Weather Regimes: Recurrence and Quasi Stationarity

PAUL-ANTOINE MICHELANGELI, ROBERT VAUTARD, AND BERNARD LEGRAS

Laboratoire de Météorologie Dynamique du CNRS, Paris, France

(Manuscript received 24 March 1994, in final form 6 October 1994)

ABSTRACT

Two different definitions of midlatitude weather regimes are compared. The first seeks recurrent atmospheric patterns. The second seeks quasi-stationary patterns, whose average tendency vanishes. Recurrent patterns are identified by cluster analysis, and quasi-stationary patterns are identified by solving a nonlinear equilibration equation. Both methods are applied on the same dataset: the NMC final analyses of 700-hPa geopotential heights covering 44 winters. The analysis is performed separately over the Atlantic and Pacific sectors.

The two methods give the same number of weather regimes—four over the Atlantic sector and three over the Pacific sector. However, the patterns differ significantly. The investigation of the tendency, or drift, of the clusters shows that recurrent flows have a systematic slow evolution, explaining this difference. The patterns are in agreement with the ones obtained from previous studies, but their number differs.

The cluster analysis algorithm used here is a partitioning algorithm, which agglomerates data around randomly chosen seeds and iteratively finds the partition that minimizes the variance within clusters, given a prescribed number of clusters. The authors develop a classifiability index, based on the correlation between the cluster centroids obtained from different initial pullings. By comparing the classifiability index of observations with that obtained from a multivariate noise model, an objective definition of the number of clusters present in the data is given. Although the classifiability index is maximal by prescribing two clusters in both sectors, it only differs significantly from that obtained with the noise model using four Atlantic clusters and three Pacific clusters. The partitioning clustering method turns out to give more statistically stable clusters than hierarchical clustering schemes.

1. Introduction

Recently, the concept of weather regimes has emerged in dynamical meteorology, as a particular representation of atmospheric dynamics in the extratropics. Although it was introduced by synopticians about half a century ago (Rex 1950), this concept has received quantitative support only recently using long archives of daily weather maps allowing statistical inference. The principle of this representation is that, to a certain degree, the number of possible states of the atmospheric large-scale circulation is finite. Of course, this is only a representation; instantaneous maps are different from one another and vary continuously with time. The task of classifying weather maps into a finite number of states is therefore not straightforward and requires a certain number of assumptions.

What do we mean exactly by weather regimes? The apparent fuzziness of the concept arises from the fact that there is no universal definition. From one definition to another, the outcoming patterns differ significantly. Thus far, the various quantitative approaches developed during the last 15 years are based on one of the following three general properties: recurrence, persistence, or quasi stationarity.

a. Recurrence

In this case, weather regimes are defined as the states of the atmosphere with the highest probability of occurrence. Accordingly, one seeks for maxima of the probability density function (PDF) of states in a certain phase space. Motivated by the results of a pioneering study of Hansen and Sutera (1986), who found significant bimodality in the planetary wave amplitude density, this direct approach has been used by Molteni et al. (1990) and Kimoto and Ghil (1993a,b; hereafter KG93) who looked at PDFs in low-dimensional phase spaces (2 to 5 variables). In higher dimensions, reliable estimate of PDFs are impossible owing to the relatively small number of data (Silverman 1986).

Cluster analysis (Gordon 1981) is a less ambitious approach that localizes high concentrations of points, called clusters, in higher-dimension spaces but does not pretend to estimate the PDF. A preliminary application of cluster analysis to the identification of weather regimes has been described by Legras et al. (1987) for the Atlantic domain. Mo and Ghil (1988) found 6–7 relevant clusters over the Northern Hemisphere. Starting from cluster "kernels" defined as density maxima,

Corresponding author address: Dr. Paul-Antoine Michelangeli, Centre National de la Recherche Scientifique, Laboratoire de Météorologie Dynamique, Université Paris VI Tour 15, 5ème étage Boite 99, 4 Place Jussieu, 75252 Paris cedex 05, France.

Molteni et al. (1990) used a clustering algorithm and found comparable results. Cheng and Wallace (1993, hereafter CW93) went further in the statistical significance of clusters using another scheme applied on the fully dimensional phase space of low-pass-filtered maps and demonstrated the stability of the results when the dimension of the phase space is varied.

b. Persistence

Weather regimes based on persistence are usually defined by searching patterns associated with persistent anomalies in some key regions (Dole and Gordon 1983). These are, for instance, geopotential anomalies that persist beyond T days above or below a given threshold. The problem when keying composites to a local domain is that one mixes anomalies originating from different large-scale patterns. Global persistence has been also taken into account by Mo and Ghil (1988) in order to refine their cluster analysis. In general, the identification of weather regimes based only on a persistence criterion is difficult since persistence itself is a subjective concept, requiring a similarity measure and a timescale.

c. Quasi stationarity

In this last case, weather regimes are defined as the states for which large-scale motion is stationary in the statistical sense. More precisely, one seeks the large-scale patterns having, on average over their possible realizations, a vanishing time derivative. In other words, the balance of forces acting on this pattern, on average, vanishes. This last definition is a natural extension of the definition of multiple equilibria (Charney and Devore 1979; Legras and Ghil 1985; Itoh 1985) where exact stationary solutions of the evolution equations are calculated. The release of the constraint of stationarity of the small scales makes the quasi-stationarity definition more realistic and intuitive: our daily life perception of weather regimes indeed corresponds to long periods (a week to a month) of similar weather type, perturbed by synoptic fluctuations. This is also more physically consistent with the fact that transient fluctuations do have a positive feedback on the planetary-scale flow (Egger and Schilling 1983; Hoskins et al. 1983). This approach, called nonlinear equilibration, has been first developed by Reinhold and Pierrehumbert (1982), who parameterized the transient feedback by the circulation induced by the fastest-growing linear mode of the current large-scale flow in a highly simplified model. The full nonlinear feedback was considered and the identification of quasi-stationary states of a two-level quasigeostrophic model was performed by Vautard and Legras (1988). Later, Vautard (1990, hereafter V90) applied nonlinear equilibration to actual data over the Atlantic domain using a set of 10-day low-pass filtered NMC final analyses of 700-hPa heights.

Yet these various approaches share common concepts. First, weather regimes are always associated with large spatial scales of the planetary size and to low-frequency variability (periods longer than, say, 10 days). Midlatitude baroclinic disturbances, having life cycles shorter than 10 days, are only considered as modulating the weather regimes. Their influence on weather regimes is due to the nonlinearity of the dynamics. Indeed, Vautard and Legras (1988) found, in a channel model without topography, that multiple weather regimes can be maintained by the transients only. Weather regimes are usually considered as arising solely from internal atmospheric processes. They differ from variability due to changes in the boundary forcing such as sea surface temperature, evolving on timescales of a month or more. A reasonable assumption that helps in separating these two sources of variability is to consider that boundary forcing is an environmental factor modulating the probability of occurrence of weather regimes but not creating regimes in its own right (Molteni et al. 1993).

Nevertheless, the three definitions of weather regimes are not equivalent. One major purpose of this paper is to compare recurrent and quasi-stationary patterns identified using the same dataset. We will not consider the persistence properties throughout the paper, still keeping in mind that a quasi-stationary pattern is more likely to persist than other patterns (see V90).

The second goal of this work is to settle technical issues that are relevant to the results of cluster analysis and that have been generally overlooked in previous work. A great uncertainty arises about the number of clusters one should use in order to classify atmospheric data. If one focuses on the wintertime season, with datasets of about 40 years, about 4000 daily maps have to be classified into k clusters. Roughly speaking, for the timescale of an event of the order of 10 days, only $400/k$ events can be detected. Standard arguments about the statistical significance of classes imply that each class should contain at least, say, 50 events. Therefore, we cannot expect to detect more than 8 clusters. Taking into account the fact that the atmosphere still spends some time during transitions between regimes, a conservative bound to the number of possible clusters is 5 or 6.

Mo and Ghil (1988) retained eight weakly populated hemispheric clusters, but their statistical significance was not assessed. By varying the parameters of the method, they found that only six clusters were stable. Nevertheless, no estimate of the stability of these clusters, when taking another sample, was made. Molteni et al. (1990) also found six hemispheric clusters, four of which were reproducible. In the application of cluster analysis, reproducibility has been quantified, in CW93, by the use of subsamples containing half of the data and comparison with results obtained from the whole dataset. Three hemispheric clusters and two sectorial clusters over the Pacific and the Atlantic sectors

were found. Using multivariate PDFs, KG93 found basically two to three robust clusters over the Pacific sector and four over the Atlantic sector, in agreement with CW93. Their statistical significance was tested based on the null hypothesis of a multinormal distribution. The difference between Mo and Ghil (1988) patterns and those obtained in KG93 and CW93 was attributed by CW93 to the use of correlation matrices in the former and covariance matrices in the latter, for the selection of the basic EOFs. V90 retained four quasi-stationary regime patterns over the Atlantic. By contrast with cluster analysis, the most sensitive parameter in the nonlinear equilibration method seems to be the number of EOFs defining the large-scale phase space, as we will show here.

We demonstrate here that the atmospheric states gather naturally into a number of well-defined, statistically robust clusters: four over the Atlantic sector and three over the Pacific sector. We will be concerned only with sectorial regimes; hence our results need to be compared with that of CW93 and KG93.

Finally, the third point addressed here is a methodological point. There exist two main categories of clustering algorithms: hierarchical and partitioning ones. In the first kind, one builds iteratively a classification tree, starting from single data points, merging them into clusters according to a similarity criterion. CW93 used such a hierarchical algorithm, called Ward's method. In partitioning methods (Legras et al. 1987), a prescribed number of clusters is chosen, and data points are agglomerated around kernels initially chosen from random seeds. The kernels are iteratively modified, and the whole procedure is designed to minimize a criterion. Therefore, for a given number of clusters, partitioning methods provide an optimal partition, which is not the case for hierarchical algorithms. The final partition corresponds to a minimum of the criterion. In practice, for large samples, several local minima can be reached, and the final partition depends on the initial seeds. This difficulty is overcome here by performing several classifications with different random seeds and choosing the partition that is the most consistent with the others. This almost entirely removes the dependence on initial seeds. We will show here that for our atmospheric data, as well as for samples of synthetic data, this partitioning method turns out to give more statistically robust results than hierarchical methods.

In order to provide quasi-stationary regimes to be compared with recurrent regimes, we follow the approach of V90, with slight modifications. With respect to this study, the differences are as follows. (i) No prior time filtering is applied here. (ii) Statistical significance tests are improved by including a reproducibility criterion. (iii) The analysis is extended to the Pacific domain. (iv) The dataset used covers 44 winters instead of 36.

The paper is organized as follows. In section 2 we describe the data and their preprocessing, consisting basically in the projection onto a few leading EOFs. In section 3, the clustering algorithm is described, and results are presented. In section 4, the nonlinear equilibration technique, used for defining quasi-stationary regimes, is briefly recalled. Results and comparisons with cluster analysis are detailed, as well as comparisons with previous studies. Section 5 contains a summary and a brief discussion.

2. Data

Like many studies of the low-frequency atmospheric variability, the 700-hPa Northern Hemisphere geopotential height maps compiled by the National Oceanic and Atmospheric Administration (NOAA) Climate Analysis Center are used. Only the winter period (15 November to 31 March) is considered. The data are provided on the diamond NMC grid (two shifted 10° × 10° regular grids). This dataset is the one used in KG93. The Atlantic (ATL) and Pacific (PAC) sectors are the same as in V90 and Plaut and Vautard (1994), covering 40° of latitude and 120° of longitude each and centered, respectively, at 50°N, 20°W and 50°N, 160°W. On each sector, 113 grid points are used. The record covers a period of 44 years, starting in January 1949 and ending in December 1992. The first and last winters are thus incomplete.

The data are sampled every other day in order to reduce redundancy. The dataset analyzed thus contains 2974 instantaneous maps. No prior time filtering is performed, unlike in CW93 and KG93. Nevertheless, a prior spatial filtering is performed here using only a small number of EOF coefficients (Preisendorfer 1988). This spatial filtering has the advantage of compressing the information along directions in phase space that retain most of the variability. Admittedly, it also induces a time filtering that eliminates the high-frequency fluctuations. The annual cycle is not removed from the data, but we verified that our weather regimes do not simply correspond to particular calendar months. No significant trend is observed in this dataset; the verification was done, however, only by visual inspection.

In principle, spatial filtering is not required for cluster analysis (CW93). On the contrary, nonlinear equilibration (Vautard and Legras 1988; V90) strongly relies on a scale separation: in this latter approach, stationarity is only required at large scale, allowing synoptic transients to evolve freely within the large-scale weather regime environment and to produce feedback interactions (Hoskins et al. 1983; Metz 1987). Accordingly, scale separation should distinguish baroclinic and barotropic eddies. When based on EOFs, scale separation is certainly not optimal but has the enormous advantage of being simple and free. A more physically meaningful EOF analysis could be carried out using the empirical normal modes (Brunet 1994), which should bring a more natural scale separation.

TABLE 1. Characteristics of the ATL and PAC EOFs sorted by increasing values of their variance ratios. For each domain, the first column contains the order of the EOF (as sorted by variance explained), the second the variance ratio, the third the percentage of variance explained, and the fourth the cumulative percentage of variance explained.

	ATL EOFs				PAC EOFs			
1	1	0.14	19	19	1	0.11	22	22
2	2	0.17	15	34	2	0.18	16	38
3	3	0.18	14	48	3	0.27	12	50
4	4	0.26	9	57	4	0.29	9	59
5	5	0.33	8	65	5	0.32	8	67
6	7	0.56	5	70	9	0.53	4	71
7	9	0.57	3	73	7	0.61	3	74
8	10	0.85	2	75	10	0.67	2	76
9	6	0.87	5	80	12	0.95	2	78
10	8	0.96	3	83	6	1.10	5	83

In each sector, the first 20 EOFs explain about 95% of the total variance. Since data are not low-pass filtered, the EOFs explaining the low-pass variance are not necessarily ranked first. In order to detect the low-pass EOFs, we seek the EOFs with the smallest associated ratio of high-pass to low-pass variance, as in Plaut and Vautard (1994). More precisely, if $a_k(t)$ denotes the kth principal component (projection coefficient onto the kth EOF at time t), $a_k^*(t)$ its centered weekly average, and $a_k'(t)$ the departure from this average, the ratio r_k of the variance of $a_k'(t)$ to the variance of $a_k^*(t)$ is calculated and the EOFs are sorted into increasing order of r_k. A subset of the first M EOFs with lowest ratio values is then retained, defining the low-pass states of the atmosphere.

Table 1 lists the values of the variance ratio r_k of the kth EOF, to increasing order of r_k, and the percentage of the total variance explained by the associated component. In both sectors, the rank of the first five EOFs is unchanged by sorting the ratio values. EOFs 6 and 8 contain more high frequencies than EOFs 7, 9, and 10. The choice of M was determined a posteriori after trying all possible values between $M = 3$ and $M = 10$. It turns out that cluster analysis is fairly insensitive to this choice. By contrast, nonlinear equilibration gives a maximal number of significant regime patterns for $M = 8$ in both sectors (see section 4). The choice $M = 8$ can also be guided by a visual inspection of the EOF patterns. Figure 1 shows the structure of EOFs 10 and 6 for the Atlantic domain and 10 and 12 for the Pacific domain, ranked at orders 8 and 9 according to their variance ratio values. In both sectors, the 10th EOFs have dipolar structures bearing similarities to the blocking pattern, hence the choice of keeping them. The dataset to be analyzed thus consists for both sectors of a series of low-pass state vectors of dimension 8. These vectors will be denoted by (X_i), $i = 1, \cdots, 2974$. In addition, the nonlinear equilibration algorithm requires a set of associated instantaneous tendency vectors T_i, also called phase space velocities (see V90). These vectors are calculated as the centered differences between the state vectors X_i at times $i + 1$ day and $i - 1$ day. Note that the values used to calculate the tenden-

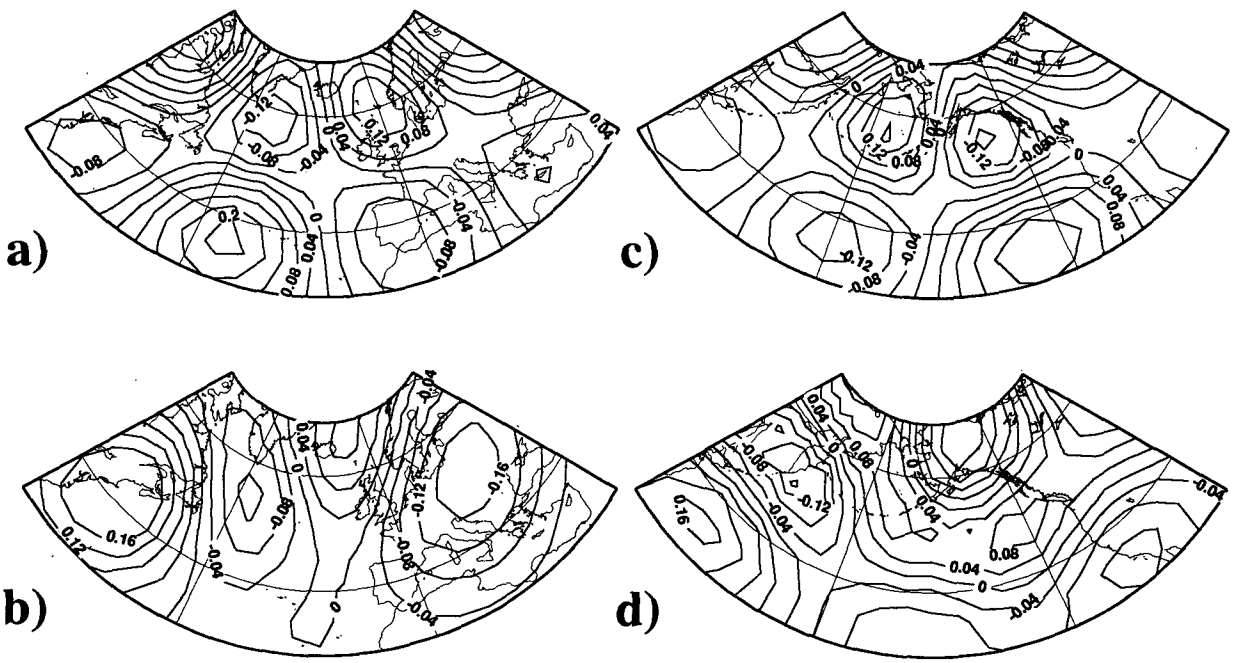

FIG. 1. Some of the sectorial empirical orthogonal functions: (a) ATL EOF 10, ranked 8th by increasing order of variance ratio; (b) ATL EOF 6, ranked 9th; (c) PAC EOF 10, ranked 8th; (d) PAC EOF 12, ranked 9th.

cies are not included in the X_i file, since the X_i are taken every other day.

In order to investigate the impact of weather regimes on near-surface atmospheric parameters, a set of temperature maps at the 850-hPa level is used. These data are extracted from the National Center for Atmospheric Research (NCAR)–National Meteorological Center (NMC) archive. Horizontal interpolation from the original octogonal grid onto a Gaussian T31 grid is applied (Brunet et al. 1995). This dataset covers a shorter period (February 1963–December 1988) than the 700-hPa heights dataset and has a lot of missing dates that are simply removed from our composites.

3. Cluster analysis

a. The dynamical clusters algorithm

The partitioning algorithm we will use here is known as the dynamic cluster method (Diday and Simon 1976). It has been used in meteorology for a long time as a pattern recognition technique in order to classify clouds from satellite images (Desbois et al. 1982). Given a prescribed number of clusters k, the overall goal of the algorithm is to find a partition P of the data points into k clusters $C_1, C_2, \ldots, C_k$ that minimizes the sum of variances within clusters,

$$W(P) = \sum_{j=1}^{k} \sum_{x \in C_j} d^2(X, Y_j), \qquad (3.1)$$

where Y_j is the *centroid* of cluster C_j. The Euclidian distance $d(X, Y)$ is used as a similarity measure between two points X and Y. The global minimum of the function $W(P)$ therefore corresponds to a partition that achieves the best separation of data points. With long data series, this optimal partition cannot be reached in practice due to the huge number of possibilities to explore. The dynamic cluster algorithm defines iterative partitions $P^{(n)}$ for which $W(P^{(n)})$ decreases with n and eventually converges to a local minimum of W that generally differs from the global one. This apparent limitation is not as serious as it seems at first. The global minimum is surrounded in phase space by many local minima that differ by the exchange of a few data points at the periphery of clusters. Such details are likely to change from one sample to another. We are rather interested in identifying reproducible partitions. The iterative algorithm proceeds as follows.

1) Initialization

First, k points are randomly chosen as initial seeds among all data. Then a small subset of the data (10% of the dataset) is randomly drawn. Elements of this subset are agglomerated to the nearest seed according to their similarity. The subset is then split into k small clusters, from which centroids are calculated, and taken as the actual initial seeds, $Y_j^{(0)}$, $j = 1, \cdots, k$. This smooth initialization avoids having initial seeds far from the density maxima of the PDF and can be shown to accelerate significantly the convergence of the process.

2) Iteration $n - 1$ to n

Given the k centroids $Y_j^{(n-1)}$ of the partition $P^{(n-1)}$, each data point is reassigned to the nearest of these k centroids. A new partition $P^{(n)}$ is thus formed, from which the centroids $Y_j^{(n)}$ are calculated, and so on. Finally, the sum of variances within clusters $W(P^{(n)})$ is diagnosed.

3) Stop criterion

The algorithm stops when $W(P^{(n)})$, the sum of variances within, has converged to a minimum—that is, when an additional iteration does not modify the partition. In our experiments, the convergence is obtained with about 50 iterations, on average, and rarely with more than 100.

b. Similarity of two partitions

The latter iterative algorithm always converges to a final partition, regardless of the distribution of the points. Nevertheless, if this distribution is uniform, for instance, one expects that the final partition depends strongly on the initial seeds. Conversely, when the dataset is distributed into marked clusters, two different pullings should give roughly the same results. The dependence on the initial seeds can therefore be used as an indicator of the degree of classifiability into k clusters. In fact, three problems arise simultaneously. (i) What is the best partition, given the number of clusters k? (ii) What is the best number of clusters k? (iii) Are the obtained clusters reproducible from another sample of data?

These three questions will be addressed by a method similar to that used in CW93. In the dynamical clusters method, the final partition is fully described by the centroids of each cluster. Therefore, the anomaly correlation coefficient (ACC, also called the pattern correlation) between the cluster centroids, used by CW93, is particularly suited as a measure of similarity between clusters. Any two partitions P and Q into k clusters of the same dataset are compared by calculating the matrix A_{ij} ($i = 1, \cdots, k, j = 1, \cdots, k$) of the ACCs between cluster i of P and cluster j of Q. In each row i of A, the maximal value $A'(i)$ is found for the cluster of Q that best fits the ith cluster of P. Finally, the minimum value of A', denoted by $c(P, Q)$, corresponds to the cluster of P that has the poorest analog in Q and is a measure of the similarity between the two partitions. When the two partitions are identical, $c(P, Q) = 1$.

c. Classifiability index

The most natural way to test the dependence on the initial seeds, and therefore the classifiability, is to com-

pare several final partitions for a given number of clusters k. The dynamical clusters algorithm is applied 50 times with a different initial random seed, providing 50 final partitions $(P_m(k), m = 1, \cdots, 50)$. The similarity between the latter is estimated by averaging the values of $c(P_m(k), P_{m'}(k))$, for all pairs (m, m'). The classifiability index $c^*(k)$, which depends on the number of clusters, is thus defined by

$$c^*(k) = \frac{1}{50(50-1)} \sum_{1 \le m \ne m' \le 50} c(P_m(k), P_{m'}(k)). \tag{3.2}$$

The classifiability index $c^*(k)$ is calculated, for the atmospheric data, for k varying from 2 to 10. For comparison we define a reference noise model, which is a first-order Markov process having the same covariance, at lag 0 and 1 (two days), as the atmospheric data (as in KG93). The model is built in the EOF space, so that only eight variables are taken into account. The PDF associated with this noise model is a multinormal PDF. We use the noise model to generate 100 samples of the same length as the atmospheric dataset, and we calculate $c^*(k)$ for each of them. The best choice of k is therefore the one that brings the largest differences between atmospheric and noise model data.

The classifiability index $c^*(k)$ is shown on Figs. 2a,b, as a function of the number of cluster k. The one-sided 90% confidence bounds of $c^*(k)$ coming from model data are also plotted. These bounds are calculated as the 11th and the 90th values of the index obtained from the 100 samples of the noise process. Values of the atmospheric index above the upper bound indicate a classifiability significantly higher than for the noise model.

For the ATL sector, $c^*(k)$ is maximum at $k = 2$, for which the ACC between centroids of different partitions always lies above 0.99: partitions are almost identical from one cluster application to another. Then, the index drops down to 0.64 at $k = 3$ and peaks again at $k = 4$, with mean ACC values of about 0.92. There, the index is significantly greater than the one obtained from noise model data. After $k = 5$, the ATL index lies within the noise confidence interval, meaning that ATL data are not more classifiable than data coming from a multinormal distribution. For the PAC sector, $c^*(k)$ only emerges from the noise confidence interval for $k = 3$ ($c^*(3) = 0.993$). Figure 2 shows clearly that the only way to distinguish atmospheric sectorial data from a first-order Markov process in terms of clusters is to choose four clusters for the Atlantic domain and three for the Pacific domain.

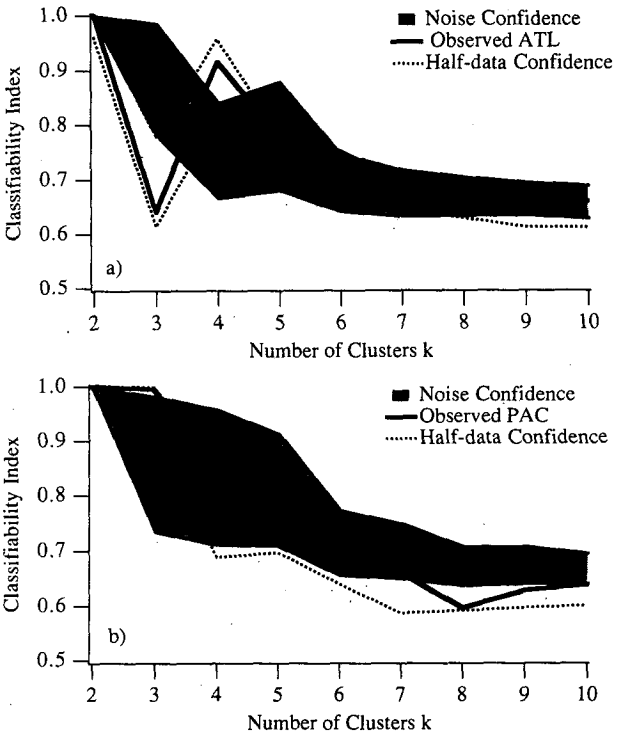

FIG. 2. The classifiability index as a function of the number of clusters k (heavy solid). Shaded area represents the 10%–90% bounds of the classifiability index distribution calculated from the first-order Markov process (noise model). Dotted lines represent the same bounds for random halves of the atmospheric data: (a) Atlantic sector; (b) Pacific sector.

d. Reproducibility

We first choose a reference partition among the 50 partitions $P_m(k)$ in order to test its reproducibility. This choice is made by selecting the partition $P_a(k)$ that exhibits the largest average similarity value $c(P_a(k), P_m(k))$ with all the other ones. This latter choice strongly reduces the dependence on the initial seeds. Indeed, several experiments (not shown) performed by recalculating reference partitions from independent sets of 50 partitions showed that they are almost identical. For instance, for the ATL sector with $k = 4$, the similarity between reference partitions is larger than 0.99.

Next, the reproducibility of the reference clusters is tested in exactly the same way as in CW93. One hundred random subsamples containing half of the atmospheric data are drawn (CW93 took 50 samples), for which reference partitions are calculated. Each cluster of the reference partition for the complete dataset (using the reference partition of the first experiment) is compared, using ACCs with the clusters of the reference partition of each of the 100 half-long samples. The average ACC with the best analogs is calculated, providing a reproducibility index for individual clusters.

Before discussing the reproducibility index, let us comment on the classifiability index obtained from these 100 half-sample cluster applications. Figure 2 shows their 90% bounds, calculated as above, from the 11th and the 90th value. The half-data classifiability index has roughly the same behavior as the whole-data

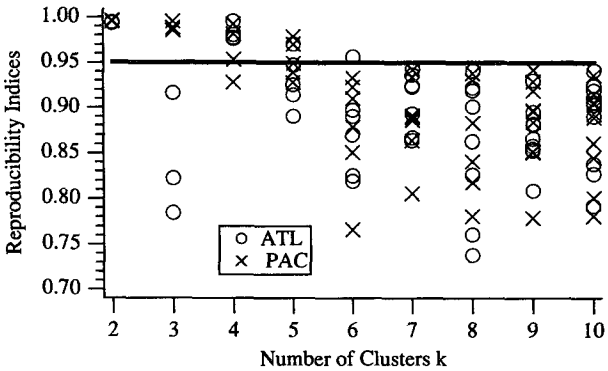

FIG. 3. Reproducibility indices of the k reference clusters as a function of k.

index. The peak at $k = 4$ for the ATL domain is reproduced faithfully, as are the values close to 1 observed for $k = 2$ in both domains. Note, however, that the half-data index is somewhat smaller, meaning that the longer the dataset, the better the classifiability, leaving us with the hope of convergence of the centroids as the sample length increases.

Figure 3 shows the reproducibility index of the clusters obtained for $k = 2, \cdots, 10$ from the atmospheric data. CW93 used a threshold of 0.89 for reproducibility of sectorial regimes. Taking into account the fact that here, both the ATL and PAC sectors cover longitudinally 120° (180° for CW93) and that our sectors are more restricted in the latitudinal direction, the equivalent threshold, estimated from CW93's formula, is about 0.95. For the ATL sector, all clusters have a reproducibility index larger than 0.99 for $k = 2$, and the reproducibility index falls below the threshold for $k = 3$. For $k = 4$, all clusters have a reproducibility index larger than 0.97. For larger values of k, most clusters are not reproducible. For the PAC sector, the sole values of k for which all clusters are reproducible are $k = 2$ and $k = 3$. Therefore, reproducibility of clusters is tightly linked with the classifiability of the set: when the dataset is classifiable, reference clusters are reproducible and vice versa.

e. *Composite patterns*

The previous analysis of classifiability and reproducibility suggests that we take $k = 4$ for the ATL domain and $k = 3$ for the Pacific domain. We will focus now on the spatial patterns of the weather regimes obtained for those values. The characteristics of the reference partitions are recapitulated in Table 2.

1) ATLANTIC SECTOR

The composite of the 700-hPa height fields associated with ATL clusters of the reference partition is displayed in Fig. 4. The first cluster is characterized by a strong positive anomaly over the Greenland area, together with a trough elongating from the southwest to the northeast Atlantic Ocean. This pattern is very close to the G+ sectorial pattern of CW93 and to the A5 pattern of KG93. When the average winter flow is added, a closed cell appears over Greenland, and the Atlantic jet is substantially displaced southward. The second pattern is characterized by an east–west dipole bearing similarities to the G pattern of CW93 and to the A2 pattern of KG93, associated with a pronounced wave structure. However, the amplitude of the positive anomaly centered over Scandinavia (about 120 m) is not large enough to produce a closed blocking cell over Europe. The third pattern, still very stable, consists of a trough centered over Iceland together with a small-amplitude positive anomaly in the southern part of the domain. It is not found in CW93 but corresponds to A6 in KG93. The structure of the anomaly extends the average jet all the way across the Atlantic Ocean. The last pattern is associated with a ridge in the mid-Atlantic. It corresponds roughly to A3 in KG93. Only half of the hemisphere is displayed on Fig. 4 since no significant anomaly is found over the Pacific area in their composites. The searching algorithm of KG93 generates two other clusters, one essentially a variant of the Greenland anticyclone cluster and the other corresponding to a southward displacement and intensification of the jet.

Note that all data maps are classified by our algorithm (unlike previous studies), and therefore averag-

TABLE 2. Characteristics of the clusters for the reference partitions over the ATL sector (left four columns) and the PAC sector (right three columns). Clusters appear in the same order as in Figs. 4 and 7. In the first line, the number of maps belonging to each cluster is listed. In the second line, we recall the values of the reproducibility index. For ATL $k = 4$; for PAC $k = 3$. The next two lines list the pattern correlation between the cluster centroids obtained using each half of the data and the corresponding centroid using all the data. The last two lines list the percentage of the number of elements in each whole-data cluster that also lies in the corresponding half-data cluster.

	ATL clusters				PAC clusters		
Number of members	630	786	832	726	1073	1086	815
Reproducibility index	0.98	0.98	0.99	0.98	0.99	0.98	1.0
Patt. corr. 1st half	0.99	0.98	0.94	0.96	0.98	0.99	0.99
Patt. corr. 2d half	0.98	0.99	0.99	0.97	0.99	0.99	0.88
Percentage 1st half	88	93	88	80	87	100	90
Percentage 2d half	95	91	90	96	97	82	86

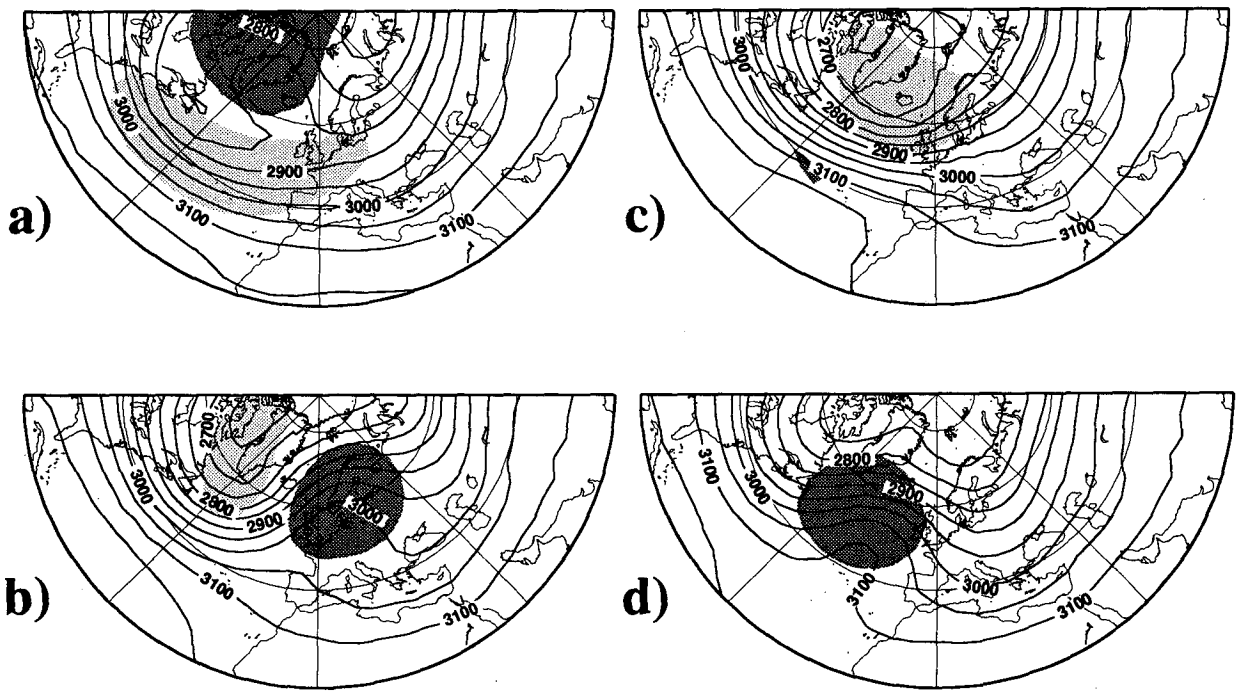

FIG. 4. Composites of the 700-hPa geopotential heights for the four clusters found over the ATL sector. Contour interval is 50 m. Dark shaded areas show areas where the anomaly of the composite with respect to the wintertime average is larger than 50 m. Light shaded areas correspond to anomalies lower than −50 m. Clusters are sorted by their consistency: (a) cluster 1; (b) cluster 2; (c) cluster 3; (d) cluster 4.

ing effects make the amplitude of the patterns relatively weak compared with instantaneous maps. This also explains the weaker amplitude found here as compared with clusters of KG93, who classified only half of the data.

Finally, the ATL classification is repeated with the first and last halves of the dataset with $k = 4$. The composites calculated from the reference partitions are displayed on Figs. 5 and 6. The obtained patterns are very similar to the patterns of Fig. 5, with pattern correlations of the order of 0.95 (see Table 2). As for membership, the percentages of the elements of the whole-data clusters also belonging to the corresponding half-data clusters are given in Table 2 as well. The lowest percentage obtained is 80%, but the average value is about 90%. We conclude that the four clusters found for the ATL area are climatic features, which are insensitive to long-term fluctuations such as climate change or interdecadal variability. The comparison with half-data samples leads us to the conclusion that cluster boundaries are not fully reproducible. Roughly speaking, 90% of the instantaneous maps can be placed within one cluster without ambiguity.

2) PACIFIC SECTOR

The reference cluster composites of the PAC sector, with $k = 3$, are displayed in Fig. 7. All clusters are very stable. The first pattern is associated with a weak anomaly somewhat reinforcing the zonal jet near its exit. It resembles the P4 pattern of KG93. The second pattern is known as the Aleutian low. It intensifies the jet across the Pacific and is found as the P1 pattern of KG93. The third pattern is associated with a blocking anticyclone centered over Alaska (regime P2 of KG93 and A of CW93). These two patterns clearly project onto the Pacific/North American pattern (PNA) but are not simply opposite phases of it. Note in particular the meridional shift between them, and the absence of the wave train characterizing the PNA. As in KG93 or CW93, the main center of the anomaly exhibits about three times the amplitude of the other centers in both phases. These features are robust even if the number of elements in our clusters is much higher than in CW93 or KG93. KG93 found four other clusters over the Pacific sector that are essentially variants of our first and third regimes in three cases, the last one being associated with an intense jet across the West Coast of the United States.

The PAC clusters are also recovered faithfully with the first and second halves of the data (not shown), with percentages of common elements of the order of 0.9. Notice, however, that the last centroid has an ACC of 0.88 with the corresponding second-half centroid. This is mostly due to the fact that during the second half of the period, this cluster is found to have a number of members substantially

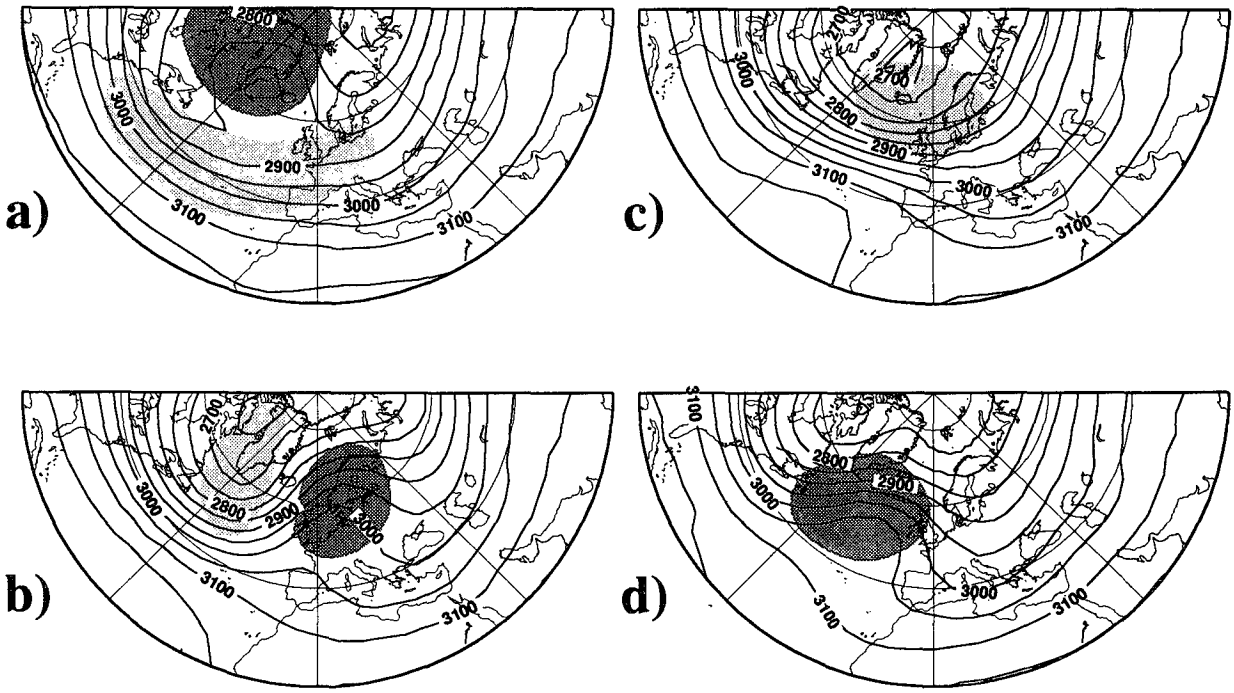

FIG. 5. Same as Fig. 4 for the clusters obtained from the first half of the data.

smaller than for the first half (489 elements in the first half versus 362 elements in the second), and therefore its centroid contains more statistical fluctuations. The Pacific blocking weather regime is therefore the regime that is the most sensitive to undersampling or natural long-term variability.

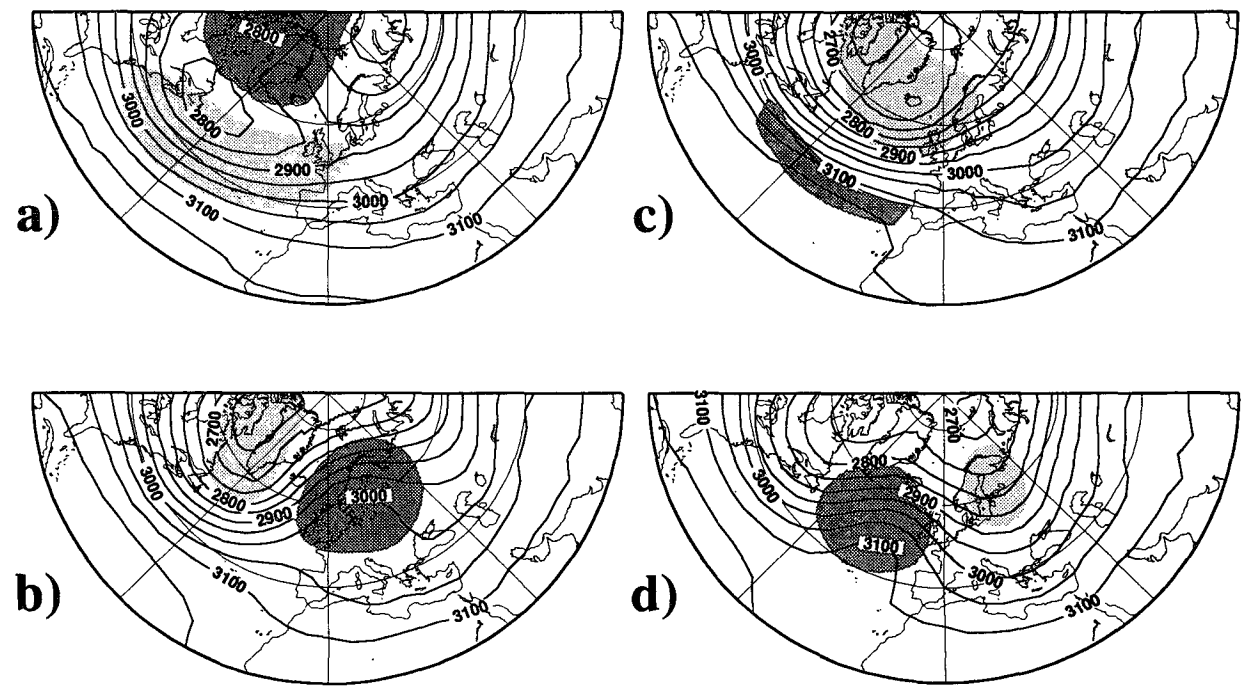

FIG. 6. Same as Fig. 4 for the clusters obtained from the second half of the data.

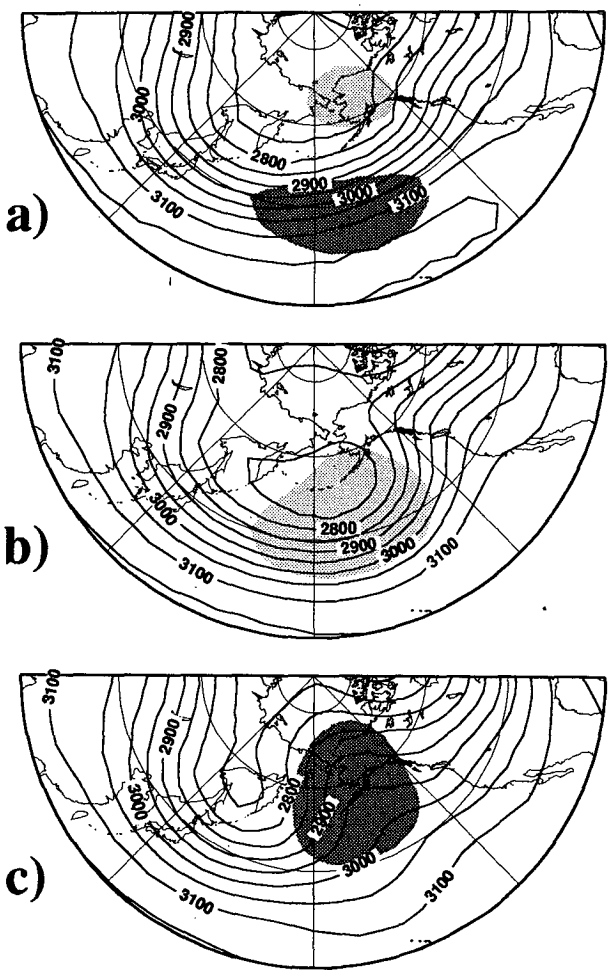

FIG. 7. Same as Fig. 4 for the three clusters of the PAC domain: (a) cluster 1; (b) cluster 2; (c) cluster 3.

3) COMPOSITES WITH TWO CLUSTERS

Figure 8 shows the composites associated with the ATL and PAC clusters obtained with $k = 2$. The anomaly patterns then consist of the positive and negative phases of the first principal components in each domain. This is further confirmed by counting the percentage of members in the first ATL cluster with positive values of the first principal component (100%), and for the second ATL cluster the percentage of elements with negative values of the first principal component (93%). The corresponding percentages for the PAC domain are, respectively, 97% and 100%. Thus, for $k = 2$, the clustering method essentially separates the cloud of data points along the first principal axis. In the results of CW93, two clusters were found over each sector. However, their approach does not suffer from this deficiency, since their clusters originate from different levels of the tree. Hence, significant departures from a simple discrimination along the first principal axis were found (see their Figs. 8 and 9).

Figure 9 shows the corresponding composites obtained for the noise model. Now, the composites are calculated directly from the cluster centroids, which are anomaly fields defined by their eight EOF coefficients. Owing to the way noise model data are constructed, it is impossible to extend, like in Fig. 8, the composite outside the PAC and ATL domains. The striking similarity between noise model and atmospheric clusters clearly demonstrates that cluster analysis does not bring here any additional information with respect to classical EOF analysis: a clustering method may give very stable clusters even for a process with a unimodal probability density distribution. This is also to be expected for the method used in KG93, since the angular PDF associated with a multinormal distribution has essentially two peaks along the first principal axis. Notice that KG93 also compared atmospheric data with the same noise process.

Next, the reference partitions for $k = 4$ (ATL) and $k = 3$ (PAC) are compared, in terms of membership, with reference partitions obtained for $k = 2$. The number of elements n belonging to the intersection between two subsets of lengths N_1 and N_2, taken at random, of the same set of length N has an hypergeometric distribution. The value of n below which 95% of this distribution lies is used here as a measure of the significance between two clusters coming from different values of k. The results are summarized in Fig. 10. From this diagram, we conclude that the clusters obtained for $k = 4$ (ATL) and $k = 3$ (PAC) are in fact subclusters of clusters obtained for $k = 2$. The former clusters are not strictly included within the latter ones, so that there is a redistribution of their members that would not occur using hierarchical schemes.

Finally, classifications have been carried out with $M = 5$ principal components, and the same results were obtained (not shown), indicating that, as in CW93, cluster analysis is fairly insensitive to the number of EOFs chosen in the definition of the phase space. Some other experiments were performed with data taken throughout the year with the annual cycle removed, and they yielded the same regime patterns.

f. Partitioning versus hierarchical cluster analysis

The main discrepancy in the results of the sectorial analysis of CW93 and ours lies in the number of reproducible patterns. We find three and four clusters, whereas CW93 found only two in each sector. The difference may be due to better stability of the dynamical cluster scheme than in the hierarchical scheme. In order to test this possibility, we perform a hierarchical analysis using Ward's method, as in CW93 with two datasets. One has a sampling rate of 10 days. The other also has a sampling rate of 10 days but is shifted by 5 days relative to the first one. Only two clusters were found with both datasets, in agreement with CW93's result.

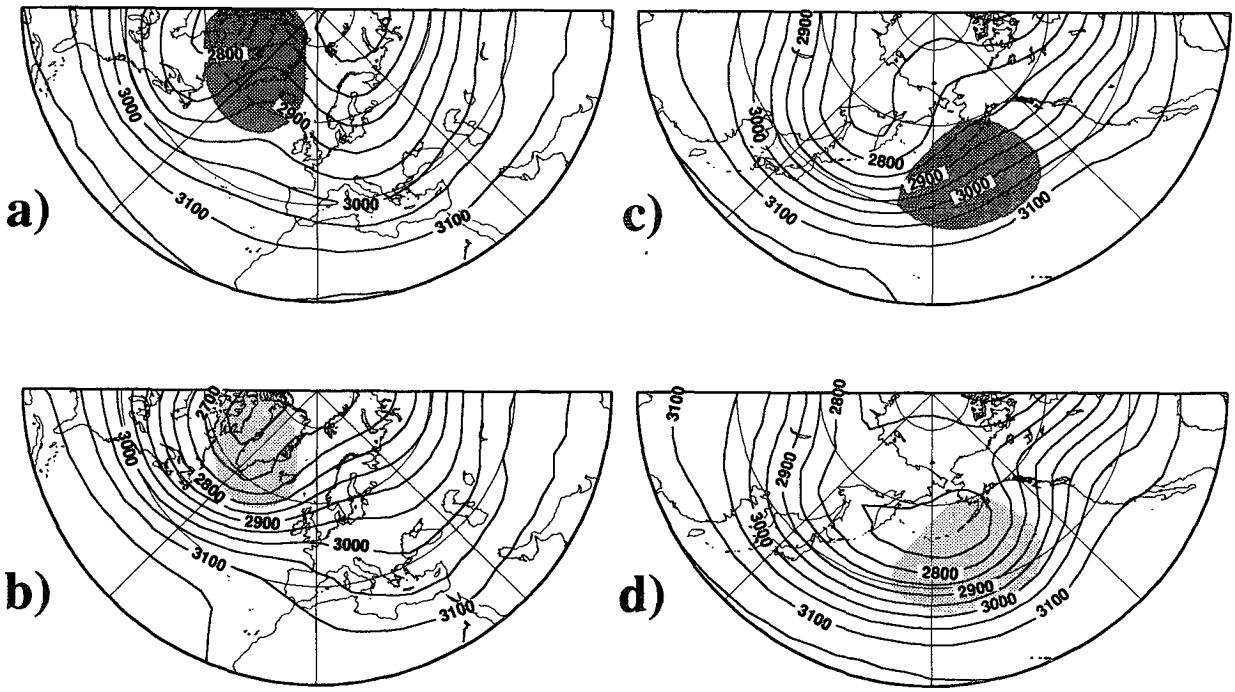

FIG. 8. Composites obtained from the analysis with $k = 2$ clusters for the ATL sector [(a) and (b)] and for the PAC sector [(c) and (d)]. Contour interval and shades are as in Fig. 4.

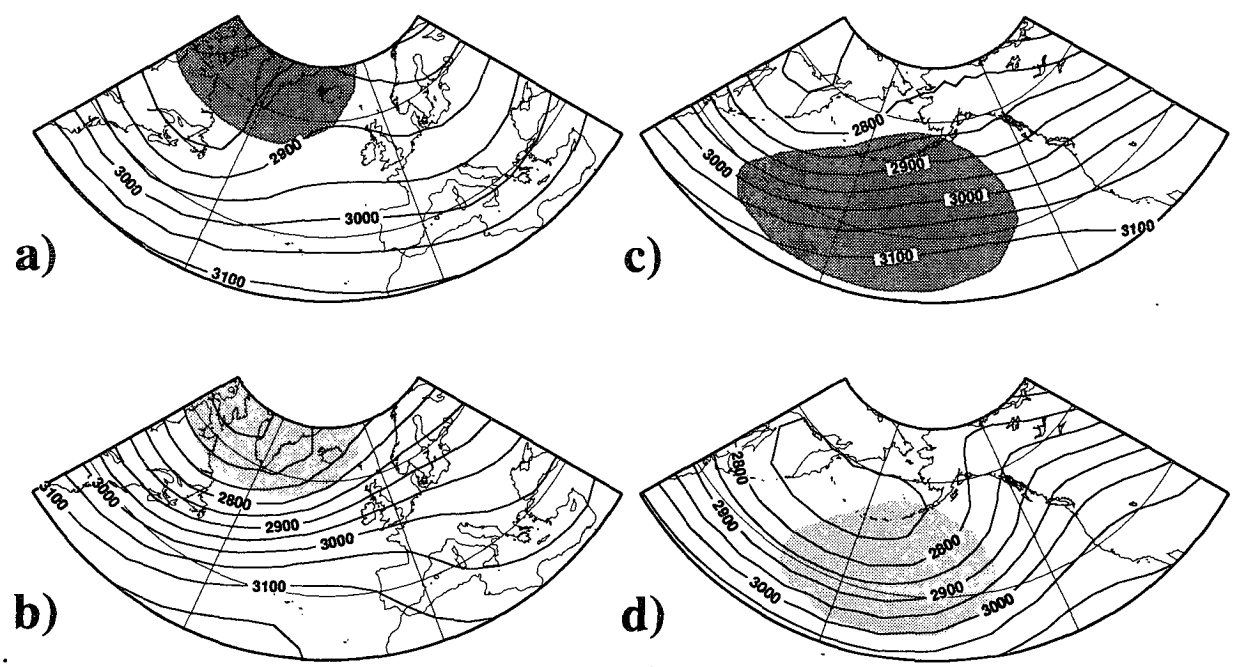

FIG. 9. Centroids of the clusters obtained from the first-order Marlov process using $k = 2$. Since these centroids are defined in the 8-dimensional EOF space, the average wintertime mean flow is added to the anomaly fields for comparison with Fig. 8. Note that it is impossible here to extend the composite outside the ATL and PAC domains, as in Fig. 8. The anomaly is defined by its 8 projection coefficients onto the EOFs.

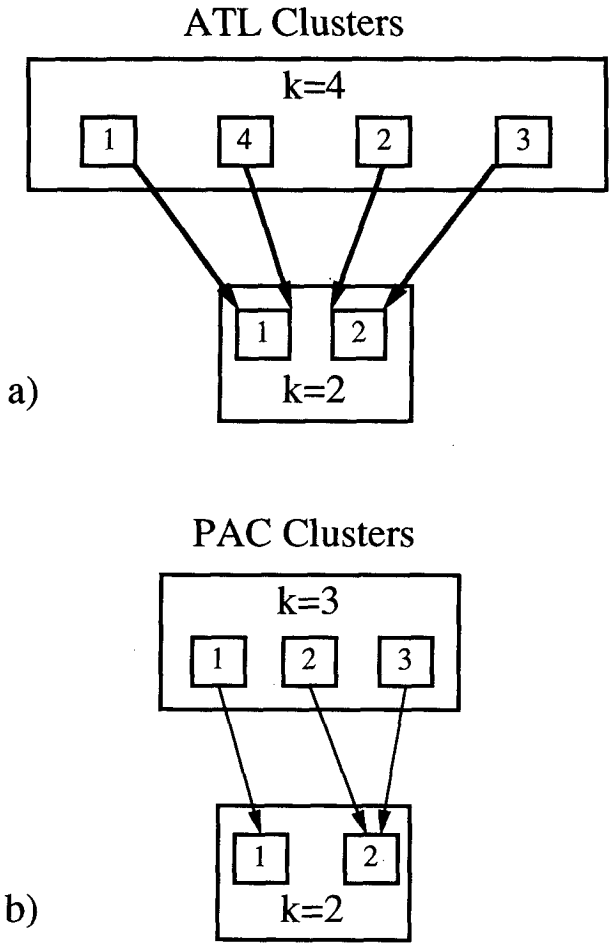

FIG. 10. Diagrams showing the significantly intersecting clusters between the reference partitions for (a) the ATL sector, with $k = 4$ and $k = 2$, and (b) the PAC sector, with $k = 3$ and $k = 2$.

Let us now justify this point more carefully. Consider, for instance, samples of data from the probability density function

$$\rho(x) = \frac{1}{3\sqrt{2\pi}}\left(e^{-(x-2.5)^2/2} + e^{-x^2/2} + e^{-(x+2.5)^2/2}\right), \tag{3.3}$$

represented in Fig. 11a. One should naturally expect to find three clusters. One hundred samples of 100 and 500 data points are drawn. The stability of Ward's method can be estimated by looking at the mean and variance, calculated from these 100 samples, of the centroids obtained by cutting the tree diagram at the level where there are three clusters. The same operation is performed with the dynamical clusters algorithm with $k = 3$. Results are summarized in Table 3a. The value of the mean centroids are almost the same for both methods, but the variance of these is much higher (one order of magnitude) for large samples (500 elements) in Ward's method than in the dynamical clusters method. We believe that this lack of stability, and maybe even lack of convergence as the sample size goes to infinity, is due to the fact that hierarchical schemes do not have a way to refine the partition at a given level and that incorporating new data might in fact cause them to deviate from the optimal partition.

Let us also consider the case of a one-dimensional nonsymmetrical process whose density is defined by (see Fig. 11b)

$$\rho(x) = \frac{1}{3\sqrt{2\pi}}\left(e^{-(x-2.5)^2/2} + 2e^{-x^2/2}\right). \tag{3.4}$$

In this case, one expects two clusters, although the density is not bimodal. The same experiments as above are performed and reported in Table 3b. Again, the variance of the estimates does not decrease significantly for Ward's method but does for the partitioning scheme. Also interesting is the fact that both methods behave quite well with regard to the probability of occurrence within clusters. About 60% of the data points fall within the first cluster and 40% fall within the other, owing to the slight shoulder of the density, instead of $2/3$ and $1/3$, as expected from the analytical expression of the density. This, in turn, shows that both methods do not have any tendency to equipartition the data among clusters. However, note that there are slight bi-

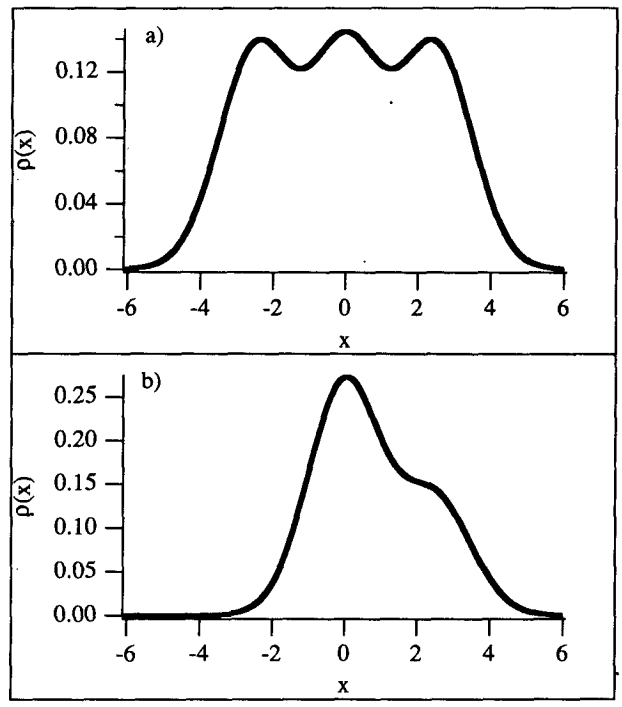

FIG. 11. Probability density functions for which the method of Ward and the method of dynamic clusters is tested at the end of section 3a.

TABLE 3. Means and variances of the three centroids and percentages of membership in corresponding clusters, calculated from 100 random samples of length 100 and 500 of the distributions shown in Fig. 11, using Ward's clustering algorithm (denoted by *W*) and the dynamic clusters method (denoted by *D*). (a) Statistics obtained for the density of Fig. 11a; (b) statistics obtained for the density of Fig. 11b.

(a)

	Mean		Variance		Percentage	
	W	*D*	*W*	*D*	*W*	*D*
			Sample of length 100			
Cluster 1	−2.59	−2.64	0.164	0.056	33	34
Cluster 2	0.01	0.03	0.363	0.102	34	35
Cluster 3	2.61	2.68	0.169	0.058	32	32
			Sample of length 500			
Cluster 1	−2.63	−2.64	0.125	0.011	33	33
Cluster 2	0.00	0.00	0.320	0.028	34	35
Cluster 3	2.64	2.65	0.119	0.015	32	32

(b)

	Mean		Variance		Percentage	
	W	*D*	*W*	*D*	*W*	*D*
			Sample of length 100			
Cluster 1	−0.22	−0.22	0.090	0.073	60	60
Cluster 2	2.46	2.41	0.183	0.105	40	40
			Sample of length 500			
Cluster 1	−0.19	−0.22	0.077	0.022	62	60
Cluster 2	2.49	2.41	0.179	0.048	38	40

ases in the estimation of the density peaks, if estimated by the cluster centroids.

4. Nonlinear equilibration

a. The equilibration problem

The methodology used in order to identify quasi-stationary regimes is very similar to that used in V90, to which the reader is referred for further details. The dataset is slightly different since we use more data (extension to December 1993, whereas in V90, 700-hPa heights are analyzed until March 1986), the data are not low-pass filtered beforehand, and they are projected onto a different set of EOFs. The statistical significance testing procedure is more restrictive here: a reproducibility test is added to the test of V90.

Following Vautard and Legras (1988), quasi-stationary regimes are defined as the solutions X^* of the nonlinear equilibration problem

$$T(X^*) = 0, \tag{4.1}$$

where X^* is a state vector containing the large-scale variables of the flow, and $T(X^*)$ is the *composite tendency* of X^*, defined as the average of the instantaneous tendency dX/dt over all possible occurrences of X^*, which can be written

$$T(X^*) = \langle dX/dt \rangle_{X=X^*}. \tag{4.2}$$

Therefore, the solutions to the nonlinear equilibration problem are, by construction, large-scale flows with no systematical drift, decay, or amplification. In other words, they are stationary on average.

In practice, the large-scale vector X is defined as the vector containing the projection coefficients onto the M first low-pass EOFs, exactly as in section 2. The composite tendency $T(X)$ of any large-scale pattern X is calculated as the average of the instantaneous tendencies T_i associated with neighbors X_i of X, found within a ball of radius D around X. The average is weighted by a function of the phase space Euclidian distance $d(X, X_i)$ between X and X_i. The weighting function decreases as the distance increases and vanishes when $d(X, X_i)$ is greater than the radius D. Here D is chosen as the value below which 10% of the distribution of the distances $d(X_i, X_j)$ between data pairs lie. The results are fairly insensitive to the values of D: taking a proportion of 5%, 15%, or even 20% instead would not affect the results presented below.

For any large-scale pattern X, the composite tendency $T(X)$ becomes a function of X only and is a vector of the same dimension as X. The nonlinear equilibration problem therefore consists in practice in solving a nonlinear set of M equations with M variables. This is done by minimizing the cost function $J(X) = |T(X)|^2/2$ using an iterative quasi-Newton algorithm (see Vautard and Legras 1988). In order to find as many solutions as possible, the iterative procedure is initiated 2974 times, starting from each data point X_i. Typically, about 5% of all these experiments diverge to phase space areas where there are no data, some others converge to a minimum located at the periphery of the attractor, and most of them converge to a solution that possesses a large number of neighbors. Overall, about 100 convergent solutions are found, among which 10–20 will be retained as significant (see below).

b. Removal of nonsignificant solutions

In order to focus on solutions that correspond to frequently observed flows, we first remove those with a number of neighbors smaller than 10% of the number of data vectors; that is, $N_m = 298$ neighbors. This is consistent with the choice of the value of D, since N_m is, by construction of D, the average number of the neighbors within balls of radius D.

Next, the statistical test described in V90 is applied to the retained solutions (or minima) in order to filter out patterns X^* for which the low-cost value may be caused by chance only. This is done by considering the

TABLE 4. Characteristics of the numbers of solutions and groups of solutions for the nonlinear equilibration experiments as a function of the number of EOFs M. For each domain, the first line shows the number of solutions that pass the tests described in section 4a, and the second line shows the number of groups of solutions.

Value of M	1	2	3	4	5	6	7	8	9	10
Significant (ATL)	0	0	0	0	7	4	6	11	7	7
Groups (ATL)	0	0	0	0	2	2	3	4	3	3
Significant (PAC)	0	0	0	0	0	4	7	5	4	5
Groups (PAC)	0	0	0	0	0	1	2	3	2	2

null hypothesis that $T(X^*)$ is a random multinormal variable and calculating the 90% upper limit, $J_{90}(X^*)$, of the associated distribution of the cost. When $J_{90}(X^*)$ exceeds a constant value J^*, the lower 10% of the distribution of the cost values $J(X_i)$, the solution X^* is rejected.

Finally, a severe reproducibility test is applied. The whole procedure above is applied using the two halves of the data. We thus get three sets of solutions: W (using the whole dataset), $H1$ (using the first half of the data), and $H2$ (using the second half). An element of W is reproducible when there are solutions in $H1$ and in $H2$ such that the maximal distance between any pair of these three solutions is smaller than D. This guarantees that the three solutions can be found within a ball of radius D. In principle, the first test should also filter out nonreproducible solutions, and it does, in general. However, this test is parametric while the reproducibility test is not; it is thus more conservative and, moreover, removes solutions that are sensitive to climate changes or interdecadal variability.

For all performed optimizations, with various values of M, the number of remaining significant solutions is about 10. As explained in V90, some solutions may be close to one another. A simple hierarchical cluster analysis using the complete linkage between groups of solutions is then applied. The relevant number of solution groups is obtained by cutting the tree at the level of distance of D. In this way, groups contain solutions that are a distance less than D apart. In practice, a visual inspection of the solution patterns shows than, in general, solutions are really close to each other within classes: they correspond to noisy fluctuations at the bottom of the wells of the cost function. This sequence of removal tests is applied with $M = 1, 2, \cdots, 10$.

This rather complex procedure of filtering nonsignificant solutions is motivated by the desire to reduce the number of solutions—or classes of solutions—obtained by nonlinear equilibration. Unlike for cluster analysis, where the testing procedure clearly identifies an optimal number of classes, the number of solutions retained here is more subjective. We have also tried

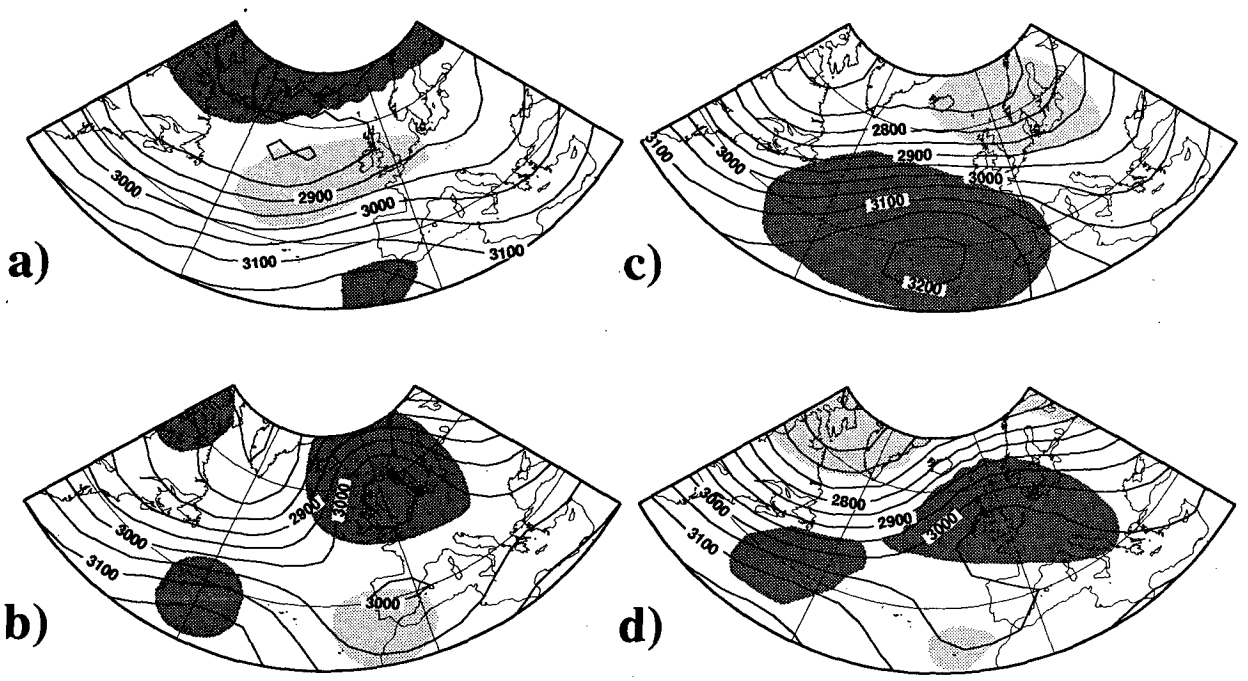

FIG. 12. The four most significant solutions of the nonlinear equilibration problem, for the ATL sector. Originally, these solutions are sought as anomalies. On this figure, the average wintertime flow is added. Contours and shades are as in Fig. 4.

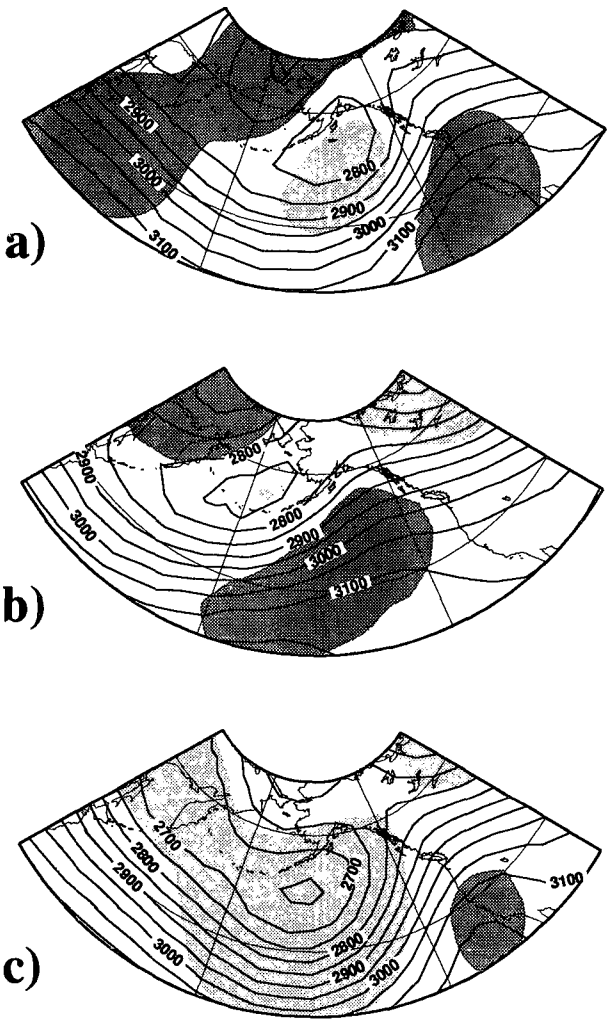

FIG. 13. Same as in Fig. 12 for the three solutions over the PAC domain.

different sequences of tests that retained about the same number of solutions; in general, the number of solutions varied by less than 1.

c. *Quasi-stationary flows*

Table 4 recapitulates the experiments performed with M varying from 1 to 10 for all experiments. First of all, for the ATL domain and $M < 5$, no significant and reproducible solutions were found; the same is true for the PAC domain with $M < 6$. This indicates that excessive filtering prevents one from finding phase space areas where the cost function is significantly lower than other areas. Tendency vectors are too homogeneously distributed. The number of clusters of significant solutions is maximum when $M = 8$ both for ATL (4 groups) and PAC (3 groups). For M greater than 8, the additional EOFs mainly represent baroclinic transient waves. As a consequence, solutions are more wavy and less reproducible. However, the patterns remain roughly the same as for $M = 8$. This justifies in fact our choice of $M = 8$ throughout this study.

For the ATL sector, 11 distinct significant solutions are found, gathering into 4 groups. The most significant solutions (in the sense of the V90 test) of each group are displayed in Fig. 12. Except the fourth solution, these patterns closely resemble the ones obtained in V90. The first solution is a high-amplitude version of the first pattern found with cluster analysis (compare with Fig. 4a). The second solution has a strong European blocking structure, with a closed anticyclonic cell over western Europe, which is responsible for the cold spells in this region. The third solution consists of an eastward extension of the Atlantic jet. Its amplitude is again larger than the corresponding pattern obtained from cluster analysis. The last pattern is a blocking ridge centered over western Europe and is not recovered by cluster analysis.

This last weather regime is the only one that does not correspond to any solution found in V90, where a regime characterized by another ridge extending from the mid-Atlantic to western Europe was found instead (regime AR). This discrepancy arises from the fact that there is, in phase space, a wide valley, containing both regimes, where the cost function is significantly small. Therefore, small variations in the values of the parameters of the method lead to significant displacements in the position of the ridge. Such is not the case for the other solutions, which are found in narrower wells of the cost function.

Over the PAC sector, five significant solutions are found, gathering into three groups, which are displayed in Fig. 13. The first two solutions are very different from those obtained by cluster analysis. The first (and most significant) solution is associated with a weak negative anomaly over the Gulf of Alaska, surrounded by two positive anomalies over the west Pacific and the California coast. This solution is rather similar to the P5 regime of KG93, although the amplitudes of the anomaly centers differ. The second one is approximately in quadrature with the first. The last solution is characterized by an extended strong negative anomaly centered near the Aleutian Islands. It corresponds to the second cluster found in section 3e(2) and the P1 regime of KG93, that is, the Aleutian low.

As shown in Table 4, the number of clusters is sensitive to the value of M. For instance, in the ATL sector with $M = 7$, the two European blocking patterns are not found, but instead one finds a solution with a weaker dipole anomaly displaced eastward. The two other flow regimes are recovered with great similarity. It can be explained by the fact that the 8th EOF, shown in Fig. 1, contributes significantly to the blocking dipole. The Greenland anticyclone is lost for $M < 7$. Using $M = 9$, all regimes displayed in Fig. 12 are recovered but the last one. Over the PAC sector, only the two most significant solutions displayed in Fig. 13 (first

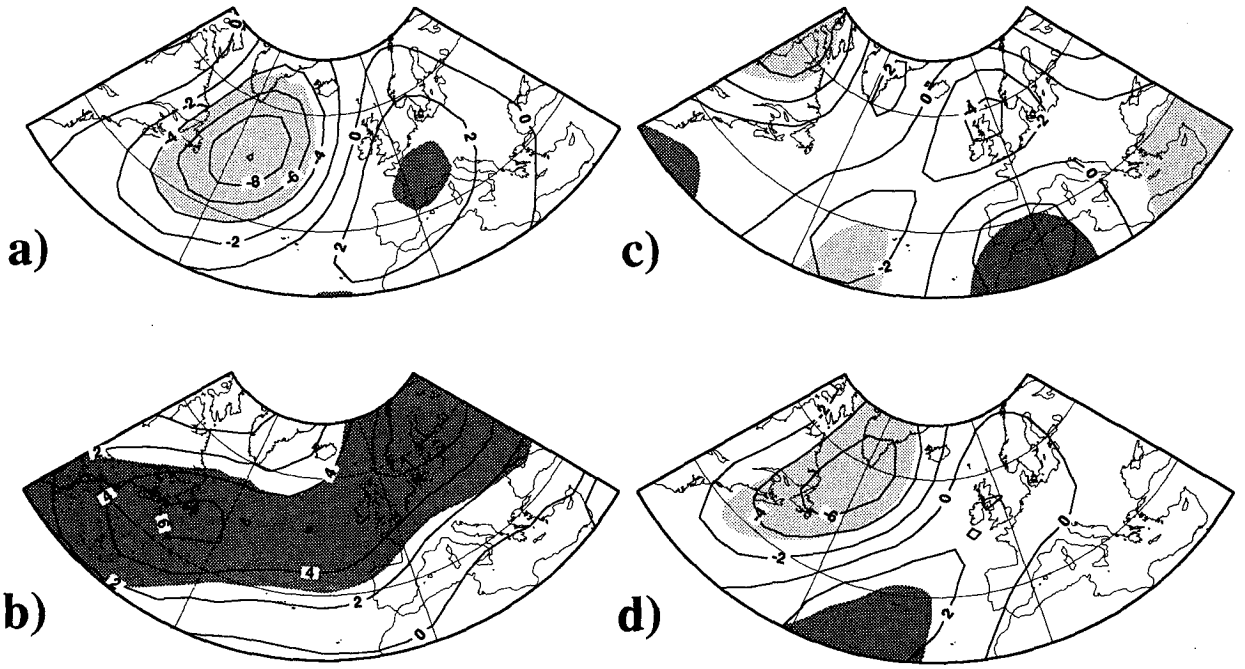

FIG. 14. Weighted composite tendency of the 700-hPa height associated with the centroids of the clusters displayed in Fig. 4. Contour interval is 2 m day^{-1}. Dark shaded areas correspond to areas where the composite is significantly greater than 0, at the 95% level. Light shaded areas correspond to significant negative tendency anomalies.

and second regime flows) are resilient to variations in M. This sensitivity of the nonlinear equilibration method to changes in the definition of the large-scale phase space is a weakness of the approach, which calls for a more objective definition of this concept. Here, objectivity is replaced by the choice of values of M that gave the best results, that is, the larger number of distinct significant groups.

d. Recurrent and quasi-stationary flows

The differences between the regime patterns obtained from cluster analysis and nonlinear equilibration result from the difference in the nature of the solved problem. Cluster analysis is designed to locate phase space regions where a higher concentration of data points exists. Nonlinear equilibration seeks the regimes for which there is a statistical balance between self-advection and transient feedback (Vautard and Legras 1988), leading to quasi stationarity. They do not need a priori to be the most recurrent. In order to show that the recurrent regimes found by cluster analysis are not in adequate balance, we display in Figs. 14 and 15 the composite tendencies associated with the centroids of the clusters found in section 3e, together with the areas where the tendency is significant at the 95% level. Since the composite tendency is a weighted average, the significance takes into account an equivalent number of degrees of freedom N_e, defined in V90 [Eq. (3.9)]. At a given physical space point x, the bounds of the 95% confidence interval of the composite tendency are given by $\pm 2s(x)/\sqrt{N_e}$, where $s(x)$ is the local standard deviation of the instantaneous tendencies, also calculated as a weighted average over neighbors of the solution at point x. We see from these two figures that the tendencies are significant in large areas. However, the standard deviation of instantaneous tendencies is of the order of 25 m day^{-1}. The tendencies displayed in Figs. 14 and 15 do not exceed 1/3 of this value, which means that at least 6 days are necessary, and much longer in general to create geopotential anomalies of the order of 50 m.

For the ATL cluster centroids (Fig. 14), only the composite tendency associated with the zonal regime is weakly significant, meaning that the centroid is the closest to being a quasi-stationary regime. Other tendencies display maximal values of the order of 5–8 m day^{-1}. The negative anomaly of the first cluster tends to be displaced northward. The positive anomaly of the second cluster tends to amplify along a ridge extending from Newfoundland to Scandinavia. The anomaly of the last cluster tends to deepen the Arctic low and to reinforce the ridge, leading to an amplification of the positive anomaly.

Over the PAC sector (Fig. 15), high tendencies are found for the Pacific blocking (cluster 3), which is not recovered among the solutions of the equilibration problem. The tendency displaces its anomaly pattern northward and creates a belt of midlatitude low geopotential. This westward drift was also observed in the

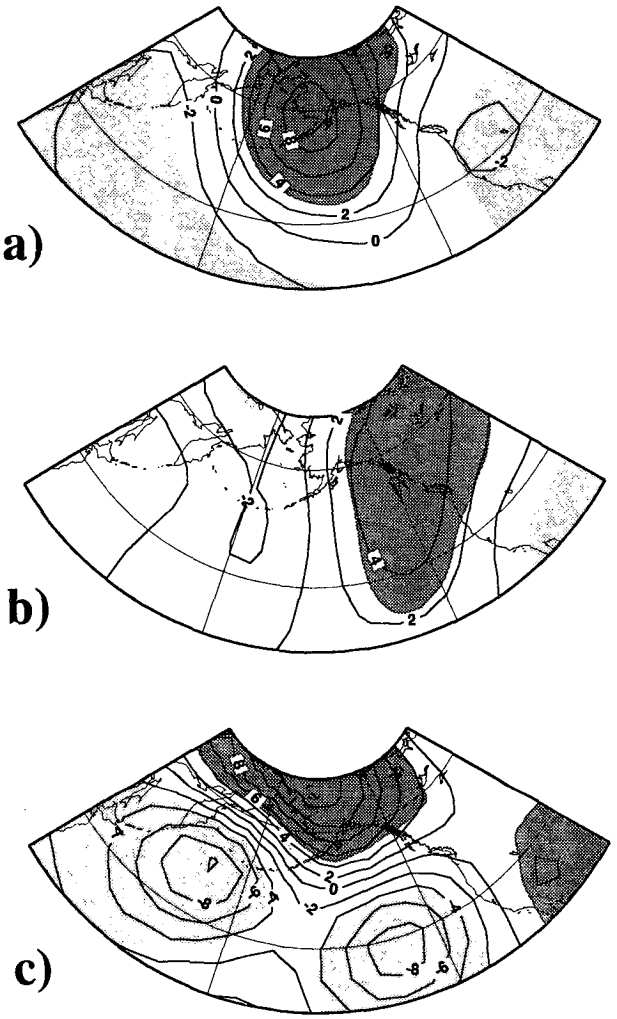

FIG. 15. Same as in Fig. 14 for the PAC cluster centroids of Fig. 7.

life cycle of a low-frequency traveling pattern by Branstator (1987) and Kushnir (1987). Not surprisingly, weak tendencies are obtained for the Aleutian low pattern, close to an equilibration regime. The zonal regime is somewhat damped and driven toward a negative PNA structure.

We now check whether cluster centroids lie near quasi-stationary states by starting the nonlinear equilibration algorithm (the quasi-Newton method) from the cluster centroids, instead of starting it from the raw observations. The first three ATL centroids converge toward their corresponding quasi-stationary states—that is, centroid 1 leads to solution 1, centroid 2 to solution 2, and centroid 3 to solution 3. The fourth cluster centroid leads to a nonsignificant solution. For the PAC area, centroid 1 leads to solution 1, centroid 2 also leads to solution 1, and the third centroid leads to a nonsignificant solution. One concludes that three of the ATL centroids and two of the PAC centroids lie in the attraction basin of quasi-stationary states, although their respective patterns differ significantly. Hence, for these latter clusters, there is a displacement of the concentration of points in phase space relative to the quasi-stationary states. By contrast, the other clusters do not lie near quasi-stationary states.

These results are further confirmed if one performs the dual analysis, that is, finds the cluster to which the quasi-stationary states belong. The first three ATL regimes belong to their respective clusters, and the fourth regime belongs to the cluster of the zonal flow. The weak and strong Aleutian low quasi-stationary regimes belong to the Aleutian low cluster, and the Pacific ridge belongs to cluster 1. The distances between quasi-stationary regimes and cluster centroids are nevertheless large.

5. Summary and further remarks

The main scope of this article was to analyze the differences between two different descriptive approaches of weather regimes. The first one relies on the definition of weather regimes as preferred, or recurrent, circulation patterns. Their identification is achieved by cluster analysis. Along this line, we present an application of a partitioning scheme called the dynamic clusters method to a 44-year set of 700-hPa geopotential heights. For a given number of clusters k, the partitioning algorithm seeks a partition that minimizes the variance within clusters. The main drawbacks of partitioning schemes are as follows.

(i) We need to explore several values of the number of clusters in order to find the most relevant one, whereas hierarchical methods are based on a tree that gives at the same time the partitions associated with all possible number of clusters.

(ii) Partitioning methods are initialized with random seeds drawn out from the data. Consequently, two calculations performed with the same sample do not necessarily yield the same partition.

We prefer, however, to use a partitioning method since the optimality property leads to more statistically stable results (see section 3). Difficulties i and ii are overcome by defining a classifiability index that is based on the dependence on initial seeds. This classifiability index is compared with the one obtained from applying cluster analysis to realizations of the same length of a multinormal noise process with the same covariance structure at lag 0 and 1 as our data. The reproducibility of clusters is tested, as in Cheng and Wallace (1993), by extracting a number of random halves of the data and comparing the results obtained with the whole dataset. Another test is to split the data into the first and last half and to perform the same comparison. Clusters that pass this test are not only robust in the statistical sense, but are also climatic features

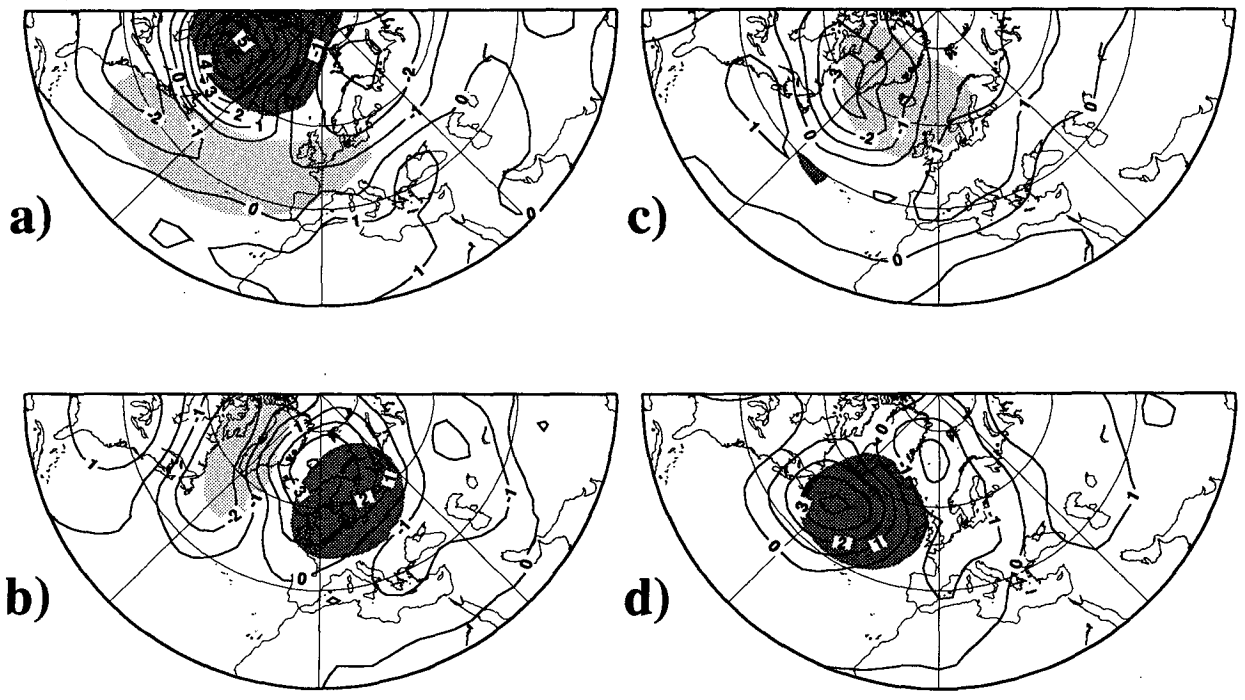

FIG. 16. Composite of the temperature anomaly at the 850-hPa level for the four ATL clusters of Fig. 4. Shaded areas are the same as in Fig. 4. Contour interval is 1°C.

unaltered by long-term climatic fluctuations such as interdecadal variability or trends.

In this study, no time filtering is applied to the dataset beforehand, so that instantaneous synoptic maps can be classified. However, data were first projected onto a reduced set of principal components that were chosen to select low-frequency variability. All data points are clustered, so that every map can be assigned to a regime. This property is convenient for long-range forecasting purposes (Maryon and Storey 1985), when a discrete representation of the atmospheric state is needed, or for studies of transitions between regimes. However, the frontiers of the clusters are not defined in a sense that is fully physical and indeed depend on the sample used as well as on the values of the parameters of the method.

The main results concerning cluster analysis can be summarized into four points.

1) The classifiability index is maximal with two clusters for both sectors but only differs significantly from that obtained from the noise model by four clusters for the Atlantic domain and three clusters for the Pacific domain.

2) The clusters associated with the above numbers are highly reproducible when the classification is conducted over the two halves of the data. Typically, 90% of the elements of one cluster obtained from the whole dataset lie within the corresponding cluster calculated from the first and second halves of the dataset. This also means that about 90% of the instantaneous maps can be classified without any ambiguity, the remaining part belonging to the less stable cluster boundaries.

3) The anomalies associated with the various regimes are consistent with the results of the sectorial analysis of Kimoto and Ghil (1993b) and Cheng and Wallace (1993). Only the number of patterns differs. We present evidence that this difference is mostly due to the difference of stability between hierarchical and partitioning methods. Partitioning methods converge toward definite clusters much more rapidly than hierarchical methods in the limit of the sample size going to infinity.

4) When imposing two clusters, the cloud of data points in phase space is essentially cut into two parts associated with positive and negative values of the first principal component. This result remains valid for the noise process. Noise cluster centroids are similar to atmospheric cluster centroids. This means that cluster analysis does not provide more information than standard EOF analysis in this case.

The second approach of weather regimes consists in finding quasi-stationary large-scale patterns. These patterns are defined as the patterns whose tendency vanishes on average over all their occurrences. Their practical determination is performed by nonlinear equilibration (Vautard and Legras 1988; Vautard 1990), applied to the same dataset as the one used for cluster analysis. The main outcome of this type of analysis and

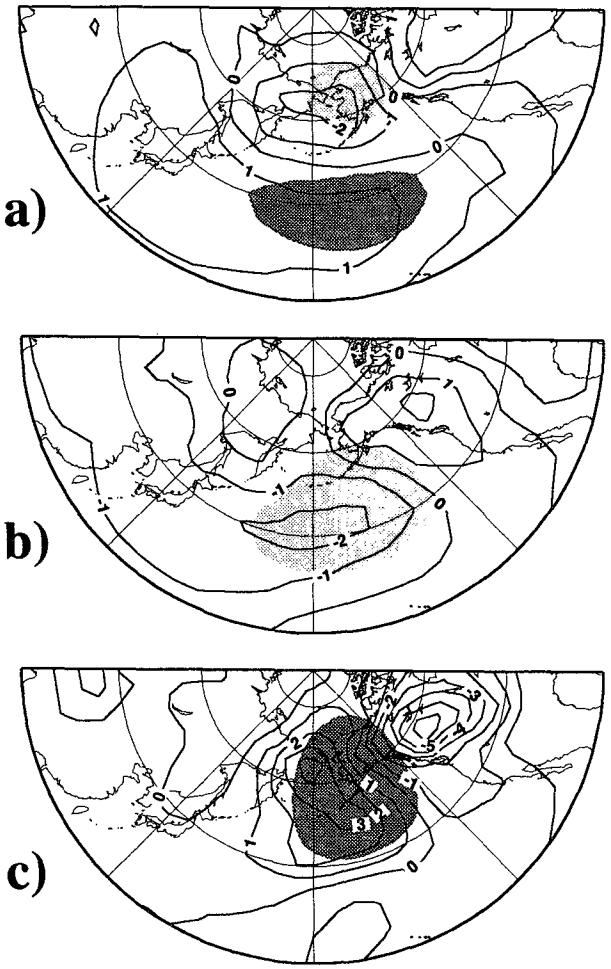

FIG. 17. Same as in Fig. 16 for the three PAC clusters of Fig. 7.

its differences with the cluster analysis approach are as follows.

1) Results are sensitive to the number of EOFs kept for the definition of the phase space spanned by large-scale patterns. A weakness lies, therefore, in this more or less arbitrary parameter. Here, eight EOFs are retained, and this choice is motivated by the fact that (i) it visually corresponds to the separation between large-scale EOFs and EOFs looking like baroclinic transients, and (ii) it gives the maximal number of weather regime patterns.

2) Four weather regimes are obtained over the Atlantic sector that have amplitudes higher than the centroids of the above clusters. Among these solutions, one finds a blocking dipole forming a closed cell over western Europe, which corresponds to synoptic situations associated with cold spells. Over the Pacific sector, three solutions are found, two of which are weak patterns corresponding to roughly opposite anomalies in the Gulf of Alaska, and a high-amplitude negative anomaly centered over the Aleutian Islands, projecting onto the positive phase of the PNA. No blocking pattern is found, meaning that Pacific blocking has a different dynamic from Atlantic blocking.

3) The relation between quasi-stationary and recurrent weather regimes has been explored by calculating the composite tendencies of the latter. It turns out that recurrent weather regimes evolve slowly when they are close to quasi-stationary ones. For instance, the Atlantic blocking cluster centroid, which exhibits much weaker amplitude than the associated quasi-stationary regime, tends to evolve toward amplification of the positive anomaly. In several instances, however—for example, the Greenland anticyclone, the Pacific zonal regime, and, above all, the Pacific blocking regime—the tendency is able to significantly alter the pattern within a few days. Legras and Ghil (1985) suggested that persistence is associated with unstable stationary solutions. We see that the real situation is somewhat more complicated. The association is seen in some cases but not in others. For some weather regimes, such as European blocking, the cluster centroid is somewhat displaced with respect to its quasi-stationary counterpart but lies in the same basin as the cost function minimized in the nonlinear equilibration algorithm. For other clusters, the centroid does not lie near any significant quasi-stationary solution.

The systematic tendencies observed presumably result in fact from the combination of self-interaction of the large-scale pattern and the systematical part of the transient feedback during occurrences of this large-scale circulation type. The dynamics acting on weather regimes or other large-scale flows cannot be studied using one-level geopotential heights but will be detailed in a future study through the use of a climatological set of potential vorticity maps on isentropic surfaces (Brunet et al. 1995).

Finally, we remark that regime anomalies do have a large influence on surface weather elements. Figures 16 and 17 show, for instance, the composites of the temperature fields at 850 hPa, obtained from the Atlantic and Pacific clusters. Although the 850-hPa temperature differs from the surface temperature because of inversion effects, it gives a qualitative idea about the expected variations at the surface from one weather regime to another. The anomalies are shifted westward slightly relative to the height anomalies, as expected from wind directions associated with the anomalies. The maximum amplitudes are in excess of about 3°C. Blocking-type regimes do not correspond to negative anomalies over western Europe. This is mostly due to the fact that during persistent anticyclonic episodes, dry conditions lead to infrared radiation usually forming a strong inversion layer near the ground, below the 850-hPa level.

Acknowledgments. This study was launched by Thierry Desponts and Bruno Piguet. It was carried on by François Bouttier and later by Thierry Mlynarz, who

helped a lot in implementing the cluster analysis algorithm into the above application. The dynamic clusters and the quasi-Newton algorithm were provided to us freely by the INRIA (Institut National de Recherche en Informatique et Automatique). We benefited from very constructive discussions with Michael Ghil. The heights data were kindly and freely provided to us by Kingtse Mo at the Climate Analysis Center, and the NCAR–NMC archive was sent by Roy Jennie at NCAR. We thank the three anonymous reviewers who helped to improve the quality of the manuscript and suggested some of the experiments. This work was partly supported by the French power company (EDF), and by the EC project "Climate of the 21st Century."

REFERENCES

Branstator, G., 1987: A striking example of the atmosphere's leading traveling pattern. *J. Atmos. Sci.,* **44,** 2310–2323.

Brunet, J. G., 1994: Empirical normal mode analysis of atmospheric data. *J. Atmos. Sci.,* **51,** 932–952.

——, R. Vautard, B. Legras, and S. Edouard, 1995: Potential vorticity on isentropic surfaces: Climatology and diagnostics. *Mon. Wea. Rev.,* **123,** 1037–1058.

Charney, J. G., and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.,* **36,** 1205–1216.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.,* **50,** 2674–2696.

Desbois, M., G. Sèze, and G. Szejwach, 1982: Automatic classification of clouds on METEOSAT imagery: Application to high-level clouds. *J. Appl. Meteor.,* **21,** 401–412.

Diday, E., and J. C. Simon, 1976: Clustering analysis. *Communication and Cybernetics 10 Digital Pattern Recognition,* K. S. Fu, Ed., Springer-Verlag, 47–94.

Dole, R. M., and N. D. Gordon, 1983: Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics. *Mon. Wea. Rev.,* **111,** 1567–1586.

Egger, J., and H.-D. Schilling, 1983: Stochastic forcing of planetary scale flow. *J. Atmos. Sci.,* **41,** 779–788.

Gordon, A. D., 1981: *Classification.* Chapman and Hall, 193 pp.

Hansen, A. R., and A. Sutera, 1986: On the probability density distribution of large-scale atmospheric wave amplitude. *J. Atmos. Sci.,* **43,** 3250–3265.

Hoskins, B. J., I. M. James, and G. H. White, 1983: The shape, propagation and mean-flow interaction of large-scale weather systems. *J. Atmos. Sci.,* **40,** 1595–1612.

Itoh, H., 1985: The formation of quasi-stationary waves from the viewpoint of bifurcation theory. *J. Atmos. Sci.,* **42,** 917–932.

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.

Kushnir, Y., 1987: Retrograding wintertime low-frequency disturbances over the North Pacific Ocean. *J. Atmos. Sci.,* **44,** 2727–2742.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking, and variations in atmospheric predictability. *J. Atmos. Sci.,* **42,** 433–471.

——, T. Desponts, and B. Piguet, 1987: Cluster analysis and weather regimes. *Proc. Workshop on the Nature and Prediction of Extratropical Weather Systems,* Vol. 2, Reading, Shinfield Park, UK, ECMWF, 123–149.

Maryon, R. H., and A. M. Storey, 1985: A multivariate statistical model for forecasting anomalies of half-monthly mean surface pressure. *J. Climate,* **5,** 561–578.

Metz, W., 1987: Transient eddy forcing of low-frequency atmospheric variability. *J. Atmos. Sci.,* **44,** 3–22.

Mo, K. C., and M. Ghil, 1988: Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.,* **93,** 10 927–10 951.

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

——, L. Ferranti, T. N. Palmer, and P. Viterbo, 1993: A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *J. Climate,* **6,** 777–795.

Plaut, G., and R. Vautard, 1994: Spells of low-frequency oscillations and weather regimes over the Northern Hemisphere. *J. Atmos. Sci.,* **51,** 210–236.

Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, 425 pp.

Reinhold, B. B., and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.,* **110,** 1105–1145.

Rex, D. F., 1950: Blocking action in the middle troposphere and its effect on regional climate. Part I: An aerological study of blocking action. *Tellus,* **2,** 196–211.

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

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

——, and B. Legras, 1988: On the source of midlatitude low-frequency variability. Part II: Nonlinear equilibration of weather regimes. *J. Atmos. Sci.,* **45,** 2845–2867.