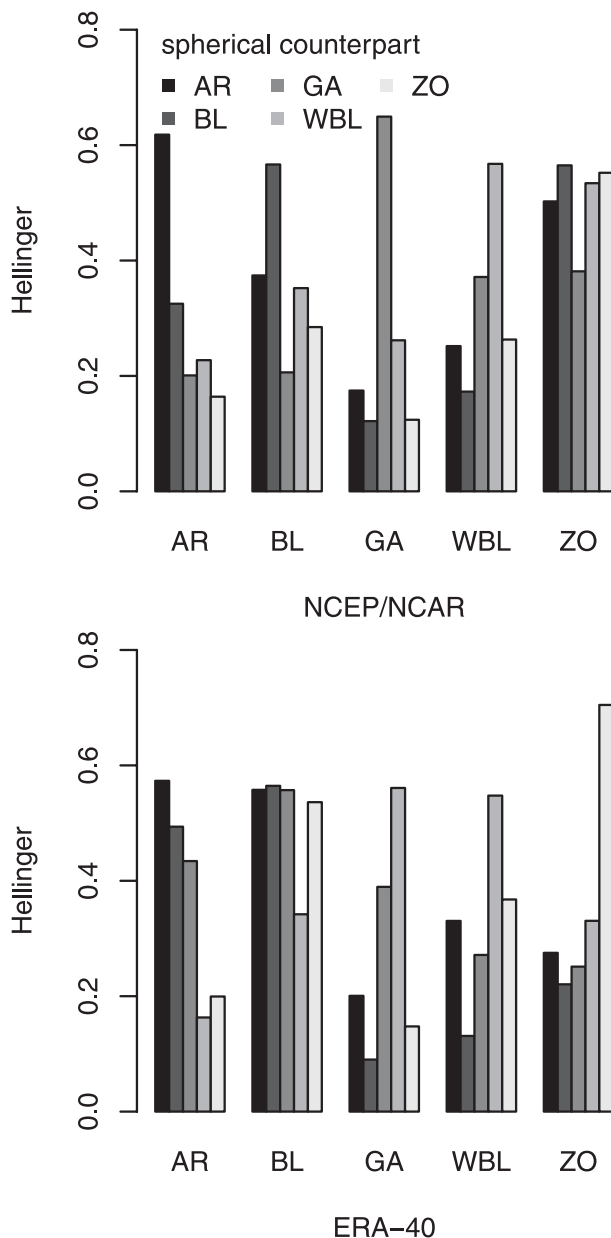
---

[1] R is a language and environment for statistical computing and graphics (R Development Core Team 2004).

FIG. 5. Comparison of spherical and nonspherical clusters for (top) NCEP–NCAR and (bottom) ERA-40 using the Hellinger similarity measure.

TABLE 4. Covariance structure with minimum BIC for the 14 GCMs using five clusters. The abbreviation for the covariance structures are given in Table 2.

| GCM | Covariance structure |
| --- | --- |
| CCCMA CGCM3.1, T47 | VEI |
| CNRM CM3.0 | EEE |
| CSIRO Mk3.0 | EEE |
| CSIRO Mk3.5 | VEI |
| GFDL CM2.0 | EEE |
| GFDL CM2.1 | EEE |
| INGV ECHAM4 | VEI |
| INM-CM3.0 | EEE |
| IPSL CM4 | VEV |
| MIROC3.2(hires) | EEI |
| MIROC3.2(medres) | EVI |
| MIUBECHOG | VEI |
| MPI ECHAM5 | EEV |
| MRI CGCM2.3.2a | EEE |

striking for AR and BL and likely to be due to sparse population. For GA, the count-based measure (Fig. 3, bottom) and the set of similarity coefficients (Fig. 5, bottom) compare well in their relative magnitudes. The nonspherical WBL shows a low relative contribution of the spherical ZO in the count-based measure, while the corresponding similarity coefficient indicates a relatively large fraction of mass shared by the two pdfs. The pdf of nonspherical WBL thus extends toward the spherical ZO, but this area is sparsely populated. For the highly populated nonspherical ZO, the count-based measure and the Hellinger similarity coefficient correspond well in relative magnitude.

This comparison of a count-based and a pdf-based measure clearly shows the limitations of discussing population (or probability) densities in a high dimensional space when it is sparsely populated.

### b. Comparison of GCM CPs with NCEP

In a similar way, we obtain CPs from the GCMs using five nonspherical clusters with the covariance structure chosen again by the BIC (Table 4). Six GCMs share the ellipsoidal clusters of EEE with the reanalysis data. Six others show a diagonal parameterization, that is, ellipsoidal clusters with axes parallel to the coordinate system (PCs) with either equal volume and shape (EEI), variable volume but equal shape (VEI), or equal volume but variable shape (EVI). In case the principal axes of the ellipsoidal clusters are not parallel to the coordinate system, more parameters are needed for their parameterization. This is the case for the Max Planck Institute (MPI) ECHAM5 (EEV), whose clusters are best described with ellipsoids of equal size and orientation but different shape. An additional variation of cluster size is required for the IPSL CM4 (VEV).

be sparsely populated because of a small cluster population but cover a significant part of the small spherical BL pdf's mass. This leads to a large Hellinger coefficient despite low counts of common elements. For GA, WBL, and particularly ZO, correspondence between Fig. 3 (top) and Fig. 5 (top) is more explicit (not in absolute numbers but in relative magnitudes).

Similarly for ERA-40, the Hellinger coefficients in Fig. 5 (bottom) indicate a common pdf mass where the count-based measure is small or even zero. This is again
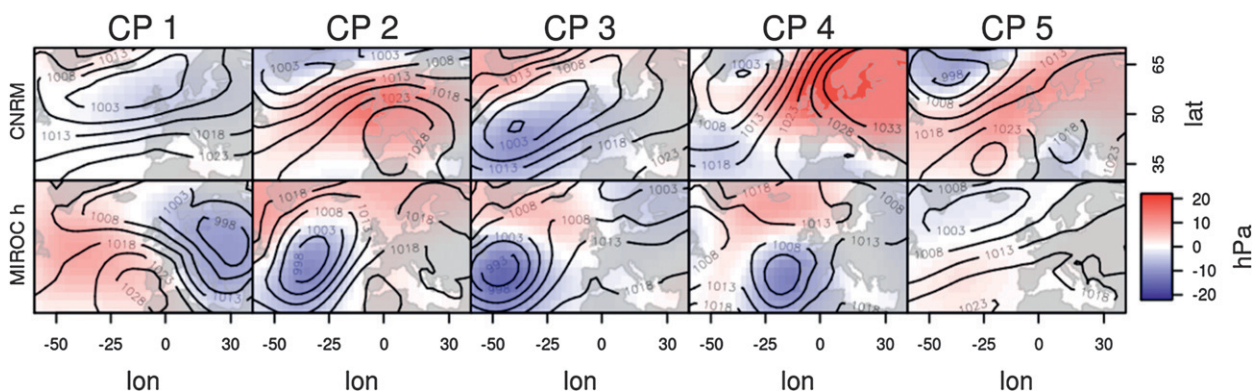
FIG. 6. Mean values (centroids) of CPs obtained from clustering (top) CNRM CM3.0 and (bottom) MIROC3.2(hires) simulations' SLP anomalies using nonspherical clusters. SLP anomalies are color coded and absolute values are given as contour lines, both in hPa. Anomaly values above 22 hPa have been cut off for reasons of visibility. This does occur only in CP4 of CNRM CM3.0 (first row, fourth column).

The GCM mean patterns for the five CPs are shown in Fig. 6 for the Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3 (CNRM CM3.0; top) and the Model for Interdisciplinary Research on Climate 3.2, high-resolution version [MIROC3.2(hires); (bottom)] and in Fig. S1 (available as supplemental material at the Journals Online Web site: http://dx.doi.org/10.1175/2010JCLI3432.s1) for the other 12 GCMs. A priori, the resulting clusters are not related to the NCEP–NCAR (or PS01) CPs and are thus enumerated as CP1 to CP5. Visual inspection shows, however, resemblance to NCEP–NCAR CPs in many cases.

We set a frame of reference by studying the internal configuration of the NCEP–NCAR CPs; Fig. 7 gives the corresponding Hellinger similarity coefficient for the CPs with themselves. Both AR and BL have a strong similarity coefficient with ZO. GA and WBL are both rather different from the other three CPs and GA from WBL. These relationships have to be kept in mind when using these five CPs in the following as a frame of reference. In particular, the Hellinger coefficient between two different CPs does not exceed 0.6; we thus consider larger values as an indication of large overlap in the present setting.

Figure 8 shows the Hellinger similarity as a bar plot for CNRM CM3.0 and MIROC3.2(hires) CPs arranged in the same order as in Fig. 6.[2]

### 1) CNRM CM3.0

The bar plot for CP1 shows qualitatively much the same characteristics as the corresponding plot for NCEP–NCAR ZO in Fig. 7. The similarity coefficient of

CP1 with ZO is not equal to 1 and the coefficients with AR and BL are slightly lower. Compared to NCEP–NCAR ZO, this pdf is thus shifted away from AR, BL, and ZO while its similarity with GA and WBL is about the same. CP2's similarity coefficients qualitatively compare well to NCEP–NCAR BL in Fig. 7. The coefficient of CP1 with BL is not 1 but about 0.5, indicating that the CP1 pdf has a significant part of its mass shifted away from NCEP–NCAR BL. It also moved away from AR and GA. The similarity coefficients of CP2 with WBL and ZO are about the same as for NCEP–NCAR BL. The plot for CP3 is almost identical to the corresponding one for NCEP–NCAR GA. This pdf is thus not only in its mean value very similar to NCEP–NCAR
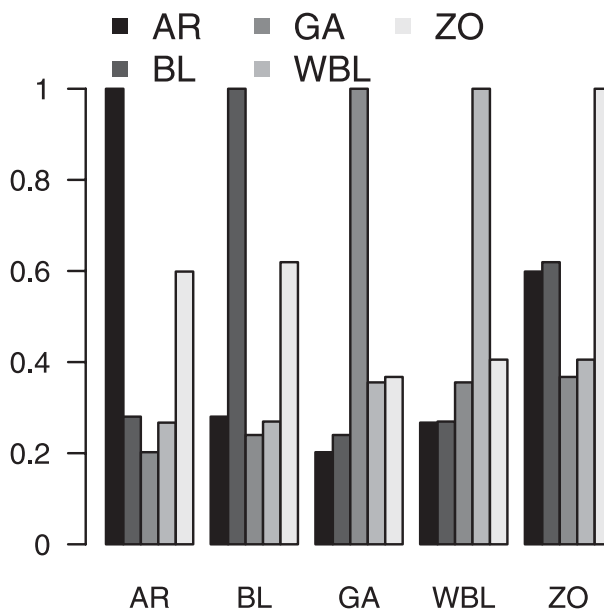
---

[2] The corresponding bar plots for the CPs of the other 12 GCM simulations are depicted in Fig. S2 (available as supplemental material at the Journals Online Web site: http://dx.doi.org/10.1175/2010JCLI3432.s2).



FIG. 7. Hellinger similarity measure for the five nonspherical NCEP–NCAR CPs with themselves.
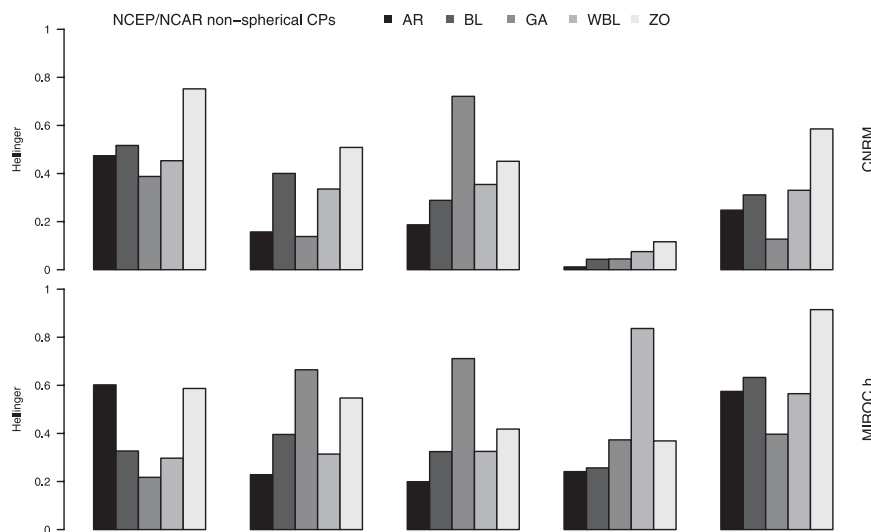
FIG. 8. Hellinger similarity coefficient calculated for the five nonspherical CPs (abscissa) from (top) CNRM CM3.0 and (bottom) MIROC3.2(hires) simulations with all nonspherical NCEP–NCAR CPs (gray shading).

GA (Fig. 6, first row, third column) but also in shape and size. Outstanding is CP4, which shows five extremely small Hellinger coefficients. It is thus not close to any of the NCEP–NCAR CPs. The centroid's dominating anticyclone shows anomalies significantly above those from all other GCM and reanalysis CPs (Fig. S1). Consisting of only 26 elements, this cluster forms a small "island" outside the usual range of the state space. CP5 shows a bar plot qualitatively similar to NCEP–NCAR's ZO. Coefficients are, however, smaller. The coefficient measuring similarity of CP5 with ZO is smaller than the similarity of CP1 with ZO, indicating that the pdf of CP5 shares less mass with ZO than the pdf of CP1.

### 2) MIROC3.2(HIRES)

The bar plot for CP1 is similar to the plot corresponding to NCEP–NCAR AR, with the coefficient for CP1 with AR itself attaining only 0.6. The mean pattern (Fig. 6, second row, fist column) reflects this strong similarity. CP2 and CP3 show a similar sequence of Hellinger coefficients as NCEP–NCAR GA, with CP2 having a slightly larger similarity with ZO. Both patterns (Fig. 6, second row, second and third column) are visually also similar to GA. The Hellinger coefficient of CP2 with CP3 is 0.54 (not shown); thus, despite their similarities, both pdfs have mass that they do not share. The bar plots for CP4 and CP5 strongly resemble the plots of NCEP–NCAR WBL and ZO, respectively. Only CP4's WBL coefficient and CP5's ZO coefficient are slightly smaller than 1.

In summary, specific CPs can be reproduced reasonably well by some GCMs, not only with respect to their mean pattern, but also regarding their defining pdfs.

MIROC3.2 medium resolution [MIROC3.2(medres)], for example, reproduces well AR, GA, WBL, and ZO, while the pdf defining NCEP–NCAR BL is better reproduced for CNRM CM3.0. Such an in-depth analysis of the CPs' configuration can be used to investigate particular state space regions and identify discrepancies to reanalysis data; we thus believe that this pdf-based analysis can be helpful for climate modelers.

### c. Comparing a specific CP across many GCMs

If a specific CP is to be compared across a large set of GCMs, the above-presented detailed analysis is not feasible. Instead, we suggest calculating a difference measure between the desired reanalysis CP and the closest GCM CP. The NCEP–NCAR nonspherical GA (Fig. 1, third row, third column) is most isolated from its accompanying CPs, as can be seen from the Hellinger similarity measure calculated for the five CPs with themselves, Fig. 7. It is less prone to be mixed with other accompanying CPs and hence particularly suitable for a comparison with its GCM counterparts. We use the GCM CP with the lowest difference to NCEP–NCAR GA measured by the $J$ coefficient for comparison.[3] In Figs. 6 and S1, this CP corresponds to CP4 for the Istituto Nazionale di Geofisica e Vulcanologia (Italy) (INGV) ECHAM4, the Institute of Numerical Mathematics Coupled Model, version 3.0

---

[3] Other difference measures would be feasible as well, depending on the intention. Here, we focus on the difference between the two pdfs, taking their two shapes into account. A choice based on the Euclidean distance, for example, would focus on the centroids.
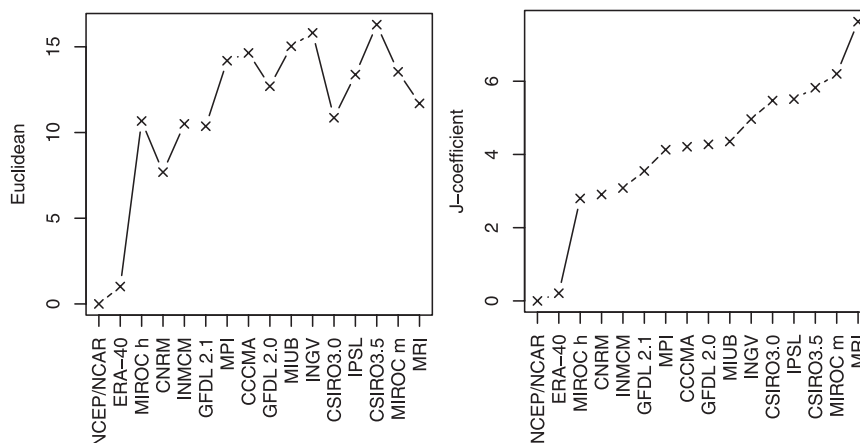
FIG. 9. (left) Euclidean distance and (right) $J$ coefficients of GCM CPs with NCEP–NCAR nonspherical GA. Chosen are the GCM CPs with the minimum $J$ coefficient with NCEP–NCAR GA. GCMs are ordered at the abscissa with increasing $J$ coefficient.

(INM-CM3.0), the Meteorological Institute of the University of Bonn, ECHO-G Model (MIUB ECHO-G), and to CP3 for all other GCMs. The $J$ coefficient values are shown in Fig. 9 (top) with the two reanalysis datasets and the GCMs on the abscissa, sorted in the ascending order of the $J$ coefficient. NCEP–NCAR is included for reference, yielding a zero $J$ coefficient. Very close to zero is the distance of ERA-40 GA to its NCEP–NCAR counterpart, defining the value for the $J$ coefficient for which CPs can still be considered as equal. Next are MIROC3.2(hires), CNRM CM3.0, and INM-CM3.0 with an almost identical $J$ coefficient. GCMs yielding larger $J$ coefficients can be read off from the abscissa in Fig. 9. The Meteorological Research Institute (MRI) Coupled General Circulation Model, version 2.3.2a (CGCM2.3.2a) centroid (Fig. S1, row 13, third column), for example, seems visually quite close to the NCEP–NCAR GA (Fig. 1, third row, third column), corroborated also by its Euclidean distance (Fig. 9, bottom); however, it exhibits the largest $J$ coefficient of all 14 GCMs and, thus, the largest distance in terms of pdfs. On the other hand, the MIUBECHOG centroid (Fig. S1, row 12, fourth column) visually appears quite different from NCEP–NCAR GA and shows a large Euclidean distance but a moderate $J$ coefficient. Although their centroids are not as similar, they share a large volume in state space.

The aim of such a comparison across many GCMs is to find one or a few GCMs that are in a certain region of the state space—defined by a CP pdf—very similar to reanalysis data. We expect this to be useful for studies that are sensitive to the model behavior in a particular state space region, such as the downscaling of precipitation or the analysis of central European heat waves related to the blocking pattern (Yiou and Nogaj 2004). As can be seen by comparing the top and bottom panel of Fig. 9, the

conventional Euclidean distance does not yield the same result. This is in general the case when pdfs' extensions (defined by their variances) are not negligible compared to the differences in the mean values. When comparing CPs from reanalysis and GCMs, one should expect that the mean values are close and the clusters' extensions play a role.

## 7. Summary and conclusions

Many objective algorithms for clustering atmospheric circulation define CPs by grouping elements into a cluster. The CPs are then usually represented using the average or composite pattern of all cluster members. Higher-order information, such as cluster size or shape in a multidimensional space, is frequently disregarded. For Gaussian mixture models clusters are not limited to spherical shapes, and taking this shape information into account can yield valuable information about the configuration of clusters—either within one dataset or between datasets. Clusters from different datasets can be compared as demonstrated here for the case of reanalysis data and GCM simulations. We defined CPs on the basis of multivariate normal probability distribution functions and the size and shape information of each CP is contained in covariance matrices. Focus of this study was on a set of difference measures, such as the Mahalanobis distance, the Kullback–Leibler divergence, the $J$ coefficient, or the Hellinger similarity coefficient to exploit this information for the comparison of CPs. With a simple simulated example, we demonstrated that these measures have the potential to add useful complementary information to the commonly used Euclidean distance of mean states.

The Gaussian mixture models were used to define five CPs in the North Atlantic region for NCEP–NCAR and

ERA-40 reanalyses. Initially, spherical clusters were employed to reproduce results obtained with $k$-means by PS01. The restriction to spherical clusters was removed and the covariance structure was selected on the basis of the BIC. Differences in the mean patterns from the reanalysis products are visible mostly for the CPs based on spherical clusters. The disparities in state space that are expected for ERA-40 and NCEP–NCAR are large enough to yield different cluster mean patterns for the two reanalysis products. Reasons for the susceptibility of cluster means to disparities in ERA-40 and NCEP–NCAR are supposed to be 1) not well-separated clusters and 2) a suboptimal cluster number. Although the CPs' mean patterns based on ellipsoidal clusters show similarities to the mean patterns of the spherical clusters, differences in clustering are evident. The population is much more unevenly distributed among clusters for the nonspherical solution. Furthermore, the relationships between nonspherical and spherical CPs were studied with conditional population histograms and the Hellinger similarity measure. Although the count-based population histograms should roughly approximate the Hellinger coefficient, differences are visible and are most likely to be the result of a sparse distribution of elements in a high dimensional space. This demonstrates also the limits of clustering and cluster configuration analysis in many dimensions.

In the same way, five CPs have been defined for twentieth-century simulations from 14 GCMs. CPs of two of the 14 GCMs are analyzed by means of the Hellinger similarity coefficient in the frame of reference set by NCEP–NCAR. The capability of a GCM to reproduce certain CPs is very dependent on the CP itself. In other words, for certain regions of the state space, GCMs reproduce the reanalysis pdf of states reasonably well. Which regions are well reproduced depends on the GCM. None of the GCMs studied here shows a good agreement with all five NCEP–NCAR CPs. MIROC3.2(hires) reproduces four CPs reasonably well.

Within the state space of NCEP–NCAR, the GA CP was found to be the most isolated, that is, its pdf showed the smallest similarity coefficient with the other pdfs. We considered it as particularly suitable for a simple comparison with GCM CPs. For every GCM, the CP closest to NCEP–NCAR GA had been chosen by means of the $J$ coefficient. The resulting distance between the GCM CP and NCEP–NCAR GA is calculated and compared for all 14 GCMs. The MIROC3.2(hires), the CNRM CM3.0, and the INM-CM3.0 yield CPs with the smallest distance to NCEP–NCAR GA. This ranking is different if only the CPs' mean patterns are compared using the Euclidean distance, emphasizing again the complementary information provided by the pdf-based difference measures.

We have mentioned the uncertainty of classification but have not addressed the GMM parameter uncertainty, that is, the uncertainty of mean patterns and covariance matrices. Although the uncertainties due to unknown cluster numbers and covariance structures are likely to be larger, this question should be investigated. More importantly, with respect to cluster comparison, is the question of significant cluster difference. The GMM approach is particularly suitable for the construction of a statistical test based on a parametric bootstrap approach. Development of such a test, as well as the investigation of other atmospheric fields than SLP, is left for future work.

With Gaussian mixture models and a suitable difference measure, CPs can be considered as volumes in state space that can differ in more than their mean states. An interpretation of their covariance structure in terms of the basis functions (PCs) can give hints on their spatial extension. Alternatively, reanalysis CPs can be used as a frame of reference for a detailed investigation of CP configuration. We consider this detailed analysis of the state space configuration by means of pdfs and the use of pdf-based distance measures as source of valuable information for climate modelers. Other than the evaluation of GCMs or a quantification of the separation of CPs (clusters) within one set of reanalysis or GCM data, the measures can be used in other climate change–related studies—for example, for obtaining weights for a model averaging in a Bayesian framework; for selecting appropriate GCMs for a CP-based downscaling scheme; to track and test for temporal changes in CPs, for example, from twentieth-century runs to future scenarios; or to investigate changes resulting from different external forcings of the GCMs.

## APPENDIX

### Closed Form for Gaussian Distributions

The KL divergence Eq. (7) for two $m$-dimensional normal distributions $p$ and $q$ is given by

$$d_{\mathrm{KL}}(p, q)$$
$$= \frac{1}{2}\left[\|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\sigma}_p^{-1}} + \mathrm{tr}(\boldsymbol{\sigma}_p^{-1}\boldsymbol{\sigma}_q - \mathbf{1}) + \log\left(\frac{|\boldsymbol{\sigma}_p^{-1}|}{|\boldsymbol{\sigma}_q^{-1}|}\right)\right],$$

$$(A1)$$

with $\mathbf{1}$ being the identity matrix in $m$ dimensions, and $\|\cdot\|_{\boldsymbol{\sigma}^{-1}}$ denoting the Mahalanobis distance, Eq. (6) (Bock and Diday 2000). The $J$ coefficient can be calculated from

this result using Eq. (8). For the Hellinger coefficient [Eq. (9)], we get in the case of two $m$-dimensional normal distributions $p$ and $q$ (Bock and Diday 2000):

$$
\begin{aligned}
d_{\mathrm{H}}^{(s)}(p, q) = {} & |s\mathbf{1} + (1-s)\boldsymbol{\sigma}_p^{-1}\boldsymbol{\sigma}_q|^{-s/2} |(1-s)\mathbf{1} + s\boldsymbol{\sigma}_q^{-1}\boldsymbol{\sigma}_p|^{-(1-s)/2} \\
& \times \exp\left\{ \frac{1}{2}\left[ \|s\boldsymbol{\sigma}_q^{-1}\boldsymbol{\mu}_q + (1-s)\boldsymbol{\sigma}_p^{-1}\boldsymbol{\mu}_p\|^2_{(s\boldsymbol{\sigma}_q^{-1} + (1-s)\boldsymbol{\sigma}_p^{-1})^{-1}} - s\|\boldsymbol{\mu}_q)\|^2_{\boldsymbol{\sigma}_q^{-1}} - (1-s)\|\boldsymbol{\mu}_p)\|^2_{\boldsymbol{\sigma}_p^{-1}} \right] \right\}.
\end{aligned} \tag{A2}
$$

## REFERENCES

Banfield, J. D., and A. E. Raftery, 1993: Model-based Gaussian and non-Gaussian clustering. *Biometrics,* **49,** 803–821.

Bárdossy, A., 1994: Downscaling from GCMs to local climate through stochastic linkages. *Climate Change, Uncertainty and Decision Making,* G. Paoli, Ed., Institute for Risk Research, 33–46.

Bock, H.-H., 1996: Probabilistic models in cluster analysis. *Comp. Stat. Data Anal.,* **23,** 5–28.

——, and E. Diday, 2000: Dissimilarity measures for probability distributions. *Analysis of Symbolic Data,* Springer, 153–165.

Brajard, J., F. Badran, M. Crépon, and S. Thiria, 2008: Validation of model simulations with respect to in situ observations by the use of probabilistic estimations. *Proc. IEEE Int. Joint Conf. on Neural Networks,* Hong Kong, China, Institute of Electrical and Electronics Engineers, 3015–3019.

Branstator, G., and F. Selten, 2009: "Modes of variability" and climate change. *J. Climate,* **22,** 2639–2658.

Carreau, J., and Y. Bengio, 2009: A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes,* **12,** 53–76.

Casado, M. J., M. A. Pastor, and F. J. Doblas-Reyes, 2009: Euro-Atlantic circulation types and modes of variability in winter. *Theor. Appl. Climatol.,* **96,** 17–29.

Casola, J., and J. Wallace, 2007: Identifying weather regimes in the wintertime 500-hPa geopotential height field for the Pacific–North American sector using a limited-contour clustering technique. *J. Appl. Meteor. Climatol.,* **46,** 1619–1630.

Cassou, C., 2008: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature,* **455,** 523–527.

Christiansen, B., 2006: Atmospheric circulation regimes: Can cluster analysis provide the number? *J. Climate,* **20,** 2229–2250.

Conway, D., and P. D. Jones, 1998: The use of weather types and air flow indices for GCM downscaling. *J. Hydrol.,* **212–213,** 348–361.

Davis, R. E., and L. S. Kalkstein, 1990: Development of an automated spatial synoptic climatological classification. *Int. J. Climatol.,* **10,** 769–794.

——, R. Dolan, and G. Demme, 1993: Synoptic climatology of Atlantic coast northeasters. *Int. J. Climatol.,* **13,** 171–189.

Dell'Aquila, A., V. Lucarini, P. M. Ruti, and S. Calmanti, 2005: Hayashi spectra of the Northern Hemisphere midlatitude atmospheric variability in the NCEP–NCAR and ECMWF reanalyses. *Climate Dyn.,* **25,** 639–652.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.,* **39B,** 1–38.

Duda, R. O., P. E. Hart, and D. G. Stork, 2001: *Pattern Classification.* 2nd ed. Wiley, 654 pp.

Fowler, H. J., S. Blenkinsop, and C. Tebaldib, 2007: Linking climate change modeling to impacts studies: Recent advances in downscaling techniques for hydrological modeling. *Int. J. Climatol.,* **27,** 1547–1578.

Fraley, C., and A. E. Raftery, 2002: Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Stat. Assoc.,* **97,** 611–631.

——, and ——, 2007: Model-based methods of classification: Using the mclust software in chemometrics. *J. Stat. Software,* **18.** [Available online at http://www.doaj.org/doaj?func=abstract&id=218544.]

Haines, K., and A. Hannachi, 1995: Weather regimes in the Pacific from a GCM. *J. Atmos. Sci.,* **52,** 2444–2462.

Hannachi, A., 1997: Low-frequency variability in a GCM: Three-dimensional flow regimes and their dynamics. *J. Climate,* **10,** 1357–1379.

——, 2007: Tropospheric planetary wave and mixture modeling: Two preferred regimes and a regime shift. *J. Atmos. Sci.,* **64,** 3521–3541.

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

——, and A. O'Neil, 2001: Atmospheric multiple equilibria and non-gaussian behaviour in model simulations. *Quart. J. Roy. Meteor. Soc.,* **127,** 939–958.

——, I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *J. Climatol.,* **27,** 1119–1152.

Hellinger, E., 1909: Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen. *J. Für Math.,* **136,** 210–271.

Hess, P., and H. Brezowsky, 1977: Katalog der Großwetterlagen Europas (1861–1976). Selbstverlag des Deutschen Wetterdienstes Bd. 15, Berichte des Deutschen Wetterdienstes, Offenbach am Main, 85 pp.

Hewitson, B., and R. G. Crane, 2002: Self-organizing maps: Applications to synoptic climatology. *Climate Res.,* **22,** 13–26.

Huth, R., 1996: An intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.,* **16,** 893–922.

——, 2000: A circulation classification scheme applicable in GCM studies. *Theor. Appl. Climatol.,* **67,** 1–18.

Jacobeit, J., 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.

Jolliffe, I. T., 2002: *Principal Component Analysis.* 2nd ed., Springer Series in Statistics, Springer, 487 pp.

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

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

Kohonen, T., 1998: The self-organizing map. *Neurocomputing,* **21,** 1–6.

Kullback, S., 1987: The Kullback–Leibler distance. *Amer. Stat.,* **41,** 340–341.

——, and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.,* **22,** 79–86.

Lamb, H. H., 1972: British Isles weather types and a register of daily sequence of circulation patterns, 1861–1971. Geophysical Memoir 116, HMSO, 85 pp.

Leloup, J., M. Lengaigne, and J.-P. Boulanger, 2008: Twentieth-century ENSO characteristics in the IPCC database. *Climate Dyn.,* **30,** 277–291.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, University of California Press, 281–297.

Mahalanobis, P. C., 1936: On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India,* **2,** 49–55.

Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.,* **48,** RG3003, doi:10.1029/2009RG000314.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.,* **88,** 1383–1394.

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

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye, 2008: Weather types and rainfall over Senegal. Part I: Observational analysis. *J. Climate,* **21,** 266–287.

Pearson, K., 1894: Contributions to the theory of mathematical evolution: Part I: On the dissection of asymmetrical frequency curves. *Philos. Trans. Roy. Soc. London,* **A185,** 71–85.

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

Plaut, G., and E. Simonnet, 2001: Large-scale circulation classification, weather regimes, and local climate over France, the Alps, and Western Europe. *Climate Res.,* **17,** 303–324.

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, 426 pp.

R Development Core Team, cited 2004: R: A language and environment for statistical computing. [Available online at http://www.R-project.org.]

Schüepp, M., 1978: Regionale Klimabeschreibung. 1. Teil Gesamtübersicht, Westschweiz, Wallis, Jura und Juranordfuss sowie Mittelland. Klimatologie der Schweiz Band 2, Schweizerische Meteorologische Zentralanstalt, 245 pp.

Schwarz, G., 1978: Estimating the dimension of a model. *Ann. Stat.,* **6,** 461–464.

Simonnet, E., and G. Plaut, 2001: Space–time analysis of geopotential height and SLP, intraseasonal oscillations, weather regimes, and local climates over the North Atlantic and Europe. *Climate Res.,* **17,** 325–342.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixtuare model clustering. *J. Atmos. Sci.,* **56,** 3704–3723.

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

Sterl, A., 2004: On the (in)homogeneity of reanalysis products. *J. Climate,* **17,** 3866–3873.

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

Vautard, R., 1990: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.,* **118,** 2056–2081.

Vrac, M., and P. Naveau, 2007: Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resour. Res.,* **43,** W07402, doi:10.1029/2006WR005308.

——, K. Hayhoe, and M. Stein, 2007a: Identification and intermodel comparison of seasonal circulation patterns over North America. *Int. J. Climatol.,* **27,** 603–620.

——, M. Stein, and K. Hayhoe, 2007b: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Res.,* **34,** 169–184.

Ward, J. H., 1963: Hierarchical grouping to optimize an objective function. *J. Amer. Stat. Assoc.,* **58,** 236–244.

Wehrens, R., and L. M. C. Buydens, 2007: Self- and super-organizing maps in R: The kohonen package. *J. Stat. Software,* **21,** 1–19.

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.,* **34,** 2995–3008.

Yarnal, B., 1993: *Synoptic Climatology in Environmental Analysis.* Bellhaven Press, 195 pp.

Yiou, P., and M. Nogaj, 2004: Extreme climatic events and weather regimes over the North Atlantic: When and where? *Geophys. Res. Lett.,* **31,** L07202, doi:10.1029/2003GL019119.