CLUSTER ANALYSIS OF MULTIPLE PLANETARY FLOW REGIMES

Kingtse Mo
Climate Analysis Center
National Meteorological Center
Washington, DC 20233

Michael Ghil
Climate Dynamics Laboratory, Department of Atmospheric Sciences,
and Institute of Geophysics and Planetary Physics,
University of California, Los Angeles, CA 90024

Draft

March ~~February~~ 1987

Abstract.

A modified cluster analysis method has been developed to identify spatial patterns of planetary flow regimes, and to study transitions between them. This method has been applied first to a simple deterministic model and second to Northern Hemisphere (NH) 500 mb data.

The dynamical model is governed by the fully-nonlinear, equivalent-barotropic vorticity equation on the sphere. Clusters of points in the model's phase space are assocaited with either a few persistent or with many transient events. Two stationary clusters have patterns similar to unstable stationary model solutions, zonal or blocked. Transient clusters of wave trains serve as way stations between the stationary ones. tr

For the NH data, cluster analysis was carried out in the subspace of the first seven empirical orthogonal functions (EOFs). Stationary clusters are found in the low-frequency band of more than 10 days, and transient clusters in the band-pass frequency window between 2.5 and 6 days.

In the low-frequency band three pairs of clust~~g~~ers determine, respectively, EOFs 1, 2 and 3. They exhibit well-known regional features, such as blocking, the Pacific/North American (PNA) pattern and wave trains. Both model and low-pass data show strong bimodality.

Clusters in the band-pass window show wave-train patterns in the two jet exit regions. They are related, as in the model, to transitions between stationary clusters.

/all caps, Center

< 1. Introduction >

It is well known that certain large-scale atmospheric circulation patterns persist for time intervals longer than those typical of midlatitude cyclones [*Baur*, 1947; *Namias*, 1982]. A few of these patterns also have a tendency to recur from time to time. To identify the patterns which tend to both recur and persist, as well as determine preferred transitions between them, can deepen our knowledge of low-frequency atmospheric variability and enhance our skill in long-range forecasting (LRF).

Recurrent and persistent patterns can be global, hemispheric or regional. Certain patterns associated with specific phases of the El Niño/Southern Oscillation are known to be global [*Rasmusson and Wallace*, 1983]. Most blocking episodes, in both the northern and the southern hemispheres, are regional in character [*Dole*, 1986; *Dole and Gordon*, 1983; *Trenberth and Mo*, 1985]. Typical of hemispheric patterns are those associated with the dominance of zonal wavenumbers three or four in the Southern Hemisphere [*Mo*, 1986; *Mo and Ghil*, 1987].

In this article, we shall concentrate on recurrent and persistent patterns of hemispheric extent, associated with the atmospheric circulation in the Northern Hemisphere (NH) extratropics. These patterns will be studied first in the solutions of a greatly simplified dynamical model, and then in an atmospheric data set.

One way to identify hemispheric patterns which persist is the pattern correlation method (PCM). A sequence of daily hemispheric weather maps is defined to constitute a persistent or *quasi-stationary* (QS) event, if the spatial correlation between any pair of maps within the sequence exceeds a given threshold p_o, say $p_o = 0.5$, and if the duration of the event so defined also exceeds a given threshold. Based on the ensemble-mean decorrelation time of daily weather maps, typical duration thresholds for QS events are seven days in the Northern, and five days in the Southern Hemisphere.

Using this criterion, *Horel* [1985a] identified 58 QS events in a set of NH winter data. These events were not easy to classify subjectively into a small number of categories, due to the apparent diversity of their spatial patterns. In the Southern Hemisphere, *Mo* [1986] classified most of 23 QS events into three major categories by visual inspection: two were dominated by a planetary wave of zonal wavenumber three, but with nearly opposite phases, one by zonal wavenumber four. The question we are asking here is to what extent can purely objective, statistical criteria be used to classify QS events into a usefully small number of categories, and how can these categories, or *flow regimes,* be used in LRF.

Recently the authors [*Ghil,* 1987; *Mo and Ghil,* 1987] have considered systematic connections between the statistical and dynamical methods of description and prediction of QS events. They found that, both in the solutions of simple dynamical models and in atmospheric data sets, the first few empirical orthogonal functions (EOFs) had patterns similar to the most frequently occurring QS events. This could be explained by the fact that these EOFs pointed to the largest concentrations of invariant measure in the system's phase space, which were also the locus of the QS events. Such a result had to be expected from the ergodic theory of dynamical systems [*Eckmann and Ruelle*, 1985; *Ghil and Childress,* 1987, Sections 6.4 and 6.6; *Ghil et al.*, 1985, pp. 14-16], but the amount of specific information extracted for a complex system like the Earth's atmosphere appears rather gratifying.

Still, the direct and exclusive use of EOFs in classifying QS events has two main disadvantages. First, the spatial orthogonality imposed on the flow patterns associated with each class, or flow regime, is an oversimplification, from relative dynamical independence to complete lack of statistical correlation. Second, the spatial patterns of EOFs become successively more complicated, with smaller and smaller scales present as the variance associated with them decreases. Higher EOFs can therefore not be expected to resemble any large-scale QS events, nor classes of such events.

The *purpose of this article* is to develop and apply an objective method for the classification of QS events into a few *planetary* flow regimes, and to examine transitions between these regimes. We develop a modified cluster analysis method and apply it to two types of data sets. One is obtained from extended integrations of a very simple, deterministic, nonlinear model of NH flow [*Legras and Ghil*, 1985; *Ghil and Childress*, 1987, Section 6.4]. The other is a set of 500 mb geopotential height maps for NH winter.

In Section 2, we describe the two data sets, and in Section 3 we present the method. Results are reported in Section 4 for the simple model and in Section 5 for the NH 500 mb data. Conclusions follow in Section 6.

2. DATA SETS AND THEIR PREPARATION

Model data

Following the approach of *Mo and Ghil* [1987], we first develop and check our statistical methodology on a data set with a simpler structure, generated by a nonlinear deterministic model. To the extent that model solutions are time-dependent and actually aperiodic, they exhibit sufficient irregularity to justify a statistical treatment, as explained by *Ghil* [1987].

The model is governed by the equivalent-barotropic form of the equation for the conservation of potential vorticity on a sphere [*Ghil and Childress*, 1987, Chapters 3 and 6; *Legras and Ghil*, 1985], truncated to 25 spherical harmonics., It has ~~a~~ forcing by a zonal jet, Ekman dissipation and a simplified topography of zonal wavenumber two representing two equal continental masses and two equal oceans. Subsequent maps of the model's streamfunction fields use polar stereographic projection from the North Pole onto a full disk. The position of the model continents is indicated by heavy lines, on the periphery of this disk. The distribution of land masses resembles, albeit schematically, more that of the Northern than that of the Southern Hemisphere, and equatorial symmetry makes this essentially a NH model.

In previous publications, the dependence of model behavior on various parameters was carefully investigated. Here it suffices to use one set of parameter values, which is both realistic [T. P. Barnett and J. O. Roads, personal communication, 1986] and at the center of the region in parameter space where interesting solution behavior obtains. The forcing parameter ρ, giving the intensity of the zonal jet, is set to the value $\rho_m = 0.211$, the dissipation parameter α to a value corresponding to the relaxation time of $\alpha^{-1} =$ 20 days, and the height of the topography to a nondimensional value of $h_o = 0.1$, relative to atmospheric scale height.

For this value of the parameters, a model integration of roughly 65 years of simulated time was used. More precisely, this corresponds to a time interval of 8000 τ, where the sampling time τ equals 1.5 nondimensional time units, which is 3.0 days at $\rho = 0.20$ and 2.83 days at our value of $\rho = \rho_m$. The first solution segment of 1000τ was omitted so as to make the results independent of initial data. The time mean was computed by averaging over $1001\tau \leqslant t \leqslant 8000\tau$, and the streamfunction anomaly at a given time is defined as the deviation from this time mean.

The persistence properties of model solutions for $\rho = \rho_m$ were discussed in *Mo and Ghil* [1987]. Pattern correlations $p(t+m\tau, t+n\tau)$ for pairs of maps were computed between each given time $\underline{t}$ and 5 consecutive sampling times after that

$0 \leq m < n \leq 5$. QS events were identified by requiring that the pattern correlations between all maps within a series of 6 or more be larger than or equal to $p_o = 0.5$. If $p[t_o+m\tau,\ t_o+n\tau] \geq p_o$ for $0 \leq m < n \leq 5$, and $p[t_o+\tau,\ t_o+6\tau] < p_o$, then the event is said to last just 5 sampling intervals. On the other hand, if $p(t_o+m\tau,\ t_o+n\tau) \geq p_o$ for both $0 \leq m < n \leq 5$ and $1 \leq m < n \leq 6$, the QS event lasts 6 intervals, and so on. Maps for all QS events were plotted.

Most events were easily classified subjectively by their similarity to one or another of the model's stationary solutions. All such solutions are unstable at this point in parameter space, but some of them generate persistent events in their phase-space neighborhood by a mechanism explained in previous publications [e.g., *Ghil*, 1987, Fig. 15]. The events with longest durations had patterns resembling either blocking (Figures 1a,c) or a zonal type of flow (Figures 1b,d). [Fig. 1 near here, please]

From the time series of streamfunction anomaly coefficients, standardized by the variance in time of each coefficient, the correlation matrix was computed, and diagonalized by EOF analysis. In contradistinction from *Mo and Ghil* [1987], EOFs will be used here only for spatial filtering purposes. As explained in the Introduction, retaining only a small number of EOFs, those associated with the highest variances, will result in smoother large-scale fields, which presumably contain the signal of the system's variability. It is these filtered fields on which cluster analysis will be performed.

Atmospheric data

The data set consists of twenty years' worth of twice-daily 500mb geopotential height maps analyzed by the U.S. National Meteorological Center (NMC) from January 1963 to December 1982. Spectral analysis was used to remove the seasonal cycle at each grid point. The seasonal cycle is defined here as the

20-year mean plus the 20-th and 40-th Fourier components of the time series. Anomalies were then computed as differences between the data and this seasonal cycle.

For greater conformity with the bulk of the existing literature on persistent anomalies in the NH extratropics, we concentrate in this article on the winter only. The winter season here is taken as the 120 days from November 15 to March 15.

Pattern correlations between pairs of maps one day apart were computed. QS events were identified by requiring that pattern correlations between the pairs of maps on 5 consecutive days be not less than 0.5.

After detrending the time series of anomalies, we filtered this series in time by using separately a low-pass filter and a band-pass filter, as designed by *Blackmon* [1976]. The low-pass filter has a large-amplitude response for frequencies $0 < f \lesssim 0.1$ day^{-1}, i.e., for periods 10 days $\lesssim T < \infty$. Due to the removal of the seasonal cycle and then retention of winter data only, the variability in this window reflects in fact two separate bands: 10 days $\lesssim T \lesssim$ 100 days and 300 days $\lesssim T < \infty$. The band-pass filter is sensitive to frequencies $0.17 \lesssim f \lesssim 0.4$ day^{-1}, i.e., for periods of 2.5 days $\lesssim T \lesssim$ 6 days. The two filters have rather sharp cutoffs and little overlap, so that the two windows of low and intermediate frequency are well separated. Cluster analysis will be carried out separately for the low-pass filtered and band-pass filtered data.

/the

The spatial filtering started by retaining only data at 358 points out of NMC's 541-point NH grid. This grid of 358 points achieves a compromise between a regular latitude-longitude distribution and an, unfortunately inexistent, uniform-spacing distribution [*Barnston and Livezey*, 1987]. Anomalies at 305 points of this grid lying between 20N and 70N were standardized by the variance in time at each grid point, and the correlation matrix for the corresponding time series was calculated.

Table 1 gives percentages of variance for each EOF, separately in the low-pass and band-pass windows. The expected error in this estimate of variance, also given in the table, was evaluated by the heuristic formula of *North et al.* [1982]

$$\delta\lambda/\lambda = [2/N]^{1/2} \quad .$$

Here $\delta\lambda$ is the standard deviation for eigenvalue λ and N is the number of independent samples. We took N=200 for the low-pass filtered data and N=400 for the band-passed time series. This is rather conservative, since the total number of samples is 2x120x20=4800, and the decorrelation time, as we shall see, is less than 10 days in the first band (Table 6) and about 3 days in the second.

[Table 1 near here, please]

Convergence of the EOF expansion is slow in both windows. For the band-pass window, 15 EOFs only give 43% of the total variance. In the low-pass window, seven EOFs give 50% of the variance, and they will prove sufficient for our analysis of low-frequency variability. The convergence for model data is much more rapid, due to their limited spatial resolution and simplified dynamics [compare Table 4 in *Mo and Ghil* , 1987].

3. METHODOLOGY

Probability density estimation

As indicated already in Sections 1 and 2, deterministic, but nonlinear dynamics can generate time series of geophysical flow fields with an appearance or randomness [*Ghil et al.*, 1985]. This randomness is associated, heuristically speaking, with a measure, or probability density function (pdf), which is invariant under the equations describing the dynamics. These equations are also said to generate a flow in the system's phase space, as each point in this space can be thought of as moving, or flowing, in time along the unique orbit, or trajectory, passing through it [*Ghil and Childress*, 1987, Section 5.4].

With this terminology, the pdf is said to be invariant under the flow. What is meant is simply that, if a set of points A in phase space is carried by the flow into a set B, then the measure, i.e., the total or cumulative probability, of the two sets is equal. For conservative dynamical systems, it is well known that such an *invariant measure* exists, is essentially unique and is just equal to the ordinary volume in the system's phase space. This result goes usually under the name of Liouville's theorem.

For the forced, dissipative systems one encounters in geophysical fluid dynamics (GFD), the situation is somewhat more complicated. The flow in phase space is volume-reducing, rather than volume-preserving, and tends in general to a strange attractor [*Lorenz*, 1963; *Ghil and Childress*, 1987, Section 5.4]. A measure on such an attractor is known to exist under certain simplifying mathematical assumptions, called Axiom A, which essentially state that for every point on the attractor the linearization of the flow has no neutrally stable directions. Requiring that the measure behave essentially like length along the unstable directions renders it also unique. Furthermore, this unique measure is ergodic for almost all points near the attractor, as well as on the attractor. That means that any physically or numerically observable time averages along trajectories starting on or near the attractor will be equal to the corresponding ensemble average with respect to the pdf on the attractor [*Eckmann and Ruelle*, 1985, pp. 639-641, and references therein; *Ghil et al.*, 1985, pp. 14-16].

The main point of our line of investigation is that this pdf is far from being either uniform or isotropic in the phase space of large-scale atmospheric flows. The most elementary form that such inhomogeneities can take is bimodality. In the case of the relatively simple phase-space flow induced by our model (Section 2), such bimodality is clearly a result of the greatly enhanced persistence near the two unstable stationary solutions with blocking and zonal-flow patterns, respectively. The approximate form of the pdf near these two generalized saddle points in phase space can actually be derived from the linearization of the equations at these points, and a small number of scaling parameters for the pdf can be determined from the data.

For the NH data set no such a *priori* form for the pdf can be derived, and one has to use the tools of *nonparametric estimation* theory. This is a particularly active field of modern statistics, which relies on an intelligent and systematic use of computer power rather than on "cookbook" formulae valid only for elementary problems involving well-known, classical pdfs. The methods of nonparametric theory permit the reliable estimation of differences between mean and median of an arbitrary distribution [*Efron*, 1982] or of the multimodality of a pdf [*Silverman*, 1986; *Tapia and Thompson*, 1978], based on samples of moderate size.

The first nonparameteric method we used is discrete maximum penalized likelihood estimation (DMPLE) [*Silverman*, 1986, Section 5.4; *Tapia and Thompson*, 1978, Chapter 5], which estimates a univariate pdf $\omega = \omega[z]$, subject to a smoothness constraint. Some such constraint, or *regularization*, is necessary for any consistent, stable estimation from noisy data. It plays a role similar to lag windows and frequency tapers in spectral analysis.

The likelihood function maximized is

Cap Greek "pi" (not l.c.)

$(\omega_{(m)})$

$$L(\omega_{(m)}) = \pi_{i=1}^{N} s(z_i) \exp\left\{-(\alpha/h) \sum_{j=1}^{m} (y_j - y_{j-1})^2\right\}, \qquad (1a)$$

$/z_i$

subject to

$$h\sum_{j=1}^{m} y_j = 1 \quad ; \quad y_j \geqslant 0 \quad , \quad j = 1, \ldots, m = 1 \ . \qquad \text{(1b,c)}$$

Here $\omega_{(m)} = \{y_o, y_1, \ldots, y_m\}$ is an approximation to the true pdf $\omega(z)$ at $m + 1$ equally spaced "nodes", or mesh points, with $y_o = y_m = 0$ since $\omega(z)$ is assumed to be zero outside the interval considered; $\underline{h}$ is the equal spacing between nodes, and $s(z_i)$ is the interpolation by linear splines, defined with respect to the $\underline{m}$ equal subintervals, of the $\underline{n}$ unequally spaced data z_i, where $n \geqslant m$ necessarily and usually $n \gg m$. The sum in the exponent is an approximation to $\int(d^2\omega/dz^2)dz$, and yields a smooth pdf estimate $\omega_{(m)}$ by minimizing the "wiggles" of $\omega(z)$ [*Tapia and Thompson*, 1978, Chapters 4 and 5].

The variable $\underline{z}$ chosen was a leading principal component of the data set of e/ int$\overset{e}{\wedge}$rest, as explained in Eq. (2) below. The maximization was carried out by the algorithm NDMPLE from the International Mathematical Statistical Library (IMSL). This algorithm was also used by *Benzi et al.* [1986] and *Sutera* [1986] on NH winter data for December 1980 - February 1984, who chose the sum of the squared amplitudes of zonal wavenumbers two, three and four as the unique variable $\underline{z}$ of their pdf $\omega(z)$.

We took $m = 40$ and $h = 0.2$, so that the total interval over which $\omega(z)$ is allowed to be nonzero equals 8 standard deviations of the variable of our choice. The smoothness parameter α was chosen by requiring the discrepancy between the estimated pdf $\omega_{(m)}(\alpha)$ and the theoretical limit pdf as $\alpha \to 0$, $\omega_{(m)}(0)$, which is an atomic measure concentrated at the m+1 equidistant mesh points, to satisfy the Kolmogorov-Smirnov (K-S) test at a confidence level of 95% [*Darling*, 1957; *Sutera*, 1986]. The K-S test is distribution-free, i.e., it is independent of the shape of the pdf $\omega(z)$ approximated, provided ω is a continuous function of

$\underline{z}$ [*Fisz*, 1963, Section 10.11]. Robust approaches to an estimation of the regularization parameter α from the data involve various resampling plans, which are much more expensive computationally [*Efron*, 1982; *Wahba and Wendelberger*, 1980].

Cluster analysis

The major drawback of the DMPLE approach to pdf estimation in phase space is that its extension to more than one variable is still prohibitively expensive. Bimodality with respect to a single dimension is an important first step in identifying multiple planetary flow regimes. But much more detail is needed to use these regimes effectively as a foundation for LRF.

We turned therewith to cluster analysis [*Anderberg*, 1973; *Silverman*, 1986, Section 6.2], which is a flexible multivariate approach. To classify QS events objectively by the similarity of their flow patterns, one needs a quantitative measure of similarity. In *Legras and Ghil* [1985] root-mean-square distance between maps was used to study the proximity of persistent events to unstable equilibria. Pattern correlations of anomalies correspond to the cosine of the angle, centered at the time mean, of two maps seen as points in phase space, rather than to their Euclidean distance. This measure is more sensitive to the meteorologically significant shape and phase of anomalies [*Horel*, 1985a; *Mo*, 1986; *Mo and Ghil*, 1987]. It was already used to identify QS events, and we used it for our cluster analysis.

We expand the time series $\phi(\underset{\sim}{x},t)$ of anomaly fields into EOFs

$$\phi(\underset{\sim}{x},t) = \sum_{\nu=1}^{\nu_0} A_\nu(t)\, E_\nu(\underset{\sim}{x}) \quad , \qquad (2)$$

where $\underset{\sim}{x}$ is the spatial coordinate, E_ν is the ν-th eigenfunction of the correlation matrix, in decreasing order of the associated eigenvalues λ_ν [see Table 4 in *Mo and Ghil,* 1987, and Table 1 here], A_ν is the corresponding *principal component* [PC], and ν_o is the truncation of the EOF expansion selected for smoothing purposes [*Barnett and Preisendorfer,* 1978]. The pattern correlation between $\phi(\underset{\sim}{x}, t_m)$ and $\phi(\underset{\sim}{x}, t_n)$, so truncated, is then given as

$$p(t_m, t_n) = \sum_{\nu=1}^{\nu_o} A_\nu(t_m) A_\nu(t_n) , \qquad (3)$$

due to the orthonormality of the EOFs.

When an anomaly (2) is small in magnitude, the pattern correlation between it and another anomaly cannot be expected to be meaningful. It was found useful therefore to define a cluster of small anomalies, for which the distance to the origin

$$d(t_n) = \left\{ \sum_{\nu=1}^{\nu_o} A_\nu^2 (t_n) \right\}^{1/2} \qquad (4)$$

is below a given threshold d_0.

For a cluster $C = \left\{\phi_1, \ldots, \phi_n\right\}$ whose elements ϕ_j are anomaly maps $\phi(\underset{\sim}{x}, t_{n_j})$, we define the center $\bar{c}$ as the arithmetic mean of its elements,

$$\bar{c} = \frac{1}{n} \sum_{j=1}^{n} \phi_j \quad . \qquad (5a)$$

Each element is interpreted as a vector in ν_o-dimensional Euclidean space, $\phi_j = \left[A_1(t_{n_j}),\ldots, A_{\nu_o}(t_{n_j})\right]$, so that Eq. (5a) is equivalent to $\bar{c} = \left[\bar{A}_1,\ldots, \bar{A}_{\nu_o}\right]$, with

$$\bar{A}_\nu = \frac{1}{n} \sum_{j=1}^{n} A_\nu(t_{n_j}) \quad . \tag{5b}$$

Clustering criteria

All clustering algorithms require two basic criteria: one to determine membership in a cluster, otherwise each point would form a cluster; the other to determine separation between clusters, otherwise all points would form a single cluster. In general, these criteria are chosen so that each point belongs to one and only one cluster - *hard* clustering, or so that each point belongs to one or more clusters- *fuzzy* clustering [*Bezdek*, 1981].

In our application to large-scale atmospheric flows, it is quite clear from synoptic experience that sizable portions of phase space are visited only very rarely, so that considerable numbers of anomaly maps will be distributed quite thinly over these portions. There is no use in trying to associate these thinly populated portions of phase space with any planetary flow regime, as they are most unlikely to recur and will not help in any substantial way to either understand or predict low-frequency variability. Eliminating thus a considerable number of points from the clustering procedure will enhance the convergence rate of any specific clustering algorithm we choose. We depart therewith from other clustering approaches by formulating a third criterion, for non-membership in any cluster. Alternatively, this can be thought of as a criterion for membership in a special, larger cluster of nonrecurrent flow anomalies, into which the really interesting clusters are embedded.

Recalling the other special cluster, of small anomalies, the five criteria we use are:

a) *Membership criterion.* The pattern correlation between the center of a cluster $\bar{c}$ and any element in the cluster ϕ_j should exceed a threshold r_1, /tr

$$p(\bar{c}, \phi_j) = \sum_{\nu=1}^{\nu_o} \bar{A}_\nu A_\nu^{(j)} \geqslant r_1 \quad . \tag{6a}$$

Remembering the interpretation of $p(\phi', \phi'')$ as the cosine of an angle in phase space, requiring for instance that $r_1 = 0.86$ means that any two elements ϕ' and ϕ'' in a given cluster form an angle smaller than 60^o with the origin, i.e., that they correlate better than 0.5. We shall use $r_1 \geqslant 0.8$.

b) *Separation criterion.* The pattern correlation between the centers of two clusters, $\bar{b}$ and $\bar{c}$, say, should not exceed a threshold r_2,

$$p(\bar{b}, \bar{c}) = \sum_{\nu=1}^{\nu_o} \bar{B}_\nu \bar{C}_\nu \leqslant r_2 \ . \tag{6b}$$

To prevent points from belonging to more than one cluster, we require that $\arccos r_2 > 2\arccos r_1$. We shall use $r_2 \leqslant 0.45$, which satisfies this requirement for the lowest value of r_1.

c] *Exclusion criterion.* If a map ϕ does not correlate sufficiently well with the center $\bar{c}_k$ of any cluster,

$$p(\phi, \bar{c}_k) < r_1 \quad , \tag{7a}$$

and it does not satisfy the separation criterion for at least one cluster, $\bar{c}_{k_o}$

say,

$$p(\phi, \bar{c}_{k_o}) > r_2 \quad , \qquad (7b)$$

then ϕ belongs to the nonrecurring cluster. Direct observational evidence for extratropical flows in either NH winter [*Charney et al.*, 1981] or SH winter [*Trenberth and Mo*, 1983] suggests that this cluster should take up about 2/3 of all anomaly maps analyzed.

d) *Small-anomaly criterion.* A map $\phi(\underset{\sim}{x}, t_n)$ belongs to the small-anomaly cluster, rather than to one of the significant clusters defined by (6a) or to the special, nonrecurrent cluster defined by (7), if its distance (4) to the origin satisfies

$$d(t_n) \leqslant d_o = \bar{d} - 1.8\ \sigma_d \ , \qquad (8)$$

where $\bar{d}$ is the mean distance of the time series of anomalies to the origin and σ_d is the variance of the distances about this mean [*Mo and Ghil*, 1987].

e) *Small cluster criterion.* Clusters with less than L_o elements are assigned to the special, nonrecurrent cluster. L_o is taken as 25 for the model results and as 8 for the NH data. This accelerates the search and eliminates nonsignificant clusters.

The schematic diagram of our clustering criteria is given in Figure 2. The exact search and clustering algorithm is given in Appendix A. [Fig. 2 near here, please]

4. MODEL RESULTS

Clusters

For the time series of 7000 streamfunction anomaly maps based on 25 spherical harmonics, ten EOFs contain 91% of the total variance. Table 4 in *Mo and Ghil* [1987] shows that the first three, in fact, already contain 65%.

Cluster analysis using ten EOFs, $r_1 = 0.85$ and $r_2 = 0.4$, yielded five clusters. Using 12 EOFs, and varying r_1 between 0.8 and 0.86, with $0.34 \leqslant r_2 \leqslant 0.45$, yielded the same number of clusters. The flow patterns of their centers stayed much the same, only the number of elements in each cluster varied from one set of criterion values to another.

The clusters for ten EOFs, $r_1 = 0.85$ and $r_2 = 0.4$, are listed in Table 2, in decreasing order of the number of elements. The distribution of persistence for passages within each cluster is given. Flow patterns of the centers of each cluster are shown in Figure 3.

[Table 2 and Figure 3 near here, please]

Clusters 1 and 2 are largest, with about 11% of the total number of points each. They are also the most stationary, being the only ones with a significant number of flow sequences persisting for longer than four sampling times within the cluster. Cluster 1 (Figure 3b) resembles clearly the model's zonal flow (compare Figure 1c), while Cluster 2 (Figure 3c) is associated with blocking (Figure 1d).

In *Mo and Ghil* [1987] we saw that the first EOF was nearly parallel to a line segment extending from the unstable blocking equilibrium E_m to the unstable zonal equilibrium B_m. This is now fully explained by the closeness of the dominant Clusters 1 and 2 to the respective unstable equilibria. Both $\bar{c}_1$ and $\bar{c}_2$ have indeed their largest components along EOF 1, with signs opposite to each other (Figure 4).

[Figure 4 near here, please]

The detailed distribution of persistence times in Clusters 1 and 2 shows that rapid passages through these stationary clusters are still the most frequent. In general, we expect the persistence time in either cluster to be most strongly correlated with distance between the points of the sequence and the unstable equilibrium nearby, cf. *Legras and Ghil* [1985, Figures 10, 13 and 16).

To verify this statement, we computed Euclidean distances between points in each flow sequence of a cluster, and the center of that cluster, $\bar{c}_1$ or $\bar{c}_2$, as the case may be. We took the minimum distance, d_{min}, corresponding to each sequence, and we averaged over all sequences with the same duration, in each cluster separately. The resulting values $\bar{d}_{min}$ are listed, as a function of sequence duration, in Table 3.

[Table 3 near here, please]

The values of $\bar{d}_{min}$ are increasing in general as the duration of the associated sequences decreases: in Cluster 1 from 0.45 for 170τ to 2.32 for 2τ, and in Cluster 2 from 0.82 for 34τ to 1.72 for 2τ. The increase is not perfectly monotonic, due to variations in the direction of approach to the unstable equilibrium and in the direction of ejection from its vicinity. But the correlation between close passage and long persistence is clearly excellent.

The pattern correlation method (PCM), as defined in Section 2, identifies only some of the most persistent passages as QS events. First, the passages have to last 5τ or longer. Second, the membership criterion of r_1 = 0.85 allows correlations between pairs within a cluster to be smaller than ρ_0 = 0.5. We shall return to a systematic comparison between the two approaches later in this section.

Clusters 3 (Figure 3d) and 4 (Figure 3e) have similar anomaly patterns, but with nearly opposite phases. Cluster 4 resembles the wave-train pattern obtained by the correlation method in *Mo and Ghil* [1987, Figure 9c]. Table 4 there indicates that this pattern has similarities with EOF 2, and Figure 4 here shows that $\bar{c}_3$ and $\bar{c}_4$ do have large components of opposite signs along this EOF, as well as along EOF 1. While EOF 1 is essentially determined by the two dominant clusters, the orthogonality constraint on EOF 2 prevents it from being uniquely determined by Clusters 3 and 4.

Cluster 5 (Figure 3f) is both the smallest and the least persistent, with only three sequences lasting 2τ and none longer. It resembles the model's unstable Zonal 2 equilibrium. All five clusters are well separated, the highest correlation between centers being $p(\bar{c}_2, \bar{c}_5) = 0.38$.

Correlating the mean $\bar{\Phi}$ of the time series with the anomaly maps of the centers of the clusters, $\bar{c}_k$, $k = 1,2,\ldots,5$, yields the largest correlation for $k = 1$, the zonal-flow cluster, $p(\bar{\Phi}, \bar{c}_1) = 0.73$. The correlations between $\bar{\Phi}$ and all the other clusters are negative, and obviously smaller. This result could explain why certain quasi-stationary wave patterns in NH winter have sometimes been interpreted as amplifications of the climatology.

Fuzziness

To study flow sequences whose patterns are more or less constant, but slowly moving in physical space, we introduce a special concept of fuzziness. This is inspired to some extent by, but distinct from the classical fuzzy clustering algorithms [*Bezdek*, 1981].

The centers $\{\bar{c}_k\}$ of the clusters are kept fixed, but the number of points belonging to each given cluster C_k is increased by relaxing the membership criterion to

$$p(\phi, \bar{c}_k) \geqslant r_3 \quad , \quad r_3 < r_1 \quad . \qquad (9a,b)$$

This will allow a flow sequence with some maps already belonging to cluster C_k

but having also maps containing a slightly displaced version of the main feature, e.g., a slowly retrogressing ridge, to belong entirely to the increased, fuzzy cluster.

We note in passing that the method of complex correlations [e.g., *Horel*, 1984] has somewhat the same purpose, and we implemented a version of it. This version only captured those flow sequences for which motion of features is strictly zonal, and we renounced developing a version which would not exhibit this shortcoming.

We used the fuzzy membership criterion (9), with $r_3 = 0.65$. This is based on the requirement that correlations be statistically significant at a 95% level. For ten degrees of freedom, a simple algebraic transformation of the classical Student t- test yields a lower bound on significant correlations of 0.63 [*Fisz*, 1963, pp. 429-430], hence $r_3 = 0.65$, by rounding up, for our ten EOFs.

The results for the fuzzy clusters, including now the small-anomaly cluster, cf. criterion (8), are given in Table 4. We concentrate on comparing the following characteristic times with those of the hard clusters in Table 2: T_d is the average duration of a passage in the cluster, while T_w is the average wandering time between exit from that cluster and entry into any other cluster. A sequence in a cluster is termed persistent if it lasts for five time units or longer, $T_d \geqslant 5\tau$. T_p is the average duration of persistent sequences.

[Table 4 near here, please]

Table 4 shows that relaxing the membership criterion has led to a total number of elements in nontrivial clusters of 62% of all points, vs. 27% before, slightly more than the double. Clusters 1 and 2 are still dominant, and most persistent. The number of elements for them has increased the least, showing that they are intrinsically stationary and well separated from all other clusters. The numbers for the smaller clusters, 3, 4 and 5, has increased more than threefold, indicating that they tend to contain sequences with slowly moving features, rather than stationary ones.

The average residence time in Clusters 1 and 2 is $T_d \cong 11\tau$, while for Clusters 3 through 6 it is $T_d \cong 2\tau$. The wandering times are $T_w \cong 5\tau$ and $T_w \cong 3.5\tau$, respectively. The wandering times are smaller for all fuzzy clusters (Table 4) than they are for the hard clusters (Table 2). This is due simply to the decrease in size of the diffuse, nonrecurrent cluster in the fuzzy formulation.

11.5τ/ The average persistence time in zonal Cluster 1 decreases from $T_d \cong 21.\overset{5}{\cancel{8}}\tau$ in Table 2 to $T_d \cong 10.5\tau$ in Table 4, while the aveage duration of persistent sequences goes from $T_p \cong 41\tau$ to $T_p \cong 24\tau$. A change in the opposite direction occurs for the blocked Cluster 2, with an increase of T_d from 6τ to 11τ, and of T_p from 10τ to 35.5τ, as cluster size increases due to fuzziness. This is in agreement with the distribution of duration of persistent sequences given in Figure 17 of *Legras and Ghil* [1985] and Figure 16a of *Ghil* [1987], if we accept the fact that the fuzzy clusters (Table 4) include a larger number of passages of the trajectory not so close to the corresponding unstable equilibrium. There are proportionately more such short events captured by an

ize/ increase in cluster size for the zonal regime, as seen also directly from the /lr

two tables.

Changing the fuzziness parameter r_3 from 0.65 to 0.7 or to 0.55 yields smaller or larger numbers of elements in each cluster. But the relative stationarity of Clusters 1 and 2, and the transient character of Clusters 3 through 6 remains the same.

Cluster Analysis and QS Events

To compare the results of the PCM method with those of cluster analysis, let Q_1 be the set of all maps belonging to a QS event and Q_2 be the set of all maps belonging to one of the five nonexceptional clusters. Let

$$T_1 = Q_1 \cap Q_2 \qquad (10a)$$

be the set of elements belonging to both Q_1 and Q_2, while

2 ∧ 1 / [illegible] /

$$T_2 = Q_{\cancel{2}\underset{1}{\wedge}} \cup Q_2 \quad \ldots \qquad (10b)$$

is the set of elements belonging to either Q_1 or Q_2 (or both).

The ratio

$$\gamma = \frac{\#(T_1)}{\#(T_2)} \qquad (11)$$

between the number $\#(T_1)$ of elements in T_1 and the number $\#(T_2)$ of elements in T_2 gives a measure of the compatibility of the two methods. For the fuzziness parameter $r_3 = 0.65$, we find $\gamma = 0.57$. This is due to the large number of elements in Clusters 3 through 5 which are *ipso facto* in Q_2, but not in Q_1.

We consider therewith Q_2' as the union of elements in Clusters 1 and 2 only, and define T_1' and T_2' accordingly by replacing Q_2 by Q_2' in Eqs. (10a,b). The corresponding $\gamma' = 0.71$ is much larger than the previous value of $\gamma = 0.57$, substantiating our designation of Clusters 1 and 2 as stationary, or persistent, while the other clusters are justifiably termed transient.

Figure 5 shows the correlation R_1 and R_2 between the time series of anomaly maps $\phi(\underset{\sim}{x},t)$, projected onto the first ten EOFs, and the centers $\bar{c}_k$, k = 1,2 of Clusters 1 and 2, respectively. The index Q(t), also shown, equals 1 if the map is either in fuzzy Cluster 1 or in fuzzy Cluster 2, and 0 otherwise.

It is clear by inspection of this figure that all major QS events, and most minor ones, are either in Cluster 1 or in Cluster 2.

[Fig. 5 near here, please]

We can conclude our intercomparison between the PCM and cluster analysis with the following two remarks:

1) Cluster analysis can identify both QS events and nonpersistent, but frequently recurring patterns. PCM, as used up to now, can only do the former.

2) PCM can take into account slowly moving features better than cluster analysis. This is especially true for a few QS events which involve gradual transition from one cluster to another, or exit and reentry into the same cluster during one QS event.

Transitions between Clusters

In *Mo and Ghil* [1987], we introduced the concept of a Markov chain of transitions between planetary flow regimes, whose flow patterns were defined there by the PCM (see also Figure 25 in *Ghil*,[1987]). It turns out that a much better description and understanding of such a Markov chain obtains when basing multiple flow regimes on cluster analysis.

Table 5 gives the number of transitions from one cluster to another for the six fuzzy clusters of Table 4. Clearly each flow sequence, or trajectory in phase space, passes through the diffuse, nonrecurring cluster. This is neither significant nor interesting and yet another reason to ignore the trivial cluster. Notice that the total number of transitions, 1030, is much smaller than the total number of elements in the six clusters, $\#(Q_2') = 4315$, since we do not count two successive maps within the same cluster as a reentry. The ratio of these two numbers is simply $T_d = 4.19\tau$ for Q_2' (see Table 4).

[Table 5 near here, please]

The transition matrix in the table is not symmetric, and not diagonally dominant. Except for Cluster 4, the largest entries in either row or column occur off the diagonal, indicating that reentry is much less likely than transition to another cluster.

In spite of the considerable dominance of Clusters 1 and 2, their diagonal entries are among the smallest. Even more strikingly, there are *no* direct transitions between the zonal-flow Cluster 1 and the blocking Cluster 2. To go from a zonal-flow pattern to a blocking pattern, or *vice-versa*, the flow has to pass through at least one, though more typically two transient, but recurring patterns. This is shown in the "flow diagram" of Figure 6.

[Fig. 6 near here, please]

In this graphic representation of our Markov chain, only those arrows have been drawn which correspond to a number of transitions larger than that given by equal probabilities, plus one standard deviation. Thus for instance there are 113 transitions out of Cluster 1, to one of six clusters, for an equal probability of 113/6 ≅ 19 transitions. Hence only the reentry arrow, with 31 transitions, and the arrow to the wave-train Cluster 4, with 42 transitions, are shown. The figure emphasizes that transitions both to and from Cluster 2 are likely only through Clusters 5 and 6, and that wave-train patterns with opposite phases account for transitions to and from Cluster 1, respectively.

Bimodality

As explained in Section 3, bimodality with respect to one variable is the simplest form that inhomogeneity of a pdf in phase space can take. Bistable solutions to simple models of planetary flow over topography were obtained by *Charney and DeVore* [1979], *Hart* [1979] *and Pedlosky* [1981]. The relevance of bistability to low-frequency atmospheric variability was often interpreted to stand or fall by the discovery of such bimodality in NH winter data. This is the approach taken in particular by R. Benzi, A. R. Hansen, P. Malguzzi, A. Speranza and A. Sutera in a number of recent publications [e.g., *Benzi et al.*, 1986; *Hansen*, 1986; *Speranza*, 1986; and references therein].

The picture of multiple planetary flow regimes which emerges from this section and the following one is considerably more complex, and potentially more applicable to LRF, than simple bimodality. But it appears interesting to verify the existence and sources of bimodality in both our model and in our NH data set.

Figure 7 exhibits the approximate pdf obtained for the first and second PC of our data, by using the algorithm NDMPLE explained in Section 3. The smoothness parameter in Eq. (1a), obtained by applying the K-S test to yield a confidence level of 95%, was $\alpha = 0.1$.

[Fig. 7 near here, please]

The pdf of PC 1 (Figure 7a) is clearly non-Gaussian with the largest peak, or *mode*, near the mean and smaller modes at approximately +1 and -2 standard deviations. The position of these smaller peaks corresponds roughly to the projection onto EOF 1 of $\bar{c}_2$ and $\bar{c}_1$, the centers of the blocking and the zonal-flow clusters, respectively (compare Figure 4). The pdf of PC 2 (Figure 7b) is significantly skewed towards positive values, but is unimodal.

The same is true of the pdf for PC 3 (not shown). This is due to the fact that Clusters 1 and 2 have small components along EOFs 2 and 3, while Clusters 3 through 6 are smaller and their distributions project without significant discontinuities onto these EOFs.

A more complete picture of the situation is given in the two-dimensional histogram of Figure 8. The points in our time series, projected onto EOFs 1 and 2, with the axes standardized as in Figure 7, were counted in boxes of

0.2x0.2 standard deviations. Two peaks on either side of the EOF 2 axis are clearly associated with Clusters 1 and 2, respectively. The peak near the origin is due mostly to the small-anomaly Cluster 6.

[Fig. 8 near here, please]

Bimodality thus results from the existence of two dominant, stationary clusters, associated with particularly persistent flow sequences. The presence of additional, more transient clusters, detectable by other means, will tend to blur this simple way of looking at multiple regimes. Still, these additional clusters, if statistically significant, can contribute to understanding, as well as predicting low-frequency atmospheric variability: they establish road posts along the preferred routes of transition between the most obvious and persistent flow patterns, such as blocked and zonal flow.

5. NORTHERN HEMISPHERE 500 MB HEIGHTS

Bimodality for Low-Pass Filtered Data

Bimodality was first illustrated in a data set of NH 500 mb data for the winter of 1981-1982 by *Benzi et al.* [1986] and for the four winters 1980-1984 in other publications of the same group [*Hansen,* 1986; *Sutera*, 1986]. They applied the NDMPLE algorithm to the univariate pdf obtained from their data set with respect to the nonlinear functional given by the sum of squares of the amplitudes of zonal wavenumbers two, three and four resulting from an average of the heights over latitudes 15N to 75N.

We start the detailed analysis of our data set of 20 NH winters (1963-1982) by obtaining smooth approximations of univariate pdfs with respect to (linear) projections onto EOFs 1, 2 and 3 of the low-pass filtered data (Figure 9). The sources of bimodality in PC 1 (Figure 9a) and of considerable skewness in

PCs 2 and 3 (Figures 9b,c) will then be further investigated by cluster analysis. [Fig. 9 near here, please]

The three leading PCs were each standardized and gridded as in Sections 3 and 4. The respective values of the smoothness parameter α given by the 95% confidence level of the K-S test are all three equal to 10. The light lines in Figures 9a-c indicate the results for the full set of low-passed data. While all three PCs show some measure of skewness, none is bimodal with any degree of significance.

For the simple model, bimodality of the first PC resulted from the persistent sequences in the dominant Clusters 1 and 2. We are thus led to consider the distribution of QS events in NH winter data. These were studied already from a slightly different point of view by *Dole and Gordon* [1983] and by *Horel* [1985a,b]. In our data set, 522 days out of a total of 2400 daily maps fall within QS events, defined as in Sections 2 and 4. The approximated pdfs for this restricted data set are shown as heavy lines in Figure 9, in terms of the original leading EOFs.

EOF 1 (Figure 9a) is now clearly bimodal, with excellent separation and a highly significant magnitude of the two peaks. Values of α both much larger and much smaller than the optimal one selected by the K-S test give the same bimodal picture. EOFs 2 and 3 (Figures 9b,c) are strongly skewed, but not significantly bimodal, as for the model (Figure 7).

We thus conclude that persistent anomalies of NH winter flow have preferred locations in phase space. The total pdf of hemispheric flows is blurred, however, by the more uniform distribution of transient sequences of maps connecting these locations. The total number of maps available did not permit us to obtain statistically significant multi-dimensional histograms, and we turn therewith to cluster analysis in order to clarify the situation further.

Clustering in the Low-Pass Filtered Data

Our low-passed data set was projected onto the seven leading EOFs, which together represent 50% of its variance (see Table 1). Cluster analysis was carried out in this seven-dimensional space, using $r_1 = 0.82$ and $r_2 = 0.34$ for the hard clusters. Eight nonexceptional clusters were obtained and they are listed in Table 6 in decreasing order of the number of elements, with the position of their centers.

[Table 6 near here, please]

These clusters were enlarged by using $r_3 = 0.65$, while keeping their centers fixed. The persistence properties given in the table refer to these enlarged, fuzzy clusters. The ninth, small-anomaly cluster, obtained by criterion (8), was not enlarged and its properties are also given for completeness.

Clustering calculations were repeated using nine EOFs, which account for about 60% of the variance of the time series. Parameters were varied in the ranges $0.8 \leqslant r_1 \leqslant 0.83$ and $0.34 \leqslant r_2 \leqslant 0.38$ for the hard clusters. Clusters 1 through 6 were all reproducible, with 7 and 8 being less stable.

Given larger data sets, techniques for objective estimation of clustering parameters from the data can be formulated, as for the univariate DMPLE procedure. For the limited set at hand, the verification of the results lies mostly in the dynamical and climatological interpretation of the flow patterns obtained by cluster analysis and their relationship to NH patterns obtained by other methods.

The flow patterns associated with fuzzy Clusters 1-6 are shown in Figures 10a-f, respectively. For each cluster, the figure shows the 500 mb height field obtained by averaging the filtered anomaly maps over all elements in the cluster. The plot thus shows the true centr of the fuzzy cluster, as opposed /e to that of the hard cluster used initially.

[Fig. 10 near here, please]

To assess the statistical significance of the features in the figure, we calculated the standard deviation of the time series of anomalous heights at each grid point, $\sigma(\underset{\sim}{x})$. The number N of independent samples for each cluster was estimated, rather conservatively, by the total number of days spent in that cluster, divided by 10 days. The latter is considerably longer than the mean duration of each sequence in any cluster, T_d (see Table 6), so that we are looking essentially at N independent passages through each cluster. Assuming a normal distribution of anomalies at each grid point, we used $\sigma(\underset{\sim}{x})$ and N to determine the points at which the mean anomaly value was different from zero at the 95% level of significance. Areas for which this statistical significance criterion is satisfied are shaded in Figures 10a-f.

Cluster 1: wavenumber-three pattern (Figure 10a). This cluster is the largest as a hard cluster (130 maps) and second largest as a fuzzy cluster (301 maps). It has a clear zonal wavenumber-three pattern. The anomaly map resembles in the Pacific sector very closely the one-point correlation map for the base point (55N, 115W), called the Pacific/North American (PNA) pattern by *Wallace and Gutzler* [1981]. In the complementary 180 degrees of longitude it resembles the wave train called the Eurasian teleconnection pattern by the same authors.

The average residence time in this cluster is $T_d \cong 6.5$ days, and the wandering time once the flow leaves this cluster and enters another one is $T_w \cong 8$ days. There are 13 persistent sequences in this cluster, with an average duration of $T_p \cong 10$ days, and they are rather evenly distributed throughout the data set. But the hard part of the cluster was much better represented during the latter ten winters of the time series.

Cluster 2: reverse wavenumber-three pattern (Figure 10b). This cluster is second largest according to hard criteria (114 maps) and third largest according to the fuzzy criterion (278 maps). Zonal wavenumber three is as prominent as for Cluster 1. Each anomalous high of Cluster 1 is matched by a low of Cluster 2, and *vice-versa.* But the features are not merely of opposite sign: they are all slightly displaced and distorted.

The North Pacific high in Cluster 2 is much more elongated and flatter than the Aleutian low in Cluster 1. The Western Canadian low is much smaller and weaker in Cluster 2 than the central high of the PNA pattern in Cluster 1. Finally, the North European high in Cluster 2 is much larger and stronger than the Scandinavian low in Cluster 1.

Clearly the dominant climatological effect associated with Cluster 1 is the extensively studied Pacific influence on North America. The dominant regional feature associated with Cluster 2 is the wave train teleconnecting the Eastern United States over Greenland to Northern Europe [*Dickson and Namias,* 1976].

Table 6 shows that the centers $\bar{c}_1$ and $\bar{c}_2$ of the two clusters have the largest components, of opposite sign, along EOF 1. The situation is thus quite similar to Clusters 1 and 2 in the model (Section 4) and to the quasi-stationary regimes found by the PCM for the Southern Hemisphere by *Mo and Ghil* [1987].

There are 12 persistent sequences in this cluster, with an average duration of $T_p \cong 8.5$ days, and they are evenly distributed throughout the time series. The hard part of the cluster is quite reproducible using first one half of the data set, then the other.

Cluster 3: wavenumber-two pattern (Figure 10c). This cluster is third largest as a hard cluster, but largest as a fuzzy cluster (323 maps). Zonal wavenumber two is dominant. Positive anomalies cover Canada and most of the North Atlantic north of 30N. A small low is centered over the Western Mediterranean and a large high is centered on the Ural Mountains. This feature is reminiscent of the regional anomaly pattern centered on the Northern Soviet Union (NSU) studied by *Dole* [1986]. Finally, a moderately large Aleutian low lies above a small high in the Central North Pacific, indicating zonal flow in the Pacific sector [*White and Clark*, 1975].

There are 12 persistent sequences in this cluster, with an average duration
11.5/ of T_p = 11~~11~~.5 days. The cluster is reproducible in both halves of the data set.

Cluster 3 appears to contain most of the interannual variability of the NH wintertime circulation. Instead of defining the seasonal cycle as in Section 2, we took out separately for each year in the data the mean of that year, as well as the annual and the semi-annual component (365 days and 182.5 days) of a Fourier expansion for that year, at each grid point. The clustering computations were then repeated for the anomaly maps so defined.

The composite anomaly map for Cluster 3 obtained in this way is given in Figure 11. All the features are much weaker, and hardly any of them are statistically significant, when compared with Figure 10c. Similar comparisons for the other clusters show only insignificant differences.

[Fig. 11 near here, please]

Cluster 4: double blocking (Figure 10d). This cluster is fourth largest as a hard, and fifth as a fuzzy cluster. Like Cluster 3, it is dominated by zonal wavenumber two, with a general aspect of phase opposition to the previous cluster. This is related to the two clusters having large components of opposite sign along EOF 2 (see Table 6). Thus EOF 2 is largely determined by the presence of Clusters 3 and 4, in the same way that EOF 1 is determined by Clusters 1 and 2.

Cluster 4 shows strong north-south oscillations in both the Pacific and the Atlantic sectors. In fact, their concomittant appearance might be related to the zonally-symmetric seesaw in sea-level pressures first noticed by *Lorenz* [1951] and discussed in the context of the North Atlantic oscillation (NAO) and the North Pacific one by *Wallace and Gutzler* [1981].

The features in the Atlantic sector resemble strikingly the Greenland - Northern Europe seesaw of *van Loon and Rogers* [1976]. In the Pacific sector, there is a dominant high centered on the northeastern tip of Siberia, accompanied by deep lows over both the Okhotsk Sea and the Central North Pacific. The former is associated with the Western Pacific teleconnection pattern of *Wallace and Gutzler* [1981]. Some of these patterns, especially NAO, have also been put in evidence by the rotated EOFs of *Barnston and Livezey* [1987]. Indeed, rotated EOFs have greater liberty to point at clusters in phase space, being less inhibited by orthogonality constraints. Oblique, or target rotation, should point even more accurately at clusters, orthogonality being replaced by statistical independence.

Cluster 4 is also reproducible in both halves of the data set. There are eight persistent sequences, with an average duration of $T_p \cong 9.5$ days. In fact, 75 days, out of a total of 86.5 days spent in the cluster, belong to persistent sequences, the largest fraction of any cluster.

All eight persistent sequences have pronounced ridges in both the Pacific and the Atlantic Oceans. Many intense, high-latitude double-blocking cases, such as those in the winters of 1963, 1968, 1977 and 1980, belong to this cluster.

Cluster 5: wave train (Figure 10e). This cluster is fifth by hard criteria (76 maps) and fourth by the fuzzy criterion (261 maps). It shows a strong high in the Gulf of Alaska, with flow parallel to the Rocky Mountains, as discussed by *Wallace and Blackmon* [1983, Figure 3.16].

The rest of the anomaly map is taken up by one huge wave train of alternating positive and negative anomalies, extending from off the southeastern coast of the United States across Eurasia, to Eastern Siberia. The strongest feature in the wave train is the Icelandic low, with the second strongest being another low over the Ural Mountains. The dipole over the Western Atlantic resembles the pattern given that name by *Wallace and Gutzler* [1981], and the entire wave train resembles their difference of composites for the positive and negative phases of the Western Atlantic teleconnection [their Figure 21c].

The center $\bar{c}_5$ of this cluster has largest components along EOFs 1 and 3, with signs opposite to those of Cluster 3. This is reflected in the dominant spatial features, south of Alaska, south of Greenland and over the Urals, having opposite signs for Clusters 3 and 5. But we saw that Cluster 3 has still mainly a wavenumber-two pattern, while wavenumber four is dominant here and for Cluster 6.

There are nine persistent sequences in this cluster, with $T_p \cong 8$ days. Of these, three are in the first half of the data set, and six in the second half.

Cluster 6: PNA pattern (Figure 10f). This is the last statistically significant cluster. It has a striking PNA pattern in its negative phase, as defined by *Wallace and Gutzler* [1981, Figure 17]. There are six persistent sequences, with $T_p \cong 10$ days, and five of them correspond to blocking in the central Pacific, as described by *White and Clark* [1975]. This cluster is also reproducible in both halves of the data set.

The features in the sector from 60W to 120E are weak and not statistically significant. This contrasts with Cluster 5, where features in this sector are very prominent. Large components of $\bar{c}_5$ and $\bar{c}_6$ with opposite signs along EOF 3 show that this pair of clusters contributes decisively to this EOF. The partial localization of spatial features for this pair is analogous to that observed for Clusters 1 and 2, determining EOF 1, and to that for Clusters 3 and 4, determining EOF 2.

The features in the Pacific sector resemble very well the positive composite of locally-defined height anomalies for Pacific (PAC) reference points of *Dole* [1986, Figure 1a]. We shall return to the complementarity of the different ways of viewing persistent anomalies, spatially, spectrally, through EOFs and through teleconnection patterns, in our concluding remarks.

Transitions between Low-Pass Clusters

From Table 6, it is clear that the relative fraction of maps in all clusters, 41 % , is much less than for the simple, low-resolution model of Section 4 (see Tables 2 and 3). As a consequence, the average time spent by the atmosphere between clusters, of $T_w \cong 9.5$ days, is almost twice as long as the time spent in the clusters, of $T_d \cong 5.5$ days. This is essentially due to the much larger number of degrees of freedom for the atmosphere's low-frequency variability. In fact, it is both surprising and encouraging that the ratio T_w/T_d is not any larger than obtained here.

Figure 12 shows each of the correlations $R_k(t)$ between the time series of NH 500 mb anomaly maps, filtered in time and space as indicated, and the cluster centers $\bar{c}_k$, k = 1,2,...,8. Also plotted is the indicator function Q(t) of the set Q_1 of QS events, i.e., Q(t) = 1 if $\phi(\underset{\sim}{x}, t)$ is part of a QS event, and Q(t) = 0 otherwise. Visual inspection of the figure clearly indicates that all major QS events are associated with passage through a cluster.

[Fig. 12 near here, please]

Following Section 4, we calculated from the obvious counterpart of Eqs. (10, 11) $\gamma \equiv \#(T_1)/\#(T_2) = 0.42$. As expected from atmospheric behavior's being more complex than that of a simple model, γ here is lower than the value of 0.57 for the model. On the whole, this still indicates good agreement between the PCM and cluster analysis in identifying preferred flow regimes. The discrepancy results on the one hand from QS events not being entirely confined to clusters, although at least some of the successive maps of a QS event typically belong to a cluster. On the other hand, considerable numbers of points in each cluster do not belong to any QS event.

For the model, many long persistences are associated with particularly close passages of the trajectory by an unstable equilibrium. To determine whether this is actually the case for large-scale atmospheric flow, models with much higher spatial resolution and greater physical realism than that of *Legras and Ghil* [1985] need to be analyzed with the same degree of care and detail. This is entirely possible on existing supercomputers and we expect to carry out such analyses in the near future.

Table 7 gives the transition matrix for the modified Markov chain of NH low-pass filtered variability, as sampled from our data set of 20 winters of 120 days each. Comparison of Tables 5 and 7 here with Tables 5 and 8 in *Mo and Ghil* [1987] shows an important advantage of cluster analysis over the PCM method of identifying preferred regimes: the number of transitions here is considerably higher, providing a much larger, although still insufficient sample for a stable estimate of the true transition probabilities between regimes. As explained in *Ghil* [1987] and *Mo and Ghil* [1987], the only way that stable, reliable estimates of such a transition matrix can be obtained in the near future is by careful and extended experimentation with general circulation models.

[Table 7 near here, please]

Thus Table 7 cannot be used for specific LRF predictions as yet. But it can be used for general guidance as to how LRF might proceed in the not-too-distant future.

As in Table 5, we notice that diagonal elements are generally small, i.e., reentry into the same cluster is rather unlikely. This confirms in a sense that our fuzzy definition of the clusters is quite appropriate: smaller clusters would show more reentries, and so would much larger clusters.

The matrix is far from symmetric: preferred paths are in evidence. These are illustrated in Figure 13. As in Figure 6, only those arrows are drawn which correspond to a probability of transition significantly higher than given by equal chances.

[Fig. 13 near here, please]

The small-anomaly Cluster 9 plays a role of crossroads even more important than for the model. Trajectories exiting from Clusters 1, 2, 3 and 5 have a relatively high likelihood of passing through Cluster 9 before continuing to Clusters 1, 2, 3 or 7(?). As explained in our previous publications, this does not indicate that the clusters represent certain linear instabilities of the time-mean flow, but rather that the time mean happens to lie close to the point where the boundaries of several attractor basins touch, permitting slow transitions between dominant clusters [*Grebogi et al.*, 1983; *Ghil and Childress*, 1987, Sections 6.4 and 6.6].

In the atmosphere, certain higher-frequency phenomena, not represented in the equivalent-barotropic model of Section 4, also play a role in the transitions between relatively stable, stationary clusters. To begin understanding this role, we turn to the band-pass window of variability.

Band-Pass Clusters

We used six EOFs, $r_1 = 0.82$ and $r_2 = 0.36$ to perform cluster analysis on the band-passed anomaly maps. Six EOFs provide only 24.2% of the variance in this window. But there is an obvious discontinuity at this level in the variance spectrum, from 3% to 2.5%, which is statistically significant by the rule of thumb of *North et al.* [1982] (see Section 2). Calculations were repeated with eight and nine EOFs and with different values of r_1 and r_2. The results below were essentially unchanged.

There are seven distinct clusters, including one of small anomalies. The number of elements in the hard clusters varies from 65 to 94 maps. Using a fuzziness criterion of $r_3 = 0.65$, the augmented clusters range from 221 to 360 elements. The size of the clusters varies much less than in the low-passed data, so we arrange them by flow patterns, rather than size.

Figure 14 shows the mean anomaly maps of the six nonexceptional band-pass clusters, as for the low-pass window (Figure 10). In agreement with the results of *Blackmon et al.* [1984a, b], all clusters in this window (called in the latter articles "short time scales", as opposed to "intermediate time scales" of 10 to 30 days, and "long time scales" of more than 30 days) show essentially wave trains elongated in the meridional direction, propagating zonally with a wavenumber of seven or eight.

[Fig. 14 near here, please]

Clusters 1 and 2 (Figures 14a, b) have a well-defined wave-train structure in the jet exit region over the Eastern United States and the Western Atlantic. The two clusters are distinguished from each other by their wave trains being roughly in quadrature.

Clusters 3 and 4 (Figures 14c, d) have the most pronounced features of their wave train over the Western Pacific. The same quadrature of phase obtains as for the Atlantic clusters. Clusters 5 and 6 (Figures 14e, f) have a well developed wave train over both oceans, being only weaker over Eurasia, but the wave activity is still strongest in the Atlantic jet exit region.

The spatial localization of baroclinic wave activity is a topic of considerable recent interest [*Brevdo*, 1987; *Merkine*, 1977; *Pierrehumbert*, 1986]. Our clustering procedure detects this localization and avoids yielding arbitrarily close successive phases of the synoptic-scale waves by the separation criterion of $p(\bar{c}_k, \bar{c}_\ell) < r_2 = 0.36$ between centers of clusters.

The role of band-pass clusters in transitions between low-pass clusters is obviously important, and will form the object of a subsequent paper. We expect them to serve as way stations on preferred transition paths within what appears merely as a diffuse, thin cloud of points in the low-pass, mostly barotropic window of variability.

6. CONCLUDING REMARKS

An Approach to Long-Range Forecasting, and a Simple Model

Our study of low-frequency atmospheric variability is guided by the practical concerns of long-range forecasting (LRF). Due to the well-known limits on detailed, pointwise predictability, one cannot expect to predict local weather with useful accuracy beyond 10 days, say, in a manner uniformly valid over all atmospheric states.

The best hope therewith for LRF is that certain large-scale atmospheric flow patterns are more persistent than others, that these patterns fall into a few identifiable classes, or flow regimes, and that these regimes exhibit well defined transition probabilities from the one to the other. The theoretical question is then to find how atmospheric dynamics generates these regimes, and their preferred transition patterns. The practical question is to extract from existing data and models the quantitative information on regime identification, expected duration and most likely successor. In the present section, we summarize our results in this perspective and sketch some promising directions for necessary research.

The first step on the proposed road to LRF is identification of multiple regimes. We prefer the term *planetary flow regime* to the earlier "weather regime" of *Reinhold and Pierrehumbert* [1982], since weather is precisely what does not persist and cannot be predicted. It clearly plays a role in maintaining the large-scale, low-frequency flow patterns, but what this role might be is one of the more difficult questions of the whole field of LRF [*Wallace and Blackmon*, 1983, pp. 89-90].

To identify these regimes, we developed a modification of standard cluster analysis methods. This modification takes into account well-known features of low-frequency atmospheric variability, not present in other applications of cluster analysis, and enhances the convergence of classical algorithms. We chose a hard clustering algorithm, based on pattern correlations as a measure of distance between points in phase space. The modification introduced allows for a thin cloud of non-classified points in which the clusters are embedded, and for a special cluster of small anomalies, based on Euclidean, or root-mean-square, distance between each point and the grand mean of all points in the data set. Our modification further enlarges the nonexceptional hard clusters so obtained by a fuzziness criterion, admitting points with a preset correlation, lower than that used in hard clustering, to the fixed centers of the hard clusters. Reasonable changes in the values of the clustering parameters did not change our results in any substantial way.

This clustering algorithm was first applied to the time series of streamfunction fields produced by an equivalent-barotropic quasi-geostrophic model with simplified Northern Hemisphere (NH) topography, zonal jet forcing and Ekman dissipation [*Legras and Ghil*, 1985]. The fields were spatially filtered from 25 spherical harmonics to ten empirical orthogonal functions (EOFs). This model time series is 65 years long, providing much higher statistical significance than available atmospheric data sets, and has the further advantage that the sources of low-frequency variability are well understood.

Six stable clusters were obtained, including that of small anomalies. They make up 62% of the data, leaving 48% for the diffuse, trivial cluster.

Clusters 1 and 2, in order of size, resemble the model's unstable equilibria termed Zonal 1 and Blocking. They contain the most persistent sequences, due to close passages near these equilibria. The first EOF is very nearly parallel to the straight line segment passing through the centers $\bar{c}_1$ and $\bar{c}_2$ of these two clusters.

Clusters 3 and 4 resemble opposite phases of the wave-train pattern also detected by the pattern correlation method (PCM) for quasi-stationary (QS) events in *Mo and Ghil* [1987]. They are less persistent and determine, subject to the usual orthogonality constraint, the direction of EOF 2. Cluster 5 in size is also the least persistent, and resembles yet another unstable equilibrium of the model, Zonal 2. EOF 3 points in the direction of this cluster, within the subspace orthogonal to EOFs 1 and 2.

The projection of the sample probability density function (pdf) of this model solution onto EOF 1 gives a univariate pdf which is clearly bimodal. The two modes are produced by the persistent sequences in Clusters 1 and 2. Univariate pdfs along EOFs 2 and 3 are strongly skewed, but unimodal.

We computed the transition matrix for the modified Markov chain whose states are defined by the six nontrivial clusters, ignoring the diffuse one. This matrix shows few reentries into any cluster. Transitions between the two dominant clusters occur preferentially through Clusters 4 and 5, in one direction, and through Clusters 5 and 3, in the other. The small-anomaly cluster is close to the boundary of the attractor basins of Clusters 1 and 2, and is also on one preferred path from Cluster 2 to 1, via Cluster 5 and a wave train.

The second and third steps in defining our LRF procedure are determining the expected residence time in each regime, and the most likely successor to each regime. These two steps have been taken above for the model. The fourth and fifth are a dynamical explanation of the results for the first three steps, and a practical verification.

For the model, we under~~t~~stand the dynamic origin of Clusters 1, 2 and 5, as related to close passages of the time-dependent solution by unstable equilibria with many directions of stability and few directions of instability. The size and mean residence time for these clusters are determined by the relative stability of the respective equilibria.

Cluster 6 is given by the slowing down of trajectories close to a complicated basin boundary, and this also explains its role in transitions between Clusters 1 and 2. The nature of Clusters 3 and 4, and their role in transitions between the dominant clusters, is more speculative. Their wave-train nature suggests a phenomenology similar to certain standing or slowly-moving Rossby waves in the atmosphere. But the phase-space structures associated with these waves require further elucidation, and we expect to do this in a more detailed and realistic model.

y/ *Clusters of Low-Frequenc~~e~~y Atmospheric Variability*

The fifth point in our LRF procedure, actual verification, only makes sense for the atmosphere itself. We have carried out therefore steps one though three of the proposed LRF procedure for a data set of 500 mb geopotential heights from 20 NH winters, January 1963-December 1982. In the atmosphere, relatively low-frequency, mostly barotropic flow structures coexist with intermediate-frequency, largely baroclinic waves. This data set was hence separated into a low-pass and a band-pass window by suitable filters [*Blackmon*, 1976].

The low-pass variability, of ten days and longer, was spatially filtered by projection onto seven EOFs, and the band-pass data by projection onto six EOFs. The low-pass data exhibit seven stable clusters, including that of small anomalies. The clusters make up only 41% of the data, compared to 62% for the model's clusters. This percentage of recognizable clustering is clearly lower in the atmosphere due to the additional degrees of freedom, but is still quite encouraging, and suggests that we might be on a promising road to LRF indeed.

Clusters 1 and 2 have a wavenumber-three pattern, with nearly opposite phases. They determine together the direction of the first EOF. Cluster 1 shows the extensively studied Pacific influence on North America, Cluster 2 a similar influence of the Atlantic on Northern Europe.

Clusters 3 and 4 are dominated by zonal wavenumber two. Subject to the well-known orthogonality constraint of classical EOFs, they determine together the second EOF. Cluster 3 represents most of the interannual variability in the data, has zonal flow over the Pacific, and a blocking high over the Northern Soviet Union. Cluster 4 exhibits a marked blocking pattern over both the Atlantic and the Pacific ocean.

Clusters 5 and 6 contain a more complicated distribution of waves, and have largest projections of opposite signs onto EOF 3. Cluster 5 has a Western

Atlantic dipole and a wave train over Eurasia, Cluster 6 a very pronounced Pacific/North American (PNA) pattern.

The result most unexpected by, and therefore most interesting to us, is the localization of features exhibited by these six clusters. While all patterns are hemispheric, there is a clear tendency for each cluster within a pair to show larger and more significant features in one of two quadrants. These quadrants, which could be called eastern and western, or Atlantic and Pacific, are separated by the polar great circle composed of, roughly speaking, the 60W and 120E meridians.

The PNA pattern in Cluster 1 is much stronger than the Eurasian wave train, while in Cluster 2 the Western Atlantic - Greenland - Northern Europe teleconnection dominates. In Cluster 3 the most important feature is the positive anomaly over the Urals, while in Cluster 4 it is the one over the Bering Sea.~~(?)~~ Finally, and most strikingly, the PNA pattern in Cluster 6 is clearly complementary to the wave train trailing off the Western Atlantic dipole in Cluster 5.

The regional, rather than hemispheric character of many persistent anomalies is subjectively well known to classical, synoptic-statistical practitioners of LRF. It provided the basis for the local definition of anomalies in the work of *Dole* [1986] and for the teleconnection approach of *Wallace and Gutzler* [1981]. The interest of the present result is that we did not build this regionality into our search for preferred patterns, but obtained it objectively and quite independently of the search procedure. It follows that partial regionality, with weaker hemispheric concomitants, is indeed a fact of large-scale atmospheric life.

The reasons for this sectorial confinement of low-frequency variability have been studied by *Held* [1983], among others. Essentially, the propagation speed of atmospheric features has to compete with the dissipation of their energy. Taking a heuristic 10 ms^{-1} for the order of magnitude of the zonal propagation velocity of the energy through a stationary or slowly-moving wave train, and 10 days for the order of magnitude of the dissipation time, one obtains about 100 degrees of longitude for the spread of a locally-generated low-frequency disturbance. The limits of the sectors indicated above strongly imply that the two jet exit regions over the western part of the NH oceans are a major localized source of energy for low-frequency variability, in agreement with the suggestions of *Green* [1977], *Kalnay-Rivas and Merkine* [1981], and *Shutts* [1983]. It is both difficult and necessary to reconcile this point of view with the spatially global one of resonant reinforcement between the flow in two sectors, i.e., what wags the jet whose exit wags a wave train?

Hence our localization result raises more theoretical questions than it answers. But from the point of view of describing, rather than explaining low-frequency variability, it is rather gratifying: the success of the local approaches of *Dole and Gordon* [1983] and *Wallace and Gutzler* [1981] appears to be less surprising, and not at all at odds with a global view of atmospheric dynamics. It also helps explain the fact that varimax orthogonal rotation of EOFs, which favors a *priori* regional patterns [*Horel*, 1981; *Barnston and Livezey*, 1987], tends to produce patterns similar to those of the clusters here. Rotated EOFs are simply less inhibited by orthogonality from pointing at the natural clusters of low-frequency variability, and obliquely rotated EOFs would essentially point straight at them, with all their intrinsic regionality.

Bimodality and Transitions between Regimes

Univariate pdfs in the first three principal components (PCs) of the low-passed NH winter data are noticeably skewed, but unimodal. Model results showed that bimodality in the first PC is induced by the highly persistent flow sequences associated with Clusters 1 and 2, respectively. Restricting the NH atmospheric data to persistent sequences only shows strong bimodality of the first PC, with very high statistical significance.

The bimodality in this case is produced by sequences in Clusters .... for the "Pacific" phase of EOF 1 and by sequences in Clusters ... for the "Atlantic" phase. We notice that separate sequences with large components of zonal wavenumber two, three and four play a role in producing *both* maxima of the univariate pdf with respect to the first PC. This result provides a view of bimodality in low-frequency atmospheric variability which is complementary to, and somewhat more complex than that of *Benzi et al.* [1986].

Our view of multiple planetary flow regimes via more than two clusters is possibly closer to synoptic experience, and hence richer in its promise for LRF. The transition matrix for the Markov chain of seven clusters, including that of small anomalies, provides useful qualitative information. The small number of reentries supports our choice of cluster size, and shows the clusters to be well separated.

Preferred paths between clusters are in evidence. The small-anomaly cluster plays an important role on some of these, indicating its position on the boundary between attractor basins of several clusters.

Additional transitions between low-pass clusters are likely to be associated with preferred patterns in the band-pass window, of 2.5 days to 6 days, roughly speaking. There are six nonexceptional clusters in this window, all showing meridionally-elongated wave trains with zonal wavenumbers of seven or eight. One pair of clusters is associated with an obviously baroclinic wave train of this type in the Atlantic jet exit region, the second pair has its strongest features in and downstream of the Pacific jet exit, the last pair has significant features in both these regions. The wave train of one cluster within a pair is in phase quadrature with the other.

These findings appear to be interesting enough to warrant further exploration of the proposed road to LRF. The basic questions that need to be answered are both quantitative and qualitative. Quantitatively, one needs stable statistics of regime persistence and of transition probabilities. These can be obtained at present only by careful and extended experimentation with general circulation models (GCMs).

First, one needs to verify that a GCM produces essentially the same clusters as found in the data, and that the persistences and transition probabilities are equal to within sampling error to those in the data. Secondly, the GCM can be run for a sufficiently long period to obtain stable statistics. Third, one has to find how these statistics change when boundary data, such as sea-surface temperatures, are changed. Finally, the statistics obtained from the GCM have to be tested in a predictive mode.

Qualitatively, one would like to know what generates the multiple regimes, and connects them by preferred paths. For instance, is it really true that particularly persistent sequences are generated by close passages near an unstable equilibrium, as the simple model here suggests? Are some of the preferred paths initiated by instabilities of such equilibria? What is the relative importance of barotropic and baroclinic instabilities in the "break" of a persistent flow pattern? We hope to find some of the answers to these questions in future work, observational, numerical and theoretical.

APPENDIX A. CLUSTERING ALGORITHM

This is the convergent version, proposed by *Anderberg* [1973, pp. 162-163], of *MacQueen's* [1967] *k-means* algorithm. It has the advantage of being relatively inexpensive computationally, while still producing a partition close, at least locally in configuration space, to an optimum. Unfortunately, both convergence and optimality proofs are only available when Euclidean distance, rather than angle, are used for membership and separation criteria. We preferred to use, cf. Section 3, a modification of this algorithm tailored to synoptic experience with the NH data set, rather than rely on theoretical results which might not provide the most significant clusters from a dynamic point of view. The proof of the pudding is in the eating, as we saw.

The algorithm proceeds in two stages: (i) finding seed points for the *k* clusters, and (ii) iterating to optimize the partition. The first stage is essentially MacQueen's original algorithm, the second is essentially Anderberg's variant.

(i) *Seed points*

Step A1. Take any map in the time series as point 1.

Step A2. Proceed through the sequence, calculating the correlations $p(\phi, \bar{c}_k)$ between any given map $\phi(\underset{\sim}{x}, t)$ and existing centers of cluster $\bar{c}_1, .., \bar{c}_m$. If $p(\phi, \bar{c}_k) \geqslant r_1$, then ϕ is assigned to cluster C_k and $\bar{c}_k$ is recomputed. If, on the other hand, $p(\phi, \bar{c}_k) \leqslant r_2$ for all $\bar{c}_k$, $k = 1, ..., m$, then ϕ is allowed to form a new cluster, $\phi = \bar{c}_{m+1}$. If the exclusion criterion (7) is satisfied, then ϕ is assigned to the special, diffuse cluster.

Step A3. Keep centers fixed, and make one pass through the data, assigning points ϕ to existing clusters if $p(\phi, \bar{c}_k) \geqslant r_1$ for some *k*, and to the diffuse cluster otherwise.

(ii) *Iteration*

Step B1. Recompute the centers of clusters using current membership.

Step B2. Compute the pattern correlations between pairs of centers. If

$p(\bar{c}_j, \bar{c}_k) \leqslant r_2$ for all pairs, the algorithm terminates. If, on the other hand, $p(\bar{c}_{j_o}, \bar{c}_{k_o}) > r_2$ for a given pair, reassign all elements in the smaller cluster, according to step A2.

Step B3. Repeat steps A3 through B2 until no more than N_o points get reassigned in the last step, and no clusters smaller than L_o elements exist.

N_o was taken equal to L_o, the small cluster criterion (see end of Section 4). The number of iterations necessary was ... for the model and ... for the NH data.

Acknowledgments. We are indebted to the late M.-J. Munteanu for communicating to us her unpublished and very interesting work on applications of fuzzy clustering to temperature and tropopause height retrieval from satellite soundings. E. N. Lorenz encouraged us in early stages of this difficult enterrise and E. Epstein gave unstinting advice on statistical significance tests. Computations were performed on the Amdahl 470/V6 at the NASA Goddard Space Flight Center. The work of one of us (M.G.) was partially supported by NASA Grant NAG-5713 and by NSF Grants ATM85-14731 and ATM 86-15424. The diligent, careful and unflappable typing of B. Gola facilitated greatly our organization of the material. Figures were proficiently drawn by M. Archie.

470/V6

[Please continue on this page w/refs.,
to avoid wasting p.55.
Please watch authors w/more than 1 paper,
——, ...]

/all caps

References

Anderberg, M. R., *Cluster Analysis for Applications*, Academic Press, New York 358 pp., 1973.

Barnett, T. P., and R. W. Preisendorfer, Multifield analog prediction of short-term climate fluctuations using a climate state vector, *J. Atmos. Sci.*, *35*, 1771-1787, 1978.

Barnston, A. G., and R. E. L~~e~~ivezey, Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Mon. Wea. Rev.*, *115*, 1987 [in press]. /i

Baur, F., *Musterbeispiele Europaeischer Grosswet**t**erlagen*, Dieterich, Wiesbaden, 35 pp., 1947. /tt

Benzi, R., P. Malguzzi, A. Speranza, and A. Sutera, The statistical properties of general atmospheric circulation: Observational evidence and a minimal theory of bimodality, *Quart. J. R. Met. Soc.*, *112*, 661-674, 1986.

Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.

Blackmon, M. L., A climatological spectral study of the geopotential height of the Northern Hemisphere, *J. Atmos. Sci.*, *33*, 1607-1623, 1976.

______________, Y.-H. Lee, and J. M. Wallace, Horizontal structure of 500 mb height fluctuations with long, intermediate and short scales, *J. Atmos. Sci.*, *41*, 961-979, 1984a.

______________, ________, ____________, and H.-H. Hsu, Time variation of 500 mb height fluctuations with long, intermediate and short time scales as deduced from lag-correlation statistics, *J. Atmos. Sci.*, *41*, 981-991, 1984b.

Brevdo, L., A study of absolute and convective instabilities with an application to the Eady model, *Phil. Trans. Roy. Soc. (London)*, submitted, 1987.

~~Charney, J. G.~~, J. Shukla, and K. C. Mo, Comparison of a barotropic blocking theory with observation, *J. Atmos. Sci.*, *38*, 762-779, 1981.

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

l.c. v

Darling, D. A., The Kolmogorov-Smirnov-Cramer-~~V~~on Mises tests, *Ann. Math. Stat.*, *28*, 823-838, 1957.

Dickson, R. R., and J. Namias, North American influences on the circulation and climate of the North Atlantic sector, *Mon. Wea. Rev.*, *104*, 1255-1265, 1976.

Dole, R. M., Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Structure, *Mon. Wea. Rev.*, *114*, 1986.

__________, and N. M. Gordon, Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: Geographic distribution and regional persistence characteristics, *Mon. Wea. Rev.*, *111*, 1567-1587, 1983.

Eckmann, J.-P., and D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.*, *57*, 617-656, 1985.

Efron, B., *The Jackknife, the Bootstrap and other Resampling Plans*, Soc. Indl. Appl. Math., Philadelphia, 92 pp., 1982.

Fisz, M., *Probability Theory and Mathematical Statistics*, Wiley, New York, 679 pp., 1963.

Ghil, M., Dynamics, statistics and predictability of planetary flow regimes, in *Irreversible Phenomena and Dynamical Systems Analysis in the Geosciences*, C. Nicolis and G. Nicolis (eds.), D. Reidel, Dordrecht/Boston/Lancaster, pp. 241-283, 1987.

Ghil, M., and S. Childress, *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics*, Springer-Verlag, New York, 479 pp., 1987.

________, R. Benzi, and G. Parisi~~s~~ (Eds.) *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, North-Holland, Amsterdam/Oxford/New York/Tokyo, 449 pp., 1985.

Green, J. S. A., The weather during July 1976: Some dynamical considerations of the drought, *Weather*, *32*, 120-126, 1977.

Hansen, A. R., Observational characteristics of atmospheric planetary waves with bimodal amplitude distributions, *Adv. Geophys.*, *29*, 101-134, 1986.

Hart, J., Barotropic quasi-geostrophic flow over anisotropic mountains, *J. Atmos. Sci.*, *36*, 1736-1746, 1979.

Held, I. M., Stationary and quasi-stationary eddies in the extratropical troposphere: Theory, in *Large-Scale Dynamic Processes in the Atmosphere*, B. J. Hoskins and R. P. Pearce (eds.), Academic Press, London/New York, ~~897 pp.,~~ pp. 127-168, 1983.

~~Horel, J. D.~~, Persistence of the 500 mb height field during Northern Hemisphere winter, *Mon. Wea. Rev.*, *113*, 2030-204~~1~~2, 1985a. /2042

___________, Persistence of wintertime 500mb height anomalies over the Central Pacific, *Mon. Wea. Rev.*, *113*, 2043-2048, 1985b.

___________, Complex principal component analysis: Theory and examples, *J. Clim. Appl. Meteor.*, *23*, 1660-1673, 1984.

Horel, ________, J. D., A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field, *Mon. Wea. Rev.*, *109*, 2080-2092, 1981.

Kalnay-Rivas, E., and L.-O. Merkine, A simple mechanism for blocking. *J. Atmos. Sci.*, *38*, 20~~8~~77-2091, 1981. /2077

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

Lorenz, E. N., Seasonal and irregular variations of the Northern Hemisphere sea-level pressure profile, *J. Meteor.*, *8*, 52-59, 1951.

———————, see Card, 1963

MacQueen, J., Some methods for classification and analysis of multi-variate observations, *5th Berkeley Symposium on Mathematics, Statistics and Probability*, *vol. 1*, 281-298, 1967.

Merkine, L., Convective and absolute instability of baroclinic eddies, *Geophys. Astrophys. Fluid Dyn.*, *9*, 129-157, 1977.

~~Mo, K. C.~~, and M. Ghil, Statistics and dynamics of persistent anomalies, *J. Atmos. Sci.*, *44*, 1987 [in press].

~~———————, and R. E. Livesey, Tropical extratropical geopotential height teleconnections during the Northern Hemisphere winter, *Mon. Wea. Rev.*, [in press].~~

Mo, K.C., Quasi-stationary states in the Southern Hemisphere, *Mon. Wea. Rev.*, *114*, 808-823, 1986.

Namias, J., *Short Period Climatic Variations.* Collected works of J. Namias, vols. I and II (1934-1974) and vol. III (1975-1982). Univ. of California, San Diego, 905 pp. + 393 pp., 1982.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions, *Mon. Wea. Rev.*, *110*, 699-706, 1982.

Pedlosky, J., Resonant topographic waves in barotropic and baroclinic flows, *J. Atmos. Sci.*, *38*, 2626-2641.

Pierrehumbert, R. T., Spatially amplifying modes of the Charney baroclinic instability problem, *J. Fluid Mech.*, *170*, 293-317, 1986.

Rasmusson, E. M., and J. M. Wallace, Meteorological aspects of the El Niño/ Southern oscillation, *Science, 222,* 1195-1201.

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

Shutts, G. J., A case study of eddy forcing during an Atlantic blocking episode, *Adv. Geophys., 29,* 135-162, 1986.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, London/New York, 175 pp., 1986.

Speranza, A., Deterministic and statistical properties of Northern Hemisphere, middle latitude circulation: Minimal theoretical models, *Adv. Geophys., 29,* 199-225, 1986.

Sutera, A., Probability density distribution of large-scale atmospheric flow, *Adv. Geophys., 29,* 227-249, 1986.

Tapia, R. A., and J. R. Thompson, *Nonparametric Probability Density Estimation,* Johns Hopkins Univ. Press, Baltimore/London, 176 pp., 1978.

Trenberth, K. E., and K. C. Mo, Blocking in the Southern Hemisphere, *Mon. Wea. Rev., 133,* 3-21, 1985.

Wahba, G., and J. Wendelberger, Some new mathematical methods for variational objective analysis using splines and cross-validation, *Mon. Wea. Rev., 108,* 1122-1143, 1980.

Wallace, J. M., and M. L. Blackmon, Observations of low-frequency atmospheric variability, in *Large-Scale Dynamical Processes in the Atmosphere,* edited by B. Hoskins and R. Pearce, Academic Press, London/New York, pp. 55-94, 1983.

______________, and D. S. Gutzler, Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Mon. Wea. Rev., 109,* 784-812, 1981.

White, W. B., and N. E. Clark, On the development of blocking ridge activity over the central North Pacific, *J. Atmos. Sci.*, *32*, 489-502, 1975.

[Eliminate this page by starting ref. on p. 49]

TABLE 1. Percentage of Variance Associated with each EOF, and its Empirical Uncertainty

/3

EOF	Low Pass (%)		Band Pass (%)	
1	10.7 ± 1.1		5.0 ± 0.35	
2	8.8 ± 0.9		4.85 ± 0.34	
3	7.4 ± 0.8		4.26 ± 0.30	
4	6.7 ± 0.7		4.09 ± 0.28	
5	6.1 ± 0.6		3.00 ± 0.21	
6	5.3 ± 0.5		3.00 ± 0.21	
7	5.1 ± 0.5		2.49 ± 0.17	(24.2%)
8	4.4 ± 0.4	(50.1%)	2.42 ± 0.17	
9	3.9 ± 0.4		2.29 ± 0.16	
10	3.4 ± 0.3	(58.5%)	2.13 ± 0.15	(31.4%)
11	3.3 ± 0.3		2.10 ± 0.15	
12	2.9 ± 0.3		1.87 ± 0.13	
13	2.5 ± 0.2		1.81 ± 0.13	
14	2.3 ± 0.2		1.73 ± 0.12	
15	2.1 ± 0.2		1.69 ± 0.11	
Total	74.9%		42.79	

/5.00

Partial totals of variance are given for subsets of EOFs used in clustering computations.

Center under '15'

TABLE 2. Clusters in Model Phase Space: Number of Elements and Persistence

Cluster	Number of Points	% of Total Data	Number of Events of Given Duration (τ)										Average Persistence Times (τ)		
			1	2	3	4	5	6	7	8	9	≥ 10	T_d	T_p	T_m
1	772	11	7	7	1	3	2	1	6	0	0	9	21.4	40.8	14
2	768	11	15	24	17	11	9	11	15	4	2	20	6.01	10.0	16
3	142	2	25	36	11	3	0	0	-	-	-	0	1.89	0	9
4	121	1.7	66	21	0	2	1	0	-	-	-	0	1.35	5	10
5	88	1.2	82	3	0	-	-	-	-	-	-	0	1.04	0	12
Total/Ave.	1891	27	195	91	29	19	11	12	21	4	2	29	6.34	10.2	12.2

The ~~the mean~~ average persistence times in the last three columns are average duration T_d of all events in each cluster, the average duration T_p of all events lasting 5τ or longer, and the average time T_w between the trajectory's leaving a given cluster and its reaching another cluster. Sequences of zeros for numbers of events are replaced by dashes for easier reading.

c.t/ TABLE 3. Nearest Distance between the Points of a Flow Sequence in a Given Cluster, and the Center of that Cluster

Cluster

1	Persistence (τ)	170	148	111	91	76	20	17	7	6	5	4	3	2
	$\bar{d}_{min}$	0.45	0.48	0.50	0.90	0.68	1.15	1.84	1.50	1.78	1.94	1.93	2.20	2.32

2	Persistence (τ)	34	33	23	19	18	17	16	14	12	9
	$\bar{d}_{min}$	0.82	0.70	0.87	1.05	1.06	1.04	0.89	1.20	1.18	1.23

~~1 (cont'd)~~	~~4~~	~~3~~	~~2~~
	~~1.93~~	~~2.20~~	~~2.32~~

	8	7	6	5	4	3	2
2 (cont'd)	1.17	1.32	1.40	1.41	1.58	1.60	1.72

9/ Values of $\bar{d}_{min}$ are nondimensional (see *Legras and Ghil* [1985], eqs. (1-4) and 9(11)).

[Put this table lying down on page to fit without continuations]

TABLE 4. Fuzzy Clusters in Model Phase Space, Including Small-Anomally Cluster

Cluster	Number of Points	% of Data	Number of Events of Given Duration (τ)										Average Persistence Times (τ)		
			1	2	3	4	5	6	7	8	9	>10	T_d	T_p	T_W
1	1222	18	31	20	11	7	9	6	1	8	5	16	10.7	24.2	4.9
2	1384	20	41	17	13	7	7	0	1	3	1	18	10.9	35.3	4.6
3	462	6.6	40	50	40	25	18	2	0	-	-	0	2.6	5.1	3.7
4	530	7.5	83	113	47	9	3	2	0	1	1	0	2.1	6.3	3.0
5	384	5.4	86	130	5	3	2	0	-	-	-	0	1.7	0	3.5
6	333	4.8	62	42	21	16	3	3	1	0	0	1	2.2	6.6	3.4
Total/Ave.	4315	62	343	372	137	67	42	13	3	12	7	35	5.03	12.9	3.85

See Table 2 and text for definition of T_d, T_p and $T_{\cancel{W}}$.

l.c. w

TABLE 5. Transition between Pairs of Model Clusters

To From	1	2	3	4	5	6	Sum	Average ± Std. Deviation
1	*31*	0	18	*42*	16	6	113	19 ± 1.33
2	0	*20*	4	14	*47*	*23*	108	18 ± 1.53
3	*54*	10	11	27	4	*69*	175	29.1 ~~2.0~~ ± 1.83
4	5	18	40	*137*	48	11	259	43.1 ± 2.78
5	2	*45*	*91*	38	30	20	226	37.6 ± 1.92
6	22	15	11	1	*80*	20	149	24.8 ± 3/7
Sum	114	108	175	259	225	149	1030	

/29.1

/ skipped 1 line

The numbers which are significantly larger than those given by equal probabilities are indicated by boldface. Significant is taken as average number of transitions, plus one standard deviation, e.g., 20.3 for transitions from Cluster 1, or 19.5 from Cluster 2.

(ital. OK, but heavy type also)

a
see Eq.(10),
p.23

TABLE 6: Statistics of Clusters in Low-Pass Filtered Data

✓Cap E

ster	No. of elements (days) and percentages (%)											Average Times	Persistence (Days)		Q_1 n Cluster	
	Hard		Fuzzy		Projection along EOF axis											
	No.	%	No.	%	1	2	3	4	5	6	7	T_d	T_p	T_w	No.	%
	65	2.68	150.5	6.2	-8.9	-1.0	2.9	-3.3	-2.6	-2.7	2.9	6.3	9.9	7.9	130	86.4
	57	2.35	139	5.7	7.4	3.2	-6.0	-1.2	-0.3	0.8	-0.5	6.3	8.6	11.6	95	68.3
	54	2.23	161.5	6.6	-6.6	4.4	-5.4	-2.4	2.1	2.6	1.6	8.1	11.3	7.7	136	84.2
	40.5	1.67	86.5	3.6	-6.3	-7.9	-0.2	7.4	-1.3	-3.4	-3.6	7.0	9.3	12.3	75	86.7
	38	1.57	130.5	5.3	6.7	-4.7	6.6	-2.4	2.2	0	-1.0	5.2	7.9	9.5	71	54.4
	36	1.48	86.5	3.6	4.3	1.5	-3.6	7.2	2.7	-8.1	1.3	5.2	9.8	14.0	59	68.2
	36	1.48	116.5	4.8	2.1	5.4	7.9	1.5	1.4	-5.0	0.9	5.1	8.9	9.5	54	46.4
	27	1.11	68	2.8	1.6	0.9	2.4	-3.1	3.9	-0.2	4.6	3.8	7.7	6.4	23	33.8
	66	2.72	66	2.7	0	0	0	0	0	0	0	2.4	6.0	6.4	18	27.3
:al	4195	17.3	1005	41.3							Ave	5.49	8.71	9.48	661	

e number of elements is listed in days, and 0.5 indicates that only one map out of two for a ven day is in the cluster. Last two columns give the number of days of QS events within each uster, and the corresponding percentage.

ORIGINAL PAGE IS
OF POOR QUALITY

Figure Captions

1 Streamfunction plots for sample persistent events (a) ~~blocking~~ zonal regime (b) ~~zonal~~ blocking regime ✓

.2 . Schematic diagram of clustering algorithm. ✓

~~2~~ 3. Center of (a) cluster, (b) cluster 2 (c) cluster 3 (d) cluster 4 and (e) cluster 5 (contour interval 5 × (0.1 standarized unit) ✓

→ 5 ~~7~~.(a) Correlations between the time series and the center of cluster 1. when $R_1 < 0.65$, we set $R_1 = 0$. ✓
(b) R_2 Same as (a) but for cluster 2
(c) Function Q(t), Q(t) = 1 when ~~at~~ time step t, the pattern is OS. Q(t) = 0, otherwise

ig ~~6~~4. positions of clusters projected onto the first three EOF axis ~~and the major transitions among them.~~ ✓

Fig. 6. Transition diagram ✓

ig 7 Density distribution functions (heavy line) fitted using the non parametric estimation theory for the ~~(a)~~ ~~cluster~~ ~~2~~ standarized (a) EOF1 amplitudes ~~(b)~~ (b) EOF2 (c) EOF3 amplitudes. The light line indicates the normal distribution function obtained using same data. ✓

'g 8 Histogram in EOF1 and 2 phase space for standarized EOF1, 2 amplitudes. ✓

g. 9. EOFs 1, 2, 3 ✓

'g ~~8~~10. Density distribution functions for amplituds of (a) EOF1 (b) EOF2 and (c) EOF3 for low pass filtered data. The light solid line is for the total data set, the ✓ heavy ~~solid~~ solid line is for points in the OS regime obtained using the pattern correlation method.

11. Composite of ~~total~~ anomalies for all members in the cluster (R1=0.82, R2=0.34 and R3=0.65). of (a) cluster 1 ~~Fo~~ (b) cluster 2 (c) cluster 3 (d) cluster 4 (e) cluster 5 (f) cluster 6 ~~(g) cluster 7 and (h) cluster 8~~ Contour interval 30 m ~~an~~. Areas where ~~which~~ the mean has the 95% confidence level ~~to~~ are shaded. ✓

'g 12. Composite of ~~total~~ anomalies of all members in cluster 3 after taking out ✓ the first and second harmonics of each individual year.

3. Correlation R_i (i=1 to 8) between the center of cluster i and the time series for low pass filtered data. where ✓ $R_i < 0.65$, we set $R_i = 0$.

* and $\theta(t)$ where $\theta(t) = 1$ if $t \geq 0.5$
$= 0$ otherwise

g.14 ~~Flow patterns of~~

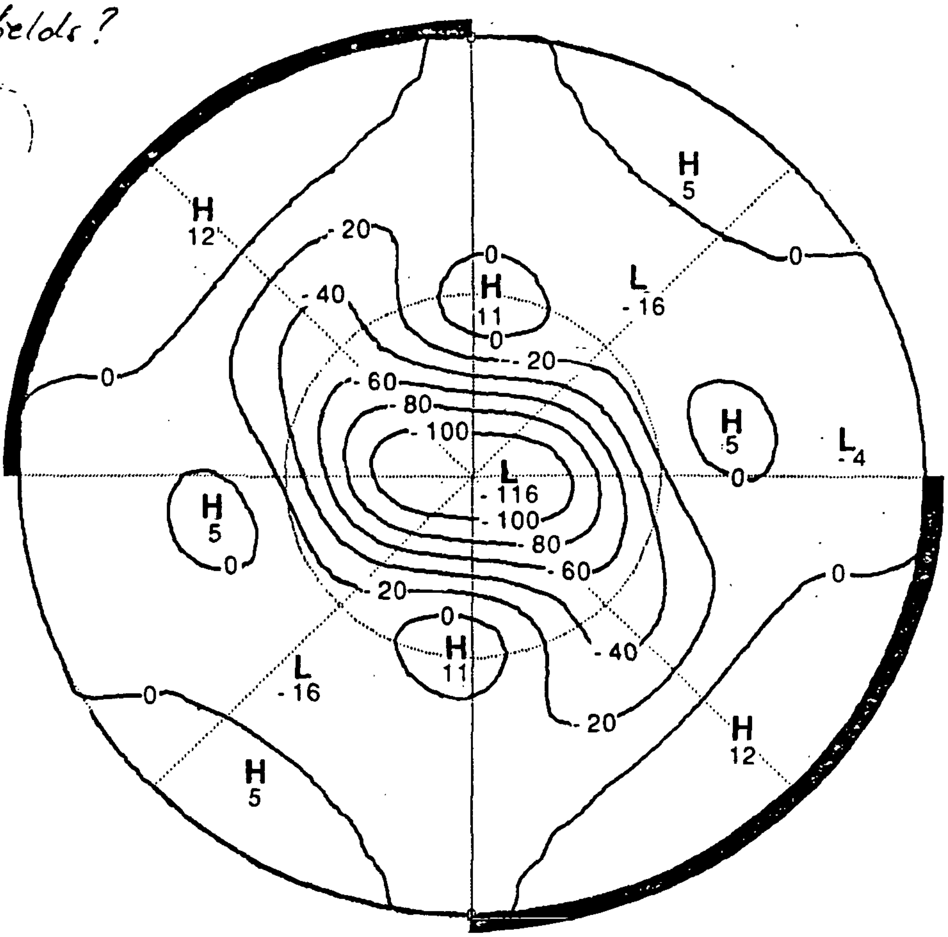
ig15 Center of band pass filtered clusters (a) cluster 1, (b) cluster 2, (c) cluster 3 ✓ (d) cluster 4 (e) cluster 5 (f) cluster 6 contour interval 20 m.

* Transition diagram.

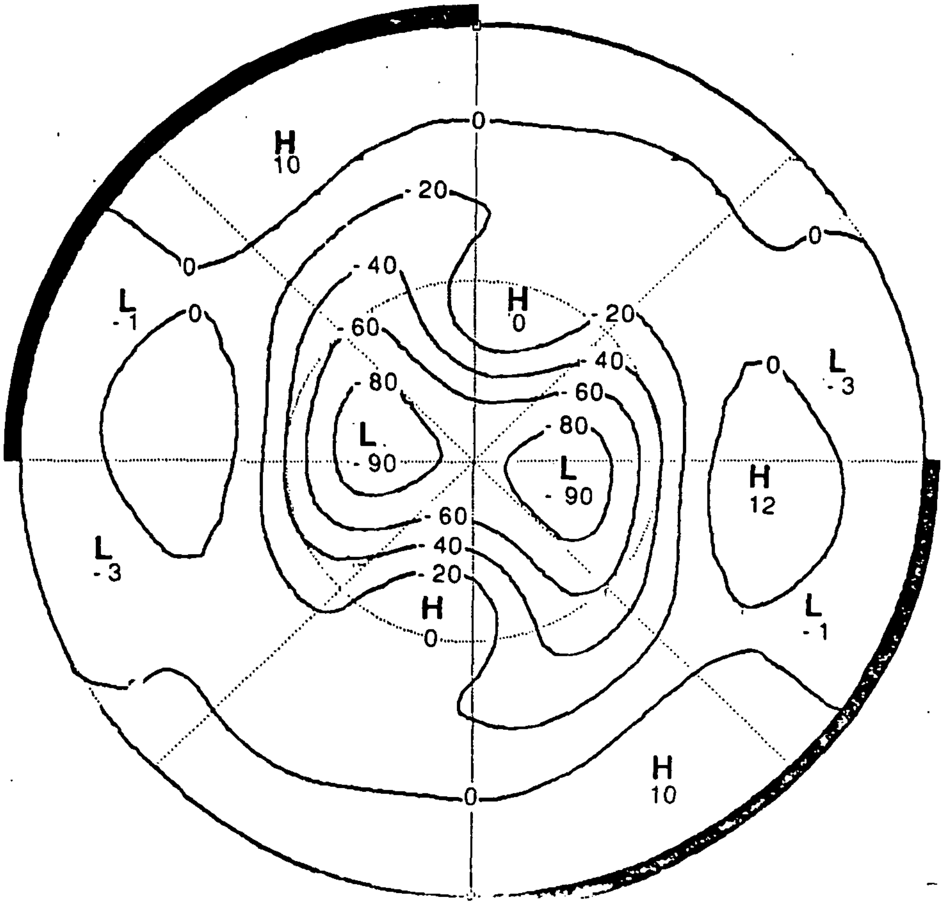
ORIGINAL PAGE IS
OF POOR QUALITY

Kingisc : Total fields?

Zonal ?
→ 1b



Blocked ?
→ 1a



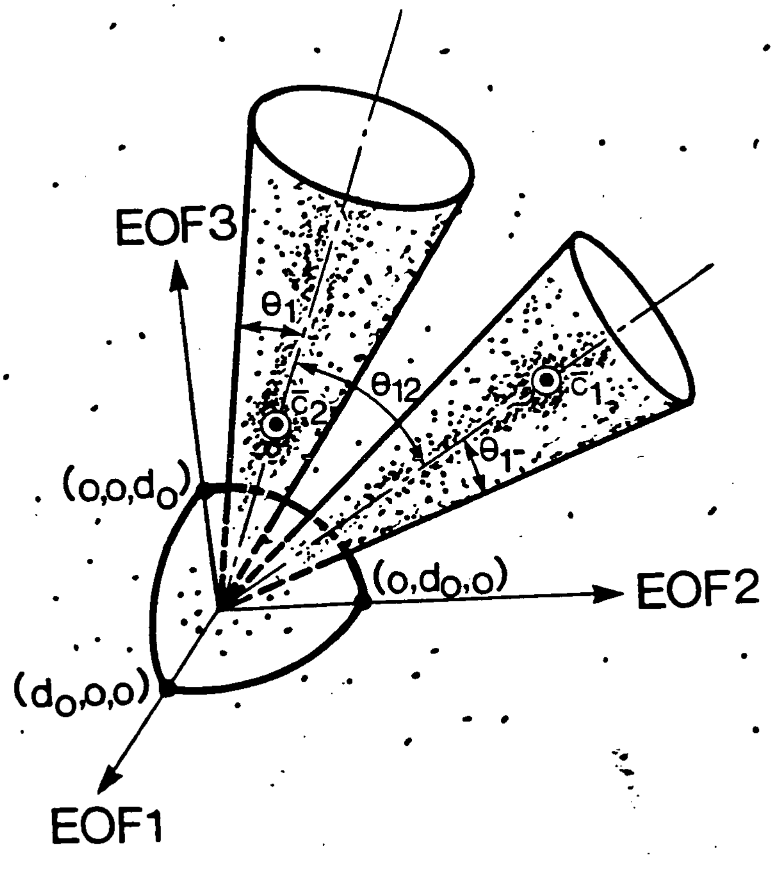
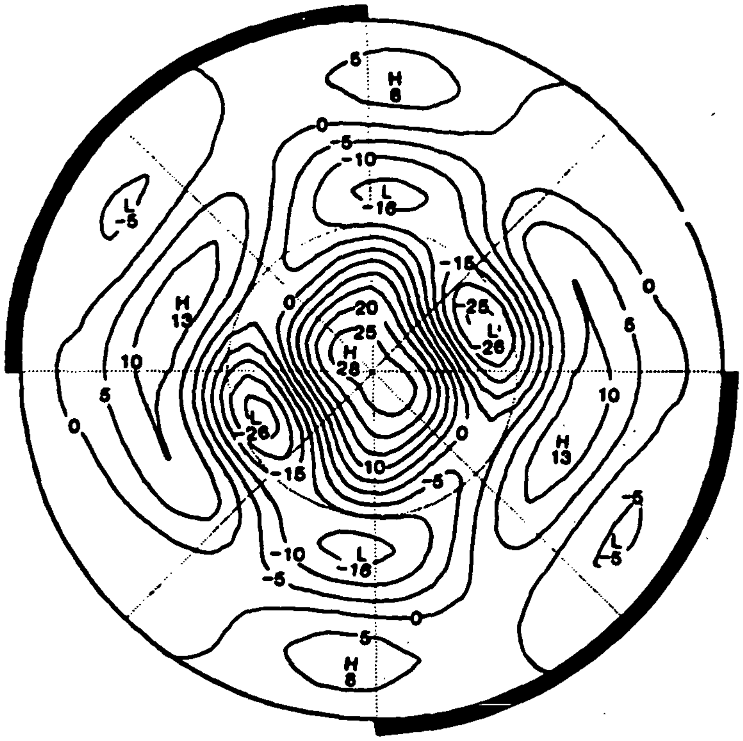


Fig. 2

nomalies

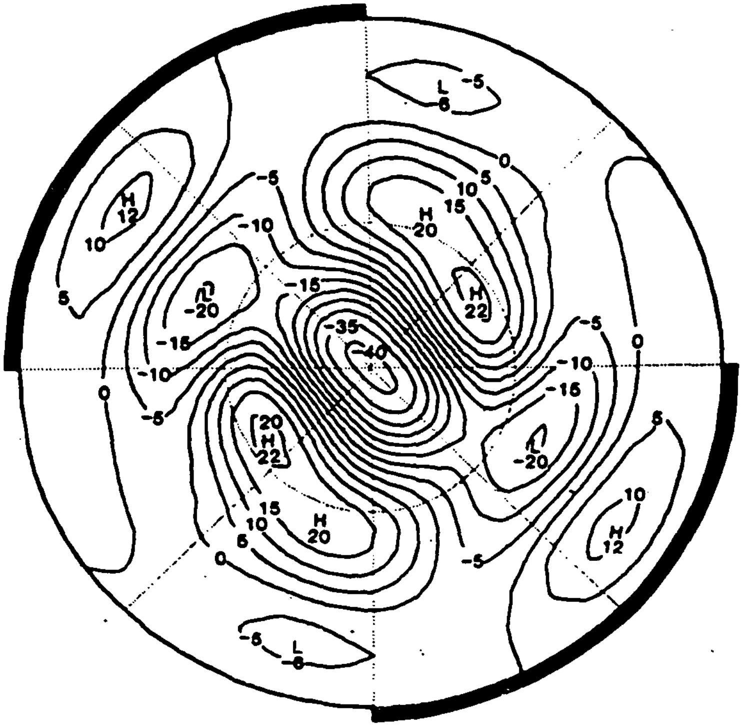
c

(blocking)



d

(zonal)



#

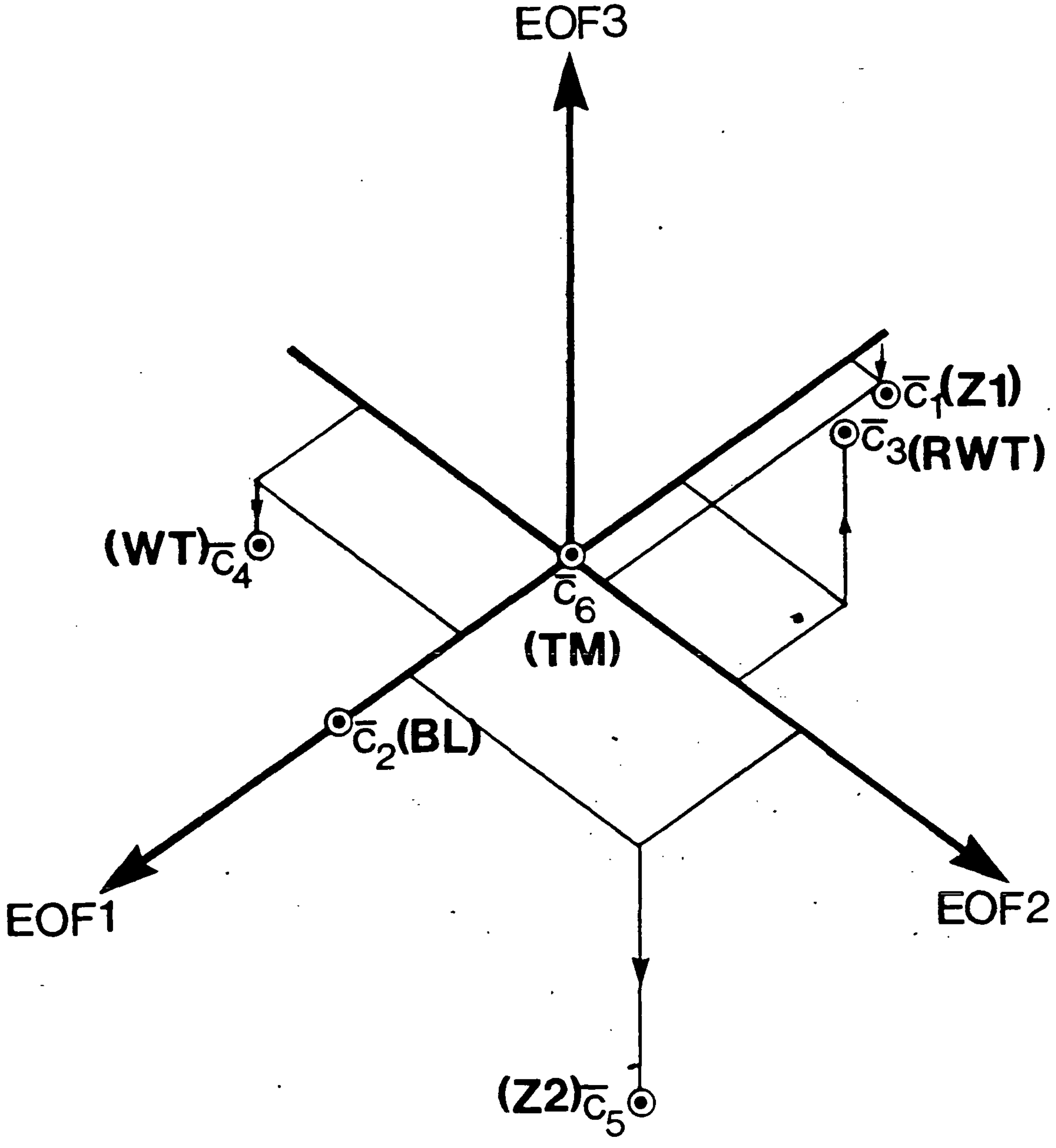

EOF3
$\bar{c}_1$(Z1)
$\bar{c}_3$(RWT)
(WT)$\bar{c}_4$
$\bar{c}_6$
(TM)
$\bar{c}_2$(BL)
EOF1
EOF2
(Z2)$\bar{c}_5$

Fig. 4

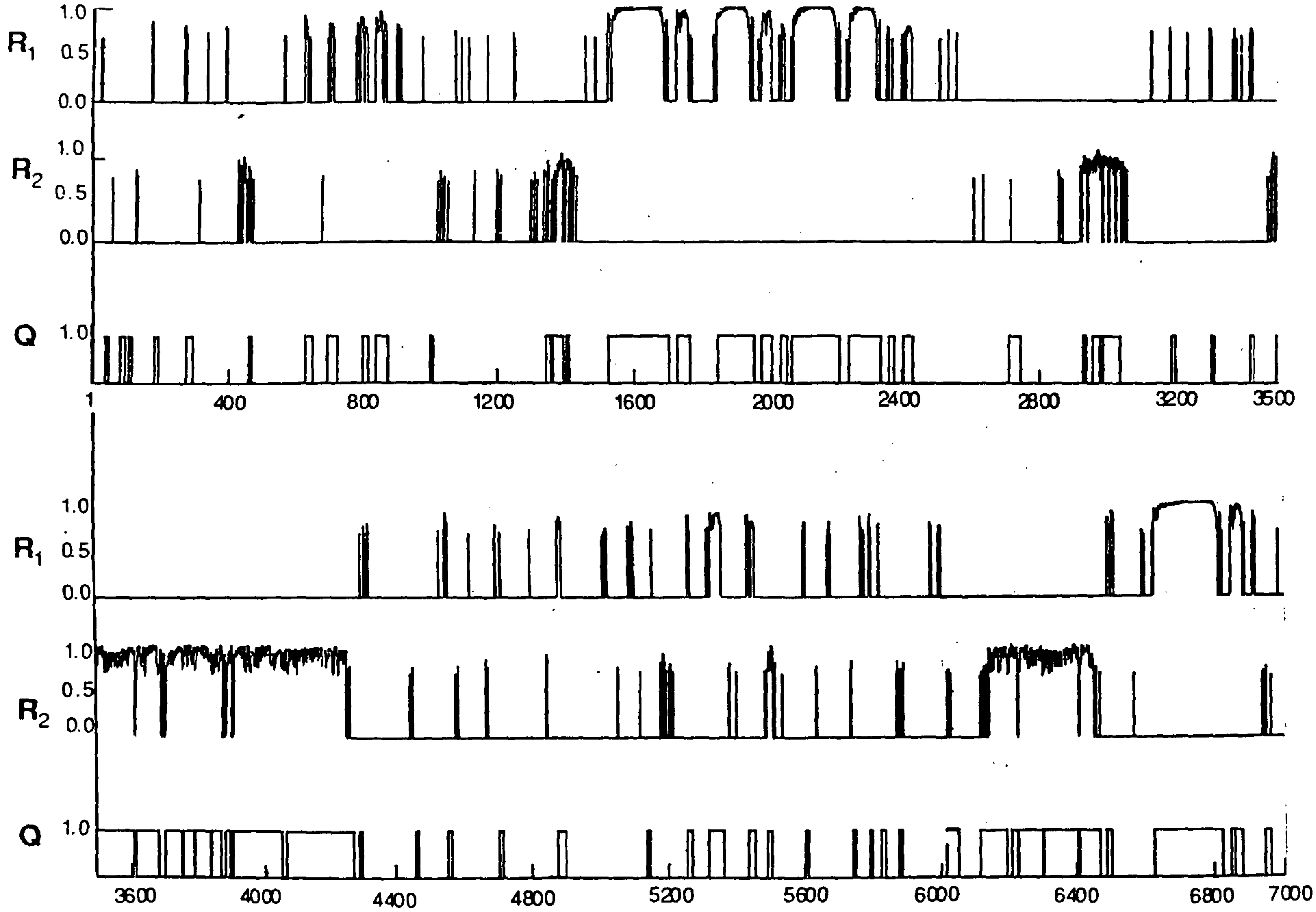

R1
1.0
0.5
0.0
R2
1.0
0.5
0.0
Q
1.0
1
400
800
1200
1600
2000
2400
2800
3200
3500
R1
1.0
0.5
0.0
R2
1.0
0.5
0.0
Q
1.0
3600
4000
4400
4800
5200
5600
6000
6400
6800
7000

Fig. 5

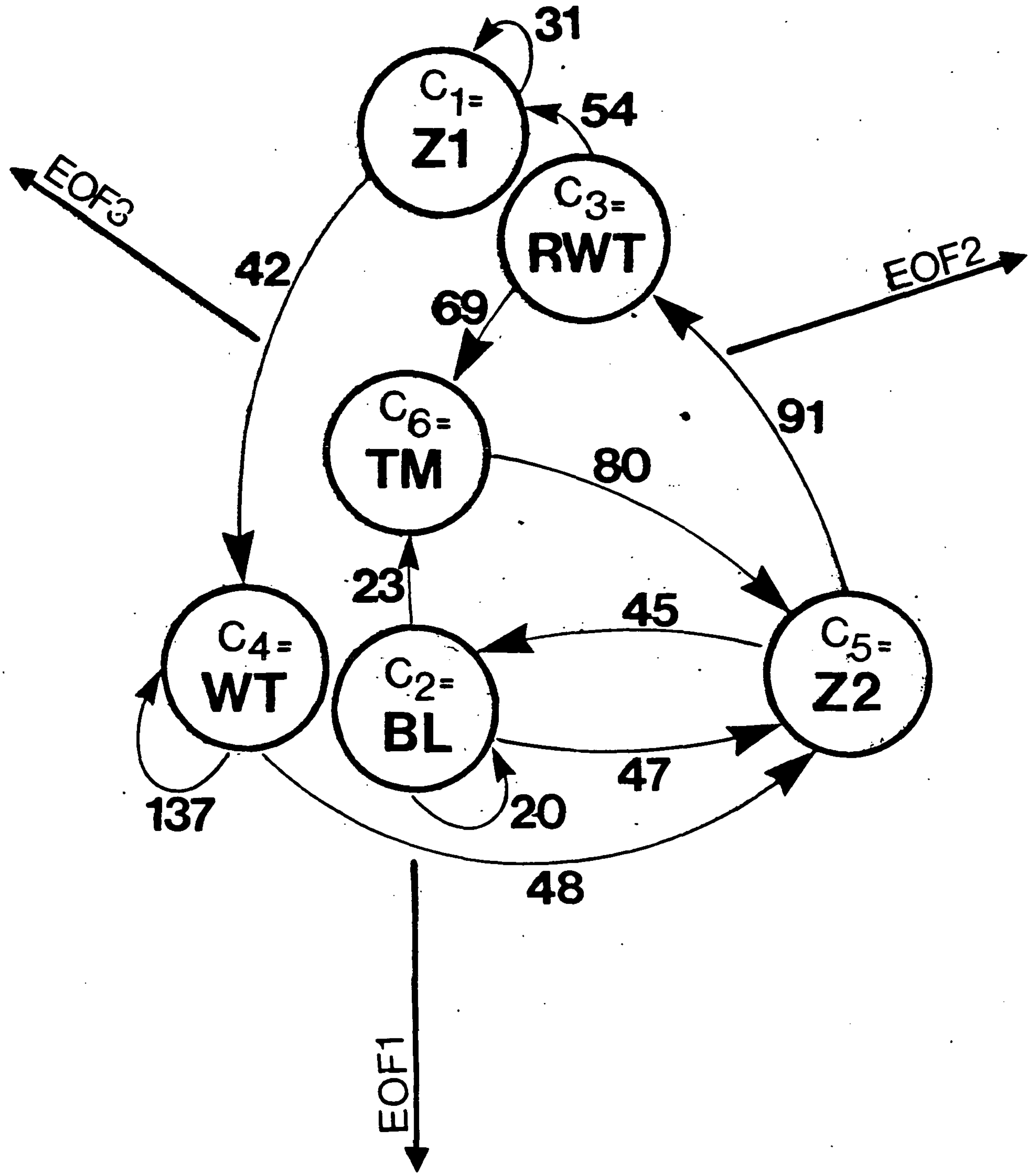
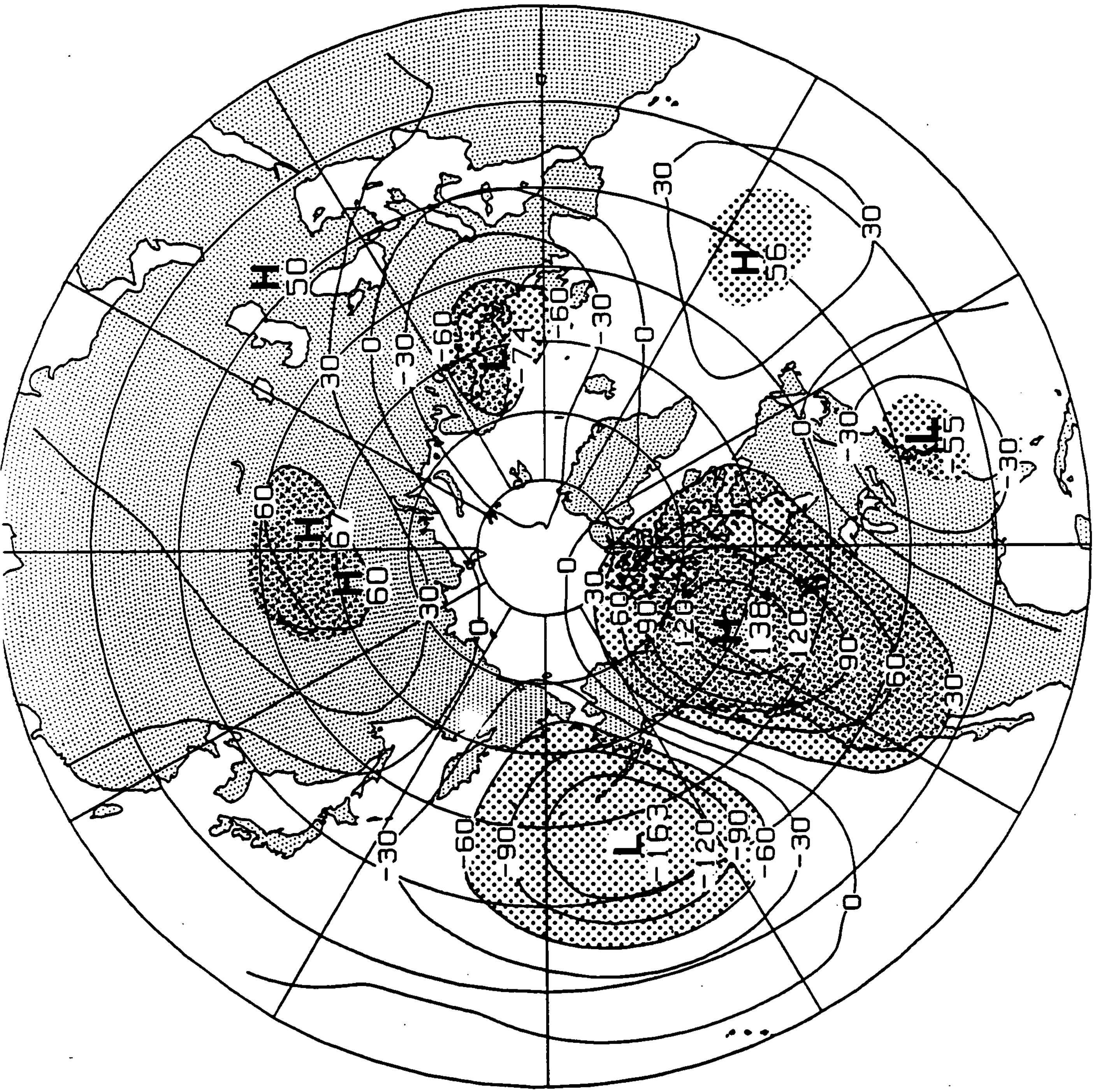
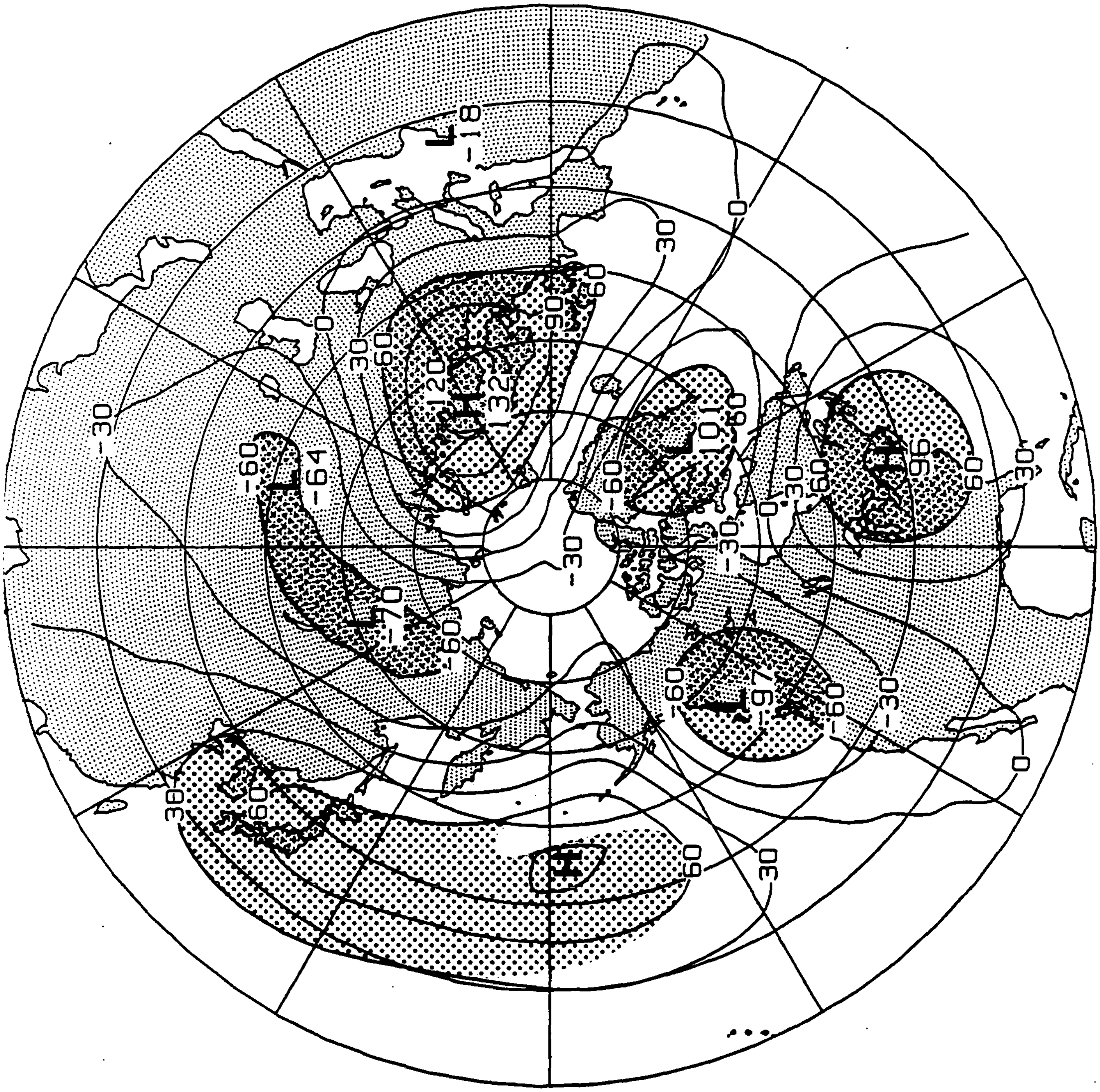
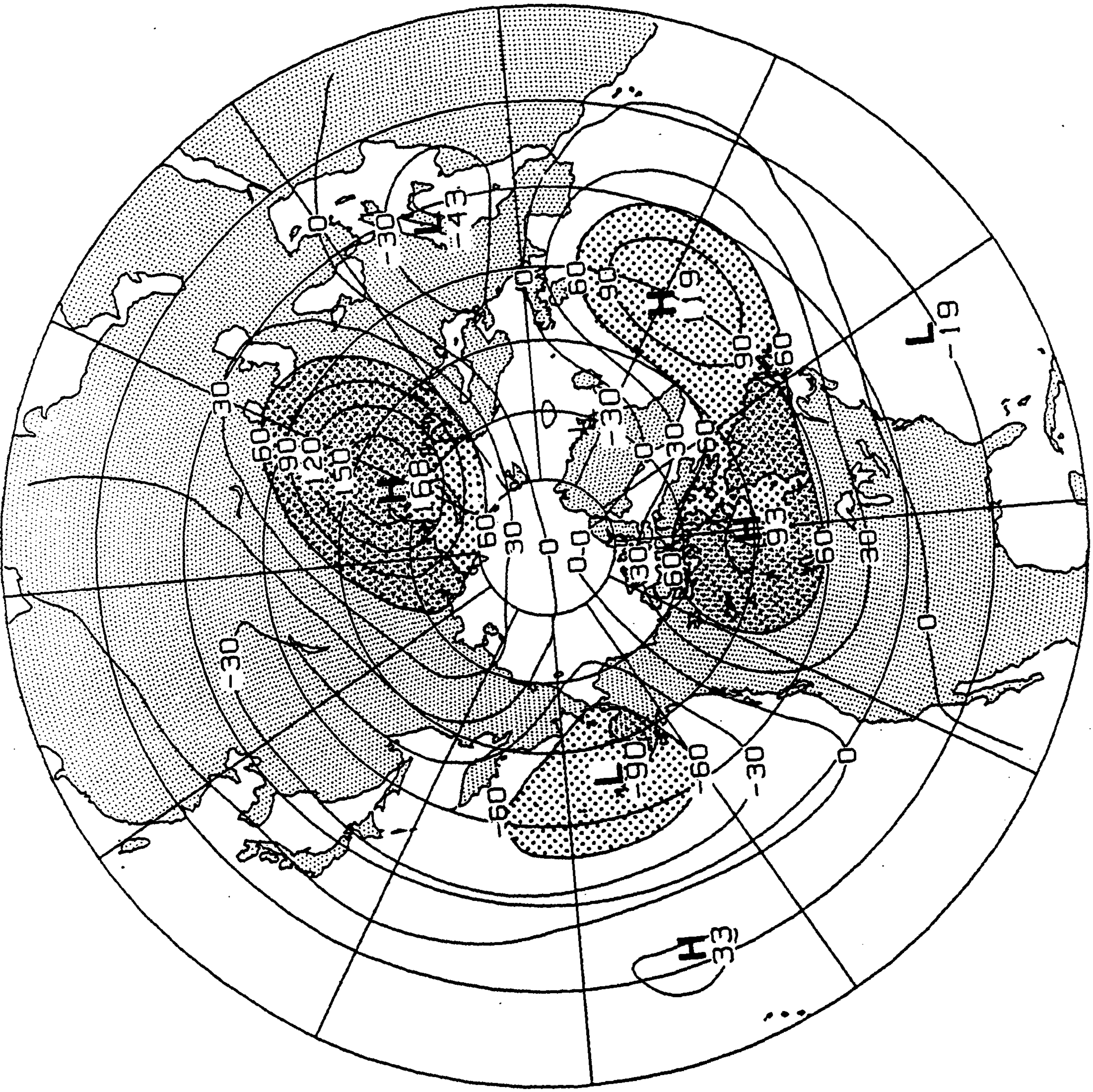

31
C1=
Z1
54
C3=
RWT
EOF3
42
69
EOF2
C6=
TM
80
91
23
45
C4=
WT
C2=
BL
C5=
Z2
137
20
47
48
EOF1

Fig. 6



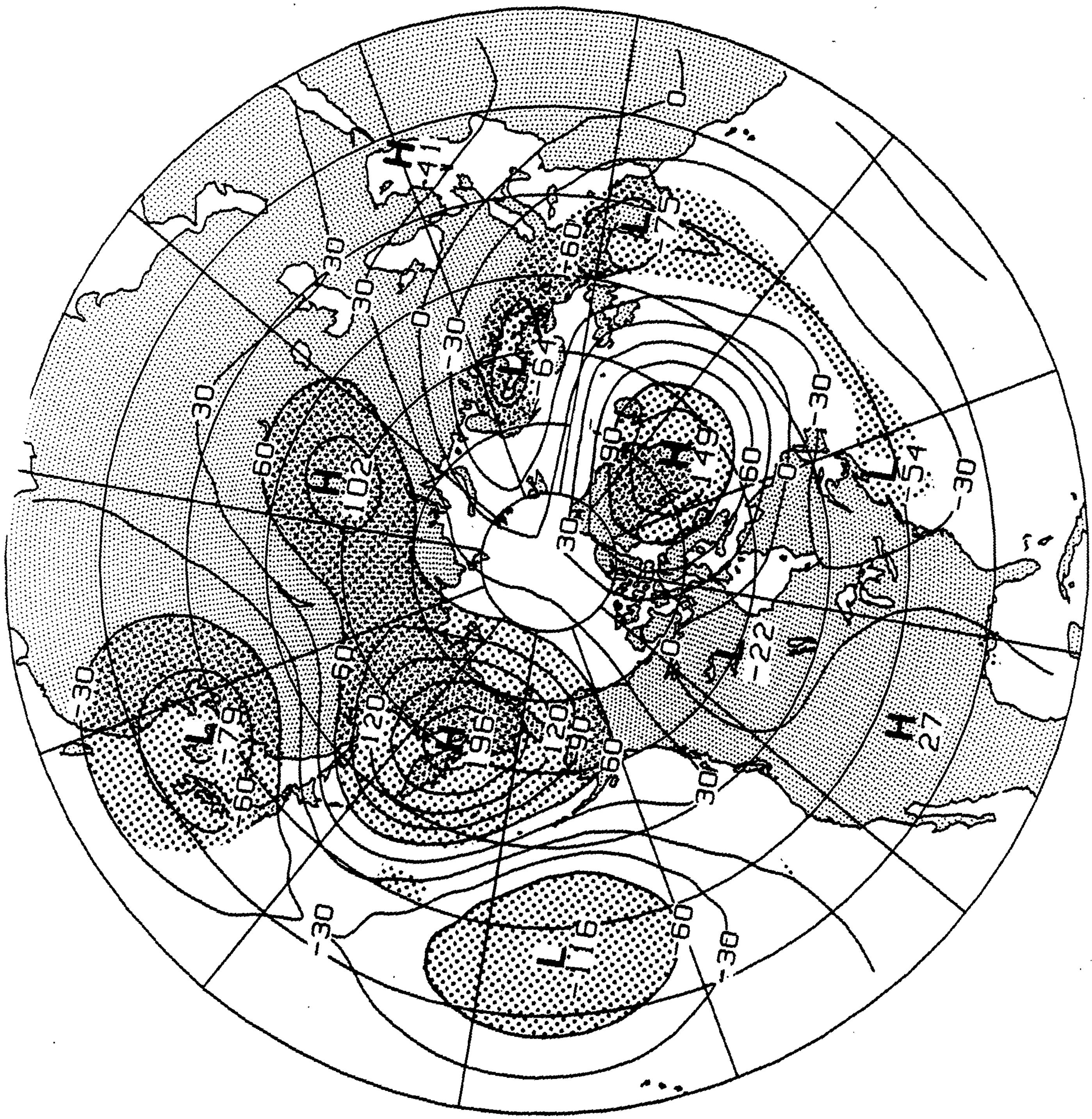
1/2

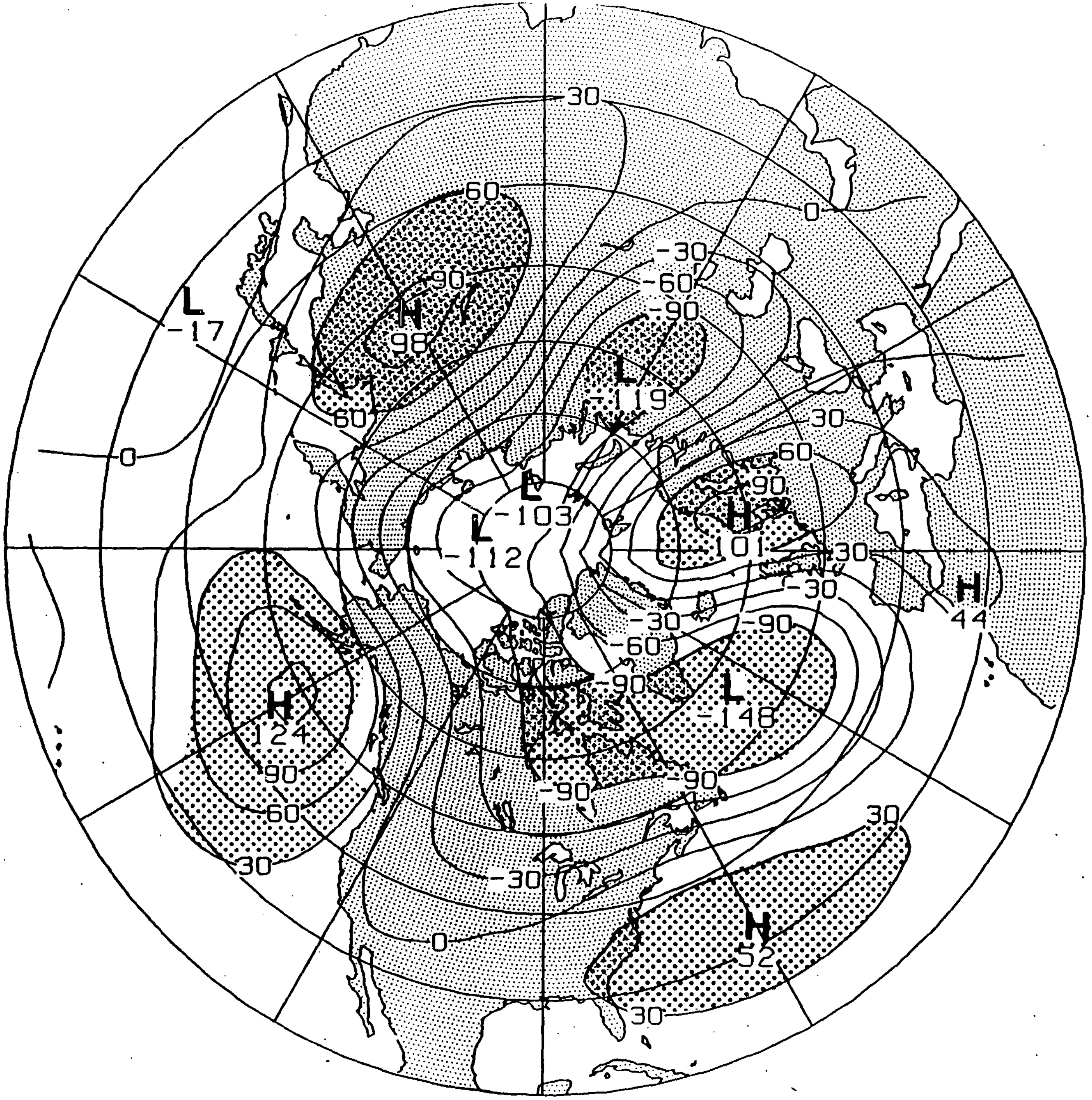

L
-18
0
30
60
90
60
H
120
132
30
60
0
-30
-60
L
-64
L
-70
-60
-30
-60
L
101
-60
0
-30
-60
H
95
60
30
L
-97
-60
-60
-30
0
30
60
H
60
30



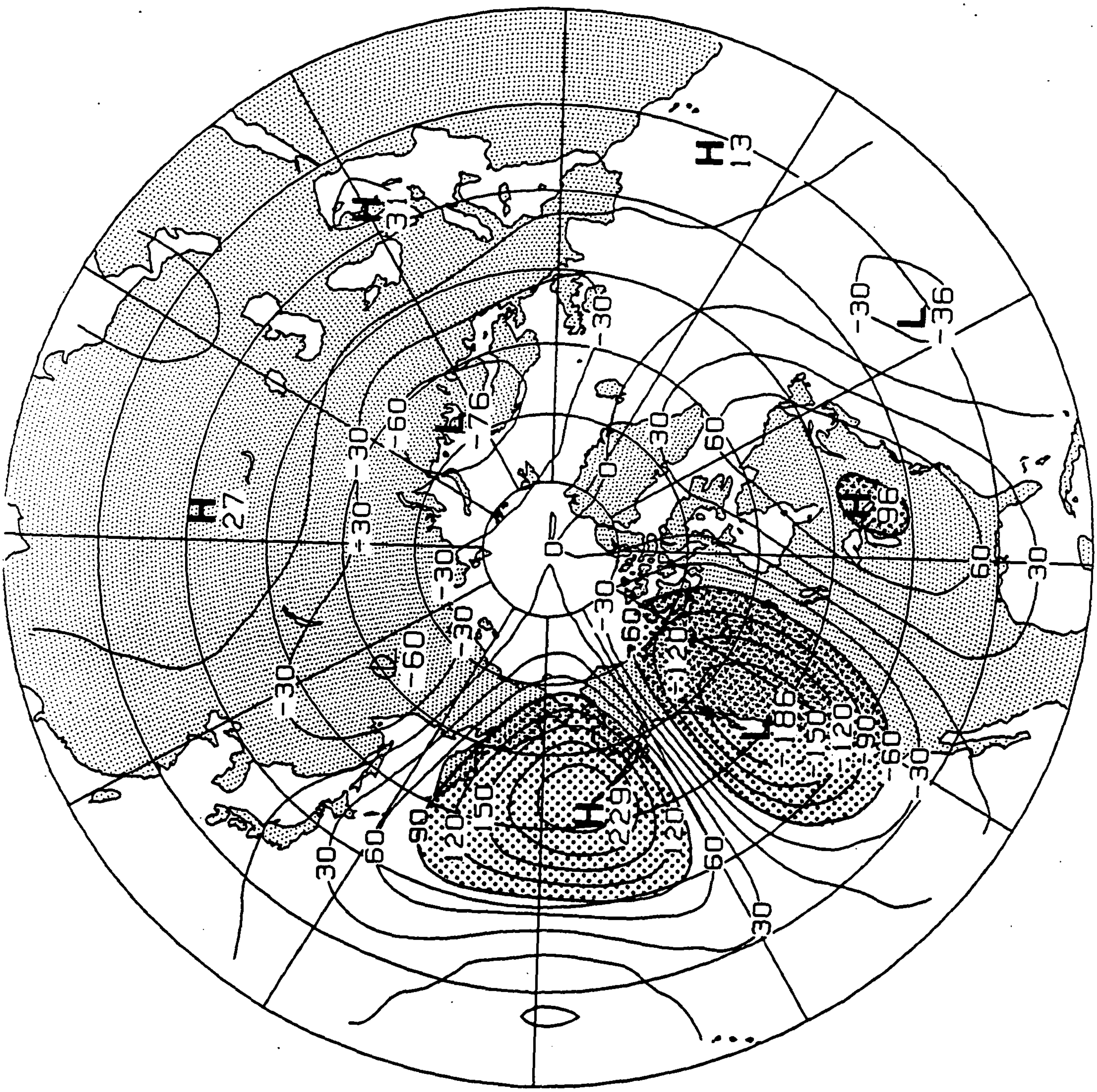
H
168
150
120
90
60
30
-30
L
-43
-30
0
H
119
90
60
30
0
-30
L
-19
H
93
60
30
0
L
-90
-60
-30
H
33

INAL PAGE IS
OF POOR QUALITY



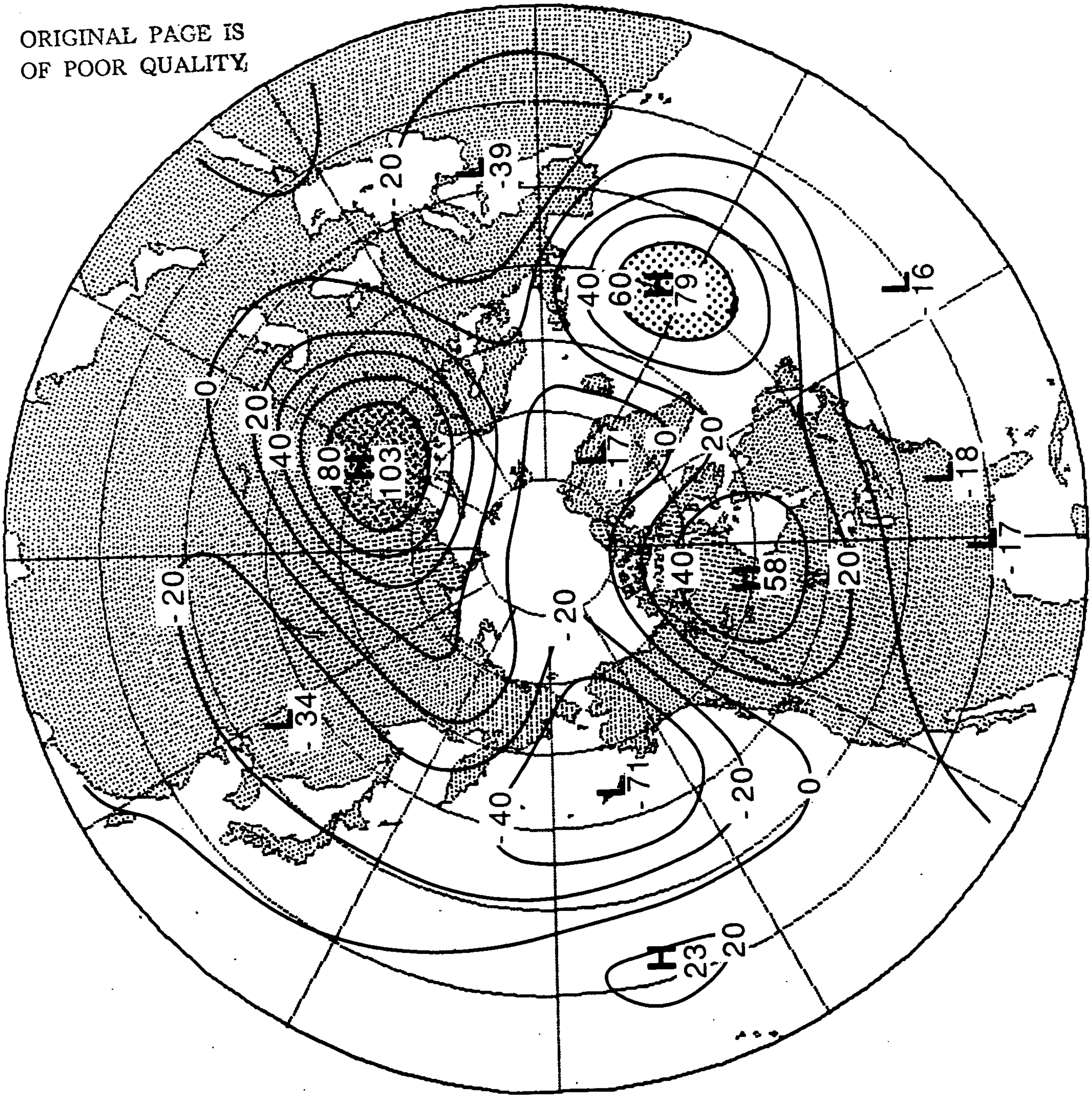

30
60
0
-30
-60
-90
L
-17
H
98
90
L
-119
60
30
0
60
90
H
101
30
-30
H
44
L
-103
L
-112
-30
-60
-90
-90
L
-148
H
124
90
60
30
-90
-90
30
-30
0
H
52
30

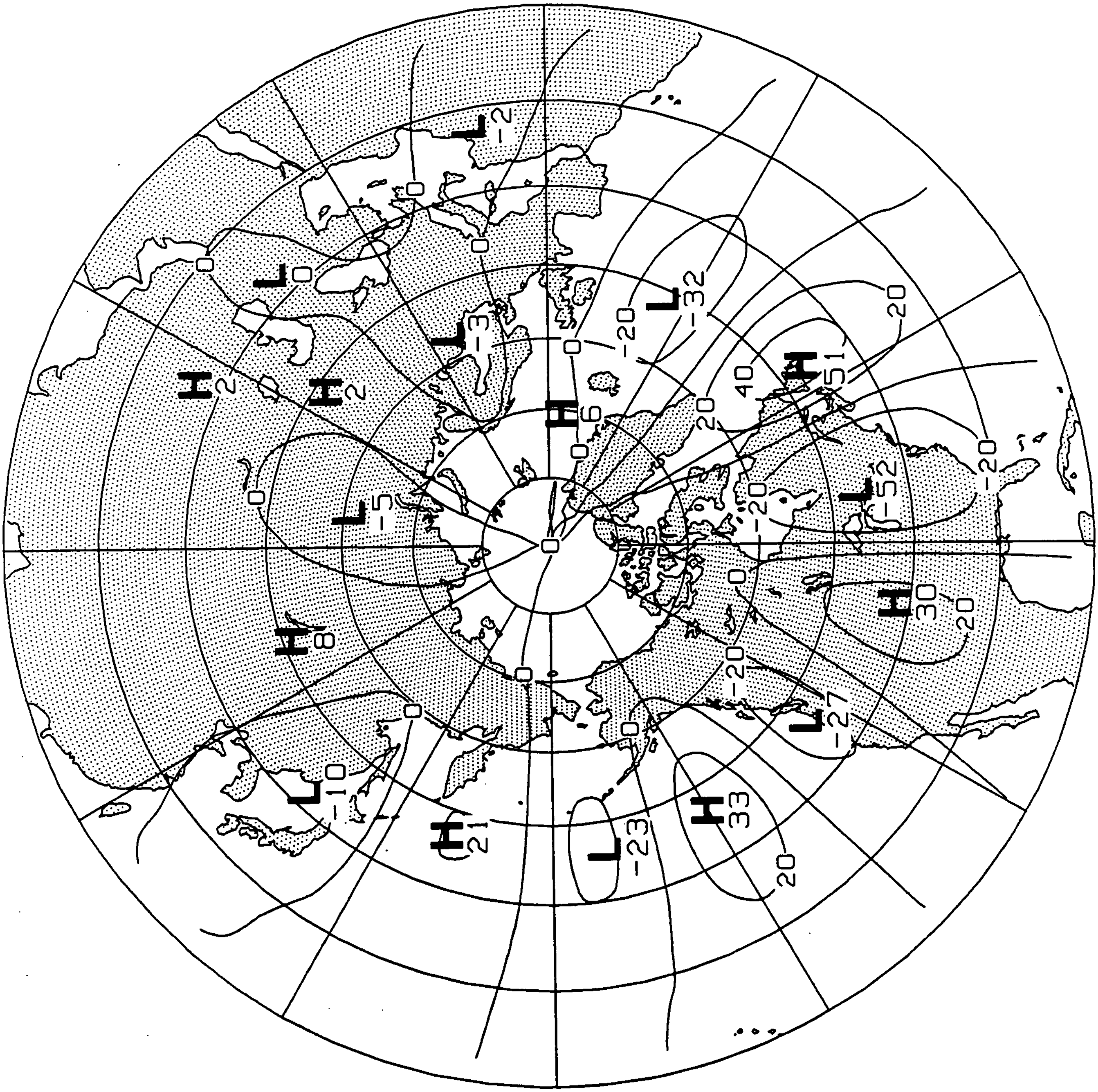
1e

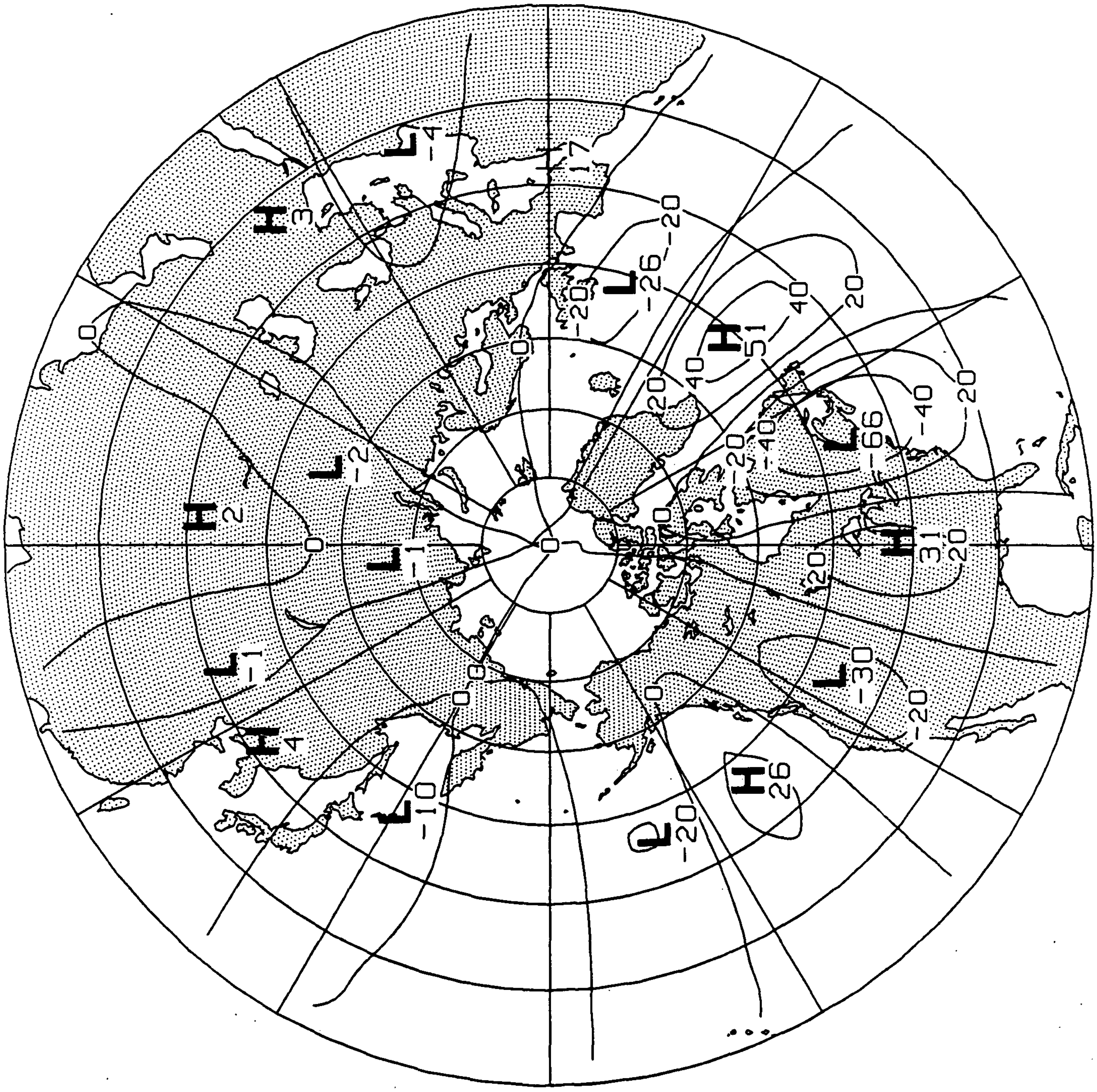


ORIGINAL PAGE IS
OF POOR QUALITY

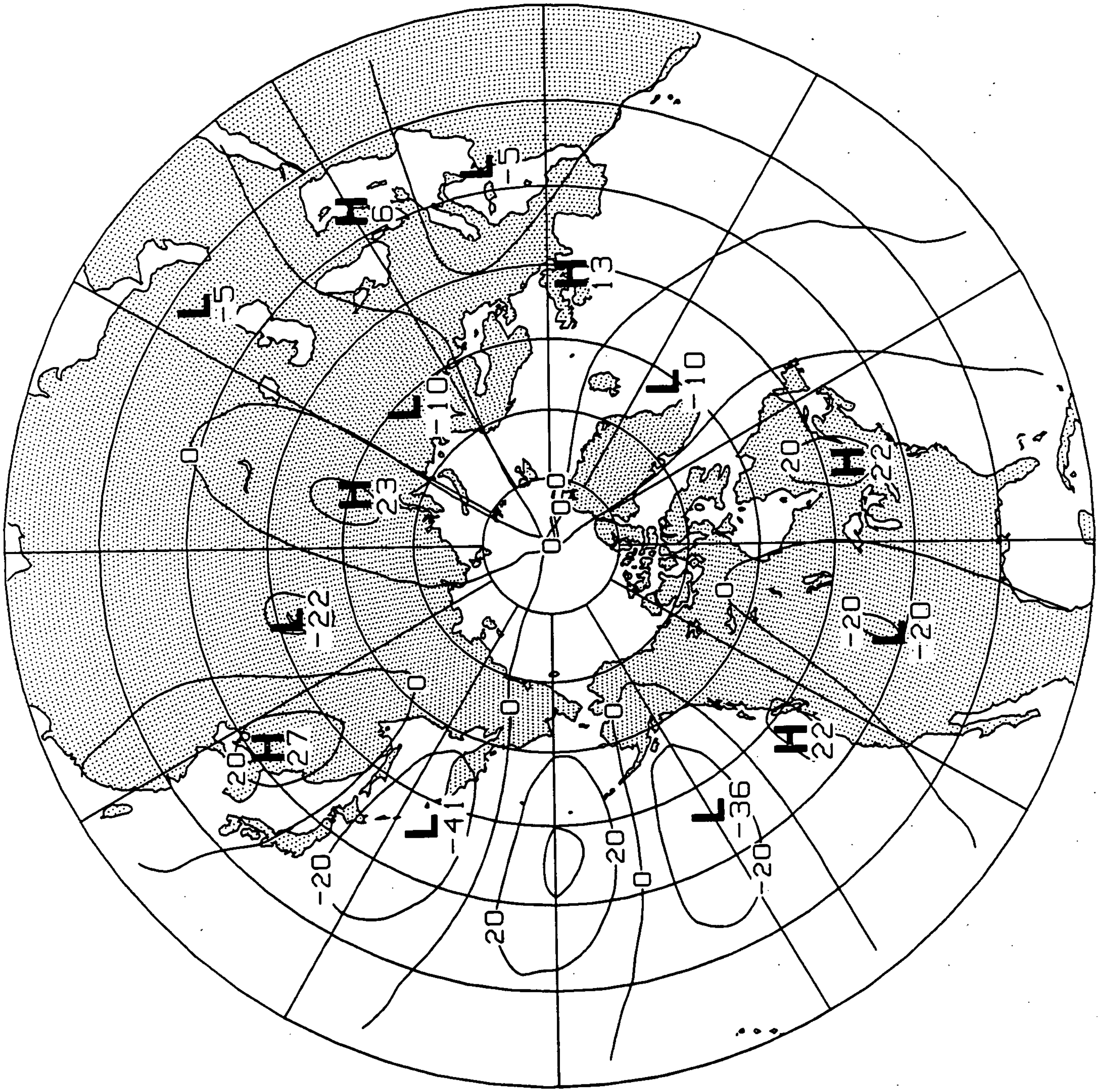
ORIGINAL PAGE IS
OF POOR QUALITY

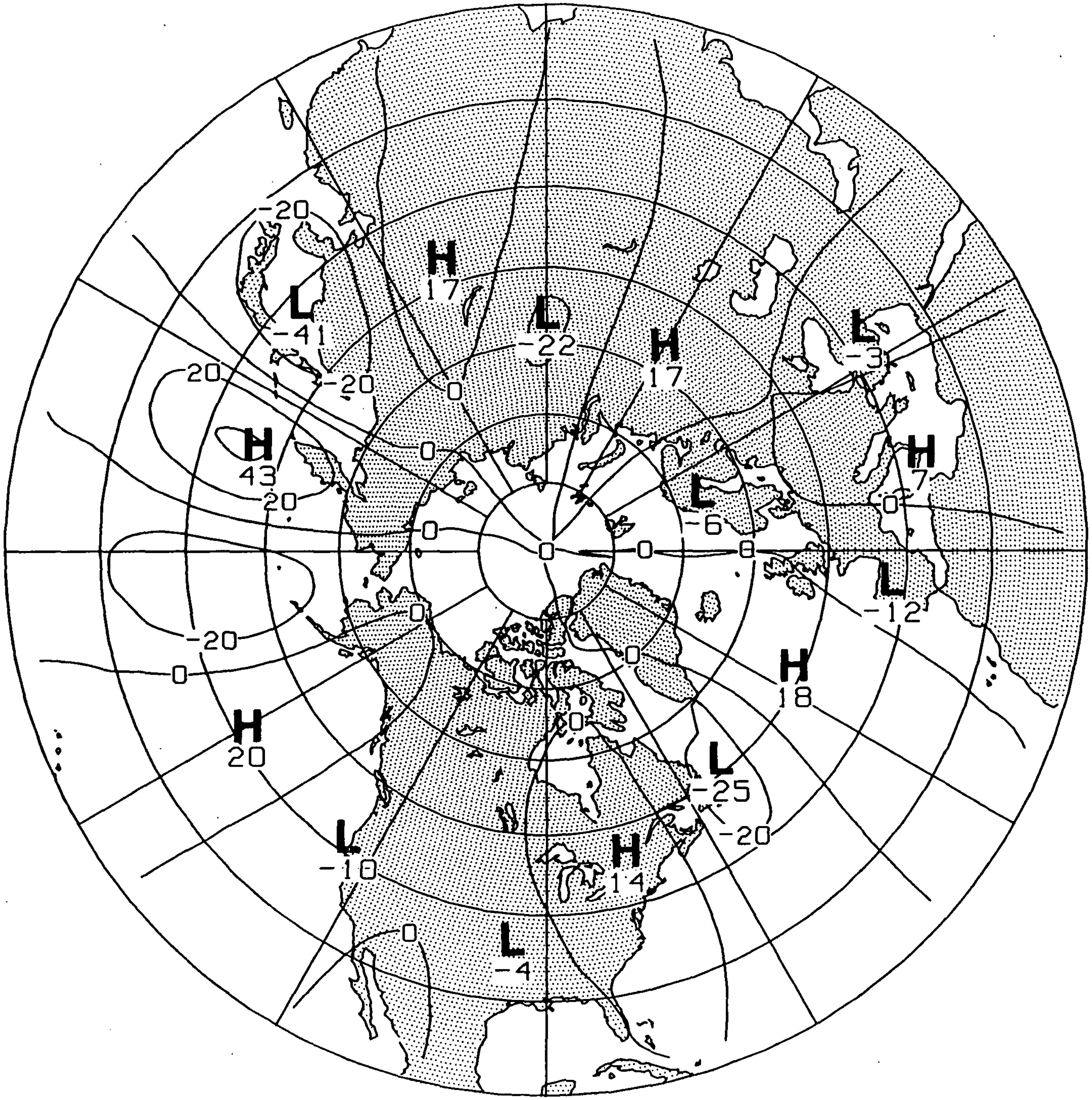



L
-2
0
L
0
0
L
-3
H
2
H
2
H
6
0
0
L
-5
0
H
8
0
L
-10
H
21
L
-23
H
33
20
L
-32
-20
20
40
H
51
20
-20
L
-52
-20
0
-20
H
30
20
-20
L
-27
20
0

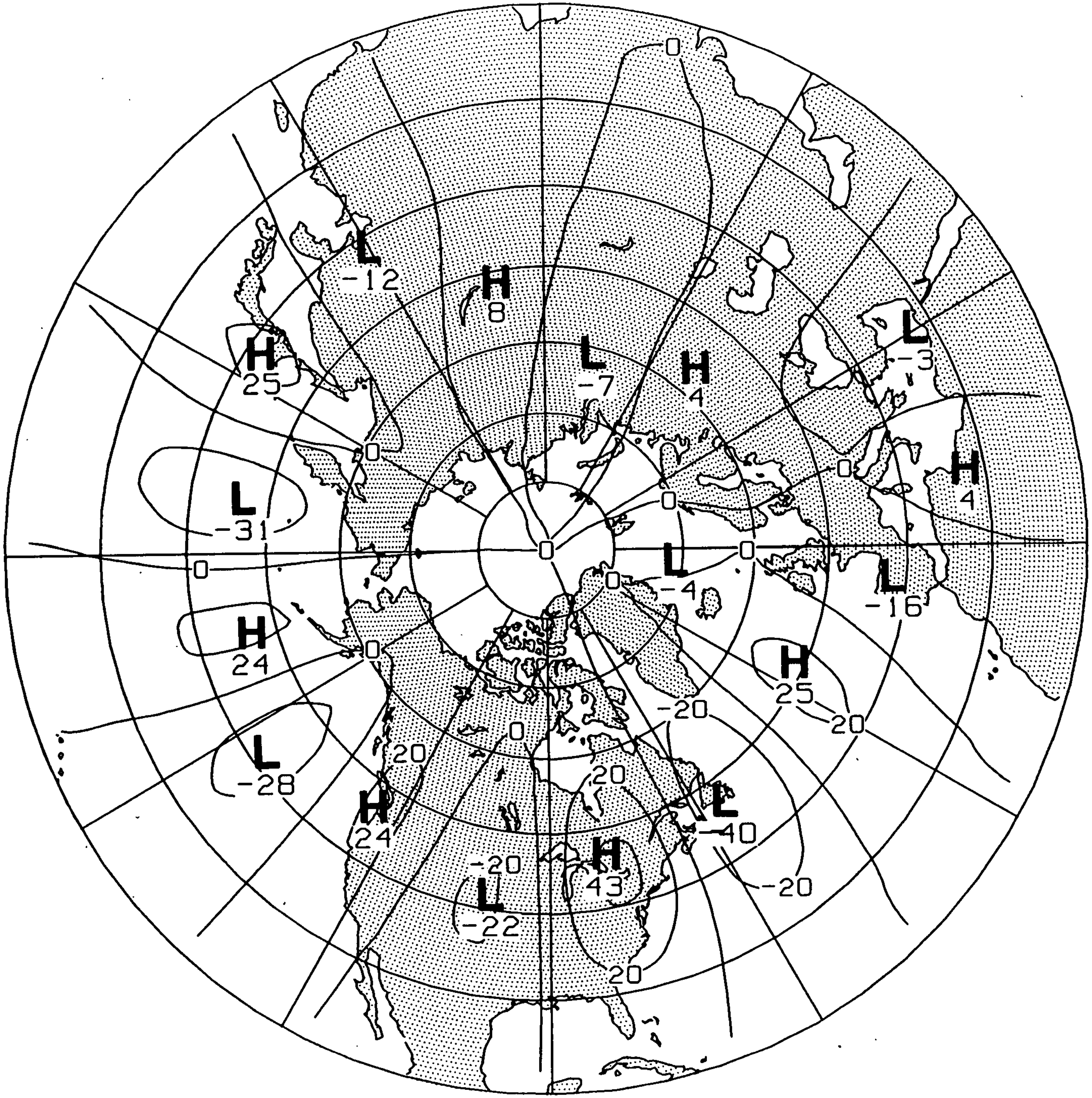


L -4
17
H 3
-20
20
L -26
-20
0
H 51
40
40
20
20
-20
-40
L -66
-40
20
0
H 2
L -2
0
L -1
0
H 31
20
20
L -1
H 4
0
0
0
L -30
-20
L -10
H 26
L -20

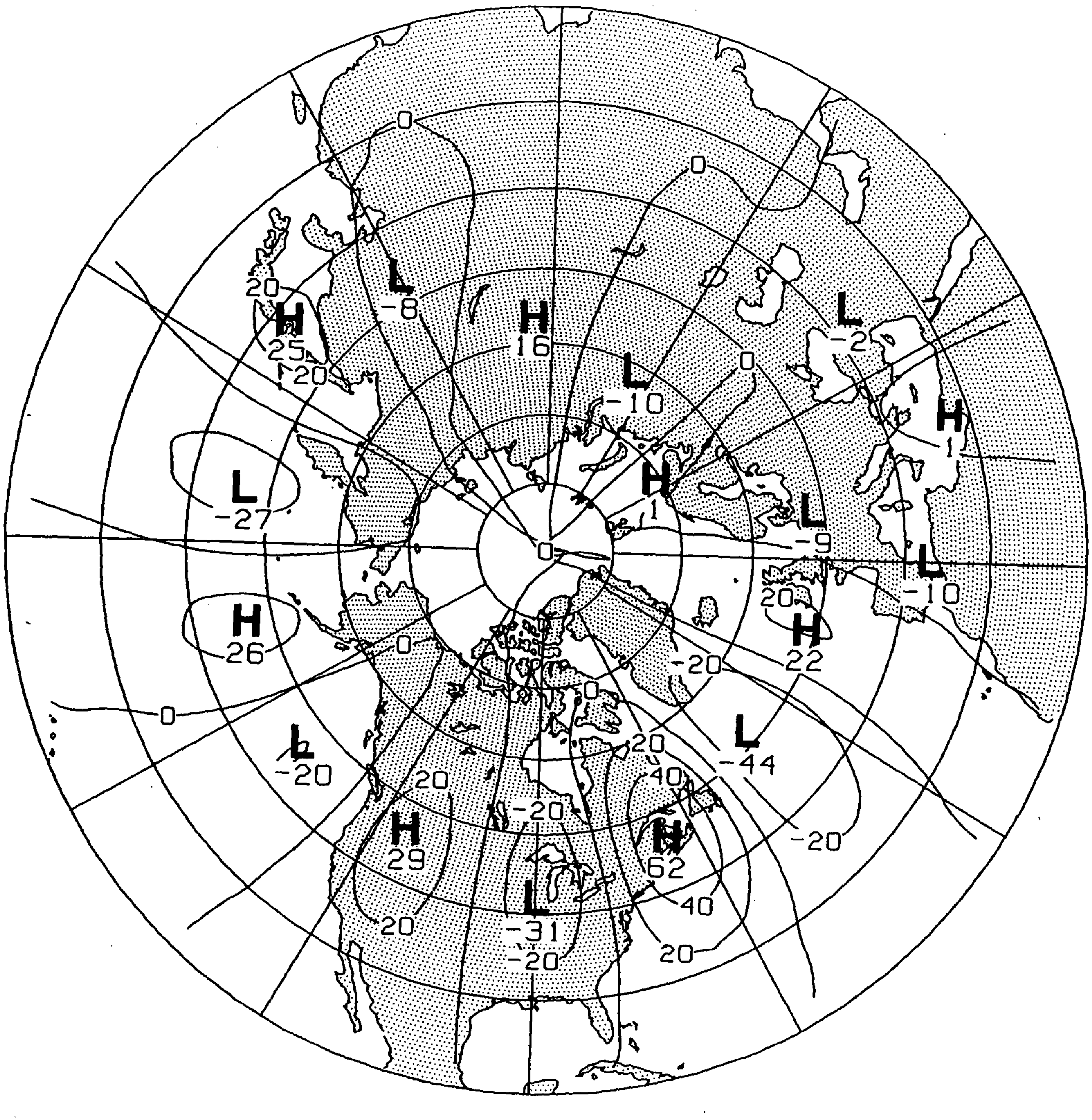

L -5
H 6
H 13
L -5
L -10
L -10
H 23
0
0
0
0
L -22
0
H 27
20
0
L -41
-20
20
20
0
L -36
-20
H 22
0
20
H 22
-20
L -20



-20
H
17
L
-41
L
-22
H
17
L
-3
-20
20
0
H
43
0
H
7
20
L
-6
0
0
0
0
8
L
-12
-20
0
0
H
18
0
0
H
20
L
-25
-20
L
-10
H
14
0
L
-4



L
-12
H
8
L
-7
H
4
L
-3
H
25
H
4
L
-31
0
0
0
0
0
L
-4
0
L
-16
H
24
0
H
25
-20
20
L
-28
20
0
20
H
24
L
-40
-20
H
43
-20
L
-22
-20
20


L
-8
H
16
L
-10
L
-2
H
1
H
25
20
20
0
0
0
H
1
L
-27
0
L
-9
L
-10
H
26
0
20
H
22
-20
0
L
-20
0
L
-44
20
40
20
-20
H
29
H
62
-20
L
-31
-20
20
40
20