Multiple Flow Regimes in the Northern Hemisphere Winter.
Part I: Methodology and Hemispheric Regimes

MASAHIDE KIMOTO
Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

MICHAEL GHIL
Climate Dynamics Center, Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California

(Manuscript received 4 April 1991, in final form 8 October 1992)

ABSTRACT

Recurrent and persistent flow patterns are identified by examining multivariate probability density functions (PDFs) in the phase space of large-scale atmospheric motions. This idea is pursued systematically here in the hope of clarifying the extent to which intraseasonal variability can be described and understood in terms of multiple flow regimes.

Bivariate PDFs of the Northern Hemisphere (NH) wintertime anomaly heights at 700 mb are examined in the present paper, using a 37-year dataset. The two-dimensional phase plane is defined by the two leading empirical orthogonal functions (EOFs) of the anomaly fields. PDFs on this plane exhibit synoptically intriguing and statistically significant inhomogeneities on the periphery of the distribution. It is shown that these inhomogeneities are due to the existence of persistent and recurrent anomaly patterns, well-known as dominant teleconnection patterns; that is, the Pacific/North American (PNA) pattern, its reverse, and zonal and blocked phases of the North Atlantic Oscillation (NAO). It is argued that the inhomogeneities are obscured when PDFs are examined in a smaller-dimensional subspace than dynamically desired.

1. Introduction and motivation

It is central in operational long-range forecasting (LRF) to understand the variability of the atmosphere beyond the time scale of daily weather disturbances. Considerable efforts have been devoted to investigating the nature of low-frequency variability (LFV), synoptically as well as theoretically.

Early in the years of upper-air observations, there already emerged a notion of Grosswetterlagen (Bauer 1951), pointing out that there are certain large-scale flow patterns that appear repeatedly at fixed geographic locations and persist beyond the life cycle of individual weather disturbances. Blocking and strong zonal flows were given as typical examples of such Grosswetterlagen.

A series of recent observational studies have described temporal as well as spatial characteristics of extratropical LFV in detail, using well-archived upper-air datasets. Wallace and Gutzler (1981; hereafter WG) examined one-point correlation maps of monthly mean heights at 500 mb. They found several types of coherent variations between widely separated regions on the globe. Following Bjerknes (1969), these are called teleconnection patterns. Most of them are now well known as dominant, geographically fixed modes of LFV.

Analyses with more emphasis on temporal characteristics (e.g., Dole and Gordon 1983, hereafter DG; Horel 1985; Mo 1986) of LFV focused on persistent and quasi-stationary flow configurations, generally called persistent anomalies (DG). These are characterized by flow patterns that differ significantly from the “normal” climatological circulation and remain stationary for more than a week. Their onsets and breakdowns are rather abrupt compared with the quasistationary periods. No preferred duration is found for these persistent events (DG).

Synoptic experience over 30 years (Namias 1975, 1982) has confirmed that the persistent flow configurations appear repeatedly in different months or years. Notwithstanding, attempts to find individual analogs, that is, weather maps that closely resemble each other, and hence might lead to particularly predictable sequences of maps, have only met with very limited success (Lorenz 1969; Gutzler and Shukla 1984).
To summarize, low-frequency atmospheric circulation patterns have three characteristics: (i) geographically fixed appearance, (ii) temporal persistence, and (iii) recurrence. These characteristics are illustrated by Figs. 1 and 2, which plot a particular contour of 700-mb height fields for successive ten-day, nonoverlapping intervals. The 2940-m contour is chosen to represent the jet-stream axis. This method of presenting the data is called a limited contour analysis; it was pioneered by Sanders and Gyakum (1980) and used by Dole (1982) and by Reinhold (1981, 1987) for LFV studies.

Figure 1 shows nine nonoverlapping ten-day intervals for the 1978/79 winter. As noted by Dole and by Reinhold, the large-scale flow remains quasi-stationary for certain time intervals. A good exam-

---

Fig. 1. Limited contour analysis of Northern Hemisphere (NH) flows. Daily contours of a prescribed height (2940 m in this case—roughly corresponding to the jet axis) are superimposed for successive 10-day intervals during NH winter 1978/79. Persistence is illustrated by some of the panels (see text for details).
ple is the set of 2940-m contours for 31 December to 9 January (panel at the top of the central column in Fig. 1), when the jet stream was displaced considerably to the north over Alaska, bringing anomalous warmth over the region. For this time interval, the contour on successive days did only undergo very small fluctuations over the entire hemisphere. Thus, the hemispheric circulation remained very quiescent during this interval. In other panels, we also see signs of similar quiescence but more regionally confined (notice, for example, the East Asian–Pacific sector in the panel for 20 January–29 January at bottom center of Fig. 1). Except during these quiescent intervals, the shape of the particular contour changes considerably from one day to another. We thus note that typical low-frequency circulation patterns are not present all the time. This is one of the major difficulties in studying persistent and recurrent flow patterns by using the limited number of sample maps at our disposal.

Temporal recurrence is illustrated in Fig. 2, which shows two limited contour analyses 13 years apart. The resemblance of the two sets of contours is striking, though a perfect repetition cannot be expected in this geophysical flow problem (Lorenz 1963; Ghil and Childress 1987, chapters 5 and 6).

With these observational facts in mind, some theoretical approaches to LFV, other than that of a pure initial value problem, that is, of extended numerical weather prediction, are also becoming available. Among these, Hoskins and Karoly (1981) shed considerable light on the dynamical mechanisms of tropical influence upon midlatitude circulation anomalies. They helped understand, for example, the influence of large sea surface temperature (SST) anomalies, as in El Niño years, on the extratropics. However, tropical influences on interannual time scales are not likely to explain most of the extratropical variability captured in Fig. 1. Indeed, tropical SST anomalies last for an entire season or more, while the typical persistencies apparent in Fig. 1 are of considerably shorter duration. Moreover, Namias and Cayan (1984) showed that El Niño winter anomalies of surface temperature and precipitation over the United States differ from each other by as much as from any non–El Niño winter.

Processes internal to the extratropics may play a more important role. Among “internal” theories of extratropical LFV, studies based on simplified models highlighted nonlinear aspects of atmospheric motions, presenting a very distinct view from the initial value or linear instability approaches. Charney and DeVore (1979) introduced the concept of multiple equilibria into atmospheric LFV studies. They related two stable stationary solutions of their barotropic model with observed blocking and zonal flows. Subsequent studies by Reinhold and Pierrehumbert (1982), Legras and Ghil (1985, hereafter LG), Vautard et al. (1988a,b), and Mukougawa (1988) emphasized more subtle in-

Fig. 2. An example of recurrent flow patterns: limited contour maps of (a) 20–29 January 1961, and (b) 31 December 1978–9 January 1979.

homogeneities in their model’s phase-space structure: they exhibited multiple regions in phase space in which chaotic model trajectories stay for a long time to form model analogues of persistent anomalies under fixed external conditions. These model anomalies had no preferred duration, as observed in the atmosphere by DG.
The aforementioned theoretical studies have ushered in the possibility of describing atmospheric LFV in terms of nonlinear dynamical systems theory. The ergodic theory of nonlinear dynamical systems (Eckmann and Ruelle 1985) tells us, in general, that chaotic systems may possess attractors much "smaller" than the phase space of the whole system. It is natural to suspect that long-lived circulation patterns, which appear repeatedly and, thus, seem to be preferred by the atmosphere, may correspond to those regions in phase space where the probability of occurrence of observed atmospheric states attains local maxima (Ghil 1987, 1988). Figure 3 illustrates our conceptual model for the way the data may be organized in phase space.

A weather map is represented as a point in phase space and the temporal evolution of the system as the trajectory of this point (Fig. 3a). When two or more independent occurrences of slow-down along this trajectory take place in a certain part of phase space, the flow pattern corresponding to such a region is recurrent and persistent. With all the available sample weather maps plotted in phase space, recurrent and persistent flow patterns give rise to local maxima in density (Fig. 3b). Preferred transition routes between such regions of phase space, called regimes, may be detected, one hopes, given a sufficient number of sample maps.

In this paper and its companion (Kimoto and Ghil 1993), we investigate the possibility that typical long-lived circulation patterns or Grosswetterlagen can be identified by systematically examining the probability density distribution in the atmosphere's phase space. Specifically, we use a 37-year dataset of Northern Hemisphere (NH) 700-mb heights and examine multivariate probability density functions (PDFs) in phase subspaces spanned by a few leading empirical orthogonal functions (EOFs) of the dataset. Throughout the study, we concentrate on wintertime anomalies since this is the season in which LFV is most active and, therefore, has been well documented.

Benzi et al. (1986) pioneered the use of univariate PDFs in investigating atmospheric LFV. They showed the first evidence of bimodality in an amplitude index of NH wintertime planetary waves. Several observational attempts to objectively classify spatial patterns into atmospheric regimes have been reported recently; these have used pattern correlations (Mo and Ghil 1987), cluster analysis (Mo and Ghil 1988; Legras et al. 1988), or minimization of phase-space velocity (Vautard 1990). Molteni et al. (1990) have used multivariate PDFs obtained with a different technique and a different dataset than Kimoto (1987, 1989), whom we follow here. As will be discussed later, it is difficult to obtain statistically reliable classifications with a limited sample size. Comparisons among different methodologies and datasets are, therefore, desirable and necessary.

In Part I of this study (the present paper), we describe and discuss the statistical methodology and investigate anomalous flow patterns of hemispheric extent. Despite the difficulty in obtaining reliable estimates of multivariate PDFs with a limited amount of data, we are able to identify interesting inhomogeneities in such PDFs that correspond to well-known circulation patterns. Part II of the study (Kimoto and Ghil 1993) is devoted to a more detailed description of the multiple flow regimes obtained, with particular emphasis on regional, recurrent flow patterns and the transitions among them.

Section 2 of the present paper describes the 700-mb dataset. Statistical methodologies are introduced in section 3; two major tools in this study, that is, EOFs and PDFs, are briefly described. Section 4 discusses temporal and spatial characteristics of the NH EOFs. In section 5, two-dimensional (2D) PDFs are investigated; subtle, but definite inhomogeneities of this 2D distribution are visually presented. In the same section, we discuss circulation patterns of the flow regimes corresponding to the relative PDF.

In Part II of the study (Kimoto and Ghil 1993), a classification of large-scale anomaly patterns in two NH sectors, that is, PAC (from 120°E eastward to 60°W) and ATL (from 60°W eastward to 120°E), is presented. Section 2 of Part II recapitulates briefly the data used. A classification algorithm is described in section 3; it is based on multivariate PDFs as in Part I, but tailored to the meteorological problem at hand by only retaining information on patterns of the anomaly maps, not on their magnitudes. Local maxima are sought by using a simple algorithm called bump hunting. The classification results are presented in section 4, in which interrelations between the two sectors are also addressed. A Markov-chain description of transitions between the sectorial flow regimes is given in section 5, while in section 6 we examine synoptically the onsets and breaks of the flow regimes by using a compositing technique. A summary of the entire study and concluding remarks follow in section 7 of Part II.

2. Data

The dataset used in this study consists of twice-daily observations of the NH 700-mb heights compiled at NOAA's Climate Analysis Center. The geopotential heights are given on a 10°×10° diamond grid north of 20°N. We use a subset of grid points selected by Barnston and Livezey (1987; BL hereafter) from the original grid by thinning out some points north of 60°. The total number of grid points used is 358. The period covered by the dataset extends over 37 winters starting from November 1949 to March 1986. This is one of the longest and most complete upper-air datasets available.

Northern Hemisphere winter is defined here as a 90-day period starting on 1 December, and a few days at the end of November and beginning of March are used for the averaging and filtering described in the next two sentences. The seasonal cycle, averaged over 37 years and subject to a weak, 5-day running average, is removed from the daily gridpoint heights to define the height anomalies; no attempt has been made to remove interannual signals. In the EOF analysis to be presented below, we apply a 10-day low-pass filter (Blackmon 1976) to the anomalies. These are referred to as low-pass (filtered) anomalies. Samples outside the 90-day period were used in this filtering so that those within the period were appropriately filtered. Anomalies without the 10-day filtering are called unfiltered. The daily samples we use in the following are those observed at 0000 UTC.

Although the data contain 90×37 = 3330 maps, these are not statistically independent of each other. Both gridpoint and hemispheric autocorrelations (not shown, but see Madden 1976; Gutzler and Mo 1983; Horel 1985) indicate the number of independent realizations to lie between 300 and 700 for both filtered and unfiltered anomalies.

3. Statistical methodology

In this section, we describe the two major statistical tools used in this study: EOFs and PDFs.

a. Empirical orthogonal functions (EOFs)

In order to examine the phase-space structure of atmospheric motions in as many as 358 dimensions, where 358 is the number of grid points, it is both convenient and necessary to reduce the dimensionality of the data. For this purpose, we first apply a linear decomposition technique called EOF analysis, also known as principal component (PC) analysis in statistics. EOF analysis has been used rather extensively in meteorology, so only a brief description suffices.

Given the height anomaly Z as a function of both space (λ, ϕ) and time (t), Z(λ, ϕ, t), we decompose it into N (= 358) linearly independent modes:

$$Z(\lambda, \phi, t) = \sum_{i=1}^{N} f_i(t)e_i(\lambda, \phi).$$  \(1\)

Here λ and ϕ are longitude and latitude and $f_i(t)$ is the PC or time coefficient at time $t$ of the corresponding spatial mode $e_i(\lambda, \phi)$, called EOF. By construction, \{f_i\} and \{e_i\} satisfy temporal and spatial orthogonalities, respectively, among the modes. In this study, we use the convention that the spatial patterns $e_i$ are scaled to have unit mean-square norm, the mean being the area integral over the NH data domain, divided by the area. Therefore, the domain-integrated variance associated with each mode, $\sigma_i^2$, is carried by the time coefficient, $f_i$; that is, the temporal standard deviation of $f_i(t)$ is $\sigma_i$. The $\sigma_i$ are the eigenvalues of the data covariance matrix $V$. They are real, positive, and, by convention, in a sequence of decreasing order in magnitude. The EOFs are the corresponding eigenvectors of $V$.

EOF analysis is a useful statistical technique for the study of multivariate time series with a large number of variables, like those arising from the observation of geofluids. Relations between dynamical systems and the EOFs derived from their data are discussed by North (1984) for linear systems and by Mo and Ghil (1987) for nonlinear ones. The latter authors argue that the leading EOFs point from the time mean to the most populated regions of the system’s phase space.
b. Multivariate probability density functions (PDFs)

A further step beyond linear decomposition is to investigate probability density distribution in phase space; in other words, to examine the shape of the attractor. A simple, nonparametric density estimation technique called the kernel method (Silverman 1986; hereafter referred to as S86) will be described in this subsection. Appendix A describes how a parameter that controls the smoothness of the density estimation can be chosen.

The most rudimentary way to estimate a PDF is to construct sample histograms. But this becomes impractical and unreliable in more than three dimensions. Kernel density estimation is not as computationally expensive as, but is also less sophisticated than, the penalized likelihood approach (Good and Gaskins 1980). Unfortunately, the latter becomes prohibitively expensive and hence impractical in dimensions larger than two.

The basic version of kernel density estimation (S86) assigns a localized kernel density function of given shape and height to each data point; the sum of these continuous functions yields an estimation of PDF at any point in phase space (see Fig. 4). The smoothness of the estimated PDF in hyperspace is controlled by a parameter $h$, which determines the radius of influence of each sample point. In an adaptive version (S86), this smoothness parameter varies according to the value of the resulting PDF at each sample point in order to achieve more reliable estimates in data-sparse regions.

Let $\hat{f}(x)$ be the estimated PDF as a function of the $r$-dimensional vector $x$, which represents the first $r$ PCs. The adaptive algorithm is given in three steps.

Step (i). Construct a pilot estimate $\hat{f}_p(x)$

$$\hat{f}_p(x) = \frac{1}{C} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h_p}\right),$$

(2)

where $X_i$ is the vector designating the $i$th sample point, $h_p$ is a smoothing parameter for the pilot estimation, $N$ is the sample size, and $C$ is a normalization constant to ensure that $\int \hat{f}_p(x)dx = 1$; $K(x)$ is the kernel function that we choose to be the so-called Epanechnikov kernel,

$$K(x) = \begin{cases} 1 - x^t x, & \text{if } x^t x < 1, \\ 0, & \text{if } x^t x \geq 1, \end{cases}$$

(3)

where the superscript $t$ denotes transpose.

Step (ii). Determine a local bandwidth $\eta_i$ for each sample point $X_i$ by

$$\eta_i = \left\{ \frac{\hat{f}_p(X_i)}{G} \right\}^{-\alpha},$$

(4)

where $\log G = (1/N) \sum_{i=1}^{N} \log \hat{f}_p(X_i)$ and $\alpha = 0.5$ is a sensitivity parameter. We fix $\alpha$ and $h_p$ since the final estimate is reported to be insensitive to the pilot estimate (S86). We use $h_p$ given by a formula based on multivariate normal distributions (S86, pp. 86–87).

Step (iii). Compute the final estimate

$$\hat{f}(x) = \frac{1}{C} \sum_{i=1}^{N} \eta_i^{-r} K\left(\frac{x - X_i}{h\eta_i}\right),$$

(5)

Fig. 4. An illustration of kernel density estimation. Thin lines represent local kernel density functions assigned to individual samples denoted by crosses. The three panels show how the window width of the smoothing parameter affect, for the same sample, the smoothness of the resulting density estimates, denoted by thick solid lines (after Silverman 1986). Window widths $h$ are (a) 0.2, (b) 0.4, and (c) 0.6.
where $h$ is the final smoothing parameter and $C$ is the normalization constant. The details of the determination of $h$, by least-squares cross validation (LSCV), are given in appendix A.

It will be seen in the following sections that the automatic LSCV procedure of the appendix only suggests a certain interval within which reasonable estimates should be made, rather than pinpointing an exact value; that is, the minimum of the LSCV score function $M_0(h)$, with respect to $h$, is rather flat. As noted by S86, examinations of several values of $h$ are desired in practice. We thus vary the parameter $h$ over the range suggested by LSCV. We found that the range is not very different from the one obtained with less elaborate methods or simply by subjective inference. Since LSCV requires expensive computations, that is, integrations over the whole phase space, we feel that subjective choice and a few trials should suffice in most applications.

4. Spectral characteristics of NH EOFs

In this section, we briefly discuss the results of the EOF analysis for our datasets. EOFs of the zonally asymmetric component of 500-mb height fields have recently been reported by Molteni et al. (1988) with a sample size of 40 years, which is comparable to ours.

Figure 5 presents percentage variances associated with NH winter EOFs. The NH EOFs are obtained by using 10-day low-pass filtered anomalies. Unfiltered anomalies give essentially the same spectra, but with slightly reduced variances due to increased high-frequency contributions. The spatial patterns of the EOFs are insensitive to the filtering. We took into account the areal factor proportional to cosine of latitude by multiplying each gridpoint anomaly by the square root of $\cos \varphi$, where $\varphi$ is the latitude. Figure 5 shows eigenvalues $\sigma_i^2$, which are scaled to represent percentages of spatially integrated variance that they accompany. Attached error bars are estimated by using a rule of thumb derived by North et al. (1982) and correspond to 68% confidence intervals. The rule is applied with an effectively independent sample size of 300, a very conservative estimate. The eigenvalues decrease slowly from the first to higher modes. No single mode is dominant, reflecting the complexity of LFV. However, there are several "gaps" in the spectrum, as judged by the separation between the error bars of neighboring eigenvalues. They are seen at the second, fifth, eighth, and eleventh mode; beyond the latter, the associated variance for individual modes becomes negligibly small.

The spatial patterns of the two leading EOFs of our dataset are shown in Fig. 6. Magnitudes of the anomalies in Fig. 6 are those attained when the corresponding PCs are equal to one standard deviation. The first mode, called EOF1, is shown in Fig. 6a and is characterized by two dominant centers of opposite sign (signs of the anomalies are immaterial in EOF analysis) over the North Pacific around 160°W in the Gulf of Alaska and over the North Atlantic around 40°W at the southern tip of Greenland. To the southwest and to the east of the Atlantic center, there exist weaker anomaly centers contributing together to the wave train–like appearance in the Atlantic–to-Eurasian sector. WG's second EOF of 500-mb monthly mean height anomalies bears a strong resemblance to this pattern. Among the EOFs presented by Schubert (1985) and by Molteni et al. (1988), analogous patterns can be found. Thus, our EOF1 appears to be a fairly robust mode of hemispheric LFV. The pattern can be viewed as a linear combination of well-known teleconnection patterns such as the Pacific/North American pattern (e.g., WG and BL, among others), the North Atlantic Oscillation (van Loon and Rogers 1978; BL), west Atlantic pattern (WG), and the Eurasian pattern (WG; BL).

EOF2, shown in Fig. 6b, is also characterized by two dominant centers, of equal sign in this case, over slightly shifted positions in the North Pacific and in the North Atlantic. This time, the Atlantic feature is more appropriately described as a "north–south dipole." In contrast to EOF1, it is difficult to find exact counterparts of this pattern in other studies, although it is possible to detect traces of the features seen in Fig. 6b. No particular significance is attached in our study to individual EOFs, only to multivariate PDFs in subspaces of leading EOFs.
are plotted in Fig. 7a, versus the natural logarithm of the frequency. This way of presenting power spectra allows better resolution at lower frequencies than the conventional, frequency-versus-power plot; note that integration of the plotted quantity with respect to the logarithm of frequency still gives the power contained in that frequency band as in the conventional plot. Figure 7b shows, for comparison, spectra of "red noise," that is, a first-order Markov process, using the autocorrelations of each PC at the lag of 1 day. The temporal coefficients used in Fig. 7 are obtained by projecting daily, unfiltered anomalies onto the low-frequency EOFs, thus examining daily changes in the low-frequency spatial modes. The winter is extended to 110 days by adding 10 days before and after the 90-day period to obtain better resolution at intramonthly frequencies.

In Fig. 7a, we note that the NH EOFs have significant variability in the so-called intraseasonal (IS) range, that is, from 10 days to 90 days. Much stronger redness is shown by the lag-one Markov processes in Fig. 7b. In Fig. 7a and Fig. 7b, we indicated the percentage ratios between interannual (IA) and IS variances, the former being the variance of 110-day means and the latter being that of the deviations from them. Our low-frequency EOFs are clearly dominated by intraseasonal rather than interannual variability.

We also see that most of the temporal variance in the IS range is contained in the first 8 EOFs. As is the case of eigenfunctions in most physical systems, the spatial scales of EOF patterns become smaller and noisier for higher modes. From Figs. 5 and 7, we conclude that the first 8, or at most 11, EOFs contain the signals in the LFV range. The rest might be needed to represent instantaneous flow patterns, but their details cannot be studied by our coarse-grained, statistical approach.

5. Two-dimensional (2D) PDFs and NH flow regimes

In this section, we investigate the 2D probability distribution of NH flows and examine the evidence for inhomogeneities. As discussed in the Introduction, such inhomogeneities in the PDF are suspected to arise from the existence of preferred flow configurations with distinct persistence and recurrence properties.

For an infinite sample of the dynamical system under statistical consideration, one should be able to compute the PDF for an arbitrary number of linearly independent directions in phase space, that is, of EOFs. Beyond a certain number of EOFs, the main features of the PDF should stabilize, providing all the desired statistical information.
In practice, we are given a finite, and rather small, sample of a trajectory of the system. Hence, a compromise has to be reached between the resolution of the inhomogeneities and their statistical significance as the number of EOFs retained increases.

One-dimensional (1D) PDFs or histograms of meteorological variables have been investigated in numerous studies. In the context of LFV of the extratropical atmosphere, White (1980) and DG examined 1D histograms of geopotential heights at individual grid points. They found statistically significant departures from normality, but failed to confirm multimodality. Recently, however, Benzi et al. (1986) and Molteni et al. (1988) found evidence of bimodality in one-dimensional PDFs of an amplitude index of planetary waves. Mo and Ghil (1988) also found evidence of bimodality in the leading PC of their NH 500-mb data, although multiple clusters—with quite different flow patterns—contributed to each mode.

As discussed by S86 (p. 94), the number of (independent) samples required in order to make accurate and robust estimates increases rapidly with dimension. For example, S86 estimates required (independent) sample size as 19, 223, and 2790 for 2, 4, and 6 dimensions, respectively, in order to have a 10% accuracy at the origin when the PDF is multivariate Gaussian. His estimate increases rapidly with dimension and is more than 40,000 for dimension 8. Thus, for the sample size of our dataset, reliable estimates are not feasible for more than four dimensions. On the other hand, the spatial and temporal characteristics of LFV seem to require investigations of eight or more spatial degrees of freedom, as we saw in the last section.

a. 2D PDFs

For ease of visual presentation, on the one hand, and for statistical reliability, on the other, we focus in this paper (Part I) on 2D PDFs in the subspace of the first two NH EOFs. This choice is further justified by the fact that there is a statistical separation between the variances associated with NH EOF 2 and 3, as seen in Fig. 5. The daily PCs were obtained by projecting unfiltered anomalies onto the low-frequency EOFs, as in the preceding section.

To reiterate, the general dilemma of small datasets, as discussed above, takes the following concrete form: the dimension of the subspace containing most of the significant LFV of NH circulation is eight, while the sample size available permits reliable identification of inhomogeneities for dimensions no larger than four. Under these circumstances, it is expected that a few leading clusters, if any, in eight dimensions may be obscured by the presence of numerous insignificant ones, which project near the origin of the small-dimensional subspace under investigation. This is illustrated for our dataset in Fig. 8, which shows the data scatter in terms of two dimensional (2D) PDFs. The abscissa and ordinate are the root-mean-
As the abscissa of Fig. 8a, we take

$$\|x\|_S = \left(\sum_{i=1}^{8} f_i^2\right)^{1/2},$$

(6a)

and as the ordinate

$$\|x\|_N = \left(\sum_{i=9}^{N} f_i^2\right)^{1/2},$$

(6b)

where $N$ is the total number of EOFs for our dataset; $N = 358$ [cf. Eq. (1)]. Thus, Fig. 8a compares the ratio between the physical “signal” (EOF modes $\leq 8$) and the “noise” (the rest). Samples falling in the lower-right octant of a line $\|x\|_S = \|x\|_N$ have signal-to-noise ($S/N$) ratio greater than unity. By visual inspection, roughly half the data have $S/N$ ratio greater than 1 when the first eight EOFs are taken to be the signal. Furthermore, we note that virtually no sample falls near the origin, that is, the climatological mean, which is actually outside the domain shown in Fig. 8a. This corresponds to the well-known synoptic fact that no instantaneous weather map resembles the climatological mean.

When examining a 2D subspace, rather than the desired eight dimensions, those samples that have small projections onto the subspace of the two leading EOFs but large projections onto the following six EOFs are relatively numerous. In Fig. 8b, the abscissa is $\|x\|$, with components projected onto EOFs 1 and 2 and the ordinate onto EOFs from 3 to 8. We can see that the data are again absent from the neighborhood of the origin, but many lie near the origin of the 2D space, that is, near the vertical axis in Fig. 8b. This observation, which appears almost trivial after the fact, should be taken into account in investigating a subspace with a lower dimension than is dynamically required (see also Fig. 9 in Mo and Ghil 1988). Presumably, it is this discrepancy between statistically practicable and physically desirable dimensions that has hindered one from detecting clear multimodality in most meteorological datasets. Note that the amplitude index of planetary waves used by Benzi et al. (1986) represents radial distributions of the data in a phase space whose origin corresponds to zonal flows. Since the phase of the spatial patterns of the weather maps, that is, solid-angular information in phase space, was discarded, it is difficult to see to which circulation pattern one can ascribe the bimodality in their case.

Based on the above arguments, we examine carefully inhomogeneities on the periphery of our 2D PDFs, rather than near the origin of the 2D space spanned by NH EOFs 1 and 2. For this purpose, we use several data subsets. For the same reason, and following also Mo and Ghil (1988) in their heuristic emphasis on solid-angle criteria for clustering, we shall

square (rms) magnitude of anomalies, $\|x\|$, computed by subsets of EOF modes, as explained in the caption and in the following. All the values are scaled by the standard deviation of PC1, $\sigma_1$ [cf. Eq. (1)]. The dimensional value of $\sigma_1$ is 25 m.
taper usual Euclidean PDFs to what we call angular PDFs, which neglect magnitude of anomalies, when classifying recurrent meteorological patterns in Part II of this study.

We call the first subset the SN subset; it consists of 1884 samples that have S/N ratio larger than unity. The signal is measured, as in Fig. 8a, by the eight leading EOFs. Thus, the SN data are those in the lower-right octant of Fig. 8a.

The second data subset consists of samples that belong to quasi-stationary (QS) events identified by the same procedures as in Horel (1985), Mo (1986), and Mo and Ghil (1987); that is, QS events are characterized by all the pairs of anomaly maps in a sequence longer than \( \tau \) days having pattern correlations greater than \( p_0 \). The pattern correlation \( p(Z_1, Z_2) \) between two anomaly maps \( Z_1(\lambda, \varphi) = Z(\lambda, \varphi, t_1) \) and \( Z_2(\lambda, \varphi) \) is defined as

\[
p(Z_1, Z_2) = \frac{\langle Z_1, Z_2 \rangle}{\langle Z_1, Z_1 \rangle \cdot \langle Z_2, Z_2 \rangle}^{1/2},
\]

where the inner product \( \langle Z_i, Z_j \rangle \) is defined as

\[
\langle Z_i, Z_j \rangle = \int Z_i Z_j \cos \varphi d\lambda d\varphi;
\]

\( p \) is the cosine of the angular distance in phase space (cf. Fig. 3a) and measures the similarity of shape between patterns. Thus, the QS dataset contains time intervals with enhanced pattern stationarity. Since the pattern correlation (7) measures only angular distance between two samples, it tends to pick up data with large radial distances from the origin (Vautard et al. 1988a). We used unfiltered daily anomalies (without EOF filtering) in computing \( p \), and the criteria \( \tau \) and \( p_0 \) were set equal to 5 days and 0.5, respectively (cf. Mo and Ghil (1987)).

We use three different QS datasets: H-QS data include 729 days identified by taking pattern correlations over the whole NH north of 20°N (H stands for hemispheric). In P- and A-QS data (Pacific and Atlantic, respectively), pattern correlations are computed in two complementary quadrants in the NH: from 120°E eastward to 60°W in the P-QS, and from 60°W eastward to 120°E in the A-QS dataset. The P- and A-QS datasets include 837 and 1216 days, respectively. Motivation for the use of these two regional QS datasets will be given in subsection 5b.

In order to estimate the appropriate range for the smoothing parameter \( h \), we compute the LSCV score \( M_0 \) given by Eq. (A.4), as shown in Fig. 9. Figure 9a shows \( M_0 \) computed using all samples, while Fig. 9b presents those computed with subsamples. The minima in the curves are rather flat, especially for the subsampled datasets. Since the use of smaller \( h \) may overemphasize insignificant details, we use a constant, rather conservative value, 0.9, throughout this section, except for Fig. 10c, where a result with \( h = 0.5 \) is presented for comparison. The smaller \( h \), the more jagged the resulting density estimates become. We only discuss inhomogeneities that appear even with a very conservative value of \( h \).

Two-dimensional PDFs of the NH data (subsets) are shown in Figs. 10 and 11. Figure 10a uses all the samples and a smoothing parameter of \( h = 0.9 \). It does not show multiple peaks, which is also the case for smaller values of \( h \) (Fig. 10c with \( h = 0.5 \)). However, indications of PDF inhomogeneity, in other words deviation from bivariate Gaussianity, appear as elongated ridges, marked by heavy dashed lines in Fig. 10a. This feature does not disappear even with values of \( h \) greater than 1.0. Figures 10b and 10d indicate the statistical significance of the PDFs shown in Figs. 10a and 10c. One hundred time series having the same covariance matrix as \( Z(\lambda, \varphi, t) \) and

![Fig. 9. The least-squares cross-validation (LSCV) scores \( M_0(h) \) using (a) all 3330 samples, and (b) various data subsets: SN, H-, P-, and A-QS. See text for a definition of these subsets.](image-url)
autocorrelations at 1-day lag equal to those of the observed PC1 and 2 were generated, and 2D PDFs were computed for each of them, defining a PDF confidence interval at each point in phase space. In Figs. 10b and 10d, numbers of random sets that have greater PDF values than those observed are plotted. In heavily shaded regions, the observed PDF is significantly greater than the random ones with 90% confidence; that is, 10 or fewer random PDFs were larger than those shown in panels (a) or (c), respectively. Similarly, light shading indicates regions where the observed PDF is significantly smaller. Note that the observed PDFs are significantly greater in regions far away from the origin, along one of four radial directions corresponding to angular maxima.

The inhomogeneities noted for Fig. 10 are highlighted in the PDFs of SN (Fig. 11a) and H-QS (Fig. 11b) data, which focus on the meteorologically intriguing periphery of the density distribution and show interesting multimodality. We label the four regions of salient PDF inhomogeneity as PNA, ZNAO, RNA, and BNAO, respectively, which are indicated by rectangles on the SN–PDF plot (Fig. 11a). These regions correspond to the statistically significant “ridges” in Fig. 10, and are subjectively defined on the 2D plane in order to facilitate further examination. The acronyms are associated with the flow patterns characterizing these four PDF features, which will be shown in Fig. 14 and discussed in the next subsection.

![Figure 10](image-url)

**Fig. 10.** Two-dimensional PDFs on a plane spanned by NH EOFs 1 and 2. Axes are scaled by the standard deviation $\sigma_1$ of PC1; the contour interval is 0.01. (a) 2D PDF using all 3330 sample days in the 37 winters, and with a smoothing parameter $h = 0.9$. (b) Numbers of random PDFs that exceed the value shown in (a). Regions smaller than 10 and larger than 90 are shaded heavily and lightly, indicating that the observed PDF is greater or smaller, respectively, with 90% confidence. Contour interval is 10, with additional, dashed contours of 5 and 95 inside the shadings. No contours are shown where the PDF in (a) is less than 0.005. (c) Same as (a) but with $h = 0.5$ and (c) same as (b) but for significance of (c).
The statistical significance of the 2D PDFs shown in Figs. 10 and 11 is further checked with half-sized data subsets: the first 19, the second 18, 19 odd, and 18 even years. In all the corresponding counterparts of Fig. 10 (not shown), ridges similar to those in Fig. 10a are found, with slight changes in relative eccentricities and statistical significance. Figures 12a–d show the four counterparts of Fig. 11a, that is, of the PDF with the SN data subset. The EOF basis and SN datasets are recomputed within each half-sized subset, 8 EOFs being used to define the “signal,” but the PDFs are plotted onto the same EOF plane as in Figs. 10 and 11 for ease of comparison. The four rectangles of Fig. 11a are duplicated in Fig. 12. Of the four regions marked in Fig. 11a, the PNA inhomogeneity is easily recognized as a local maximum in Fig. 12d and as a ridge in Figs. 12a,b. Slight positional changes are noted for the peak corresponding to ZNAO in panels (b) and (c) of Fig. 12. Ridges, rather than isolated peaks correspond to RNA and BNAO in panels (b) and (d). In summary, we conclude that the inhomogeneity represented by the four regions is significant, although no well-defined multimodality appears in this low-dimensional subspace.

Next, we examine persistence characteristics of the above four regions in 2D phase space. In Fig. 13, phase-space averages of a measure of day-to-day circulation change are shown. The Euclidean distance $\|\delta x\|$ between maps at day $(t + 2)$ and $(t - 2)$ is computed for each sample day $t$ using the first eight EOFs. Weighted averages of the phase-space speed $\|\delta x/\delta t\|$, defined as

$$\mathbf{\langle \|\delta x/\delta t\| \rangle (x) = \frac{1}{\hat{f}(x)} \sum_{i=1}^{N} \eta_i^{-r} K \left( \frac{x - X_i}{\eta_i} \right) \cdot \|\delta x/\delta t\| (X_i) }$$

at points in the 2D plane, are then computed in the figure; $C$, $\eta_i$, $K$, and $\hat{f}$ are those defined in section 3. Here, the dimension $r$ is equal to 2. Relative smallness of $\|\delta x/\delta t\|$ implies slowness of the phase-space trajectory or persistence of the circulation (LG; Ghil and Childress 1987, section 6.4). We see that the PNA and ZNAO regions are characterized by local minima in $\|\delta x/\delta t\|$. Persistence in terms of the Euclidean speed in the other regions is not apparent. Pattern stationarity, that is, persistence in terms of angular speed, as seen in the QS-PDF (Fig. 11b), and recurrence, as exemplified by the relative maxima in Figs. 10a,c and 11a are, however, significant for all four regions.

To examine the existence of preferred transition routes, one can also average the 2D projections of velocity vectors along the system's phase-space trajectory. Plots of such “streamlines” do exhibit inhomogeneities suggestive in part of 2D dynamics, but not with enough statistical significance and, therefore, are not shown.

b. Anomaly patterns and discussion of NH flow regimes

Having seen indications of statistical and dynamical significance for the four NH flow regimes (cf. Fig. 11a) in the reduced phase space of the two leading EOFs, we next examine them in physical space. Figure 14 presents composite maps of unfiltered anomalies of the NH regimes. Samples falling within the
Fig. 12. Sensitivity to sampling of the 2D PDF shown in Fig. 11a. Four half-sized subsamples are used to define EOFs and the corresponding SN data, based on the same procedures as for the whole dataset: (a) the first 19 and (b) the second 18 years, (c) the odd 19, and (d) the even 18 years. PDFs are plotted on the same plane as in Figs. 10 and 11, with $h = 0.9$. The number of SN sample maps is: (a) 935, (b) 992, (c) 969, and (d) 970, respectively. The four rectangles in each panel are duplicated from Fig. 11a.

Four rectangles in Fig. 11a are collected and arithmetically averaged. Shading in the composite maps indicates regions statistically significant at the 99% level by a Student's t-test applied at each grid point in physical space. The number of degrees of freedom, or effective sample size, is set conservatively equal to the number of collected samples divided by 10.

The composite for PNA (Fig. 14a) is characterized by a large negative anomaly over the North Pacific with a wave train–like feature over the North American continent. This composite bears considerable resemblance to the Pacific/North-American pattern, with the same acronym, rendered famous by WG, and studied by many other authors. The wave train of our composite is not as pronounced as WG's corresponding pattern (their Fig. 17). Another wave train–like feature is also visible in the Atlantic-to-Eurasian sector; while it is less statistically significant here, it is consistent with features in the leading NH cluster of Mo and Ghil (1988).

The RNA, that is, reverse PNA pattern (Fig. 14b), is roughly opposite to PNA in physical space and in phase space, characterized by a positive anomaly over the North Pacific. The Atlantic-to-Eurasian wave train is slightly more significant than in the PNA case (compare also Mo and Ghil 1988, Figs. 10a,b, and f).

The composite for ZNAO (Fig. 14c) is characterized by a north–south dipole over the North Atlantic with a negative anomaly centered over Greenland. This pattern corresponds to a phase of the extensively
studied North Atlantic Oscillation (NAO; e.g., van Loon and Rogers 1978; BL; WG) characterized by a strong westerly jet over the region.

Similarly, the BNAO regime (Fig. 14d) is opposite to ZNAO with a positive center over (southern) Greenland, and corresponds to a blocked or split jet over the North Atlantic. Thus, B stands for the blocked phase of the NAO, and Z for its zonal phase.

It is intriguing to note that the major features of the four patterns describe two separate circulation regimes, zonal and blocked, in the North Pacific and the North Atlantic regions, respectively. Each is associated with the end of one of the two dominant storm tracks (Blackmon et al. 1977), where the climatological jets are decelerated (cf. Blackmon et al. 1984a,b; Miyakoda and Sirutis 1985).

Two-dimensional PDFs using P- and A-QS datasets are presented in Fig. 15. They show clear bimodality between the two regimes of NH modes, that is, PNA–RNA and ZNAO–BNAO. The P- and A-QS sets overlap only over 379 days. This amounts to 45% and 31% of the two regional QS datasets, respectively. It is tempting, therefore, to assume that LFV in each of the two regions is approximately independent, with weak interactions between the two. Kimoto (1987) showed that it is possible to form two sets of varimax-rotated EOF patterns that have significant features in either one of the two regions. Since the rotated EOFs are mutually uncorrelated in time, at zero lag, approximate statistical decoupling can be demonstrated. The regionality in LFV is assumed a priori in some studies (e.g., Horel 1981; DG; BL). Mo and Ghil (1988) also note it in their hemispheric analysis. This motivates us to examine regional flow regimes in Part II of this study, where partial interdependence between the regions is also addressed.

6. Concluding remarks

Whether the extratropical atmosphere possesses distinctly defined multiple flow regimes, and whether observed sequences of maps can be examined in terms of sojourning within and transiting among these regimes, has been a subject of considerable theoretical as well as practical interest. The notion of Grosswetterlagen emerged as early as the 1940s (Bauer 1951; Namias 1953). However, until recently few systematic efforts have been made to substantiate this idea. In the present study, we argued that the paradigm of multiple flow regimes arises naturally when the atmosphere is viewed as a nonlinear dynamical system that exhibits chaotic, but not totally random, behavior.

The search for persistent and recurrent flow regimes can be objectively defined as looking for—possibly multiple—local maxima in probability density evaluated in the system's phase space (Benzi et al. 1986; Ghil 1987, 1988; Kimoto 1987, 1989; Mo and Ghil 1987, 1988; Molteni et al. 1990). We studied this possibility first by defining a low-dimensional subspace—in which most of the attractor's variability is effectively embedded—through the use of EOFs, and second by evaluating multivariate PDFs on this subspace. Inherent difficulties of such an ambitious approach with a limited sample size were discussed in terms of a discrepancy between statistically feasible and dynamically desired dimensions (four and eight, respectively). The limited success of any reasonable search for multimodality may be related to this difficulty.

It is possible, however, with the help of physical and synoptic intuition, to pin down inhomogeneities on the periphery of PDFs supported near the origin of low-dimensional subspaces. Such inhomogeneities—which we found to appear as radially oriented ridges—indicate, presumably, an underlying multimodality in higher dimensions. In view of the above-mentioned difficulties, we are encouraged by the fact that Cheng and Wallace (1993) have recently obtained results consistent to a large degree with Kimoto's (1987) and the present ones. Without claiming necessarily true multimodality of PDFs in the NH atmosphere's phase space, we conclude that a macroscopic, statistical—dynamical view is a promising approach in understanding the large-scale atmosphere's low-frequency variability (LFV).

The subtle, but definite inhomogeneities detected by our analysis of two-dimensional (2D) PDFs can
be related to well-known, anomalous circulation patterns: PNA, RNA, BNAO, and ZNAO. These may be understood as zonal and blocked flow patterns in the two oceanic sectors, near the exit region of either one of the two NH jets. The possible regionality of these patterns is suggested by our analysis without any a priori assumption.

The two patterns in either sectorial pair are found to lie approximately opposite to each other with respect to the origin in phase space, that is, to the
in Part II on the description of LFV's synoptic features.

Acknowledgments. It is a pleasure to thank A. Arakawa, K.-C. Mo, H. Mukougawa, J. Namiot, P. H. Roberts, G. Schubert, R. Vautard, J. M. Wallace, M. Yanai, and S. Yoden for their comments and encouragement. D. Rodenhuis's continued interest was particularly helpful. Three reviewers (S. L. Mullen and two anonymous ones) provided detailed, constructive criticism. This work was supported by NSF Grant ATM90-13217, by NASA Grant NAG-5713, by NOAA Grant DOC A90AA-D-MC076, and by a Guggenheim Fellowship to M.G.

APPENDIX A

Least-Squares Cross Validation (LSCV) to Choose Smoothing Parameter

S86 describes a least-squares cross validation (LSCV) technique to choose the smoothing parameter $h$. LSCV minimizes a score function $M(h)$ defined as

$$M(h) = \int (f - \hat{f})^2 \, dx,$$

(A.1)

where $f$ is the "true" density, and $\hat{f}$ is the estimated PDF; the integration is over the whole phase space. $M(h)$ can be rewritten as

$$M(h) = \int \hat{f}^2 - 2 \int f \hat{f} + \int f f$$

$$= \int \hat{f}^2 - 2 \int f \hat{f} + \text{const.}$$

(A.2)

Although we do not know the "true" density $f$, the second term on the right-hand side can be evaluated by using the following ensemble-mean relationship:

$$E\left[ \frac{1}{N} \sum_{i=1}^{N} \hat{f}_{-i} \right] = E[\hat{f}_{-i}] = \int \hat{f}_{-i} f = \int \hat{f} f,$$

(A.3)

where $\hat{f}_{-i}$ is the estimated PDF at $X_i$ by discarding the $i$th data point; $E[\cdot]$ represents the expected value with respect to the true PDF $f$. It is expected that the cross-validating use of $\hat{f}_{-i}$ instead of $\hat{f}$ in (A.3) leads to a better estimate of $\int \hat{f} f$ from a finite sample.

We compute, therefore, the score function $M_0(h)$, given in terms of data only, and search for $h$ that gives its minimum, where
\[ M_0(h) = \int f^2 - \frac{2}{N} \sum_{i=1}^{N} \hat{f}_{-i} (X_i) . \]  
\[ (A.4) \]

The integral in the first term on the right-hand side is evaluated numerically.

We use daily values of PCs in PDF estimates. Since the time series possesses temporal persistence, we discard samples from \((i-10)\) days to \((i+10)\) days in computing \(\hat{f}_{-i}\) in \((A.4)\). Discarding twice the number of samples did not affect the results.

REFERENCES


