Observed Nondiffusive Dynamics in Large-Scale Atmospheric Flow

D. T. CROMMELIN*

Royal Netherlands Meteorological Institute (KNMI), De Bilt, and Department of Mathematics, Utrecht University, Utrecht, Netherlands

(Manuscript received 12 December 2002, in final form 4 May 2004)

ABSTRACT

The dynamics of large-scale atmospheric flow is characterized by both noisy, diffusive behavior, stemming mainly from small-scale, high-frequency atmospheric processes and by more structured, nondiffusive aspects, coming from, for example, the internal nonlinear dynamics of the flow system. In this study an attempt is made to isolate the latter influence from the former in an analysis of observational data of the atmosphere system during winter. The data are projected onto the unit sphere in a reduced three-dimensional phase space of dominant variability patterns. The probability density function associated with the data on the unit sphere shows clear maxima that correspond with flow regimes also found in previous studies. Concepts from the theory of Markov chains and phase-space partitions are used to bring out the nondiffusive, conservative dynamics of the large-scale atmospheric flow. By inspecting the asymmetries of transitions between different phase-space cells, a preferred, closed path, with a preferred direction, over the unit sphere of the reduced phase space is detected. This path, or cycle, connects episodes of zonal and blocked flow in the Atlantic sector. It is reminiscent of the remnants of a heteroclinic cycle found to guide regime transitions in a model study by the author.

1. Introduction

The old topic of large-scale atmospheric flow dynamics and its relation the phenomenon of circulation regimes does not cease to be a field of active research in the age of climate change studies. Indications (Palmer 1999; Corti et al. 1999) exist that climate change causes changes in the occupation statistics of naturally occurring circulation regimes, rather than inducing new patterns of atmospheric circulation. Compared with the 1950s and 1960s, some regimes have been visited more frequently in the 1970s and 1980s, others less frequently. This change is thought to be due to the overall change of climate over the last 50 yr. Such a perspective on climate change clearly calls for a better understanding of the dynamics of large-scale flow and of regime transitions. It is the sensitivity of those dynamics to changing climate parameters that will eventually determine which flow patterns will be realized more or less often.

The study of large-scale flow dynamics from observational data is particularly difficult. A range of scales of motion, many of them highly turbulent, interacts with the large-scale flow, burying whatever structured dynamics the large scales may possess under great amounts of noisy behavior. Such structured behavior may come, for example, from the internal nonlinear dynamics of the large-scale flow system. For the flow statistics, techniques are known that can determine whether the anomalously frequent realization of a flow pattern is a structural phenomenon or rather the accidental aggregation of realizations of an essentially random system. An example is the estimation of probability density functions (PDFs) in combination with Monte Carlo simulations.

The existence of atmospheric circulation regimes is a somewhat controversial issue, and in particular the question whether or not large-scale flow data are characterized by multimodal distributions is intensely debated. Compare, for instance, studies by Mo and Ghil (1988), Molteni et al. (1990), Kimoto and Ghil (1993a,b), Corti et al. (1999), and Smyth et al. (1999) with those by Wallace et al. (1991), Toth (1991), Nitsche et al. (1994), Hannachi and O'Neill (2001), and Stephenson et al. (2004). It needs to be stressed that an unimodal distribution of data does not preclude the possibility of preferred circulation patterns. As is pointed out by, for instance, Wallace et al. (1991) and Stephenson et al. (2004), a distribution can be unimodal but skewed (or kurtotic), giving rise to preferred patterns.

In this study we use 500-mb geopotential height winter data from the National Centers for Environmental Prediction (NCEP) reanalysis dataset (1948–2000), not only to calculate (spherical) PDFs, but also to extract information about structured phase-space motion that may be present in the data. This is done by dividing a

* Current affiliation: Courant Institute of Mathematical Sciences, New York University, New York, New York.

Corresponding author address: D. T. Crommelin, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012.
E-mail: crommelin@cims.nyu.edu

reduced phase space (that of the leading three EOFs) in cells and studying the transitions, observed in the data, between the various cells. The transition probability matrix obtained by the division in cells and the counting of transitions between cells determines a Markov chain, and its symmetric and antisymmetric components can give information about the probability flow through the reduced phase space.

The analysis technique is inspired by the work of Pasmanter and Timmermann (2003, hereafter PT), who use phase-space partitions and Markov chains to study ENSO dynamics. This approach allows us to distinguish between the purely diffusive (irreversible) dynamics in a system that has an inherently stochastic nature and the nondiffusive (conservative, reversible) dynamics in the system. The influence of dynamical structures such as periodic orbits, or preferred paths due to heteroclinic or homoclinic connections, should be visible in the latter part of the dynamics. We use this approach to study the observational data of the Northern Hemisphere (NH) flow.

In section 2 the preparation of the NCEP observational data, to make it suitable for analysis, is summarized. In section 3 the calculation of the PDF is described. The data points in the normalized reduced phase space of the first three principal components (PCs) are projected onto the unit sphere. The PDF of the projected data is calculated on the sphere, making the non-Gaussianity of the data distribution, and the flow regimes present, clearly visible. In section 4 we review a theory on Markov chains and phase-space partitions and apply it to some simple examples. In section 5 these concepts are used to analyze the flow of the atmospheric system through the reduced phase space, or more precisely, on the unit sphere. Conclusions follow in section 6.

2. Data

The data used are the 1948–2000 NCEP reanalysis data for the NH (17.5°–87.5°N). Since we are interested in large-scale flow characteristics, a 5° resolution in the latitude–longitude grid will be sufficient, giving 1080 grid points in total. The 500-hPa geopotential height field (Z_{500}) is a good representative of the planetary-scale midlatitude flow. As regime behavior mainly plays a role during the winter season, we perform our analysis on the data from the winter half year, which we define to be the 181 days starting at 1 November. Thus, the dataset used consists of 52 winters, each 181 days long, with the first winter starting on 1 November 1948 and the last winter ending on 29 April 2000. To prepare the data for analysis we remove the seasonal cycle: if $Z(d, i)$ is the field at day d of winter season i, we subtract from it the average $Z_{\text{av}}(d)$,

$$Z_{\text{av}}(d) = \frac{1}{31 \times 52} \sum_{d'=d-15}^{d+15} \sum_{i=1}^{52} Z(d', i). \qquad (2.1)$$

This gives the anomaly for each day with respect to the long-term mean of the 31-day time segment centered around day d of the season. At the beginning and end of each winter, 15 days are lost this way, leaving us with 151 daily data points for each winter. From the resulting dataset, empirical orthogonal functions (EOFs) and their corresponding PCs are calculated. The increasing density of grid points at higher latitudes is taken into account in the EOF calculation by multiplying each gridpoint value with the square root of the cosine of the gridpoint latitude. Note that no detrending was performed on the data.

In Fig. 1 the flow patterns corresponding to the first three EOFs are shown, with the standard deviation of their respective PCs as amplitudes. EOFs 1 and 2 closely resemble those of Kimoto and Ghil (1993a, hereafter KG1). The patterns of the first three EOFs are hardly different when calculated from 3-, 5-, or 10-day means, but the variance spectrum changes. The variance associated with EOFs 1, 2, and 3 together ranges from 22.0% for daily data to 31.2% for 10-day means.

We restrict our analysis to the reduced phase space spanned by the first three EOFs, instead of the full 1080-dimensional phase space. A justification for this restriction comes from the fact that the low-frequency behavior of the large-scale atmospheric flow is dominated by the leading EOFs. In KG1 it is found that this low-frequency behavior is largely captured in the leading eight EOFs. However, as is also discussed there, the limited amount of data in practice prevents a study of the dynamics in the subspace of the first eight EOFs. For example, for a reliable PDF estimation in eight dimensions, over 40 000 (independent) data points are necessary. Even more data will be needed for a study of transitions and phase-space motion in eight dimensions. The restriction to the first three EOFs is a compromise between dimensional requirements on the one hand and limited sample size on the other hand. The effect of the addition of more dimensions can only be studied thoroughly with a long model dataset.

In what follows we mainly deal with the time series of the PCs normalized by their respective standard deviations. These normalized PCs are denoted by a_1, a_2, and a_3. The corresponding normalized reduced phase space will be denoted by $\mathcal{P}$.

3. A spherical PDF

In studies of atmospheric regime behavior it is rather common to focus on the projection of the data (either from observations or from model simulations) onto the plane spanned by the first two EOFs, that is, the (a_1, a_2) plane. Calculations of PDFs in this plane typically use Euclidean distance as a measure of similarity (e.g., KG1). However, one can also use angular distance in phase space (which equals the arccosine of the pattern correlation) as a measure of similarity between two flow patterns. This was done by Kimoto and Ghil (1993b, hereafter KG2) for the calculation of PDFs of the large-

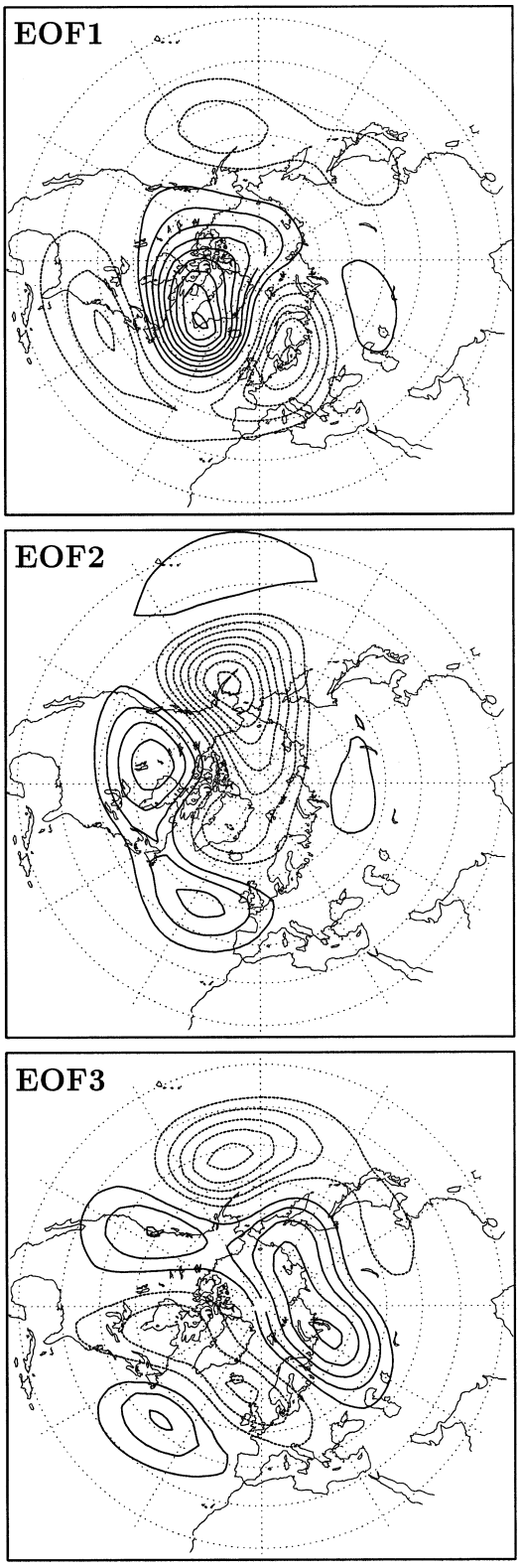

FIG. 1. EOFs 1, 2, and 3 of the 500-mb geopotential height field, calculated from NCEP reanalysis data (1948–2000, Nov to Apr) of the NH. The EOFs have been multiplied by the standard deviation of their respective PCs. Contour interval is 15 m: solid contours are positive, and dashed contours are negative (zero contour not drawn). The variance associated with EOFs 1–3 is 22.0%.

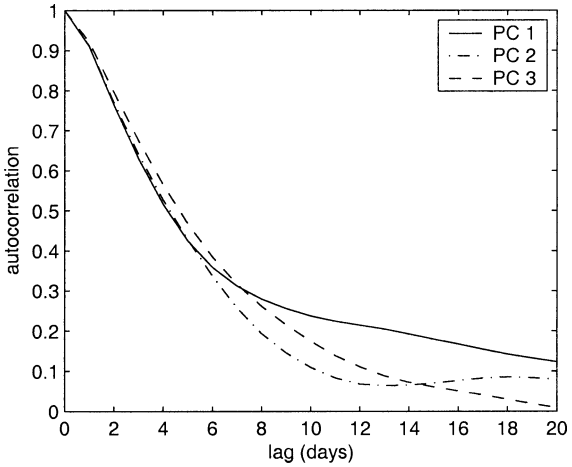

FIG. 2. Autocorrelation functions for PCs 1–3 projected onto the unit sphere in the reduced phase space $\mathcal{P}$.

scale NH flow in the Atlantic and Pacific sectors. They remark that ''the larger significance attached to the pattern relative to the magnitude of the anomaly is in concert with usual meteorological intuition.'' Prior to KG2, Horel (1985) and Mo and Ghil (1987, 1988) used spatial correlation of flow patterns as a distance measure for cluster analysis of NH atmospheric flow data.

In this study we project the data onto the unit sphere in the reduced phase-space $\mathcal{P}$ and calculate the PDF on the sphere. This involves a coordinate transformation from the Cartesian coordinates (a_1, a_2, a_3) to the spherical coordinates (ρ, θ, ϕ) according to

$$a_1 = \rho \cos\theta \sin\phi,$$
$$a_2 = \rho \sin\theta \sin\phi,$$
$$a_3 = \rho \cos\phi, \tag{3.1}$$

with $0 \leq \rho$, $0 \leq \theta < 2\pi$, $0 \leq \phi \leq \pi$. By projecting onto the unit sphere, more information about the structure of the flow patterns is retained than in the (a_1, a_2) projection. However, information on the amplitudes (ρ) of flow patterns is ignored: two flow patterns that have spatial correlation 1 (as far as the restriction to the first three EOFs is concerned), but different amplitudes—that is, $\rho_1 \neq \rho_2$—project onto the same point (θ, ϕ). We use angular distance as a measure of distance between two points on the sphere: the angle between the two lines that connect those two points with the center point of the sphere (i.e., with the origin of $\mathcal{P}$). As mentioned, the cosine of this angle equals the correlation between the two points.

Before calculating the actual PDF, let us consider the time series of the PCs projected onto the unit sphere. Their autocorrelation functions are shown in Fig. 2. After 7 days, the correlations have largely disappeared, since they have all dropped below e^{-1}. Since the use of Markov chains in later sections of this study would be problematic for time series with highly correlated

subsequent data points, we will use 7-day means of the PCs described in the previous section. This gives us 1 data point every 7 days, or 21 data points for every winter. The 7-day means are again projected onto the unit sphere.

The spherical PDF for the 7-day means is estimated using kernel density estimation with an Epanechnikov kernel (Silverman 1986; KG1). The estimated PDF $f(\theta, \phi)$ is a summation of the contributions $f^i(\theta, \phi)$ from each data point i with coordinates $(\rho^i, \theta^i, \phi^i)$, or equivalently (a_1^i, a_2^i, a_3^i). Each contribution $f^i(\theta, \phi)$ is an Epanechnikov kernel on the sphere, centered at (θ^i, ϕ^i), with smoothing parameter σ. The total PDF $f(\theta, \phi)$ gives the probability per unit area $\sin\phi d\theta d\phi$. Thus,

$$f(\theta, \phi) = \frac{1}{c} \sum_{i=1}^{N} f^i(\theta, \phi)$$

$$f^i(\theta, \phi) = \begin{cases} 1 - (\alpha^i/\sigma)^2 & \text{if } (\alpha^i/\sigma)^2 < 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

in which α^i is the angle spanned by the points (θ, ϕ), the center of the sphere, and (θ^i, ϕ^i). It is easily evaluated as $\alpha^i = \arccos(\mathbf{v}^T\mathbf{v}^i)$ with $\mathbf{v} = \mathbf{a}/\rho$ and $\mathbf{v}^i = \mathbf{a}^i/\rho_i$, the unit length vectors in $\mathcal{P}$ pointing to (θ, ϕ) and (θ^i, ϕ^i) on the sphere. The normalization constant c guarantees that the integral of the PDF over the sphere equals one: $\int d\theta d\phi f \sin\phi = 1$.

We do not use an adaptive kernel estimator (as is done in KG1), since the projection on the sphere does not give a distribution with low-probability tails, as one would have when considering the distribution in $\mathcal{P}$. As we will see, the difference between the minimum and the maximum values of the estimated spherical PDF is less than a factor of 2 for our choice of σ. An estimation method designed to handle long tails is therefore not considered necessary here.

In Fig. 3, the estimated PDF of the 7-day means is shown with bandwidth parameter $\sigma = \pi/6$. The value of the PDF at each point was divided by $f_G(\theta, \phi) = 1/(4\pi)$. This is the value of the PDF all over the sphere in case of a perfectly spherically symmetric distribution (e.g., a Gaussian distribution, either univariate or multivariate; recall that the PCs were normalized) of data points in $\mathcal{P}$. Thus, it is very easy to see that the PDF in Fig. 3 is not spherically symmetric. Several areas of maximal probability can be identified, as well as regions of minimal probability. Figure 3 shows a rectangular projection of the sphere, as well as ''north pole'' and ''south pole'' stereographic projections (hereafter NPS and SPS, respectively) of the upper half and the lower half of the sphere. We define the north pole of the sphere to be the point $\phi = 0$ and the south pole the point $\phi = \pi$.

The statistical significance of the regions of increased and decreased probability (compared to a Gaussian distribution) was tested by calculating 10 000 random PDFs from Gaussianly distributed datasets, each 1092 data points long. Regions where 80% or more of the random Gaussian PDFs have lower probability than the PDF of Fig. 3 are shown in Fig. 4. Five regions of significantly increased probability are present in the spherical PDF; they are indicated with letters A–E. PDF maxima A, C, and D are significant at the 95% level, B at 90%, and E at 85%. From now on we refer to these maxima as regimes.

It must be stressed that the calculation of the spherical PDF and the statistical significance levels of its maxima do *not* provide any proof of multimodality of the distribution of states in the reduced phase-space $\mathcal{P}$ (nor do they provide proof of unimodality). Since the data points were projected onto the unit sphere in $\mathcal{P}$, all that can be concluded here is that the 7-day means have several preferred *patterns* (regardless of their amplitudes) but not per se preferred *states* (which would require specification of amplitudes). The presence of preferred patterns may reflect multimodality of the dataset, but can equally well be due to the dataset being non-Gaussian yet unimodal. It rules out multinormality (i.e., Gaussianity) of the data, however.

The flow patterns corresponding to the five maxima visible in the spherical PDF are shown in Fig. 5. Their amplitudes correspond with points on the unit sphere in $\mathcal{P}$, that is, $\rho = 1$. The maxima of the PDF were identified by visual inspection.

The five flow regimes in Fig. 5 are in reasonable correspondence with the results of earlier studies, such as Mo and Ghil (1998, hereafter MG88), Cheng and Wallace (1993, hereafter CW), KG1, Smyth et al. (1999, hereafter SIG), Corti et al. (1999, hereafter CMP), and Monahan et al. (2001, hereafter MPF). Our regime A resembles the R regime of CW, also found by MPF and SIG, the Pacific–North American (PNA) regime of KG1, and MG88's cluster 1. Regime B, with zonal North Atlantic Oscillation (NAO) character, does not have a counterpart in the A–R–G classification of CW, but corresponds with the zonal NAO (ZNAO) regime of KG1. CMP's cluster A is in between our A and B. Our regime C can be seen to correspond well with SIG and MPF's regime A, MG88's cluster 2, and CMP's cluster B. CW's regime A looks more like our D. KG1's reversed PNA (RNA) pattern as well as CMP's cluster C resemble both C and D. Finally, regime E can be identified as the G (CW, SIG, MPF) or blocked NAO (BNAO) regime (KG1), also dubbed cluster D by CMP.

As a final remark, it must be said that the focus in this study is on hemispheric flow, and thus the regimes studied and discussed here are hemispheric flow regimes. This is not meant to imply that the notion of sectorial regimes (i.e., flow patterns restricted to the Pacific or Atlantic sectors) would be less meaningful or less reliable than the hemispheric version. We end up with hemispheric regimes simply because we study hemispheric flow patterns. The analysis could be carried

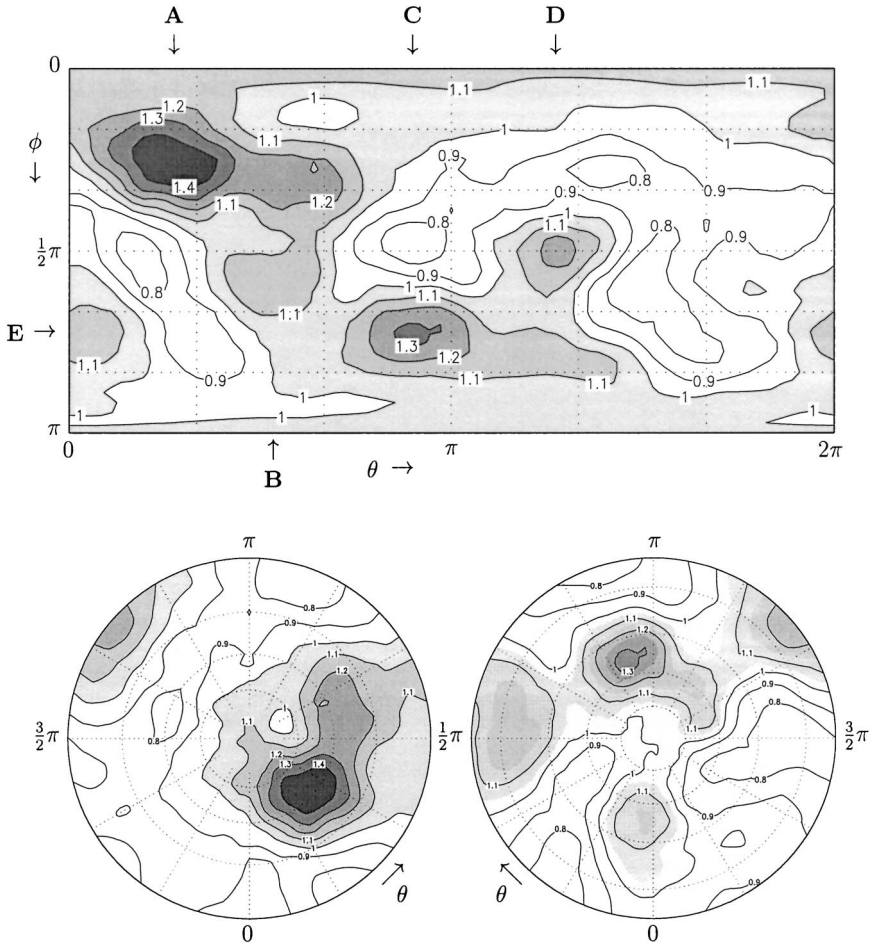

FIG. 3. Spherical PDF of NH atmospheric flow data (7-day means) in normalized reduced phase-space $\mathcal{P}$, calculated with bandwidth parameter $\sigma = \pi/6$. Shown is $f(\theta, \phi)/f_G$, the PDF divided by the (constant) probability density $f_G(\theta, \phi) = 1/4\pi$ of a spherically symmetric (e.g., Gaussian) distribution. A value higher (lower) than 1 indicates higher (lower) probability than Gaussian. (top) Rectangular projection, $(\theta, \phi) \in [0, 2\pi) \times [0, \pi]$. (bottom left) NPS and (bottom right) SPS, $(\theta, \phi) \in [0, 2\pi) \times [0, \pi/2]$ and $(\theta, \phi) \in [0, 2\pi) \times [\pi/2, \pi]$, respectively. Contour interval is 0.1. The letters A–E indicate five main maxima in the PDF; their corresponding flow patterns are shown in Fig. 5. Significance levels are shown in Fig. 4.

out using data of sectorial flow patterns as well, but this would be a subject for a different study.

4. Phase-space partitions and Markov chains

The PDF reveals the preferred flow patterns, or more generally the regions of increased and decreased probability. However, it does not tell much about the dynamical behavior of the system under consideration, since the PDF presents a rather static view. It shows the stationary distribution of states (assuming stationarity of the system), but not the flow of the system through phase space. In order to get an idea of this phase-space flow, we divide the normalized, reduced phase-space $\mathcal{P}$ in cells and consider the probabilities of transitions between the various cells.

The description of the time evolution of a system as a probability vector $\boldsymbol{p}(t)$ (assigning a probability to each cell) at time t, which is multiplied by a matrix $\mathbf{M}$ of transition probabilities to arrive at the next time step—that is, $\boldsymbol{p}(t + \Delta t) = \mathbf{M}\boldsymbol{p}(t)$—is known as a Markov chain description. It has been used in the past to model atmospheric regime behavior by Spekat et al. (1983), Mo and Ghil (1987, 1988), Molteni et al. (1990), and Kimoto and Ghil (1993b). These studies determine a number of regimes and calculate transition matrices by counting regime transitions. During the transitions the system is assumed to be in a transient state that is not classified. Thus, not the entire (reduced) phase space is studied, but (transitions between) a few bounded regions of it. In this manner, certain preferred orders of transitions between regimes are found (see Mo and Ghil 1987, 1988; Kimoto and Ghil 1993b).

In this section, we do not limit ourselves to specific

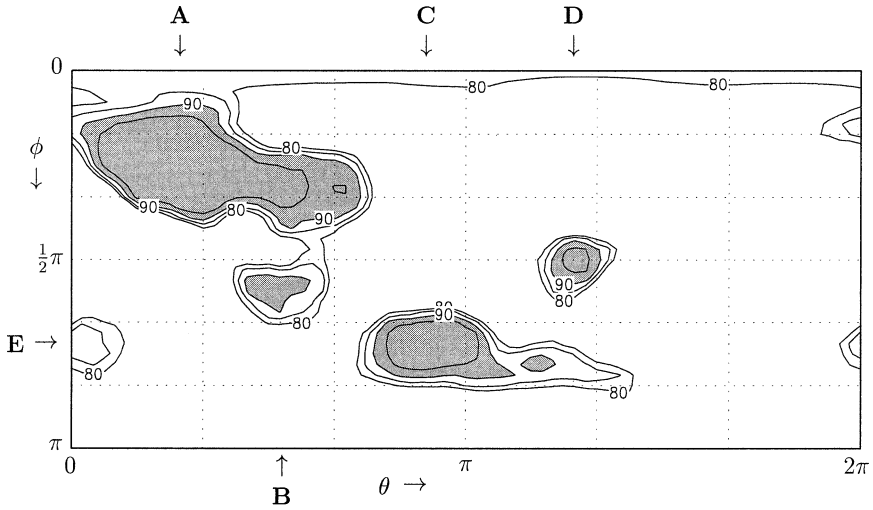

FIG. 4. Regions of the spherical PDF of Fig. 3 with significantly increased probability compared to a Gaussian PDF. Shown are regions where 80% or more of 10 000 random Gaussian PDFs have lower probability than the PDF of Fig. 3. Contour interval is 5%. Regions with a statistical significance level of 90% or higher are shaded.

phase-space regions identified as regimes, but rather we divide all of $\mathcal{P}$ into cells. The transition matrix associated with such a partition can reveal information on preferences and inhomogeneities of the flow of the system through $\mathcal{P}$. More specifically, we partition $\mathcal{P}$ into k cells all containing an equal number of data points, as was done in, for example, PT. We shall refer to such a partition into cells all with an equal amount of data as an *equipartition*. The transition probability matrix $\mathbf{M}$ associated with an equipartition has the useful property that the sum over each row or column of the matrix equals one. Thus, if the Markov chain description is

$$p_i(t + \Delta t) = \sum_j M_{ij} p_j(t), \tag{4.1}$$

the matrix $\mathbf{M}$ satisfies

$$M_{ij} \geq 0 \ \forall i, j, \tag{4.2}$$

$$\sum_i M_{ij} = 1 \ \forall j, \tag{4.3}$$

$$\sum_j M_{ij} = 1 \ \forall i. \tag{4.4}$$

Matrices satisfying all three properties are called *doubly stochastic* matrices. A detailed treatment of equipartitions and (doubly) stochastic matrices in a geophysical context can be found in PT.

In PT, the matrix $\mathbf{M}$ is split up in its symmetric part $\mathbf{M}^S$ and antisymmetric part $\mathbf{M}^A$ in order to distinguish between the diffusive or dissipative part of the dynamics (called irreversible) and the nondiffusive or conservative part (called reversible); $\mathbf{M}^S$ is again doubly stochastic, and $\mathbf{M}^A$ is not. Pasmanter and Timmerman (2003) calculate $\langle \Delta I \rangle$, the average information loss, or entropy production, associated with the system $\mathbf{p}(t + \Delta t) = \mathbf{M}p(t)$ and distinguish two contributions to the information loss:

$$\langle \Delta I \rangle = -\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} M_{ij} \ln M_{ij} = \langle \Delta I \rangle_S + \langle \Delta I \rangle_{A/S},$$

$$\langle \Delta I \rangle_S = -\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} M^S_{ij} \ln M^S_{ij},$$

$$\langle \Delta I \rangle_{A/S} = -\frac{1}{k} \sum_{i=1}^{k} \sum_{j>i}^{k} M^S_{ij} \ln\left[1 - \left(\frac{M^A_{ij}}{M^S_{ij}}\right)^2\right] + M^A_{ij} \ln\left(\frac{M^S_{ij} + M^A_{ij}}{M^S_{ij} - M^A_{ij}}\right), \tag{4.5}$$

where $\langle \Delta I \rangle_S$ is the information loss one would have in a system $\mathbf{p}(t + \Delta t) = \mathbf{M}^S p(t)$. Both $\langle \Delta I \rangle \geq 0$ and $\langle \Delta I \rangle_S \geq 0$. In contrast, $\langle \Delta I \rangle_{A/S} \leq 0$, so the total information loss in a system is decreased by the presence of a nonzero $\mathbf{M}^A$. These observations led PT to associate $\mathbf{M}^A$ with the conservative, or nondiffusive, part of the dynamics of the system $\mathbf{p}(t + \Delta t) = \mathbf{M}p(t)$. A system with purely diffusive dynamics has zero $\mathbf{M}^A$; in contrast, $\mathbf{M}^S$ is still nonzero in case of purely conservative dynamics. Pasmanter and Timmerman (2003) present $\mu = 1 - k^{-1} \Sigma_{i,j} |M^A_{ij}|$ as a measure of the purely diffusive dynamics inherent to $\mathbf{M}$, ranging from 1 (totally diffusive; all M^A_{ij} are zero) to 0 (totally conservative; $\mathbf{M}$ is a permutation matrix).

It is not obvious that the atmospheric flow system can be represented by a Markov process. The Markovian assumption implies that the probability of the transition $i \to j$ does not depend on the state of the system previous to arriving at state i. A Markovian system has no memory of its history. On a day-to-day time scale, the large-

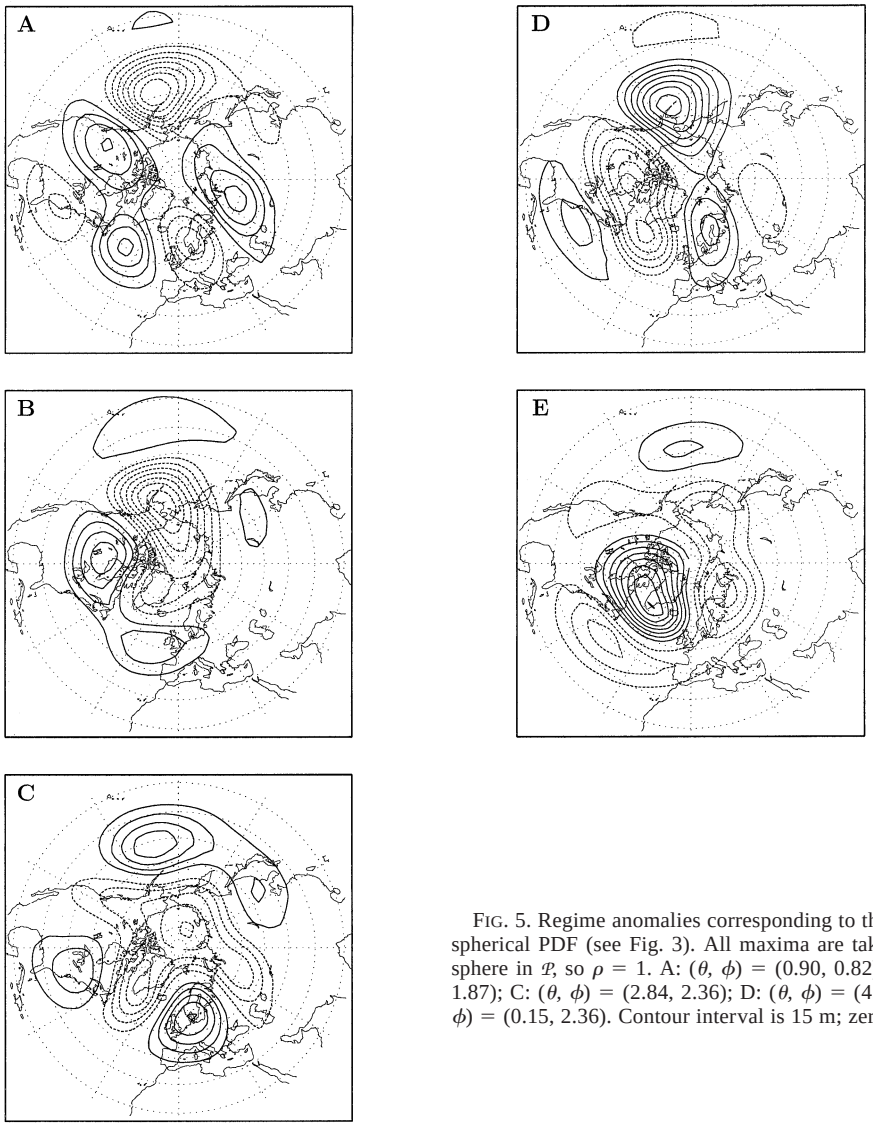

FIG. 5. Regime anomalies corresponding to the five maxima in the spherical PDF (see Fig. 3). All maxima are taken to be on the unit sphere in $\mathcal{P}$, so $\rho = 1$. A: $(\theta, \phi) = (0.90, 0.82)$; B: $(\theta, \phi) = (1.65, 1.87)$; C: $(\theta, \phi) = (2.84, 2.36)$; D: $(\theta, \phi) = (4.04, 1.57)$; and E: $(\theta, \phi) = (0.15, 2.36)$. Contour interval is 15 m; zero contour not drawn.

scale atmospheric flow system may well be non-Markovian, since the large-scale flow is highly correlated at a 1-day time lag. For this reason we use 7-day means instead of daily data in our analysis. The consecutive data points of the 7-day means are largely uncorrelated, so it is reasonable to assume that the atmospheric system has no memory of its history after 7 days.

A study that uses techniques similar to the analysis used here, drawn to our attention by one of the reviewers, was carried out by Egger (2001). He determines a master equation (also based on the Markovian assumption) from data and uses it to obtain information about a transition velocity vector (field) that relates to the mean motion in phase space. This vector also stems from asymmetries in transition rates. Egger uses 6-hourly data of angular momentum of the atmosphere and finds oscillating motion with a drift toward the origin of phase space.

a. Examples

It will be instructive to look at some simple examples to see how the information obtained by inspecting the transition matrix and its antisymmetric part relates to the behavior of a system. As a first example we consider a system with four different states (or four cells). It is governed by completely periodic behavior, since it always follows the trajectory $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1 \rightarrow \ldots$. The transition probability matrix (in this case a permutation matrix) is

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}; \tag{4.6}$$

its symmetric and antisymmetric parts are

$$\mathbf{M}_1^S = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{M}_1^A = \frac{1}{2}\begin{pmatrix} 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \end{pmatrix}. \tag{4.7}$$

Note that $\mathbf{M}_1^S$ is nonzero even though this system is periodic. The matrix $\mathbf{M}_1^A + |\mathbf{M}_1^A|$ recovers the periodic motion of the system:

$$\mathbf{M}_1^A + |\mathbf{M}_1^A| = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.8}$$

For the second example we modify the previous system slightly: at each step, the system has a probability of 0.7 to continue along the cycle of the previous system and probabilities of 0.1 to go to each of the other the cells (including itself). In other words, the transition matrix is now

$$\mathbf{M}_2 = \begin{pmatrix} 0.1 & 0.1 & 0.1 & 0.7 \\ 0.7 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \end{pmatrix}, \tag{4.9}$$

resulting in

$$\mathbf{M}_2^A = \begin{pmatrix} 0 & -0.3 & 0 & 0.3 \\ 0.3 & 0 & -0.3 & 0 \\ 0 & 0.3 & 0 & -0.3 \\ -0.3 & 0 & 0.3 & 0 \end{pmatrix},$$

$$\mathbf{M}_2^A + |\mathbf{M}_2^A| = \begin{pmatrix} 0 & 0 & 0 & 0.6 \\ 0.6 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \end{pmatrix}. \tag{4.10}$$

The matrix $\mathbf{M}_2^A + |\mathbf{M}_2^A|$ recovers the preferred cycle still present in the system. Note that direct transitions between cells 1 and 3 and between cells 2 and 4 are possible in this system. However, they do not appear in $\mathbf{M}_2^A + |\mathbf{M}_2^A|$, because each of these transitions is as probable as its reverse and is thus due to diffusion.

The last example is a purely diffusive system, with transition matrix

$$\mathbf{M}_3 = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}. \tag{4.11}$$

The absence of periodic motion is reflected in the fact that $\mathbf{M}_3^A = 0$.

The average information loss and measure μ of diffusive dynamics confirm, for each of these examples, the conclusion drawn from $\mathbf{M}_i^A + |\mathbf{M}_i^A|$. The first system has $\mu = 0$ and $\langle \Delta I \rangle_S = -\langle \Delta I \rangle_{A/S} = \ln 2$, resulting in zero information loss, or $\langle \Delta I \rangle = 0$. This reflects the completely periodic, nondiffusive character of the first system. For the second system, $\mu = 0.4$, $\langle \Delta I \rangle_S = 1.19$, $\langle \Delta I \rangle_{A/S} = -0.25$, and thus $\langle \Delta I \rangle = 0.94$. This system has significant diffusivity (nonzero μ) but also still considerable nondiffusive structure: $0 < \langle \Delta I \rangle < \langle \Delta I \rangle_S$. The third system is completely diffusive, so the upper bound of μ is saturated: $\mu = 1$. The information loss is not suppressed by any nondiffusivity, that is, $\langle \Delta I \rangle_{A/S} = 0$ and $\langle \Delta I \rangle = \langle \Delta I \rangle_S = \ln 4$.

b. *Significance test*

The examples made clear that the matrix $\mathbf{M}^A + |\mathbf{M}^A|$ reflects the structured, conservative dynamics present in the system described by the transition probability matrix $\mathbf{M}$. When calculating $\mathbf{M}$ from data, it will be necessary to estimate which elements in the matrix $\mathbf{M}^A$ are statistically significant (implying some presence of conservative dynamics) and are unlikely to be the result of finite sample size. Calculating $\mathbf{M}$ from data can hardly be expected to give a precisely symmetric matrix, even if the underlying system is purely diffusive. Thus, a test is needed to determine which elements of $\mathbf{M}^A$ are significantly different from zero.

From the equipartition of $\mathcal{P}$ in k cells with each N_k data points, the transition matrix $\mathbf{T}$ can be calculated, counting the number of transitions between cells. The transition probability matrix $\mathbf{M}$ is derived from $\mathbf{T}$ by $\mathbf{M} = N_k^{-1}\mathbf{T}$. Their respective symmetric and antisymmetric parts have the same relation: $\mathbf{M}^S = N_k^{-1}\mathbf{T}^S$, $\mathbf{M}^A = N_k^{-1}\mathbf{T}^A$. An element M_{ij}^A is significant if the corresponding element T_{ij} differs significantly from T_{ij}^S. That is, T_{ij}^A should lie outside, say, one or two standard deviations associated with T_{ij}^S.

If the system dynamics is driven by $\mathbf{M}^S$, we expect the N_k transitions from cell j to other cells to result in $T_{ij}^S = N_k M_{ij}^S$ transitions to cell i, where T_{ij}^S is the expectation value from N_k independent binomial Bernoulli trials, with at each trial a probability M_{ij}^S of ''success'' (transition to i) and a probability $1 - M_{ij}^S$ of ''no success'' (transition to any other cell). The standard deviation associated with the binomial distribution is $S_{ij} = [M_{ij}^S(1 - M_{ij}^S)N_k]^{1/2}$. For large enough N_k the binomial distribution can be approximated by the normal distribution (Feller 1968), and T_{ij} will therefore with 68.3% certainty lie in the interval $(T_{ij}^S - S_{ij}, T_{ij}^S + S_{ij})$, with 95.4% certainty in the interval $(T_{ij}^S - 2S_{ij}, T_{ij}^S + 2S_{ij})$, etc. We will require $T_{ij}^S \geq 10$ for T_{ij}^A to be at all potentially significant.

The value T_{ij} lies with 95% certainty between $T_{ij}^S \pm 1.96\, S_{ij}$, so T_{ij}^A is significant at the 95% confidence level if $|T_{ij}^A| > 1.96\, S_{ij}$. To be significant at the 90%, 85%, 80%, and 75% confidence level, $|T_{ij}^A|$ must exceed $1.64\, S_{ij}$, $1.44\, S_{ij}$, $1.28\, S_{ij}$, and $1.15\, S_{ij}$, respectively. Note that it is $\mathbf{T}^A$ rather than $\mathbf{M}^A$ that is tested, since a small

but nonzero element of $\mathbf{M}^A$ may very well be significant if N_k is large enough, whereas a large element of $\mathbf{M}^A$ can be insignificant if N_k is small.

5. Results

We partition $\mathcal{P}$ such that the cell boundaries only depend on θ and ϕ. The partition can be drawn on the unit sphere on which the PDF was calculated, thereby enabling an easy visualization and interpretation of probability flows. First we draw boundaries of constant ϕ, dividing the unit sphere in $\mathcal{P}$ into zonal bands. These bands are then each divided into cells, by drawing boundaries of constant θ within each band. The total number of cells is k; the number of data points used is kN_k. A possible remainder of $1092 - kN_k$ points at the end of the dataset is not used. Once the partition is made, $\mathbf{M}$ is calculated by counting transitions. Transitions from the last data point of one winter to the first data point of the next winter are left out, thereby introducing an unavoidable but small deviation of $\mathbf{M}$ from the precise properties of doubly stochastic matrices [Eqs. (4.3) and (4.4)]. This deviation will be ignored.

We made four different equipartitions, whose structures are visible in Fig. 6. The associated matrices $\mathbf{T}$ and $\mathbf{M}^A + |\mathbf{M}^A|$ are given below. The nonzero elements of $\mathbf{M}^A + |\mathbf{M}^A|$ show their significance level using typographic means (boldface, underline, italic): $\geq$**95%**, **95%–90%**, **90%–85%**, 85%–80%, 80%–75%, $<$*75%*.

The first partition gives

$$\mathbf{T} = \begin{pmatrix} 123 & 48 & 64 & 22 \\ 76 & 98 & 42 & 46 \\ 43 & 48 & 100 & 71 \\ 20 & 66 & 50 & 123 \end{pmatrix},$$

$$\mathbf{M}^A + |\mathbf{M}^A| = \frac{1}{100}\begin{pmatrix} 0 & 0 & \mathbf{7.7} & \mathit{0.7} \\ \underline{\underline{\mathbf{10.2}}} & 0 & 0 & 0 \\ 0 & \mathit{2.2} & 0 & \mathbf{7.7} \\ 0 & \mathbf{7.3} & 0 & 0 \end{pmatrix}, \tag{5.1}$$

the second partition

$$\mathbf{T} = \begin{pmatrix} 62 & 33 & 9 & 24 & 33 & 10 \\ 43 & 55 & 28 & 6 & 28 & 16 \\ 21 & 14 & 70 & 32 & 6 & 34 \\ 21 & 14 & 29 & 53 & 34 & 24 \\ 22 & 30 & 7 & 29 & 56 & 26 \\ 5 & 29 & 28 & 28 & 15 & 66 \end{pmatrix},$$

$$\mathbf{M}^A + |\mathbf{M}^A| = \frac{1}{100}\begin{pmatrix} 0 & 0 & 0 & \mathit{1.6} & \mathit{6.0} & \mathit{2.7} \\ \mathit{5.5} & 0 & \mathbf{7.7} & 0 & 0 & 0 \\ \mathbf{6.6} & 0 & 0 & \mathit{1.6} & 0 & \mathit{3.3} \\ 0 & \underline{4.4} & 0 & 0 & \mathit{2.7} & 0 \\ 0 & \mathit{1.1} & \mathit{0.5} & 0 & 0 & \underline{6.0} \\ 0 & \mathbf{7.1} & 0 & \mathit{2.2} & 0 & 0 \end{pmatrix}. \tag{5.2}$$

For the third partition we find

$$\mathbf{T} = \begin{pmatrix} 117 & 70 & 43 & 33 \\ 62 & 105 & 33 & 57 \\ 64 & 33 & 116 & 44 \\ 19 & 47 & 67 & 130 \end{pmatrix},$$

$$\mathbf{M}^A + |\mathbf{M}^A| = \frac{1}{100}\begin{pmatrix} 0 & \mathit{2.9} & 0 & \mathbf{5.1} \\ 0 & 0 & 0 & \mathit{3.7} \\ \mathbf{7.7} & 0 & 0 & 0 \\ 0 & 0 & \underline{\mathbf{8.4}} & 0 \end{pmatrix}, \tag{5.3}$$

and for the fourth partition

$$\mathbf{T} = \begin{pmatrix} 112 & 74 & 42 & 34 \\ 49 & 119 & 25 & 65 \\ 69 & 18 & 133 & 42 \\ 32 & 44 & 58 & 124 \end{pmatrix},$$

$$\mathbf{M}^A + |\mathbf{M}^A| = \frac{1}{100}\begin{pmatrix} 0 & \underline{\mathbf{9.1}} & 0 & \mathit{0.7} \\ 0 & 0 & \mathit{2.6} & \mathbf{7.7} \\ \underline{\underline{\mathbf{9.9}}} & 0 & 0 & 0 \\ 0 & 0 & 5.9 & 0 \end{pmatrix}. \tag{5.4}$$

In Table 1, the values of $\langle \Delta I \rangle$, $\langle \Delta I \rangle_S$, $\langle \Delta I \rangle_{A/S}$, and μ are shown for these four partitions. Diffusive dynamics clearly dominates the behavior of all four systems. However, some structured, nondiffusive behavior is also present. The partitions in four cells all possess a preferred cycle, with all or most of its segments significant at more than 85%, and some segments significant at 90% (partition 3, $3 \rightarrow 4$; partition 4, $2 \rightarrow 1$) and 95% (partition 1, $1 \rightarrow 2$, partition 4, $1 \rightarrow 3$). Because of the differences between the partitions, the cycles detected in the various partitions are not identical. However, their similarities are remarkable and are an indication of the robustness of the phenomenon. The partition in six cells also recovers the preferred cycle, but significance levels are lower for some of its segments (in particular transition $5 \rightarrow 1$) because of the limited amount of data available.

Shortage of data makes it very difficult to reach higher significance levels in the determination of the cycle segments, or to explore the structure of the cycle in more detail using a finer partition with more cells. It is unlikely one can get around this by using other ways of partitioning. For instance, defining cells to be centered around the regions of observed regimes will run

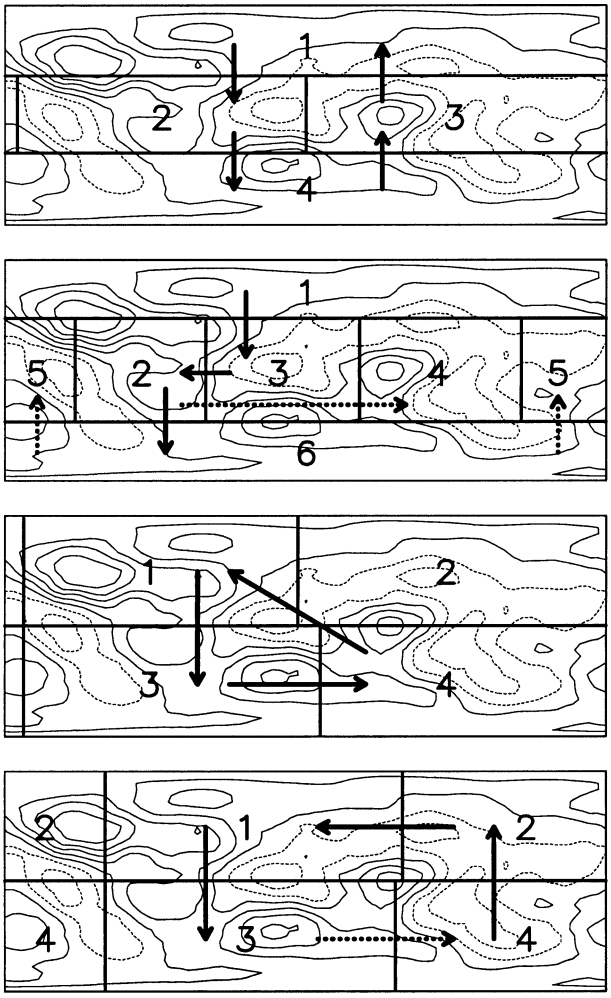

FIG. 6. Four different equipartitions of the data. Solid arrows indicate transition asymmetries significant at 85% or more (several are significant at 90% or 95% or more; see text). Dotted arrows indicate asymmetries significant at 80%–85%. Thin lines are contours of the scaled spherical PDF $f(\theta, \phi)/f_G$ of the 7-day-mean data (see Fig. 3), with contour interval 0.1, solid contours for values ≥ 1, and dashed contours for values <1.

into the same problem: the cells have to contain enough data points and will therefore in practice be so large that not all data points in one cell can really be classified into one regime anymore. It will also make the partitioning into equally probable cells much more complicated. Lifting the constraint of equally probable cells may only worsen the problem, as part of the data could end up in cells having too few data points to yield any statistically significant results. That part would then effectively be lost for the analysis. In this respect, an equipartition is an efficient way of making use of data.

In Figs. 7 and 8 the evolutions of the atmospheric flow along the preferred paths of partitions 1 and 4 are shown. For θ and ϕ we picked the mean values within each cell, and the average of these mean values for the steps in between the cells. The flow patterns are 500-mb geopotential height anomalies, with $\rho = 1$ chosen as amplitude.

TABLE 1. Average information loss $\langle \Delta I \rangle$, its constituents $\langle \Delta I \rangle_S$ and $\langle \Delta I \rangle_{A/S}$, and μ (a measure of the relative importance of diffusive dynamics) for four equipartitions.

Partition	$\langle \Delta I \rangle$	$\langle \Delta I \rangle_S$	$\langle \Delta I \rangle_{A/S}$	μ
1	1.260	1.268	−0.008	0.91
2	1.590	1.603	−0.013	0.90
3	1.246	1.253	−0.007	0.93
4	1.224	1.233	−0.009	0.91

The two evolutions are quite similar, the main difference being that in Fig. 7 the Pacific sector sees more activity than in Fig. 8. The positive and negative phases of the Artic Oscillation (AO) visible in Fig. 8 (steps 1–3 and 4–2) get more Pacific emphasis in Fig. 7 (steps 2 and 3). Positive and negative phases of the NAO are present in evolution 1 (Fig. 7) but not as pronounced as in evolution 4 (Fig. 8). It is difficult to say which of the two evolutions is the more truthful. Results from a preliminary version of this study (Crommelin 2003a), as well as the results from the partition in six cells (Fig. 6), suggest that the transitions between the upper and lower halves of the sphere (and back) take place at values of θ, most in agreement with the evolution of Fig. 8. Partition 1 locates these transitions at somewhat too small values of θ.

The evolution of partition 4 (Fig. 8) shows an alternation between episodes of mainly zonal anomalies (1–3, 4–2, $\phi \approx \pi/2$) and episodes of more meridionally oriented anomalies (2–1, 3–4, $\phi \approx 0, \pi$). The negative anomalies over the North Atlantic and the Pacific (step 2–1) start to merge and form one large negative arctic anomaly (step 1–3). Farther down the path, this negative anomaly gets elongated meridionally over the Eurasian continent and the west coast of North America. Over the North Atlantic and the North Pacific, it is replaced by positive anomalies. Reaching the bottom of the unit sphere (step 3–4), the pattern consists again of mainly meridionally oriented anomalies. The second half of the cycle shows a reverse evolution: positive North Atlantic and Pacific anomalies merge, forming a positive Arctic anomaly with negative anomalies surrounding it (step 4–2). In the last part of the cycle, the Arctic anomaly again stretches southward, and at 2–1 the anomaly pattern is back in its meriodional state.

Interpreted in terms of phases of the AO and the NAO, the preferred cycle of partition 4 shows the positive phase of the NAO (step 1) preceding the positive phase of the AO (step 1–3) and negative NAO (step 4) preceding negative AO (step 4–2). A similar preferred path was found by Crommelin (2003b) using a barotropic model of NH flow. That preferred path was interpreted as the remnant of a heteroclinic cycle, a cycle of phase-space connections back and forth between regime-like steady states.

The cause of the preferred cycle(s) showing up here

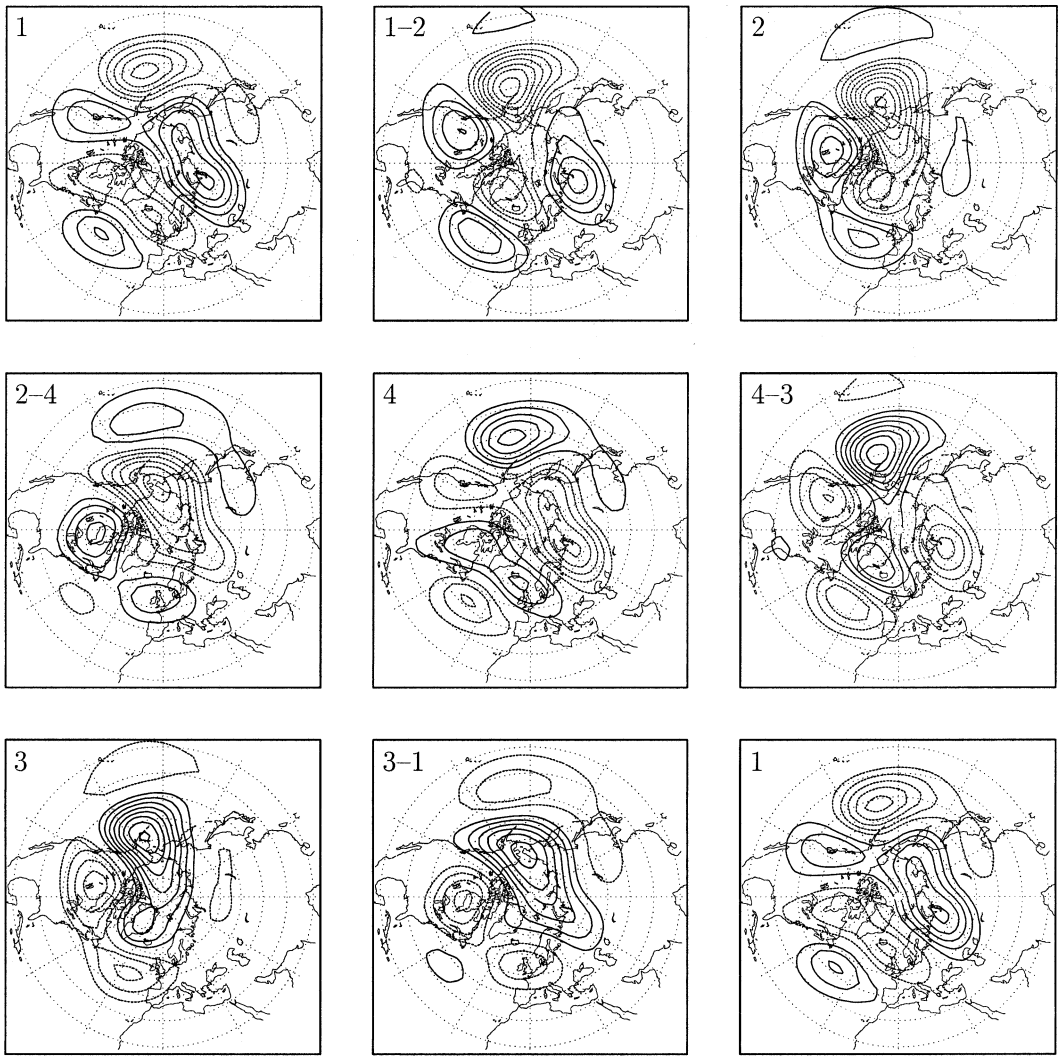

FIG. 7. Evolution of the atmospheric flow (500-mb geopotential height anomaly) along the preferred path (partition 1) through the reduced phase-space $\mathcal{P}$ (see Fig. 6). Contour interval is 15 m.

remains elusive. It is tempting to associate the nondiffusive part of the dynamics with the internal nonlinear dynamics of the large-scale atmospheric flow and to relate the diffusivity to the noise caused by small-scale, high-frequency atmospheric processes. However, as is shown by Majda et al. (1999, 2003), noise processes can induce systematic corrections in the large-scale flow system. Not only does the noise tend to ''smooth out'' the structured behavior of the large scales, it can also alter the structured behavior itself.

If, nevertheless, we hypothesize about interpreting the nondiffusive behavior using concepts from dynamical systems theory, the first possibility that comes to mind is that of an unstable periodic orbit. The system can be captured by the orbit for a (short) while, following (part of) the orbit, can leave it, return for a while, and so on. One can see some traces of the periodic behavior, but the system never follows the periodic orbit for long. Selten and Branstator (2004) detect a preferred cycle in a three-layer quasigeostrophic model of NH flow and interpret this cycle in terms of an unstable periodic orbit. As mentioned previously, in Crommelin (2003b) a preferred cycle, similar to the one found here, was detected and related to remnants of a heteroclinic cycle connecting regimes. Some phases of the preferred cycle detected here resemble regimes shown in Fig. 5: 3 (Fig. 8) resembles C well; some, but not total, resemblance can be seen between 2–1 and A, 1–3 and B, 4–2 and E. Unfortunately, the partitioning carried out here is too coarse to see whether the preferred cycle really connects (some of) the regimes. An interpretation along the same lines as in Crommelin (2003b) must therefore remain speculative at this point.

It must be stressed that although the structure detected and discussed here is one of a preferred cycle, this does not imply that the atmosphere often, or even occas-

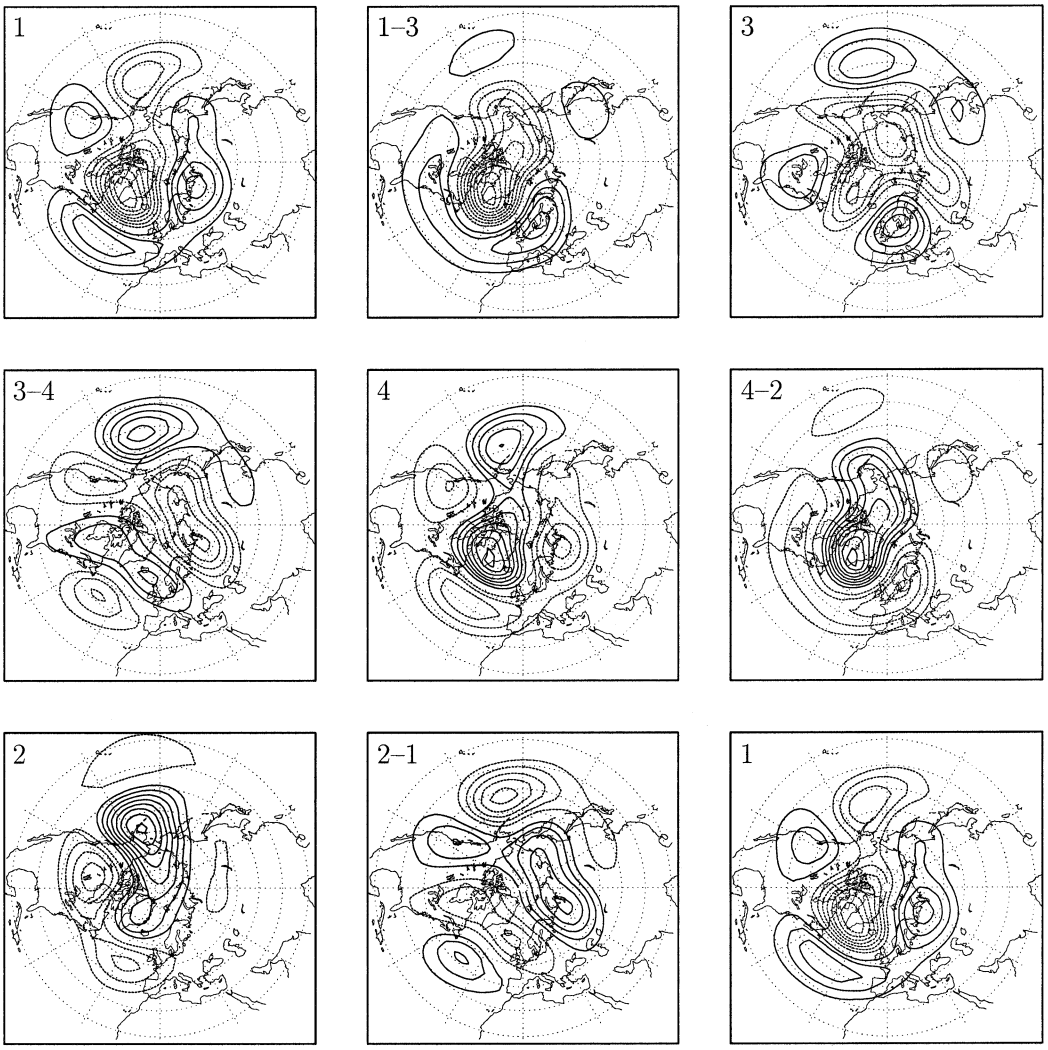

FIG. 8. Evolution of the atmospheric flow (500-mb geopotential height anomaly) along the preferred path (partition 4) through the reduced phase-space $\mathcal{P}$ (see Fig. 6). Contour interval is 15 m.

sionally, completes one or more full cycles. At each step, the chances of evolving into another direction than along the cycle are high (due to high diffusivity); therefore, the chances of completing several steps along the cycle are very small. This is even the case in the second example of section 4, where a preferred cycle dominates the behavior: the chance of completing the cycle in four steps is only 0.24.

6. Conclusions

It has been known for some time that the distribution of observational atmospheric flow data, in a reduced phase space spanned by the leading EOFs, tends to show preferred phase-space regions. Such regions, or regimes, can be hard to make visible. In this study, the projection of the data onto the unit sphere in the space of the leading (normalized) three PCs resulted in a PDF in which regions of increased probability were clearly visible. The atmospheric circulation regimes associated with these regions correspond with earlier found flow regimes. All the regimes presented in previous studies such as CW, KG1, SIG, CMP, and MPF are found in the spherical PDF.

Since a probability distribution does not provide information on the dynamics of the system under study, the reduced phase space was divided into cells, all with an equal number of data points. The transition probability matrix **M**, describing (probabilities of) transitions between cells, reflects the effect of the dynamical processes governing the system behavior, in particular the regime transitions. Contributions to **M** come from both diffusive, irreversible dynamics (e.g., noise-driven motion in a potential field) and conservative, reversible dynamics (generated by, e.g., periodic orbits or tori in phase space). The diffusive dynamics add to the symmetric part $\mathbf{M}^S$ of **M**, whereas the antisymmetric part $\mathbf{M}^A$ is in principle generated by the conservative dynamics.

A statistical significance test of the elements of $\mathbf{M}^A$ (to select elements unlikely to be due to finite sample size) left us with a set of transition asymmetries forming a closed path or cycle over the unit sphere in the reduced phase space.

No pertinent conclusions can be drawn about the dynamical origin of the preferred cycle on the basis of the results presented here. Concepts from dynamical systems theory, such as unstable periodic orbits or heteroclinic cycles, could provide an explanation. But noise, stemming form the small, turbulent scales of motion, can also cause structured behavior of the large scales, as was shown by Majda et al. (1999, 2003). Unambiguously separating ''noise'' from ''structure,'' if possible at all, probably needs a more sophisticated analysis than was used here. An approach that goes beyond the limitations of the Markov chain modeling may provide better possibilities in this respect. The Markovian assumption, justified in this study by the use of 7-day means leading to largely uncorrelated consecutive data points, restricts the kind of information that can be extracted from the data.

Interpreted physically, the evolution of the atmospheric flow pattern along the preferred path shows an alternation between anomaly states with considerable zonal symmetry and states with mainly meridionally oriented anomalies. In particular the preferred cycle of partition 4 shows that positive and negative phases of the NAO are followed (or rather *concluded*) by positive and negative phases of the AO, respectively. It resembles, in this respect, results from an earlier study by the author (Crommelin 2003b), obtained using a barotropic model of NH flow.

Acknowledgments. Stimulating discussions with R. A. Pasmanter, as well as valuable comments from J. D. Opsteegh, F. Verhulst, and F. M. Selten, have helped in the investigation and preparation of this paper. Comments and criticisms by Michael Ghil and by three anonymous reviewers helped to improve the quality of the paper significantly. Financial support for this study came from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.,* **50,** 1205–1216.

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

Crommelin, D. T., 2003a: *Nonlinear dynamics of atmospheric regime transitions.* Ph.D. thesis, Utrecht University, 127 pp.

——, 2003b: Regime transitions and heteroclinic connections in a barotropic atmosphere. *J. Atmos. Sci.,* **60,** 229–246.

Egger, J., 2001: Master equation for climatic parameter sets. *Climate Dyn.,* **18,** 169–177.

Feller, W., 1968: *An Introduction to Probability Theory and Its Applications.* Vol. 1, 3d ed., John Wiley and Sons, 509 pp.

Hannachi, A., and A. O'Neill, 2001: Atmospheric multiple equilibria and non-Gaussian behavior in model simulations. *Quart. J. Roy. Meteor. Soc.,* **127,** 939–958.

Horel, J. D., 1985: Persistence of the 500 mb height field during Northern Hemisphere winter. *Mon. Wea. Rev.,* **113,** 2030–2042.

Kimoto, M., and M. Ghil, 1993a: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.,* **50,** 2625–2643.

——, and ——, 1993b: Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions. *J. Atmos. Sci.,* **50,** 2645–2673.

Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 1999: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA,* **96,** 14 687–14 691.

——, ——, and ——, 2003: Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.,* **60,** 1705–1722.

Mo, K. C., and M. Ghil, 1987: Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.,* **44,** 877–901.

——, and ——, 1988: Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.,* **93,** 10 927–10 952.

Molteni, F., S. Tibaldi, and T. N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Quart. J. Roy. Meteor. Soc.,* **116,** 31–67.

Monahan, A. H., L. Pandolfo, and J. C. Fyfe, 2001: The preferred structure of variability of the Northern Hemisphere atmospheric circulation. *Geophys. Res. Lett.,* **28,** 1019–1022.

Nitsche, G., and J. M. Wallace, 1994: Is there evidence of multiple equilibria in planetary wave amplitude statistics? *J. Atmos. Sci.,* **51,** 314–322.

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

Pasmanter, R. A., and A. Timmermann, 2003: Cyclic Markov chains with an application to an intermediate ENSO model. *Nonlinear Proc. Geophys.,* **10,** 197–210.

Selten, F. M., and G. Branstator, 2004: Preferred regime transition routes and evidence for an unstable periodic orbit in a baroclinic model. *J. Atmos. Sci.,* **61,** 2267–2282.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

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

Spekat, A., B. Heller-Schulze, and M. Lutz, 1983: Über Grosswetter und Markov Ketten. *Meteor. Rundsch.,* **36,** 243–248.

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.

Toth, Z., 1991: Circulation patterns in phase space: A multinormal distribution? *Mon. Wea. Rev.,* **119,** 1501–1511.

Wallace, J. M., X. Cheng, and D. Sun, 1991: Does low-frequency atmospheric variability exhibit regime-like behavior? *Tellus,* **43AB,** 16–26.