Climatology of the Stratospheric Polar Vortex and Planetary Wave Breaking

MARK P. BALDWIN AND JAMES R. HOLTON

Department of Atmospheric Sciences, University of Washington

(Manuscript received 1 May 1987, in final form 22 October 1987)

ABSTRACT

We use the distribution of Ertel's potential vorticity (PV) on the 850 K isentropic surface to establish a climatology for the transient evolution of the planetary scale circulation in the Northern Hemisphere winter midstratosphere. We compute PV distributions from gridded NMC daily temperature and height maps for the 10 and 30 mb levels, and show that a very good approximation for 850 K PV can be derived from 10 mb heights and temperatures alone.

We assume that reversals of the latitudinal gradient of PV, localized in longitude and latitude may be regarded as signatures of planetary wave breaking. Wave breaking identified by such signatures tends to occur mainly in the vicinity of the Aleutian anticyclone, with a secondary maximum over Europe. The area of the polar vortex, defined as the area enclosed by PV contours greater than a certain critical value, is strongly influenced by wave breaking. Erosion of the polar vortex due to transport and mixing of PV leads to a preconditioned state, when defined in terms of vortex area, that always occurs prior to major stratospheric warmings.

During winters with little PV transport or mixing, the vortex area evolves rather uniformly in response to radiative forcing. During winters with major sudden warmings, the wave breaking signature as defined here first appears at low values of PV, then rapidly moves toward higher values as the vortex area is reduced and the "surf-zone" structure becomes well defined.

1. Introduction

Matsuno (1971) showed by numerical experiment that sudden stratospheric warmings, which occur in the Northern Hemisphere winter stratosphere, can be largely explained on the basis of mean zonal flow decelerations driven by planetary waves propagating from the troposphere into the polar stratosphere. Although Matsuno's explanation quickly became the accepted theoretical paradigm for the sudden warming phenomenon, it was recognized that not every case of enhanced wave propagation into the stratosphere leads to a sudden warming, so that relating the occurrence of warmings to particular tropospheric precursors has remained a rather elusive goal.

An appealing explanation for the observed weak relationship between enhanced wave propagation into the stratosphere and the development of sudden warmings was offered by McIntyre (1982) in a provocative essay on the nature of sudden warmings, bringing together suggestions by Quiroz et al. (1975), Dunkerton et al., 1981, Kanzawa (1982) and Palmer (1981). McIntyre suggested that for a reversal of the mean zonal wind (and related warming) to occur in the polar region it was necessary not only to have enhanced planetary wave propagation into the stratosphere, but to have the stratospheric vortex in an anomalous "preconditioned" state that would cause the wave EP flux to be focused into high latitudes. Such a preconditioned state is characterized by a tight polar vortex (not necessarily centered on the pole) marked by strong potential vorticity (PV) gradients, surrounded by a broad latitudinal belt of weak PV gradients. McIntyre also pointed out that the latter could act as a reflector, as suggested by nonlinear critical layer theory, despite the breakdown of linear Rossby wave theory (and associated concepts like the refractive index). Such reflection gives rise to the possibility not only of focusing but also of self-tuning resonant wave amplification in some ways similar to the mechanism suggested by Plumb (1981). This suggests that analysis of the distribution of PV and its temporal evolution is a logical approach to the diagnosis of sudden warmings.

The use of PV as a stratospheric diagnostic is based mainly on two principles. The first is that PV is a quasi-conservative tracer of air motion along isentropic surfaces and the second is that the global PV distribution, together with the stipulation of an appropriate basic static stability, balance conditions and boundary conditions, completely defines the flow. (See Hoskins et al., 1985, for details.)

Potential vorticity is defined as

$$P = \frac{1}{\rho}(\nabla \times \mathbf{u} + 2\Omega)\cdot\nabla\theta \qquad (1)$$

where ∇ is three-dimensional, ρ is air density, Ω is the earth's rotation, and θ is any function of specific en-

Corresponding author address: Dr. Mark P. Baldwin, Northwest Research Associates, Inc., P.O. Box 3027, Bellevue, WA 98009.

tropy (taken here as potential temperature). The approximations and methods used in calculating PV from observational data are discussed in section 2.

McIntyre and Palmer (1983, 1984) were the first to document the temporal changes in Ertel's PV during a sudden warming by displaying daily maps of PV for the 850 K isentropic surface. Using Stratospheric Sounding Unit (SSU) data they created coarse-grain (up to zonal wave number 12) maps of PV on the 850 K surface for the late winter of 1978/79, which included the spectacular wave-2 major warming. By calculating PV on the 850 K surface, McIntyre and Palmer were able to show a coarse-grain view of what appeared to be "breaking" planetary waves, which apparently eroded the main vortex during the course of the winter and eventually caused it to break up.

In their conceptual model, "the extratropical middle stratosphere is divided into two sharply defined, zonally asymmetric regions, a polar main vortex, characterized by steep gradients of PV at its edge, surrounded by a broad "surf zone" within which systematic, large-scale gradients of PV are comparatively weak." Wave breaking is a vivid term to describe the process by which potential vorticity undergoes rapid irreversible transport and mixing when planetary waves encounter the surf zone. Although the appropriateness of the term "wave breaking" has been questioned by Rood (1985) we shall use the wave breaking terminology as a matter of convenience. Its justification and theoretical significance in terms of wave–mean flow interaction theory has been elaborated on by McIntyre and Palmer (1984, 1985).

The study of McIntyre and Palmer (1983, 1984) indicated that as the waves break, high potential vorticity air is stripped off the edge of the main vortex, and is mixed with the surrounding low PV air. Unfortunately, the observations allow only a low-resolution, blurred view of process, so it is unclear, for instance, how perfect or imperfect the mixing might be.

Clough et al. (1985) used SSU data truncated at zonal wave 12 to compute the PV distribution for December 1981 with great care over data processing validation. Their results essentially confirmed the wave breaking hypothesis of McIntyre and Palmer using five years of SSU data (most of which was not shown). They concluded that wave breaking appears to be happening nearly all the time in the winter stratosphere of both hemispheres. In the Northern Hemisphere the most pronounced breaking occurs in a preferred geographic location, and a monthly mean isentropic map of PV in midwinter shows distinct tongues of high and low PV in the vicinity of the Aleutian anticyclone.

Butchart and Remsberg (1986) calculated diagnostics based on the vortex area for the year 1978/79. Using Limb Infrared Monitor of the Stratosphere (LIMS) data and retaining only six zonal harmonics, they defined the vortex areas as the fractional area of the Northern Hemisphere for which PV is greater than or equal to a given critical value. That critical value, which varied throughout the winter, essentially marked the boundary between the regions of strong PV gradients and weak PV gradients. Butchart and Remsberg's results showed that in 1978/79 the surf zone/main vortex structure became evident after mid-December (the time of a Canadian warming) and became very clear during mid-to-late winter. Moreover, the area of the vortex thus defined generally decreased in time owing to two factors: the area enclosed by a given contour tended to decrease and the area defined as the vortex (i.e., the area with large PV gradients) involved only higher and higher values of P as the winter progressed. The breaking planetary waves caused higher and higher PV contours to be drawn out into the surf zone as the winter progressed. These area diagnostics showed that, at least for the 1978/79 winter, the area of the vortex decreased. It may be hypothesized that this preconditioned the vortex so that upward and poleward propagating planetary waves were later able to disrupt the entire vortex in late February.

Dunkerton and Delisi (1986) used the LIMS data for the 1978/79 winter to calculate daily PV maps at 850 K and 1300 K and to do trajectory calculations. They truncated the data using spherical harmonics keeping zonal wave 6 and total wavenumber up to 24. In addition to presenting daily maps documenting the evolution of PV at 850 K, they concluded that the vortex shape and orientation are both important attributes of the vortex. The presence of traveling waves was shown to be likely to affect the timing of the warming.

Juckes and McIntyre (1987) used an ultrahigh resolution barotropic model to simulate planetary wave breaking in the stratosphere. The model developed breaking planetary waves which "rolled up" into small vortices. The wave breaking process appeared to be one-sided, with high PV air being drawn out into lower latitudes, but low PV air not observed to be drawn into the main vortex. This suggests that the main vortex may act as a material entity, the interior of which is isolated from the surrounding surf zone.

A number of questions still remain about the use of PV diagnostics and the climatology of PV. Clough et al. found, by inspection of a five-year dataset, that the observably resolved wave breaking tended to occur near the Aleutian high. Can their findings be verified objectively in a long-term dataset? How does the vortex/surf zone structure develop, if it develops at all, in a typical year? How do warming years differ from less disturbed years? Is the vortex always eroded or otherwise reduced in size prior to a major warming, or can a warming occur without any such preconditioning? In order to answer these questions, we have developed a climatology of PV at the 850 K level.

2. Data analysis

All results were calculated from NMC data for the years 1964–82, consisting of geopotential height and

temperature at 10 and 30 mb. The data were transformed onto a 4° latitude by 5° longitude grid from 18°N to the pole.

All grids were checked to make sure that the average, high and low values were within reasonable limits. Grids with suspect data were plotted and checked by eye. This procedure resulted in the deletion of about ten days of data. The dataset contains several gaps ranging from one day to eight months. Linear interpolation in time was used to fill gaps of up to one week. Two longer gaps affected the calculations. These periods, from May 1972 through 7 January 1973 and March 1981 through December 1981, were simply omitted.

One problem with the use of satellite data is that the data have less resolution zonally than meridionally. This is particularly true of the LIMS data which contain only six zonal harmonics. This truncation emphasizes tongues of PV elongated in the zonal direction. In this respect it is interesting to compare maps produced for the same days by McIntyre and Palmer (SSU, wave 12) and Dunkerton and Delisi (LIMS, wave 6). The latter show much less structure in the zonal direction. The ideal truncation would be isotropic and at a resolution sufficient to capture the blobs of PV debris streaming off the main vortex, while eliminating spurious, small-scale noise. If some or all of the blobs are real, then they may be smaller vortices themselves. All grids were transformed into spherical harmonic coefficients by the method of Blackmon (1976). After trying many different truncation schemes, triangular truncation at wave 15 was chosen. The NMC dataset quality, judged from the apparent level of spurious noise in the height fields, varies somewhat in time, and in particular is worse after 1975 when satellite observations were more heavily relied upon. Truncation at wave 15 seems to provide the necessary smoothing for the data during the late 1970s while not substantially affecting the earlier data. Triangular truncation has the additional advantage that it is an isotropic filter on the sphere (Sardeshmukh and Hoskins, 1984). Thus there is no artificial emphasis by the filter on features that are elongated in the zonal, or any other, direction.

Since observed winds were not available, winds were calculated by the linear balance method (Robinson, 1986; Hitchman et al., 1987). The commonly used geostrophic approximation has been shown to be a mediocre approximation to the actual winds in a general circulation model (Boville, 1987). The linear balance method produces winds which are qualitatively similar to gradient winds but which take into account the deceleration correction (McIntyre and Palmer, 1984) and thus are allowed to have a cross-isobar component. Randel (1987) has demonstrated that the linear balance winds are an improvement over geostrophic winds, which overestimate the speed of the polar night jet.

Following McIntyre and Palmer (1984) the high Richardson number approximation to Ertel's PV (Hartmann, 1977) was used

$$P = -g(\zeta + f)\frac{\partial \theta}{\partial p} \tag{2}$$

where relative vorticities, ζ, are computed on constant pressure surfaces. The pressure of the 850 K surface was computed using the 10 mb temperature and the 10–30 mb temperature difference. The 850 K vorticity was then interpolated (and sometimes extrapolated slightly above 10 mb) from that on the 10 and 30 mb surfaces. In addition, P was scaled by the method Dunkerton and Delisi (1986). Here P was multiplied by

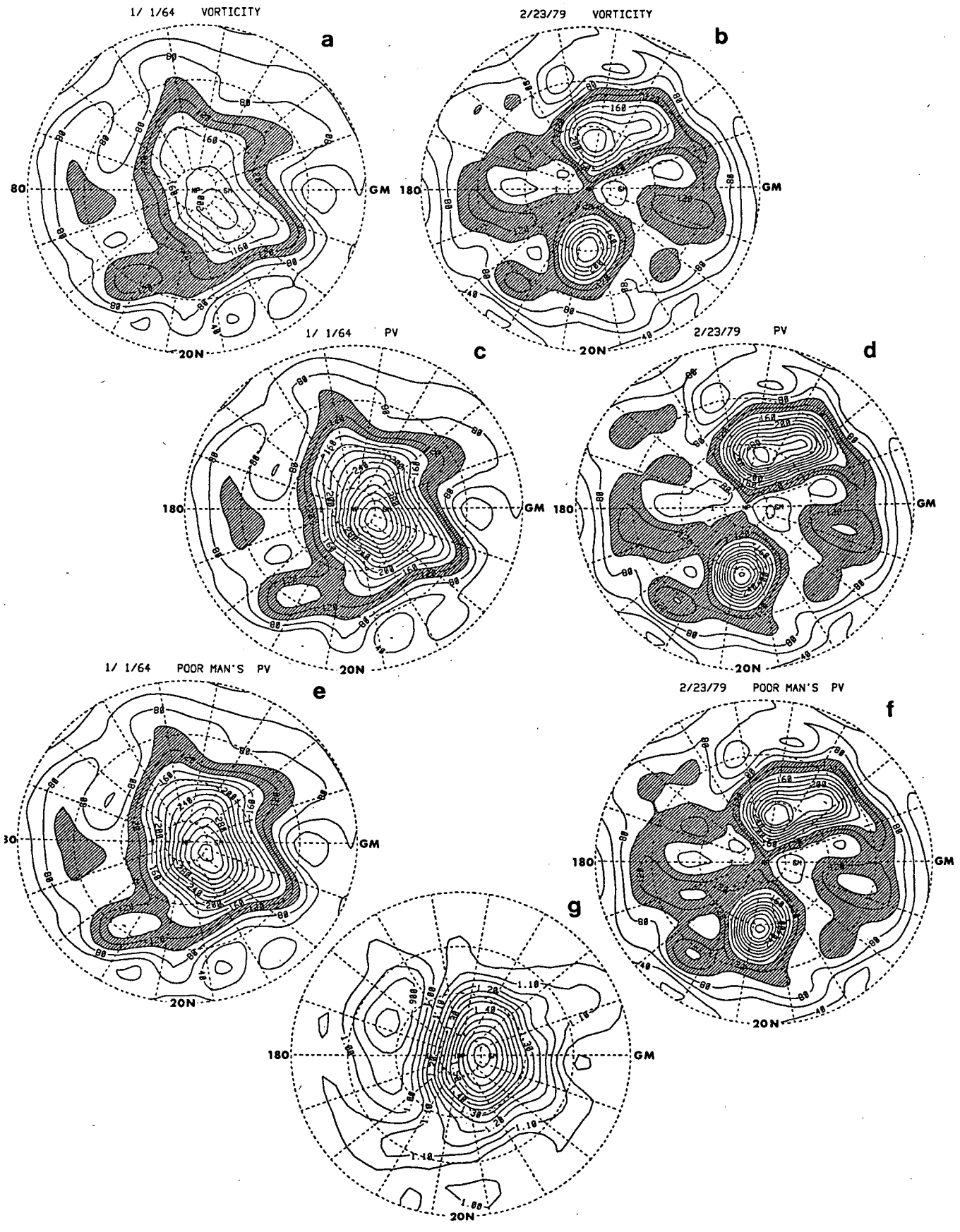
$$-1/[g(\partial\theta_0/\partial p)] \tag{3}$$

where $\partial\theta_0/\partial p$ is the standard atmosphere value of $\partial\theta/\partial p$ at the standard height of the 850 K isentropic surface. The numerical value is -26.7 K mb^{-1} and is constant on the 850 K surface. The P is then expressed in vorticity units (s^{-1}). This scaling simplifies comparisons of P with absolute vorticity.

All height and temperature fields were interpolated onto a 4° latitude by 5° longitude grid for the calculations. Although the height and temperature fields did not show any spurious boundary features, the differentiations required to calculate vorticity revealed some obviously spurious features at the southern boundary of the plotted fields (20°N). The actual grid ends at 18°N and one-sided differences may be the cause of some of these features. Such spurious features are most prominent during the late 1970s (in particular, after 16 December 1975) when the data are noisiest. To reduce the influence of this boundary noise, the wind and vorticity fields were truncated at zonal wavenumbers 6, 8 and 10 at 18°, 22° and 26°, respectively. This procedure largely eliminated the smaller-scale spurious features on the worst grids while not significantly affecting the grids that were already relatively smooth. Some of the increased "noise" may actually represent real features made visible by the increased data coverage afforded by the satellite coverage. However, judging from its temporal inconsistency and comparison with "pure" satellite data (e.g., Dunkerton and Delisi, 1986), most of these features appear to be spurious.

3. Approximations to potential vorticity

In the middle stratosphere, where the isentropic surfaces coincide fairly well with pressure surfaces, how well does absolute vorticity alone represent the features seen on PV maps? To answer this question 10 mb vorticity and 850 K PV maps were compared for a large number of days. The 10 mb level was chosen because it is close to the 850 K surface. Two representative days (shown as Fig. 1) were selected: 1 January 1964 because a wave breaking event is visible early in the


1/ 1/64 VORTICITY
a
2/23/79 VORTICITY
b
1/ 1/64 PV
c
2/23/79 PV
d
1/ 1/64 POOR MAN'S PV
e
2/23/79 POOR MAN'S PV
f
g
GM
20N

data record and 23 February 1979 because it represents an extremely disturbed state late in the dataset and during the often-cited 1979 warming. A comparison of (a) vs (c) and (b) vs (d) in these figures indicates that the surf zone is well represented by the vorticity field but that the relative intensity of the polar vortex is greatly diminished. If one were interested only in the outer contours, which define tongues of high PV air and PV debris (with enough vertical coherence to be seen), then vorticity would be an adequate diagnostic. With coarse vertical resolution, shallow anomalies in the stability field are not resolved. To capture the intensity of the polar vortex, the static stability must be taken into account. Figure 1g illustrates the static stability for 1 January 1964 divided by its standard value. The contours of static stability coincide fairly well with the vorticity contours, and by far the largest effect from including the static stability is an amplification of the vorticity field in the polar vortex.

The vorticity field is calculated only from the height field without reference to the temperature field. In order to see how the static stability affects PV, static stability may be rewritten as

$$\frac{\partial\theta}{\partial p} = \frac{-\theta R}{p_\theta g}\cdot\left(\frac{dT}{dZ} + \frac{g}{C_p}\right) \tag{4}$$

where p_θ is the (spatially variable) pressure of the isentropic surface. Note that the only spatially variable quantities are pressure and dT/dZ on a fixed isentropic surface. Typically dT/dZ is an order of magnitude less than g/C_p and thus contributes very little to $\partial\theta/\partial p$. It is the pressure of the isentropic surface that provides most of the variation in the stability. The isentropic surfaces deviate most from constant pressure surfaces in the polar vortex.

In the present case, the 850 K surface lies near the 10 mb level, so that the vorticity at 850 K is nearly the same as that at 10 mb. It is possible to obtain a much closer approximation to 850 K PV than 10 mb vorticity by using the 10 mb temperature field to estimate the pressure of the 850 K surface. If dT/dZ is assumed to have the standard atmosphere value of $1\,^\circ\text{C km}^{-1}$, then it is possible to integrate hydrostatically to compute the approximate pressure of the 850 K surface. If 10 mb vorticity is substituted for 850 K vorticity, and dT/dZ is assumed to be $1\,^\circ\text{C km}^{-1}$, then

$$P^* = -g(\zeta_{10\,\text{mb}} + f)\left(\frac{-\theta R}{p_{850} g}\right)\left(\frac{dT}{dZ_{\text{std}}} + \frac{g}{C_p}\right) \tag{5}$$

with

$$p_{850} = 10\ \text{mb}\left(\frac{850\ \text{K}}{\theta_{10\,\text{mb}}}\right)^{\alpha} \tag{6}$$

where

$$\alpha = \frac{-R}{g}\left(\frac{dT}{dZ_{\text{std}}} + \frac{g}{C_p}\right)$$

defines a "poor man's" PV. The 10 mb vorticity replaces 850 K vorticity. The dT/dZ is that of the U.S. Standard Atmosphere and only 10 mb data are used. If dT/dZ were everywhere equal to the U.S. Standard Atmosphere value ($1\,^\circ\text{C km}^{-1}$) then p_{850} would be exact. Figures 1e and 1f show this poor man's PV for the previously discussed two days. The match with PV, in both cases, is remarkable. Since the match is so close, this approximation to PV may be useful in those cases when data from only one level are available or when computational costs might be prohibitive.

The above examples, and many not shown, indicate that this approximation to PV is very good in the middle stratosphere, given the coarse vertical resolution of the data. We would expect that poor man's PV would work at least as well in the Southern Hemisphere as in the two examples in Fig. 1. In particular, 23 February 1979 (Fig. 1f) was chosen because the stratosphere was extremely disturbed—far more than has been observed in the Southern Hemisphere winter. It would not be as useful in the troposphere, near the tropopause or near the stratopause, where dT/dZ varies greatly on an isentropic surface or over the vertical interpolation interval.

4. Climatology of Rossby wave breaking

Most of the work done on Rossby wave breaking in the middle stratosphere has concentrated on the winter of 1978/79 and its major stratospheric warming. Clough et al. used five years of SSU data but showed results from the 1981/82 winter only. The present NMC dataset is used here to establish a climatology based on 19 years of data.

In order to establish a climatology, using such a large dataset, some kind of objective definition of the wave breaking signature is necessary. Wave breaking, as during the 1978/79 winter or as simulated numerically by Juckes and McIntyre (1987), is typically characterized by tongues and blobs of high PV air streaming off the main vortex. It will be assumed that the presence of these PV features represents the results of breaking planetary-scale Rossby waves. Since this phenomenon

FIG. 1. Comparison of 10 mb vorticity, the distributions of 850 K potential vorticity and "poor man's" isentropic potential vorticity for 1 January 1964 and 23 February 1979. Contour interval is $2 \times 10^{-5}\ \text{s}^{-1}$. The shaded region lies between 1.0×10^{-4} and $1.4 \times 10^{-4}\ \text{s}^{-1}$. The outermost contour is at 20°N. The shaded region lies between 1.0×10^{-4} and $1.4 \times 10^{-4}\ \text{s}^{-1}$. (a) Vorticity for 1 January 1964; (b) vorticity for 23 February 1979; (c) potential vorticity for 1 January 1964; (d) potential vorticity for 23 February 1979; (e) "poor man's" potential vorticity for 1 January 1964; (f) "poor man's" potential vorticity for 23 February 1979; (g) static stability divided by its U.S. Standard Atmosphere value at 850 K for 1 January 1964.

tends to occur in the midlatitudes, it is not necessary to confront the geometric problems of searching for such events over the polar cap. If one were to plot a meridional profile of PV, at constant longitude, through a tongue of high PV air, it might look like Fig. 2. The key feature is the reversal of the PV gradient on the poleward side of the tongue or blob.

In order to compute the diagnostic shown in Fig. 3a a search was made, south of 60°N, for the type of latitudinal profile shown in Fig. 2—involving a negative PV gradient. In order to isolate tongues and blobs of high PV air, the maximum value equatorward of the region of reversed gradient was required to be 10^{-4} s^{-1}, as shown in Fig. 2 (which is also the lowest value in the shaded regions in Fig. 1). The use of this value, as well as using any cutoff at all, is arbitrary and was made after comparing the results from using several different limits. For each day in December, January and February this search was made every 5° in longitude, and a (percent) frequency of occurrence of the reversed gradient pattern was assigned to each longitude. Figure 3a clearly shows that wave breaking does occur most frequently in the vicinity of the Aleutian high. In addition, there is a secondary maximum over Europe. Because this diagnostic counts blobs and tongues of PV debris, there is a fairly high signal (at least 25%) everywhere. This diagram confirms the tentative results of Clough et al. (1985) based on five winters of SSU data.

In order to extend this type of diagnostic to the latitude–longitude plane a similar search of winter data was made for local maxima in the meridional direction. Figure 3b shows the frequency of occurrence (arbitrary units) of a local maximum in the meridional direction at each grid point. The contour interval is entirely dependent on the grid spacing (4° latitude), but the shapes of the contours would be nearly identical if a finer grid spacing had been used. The contours near the pole represent the migration of the polar vortex, which appears as a meridional maximum when displaced from the pole. South of 50°N the spiraling pattern represents high PV air as it separates from the main vortex and streams westward and slightly southward.

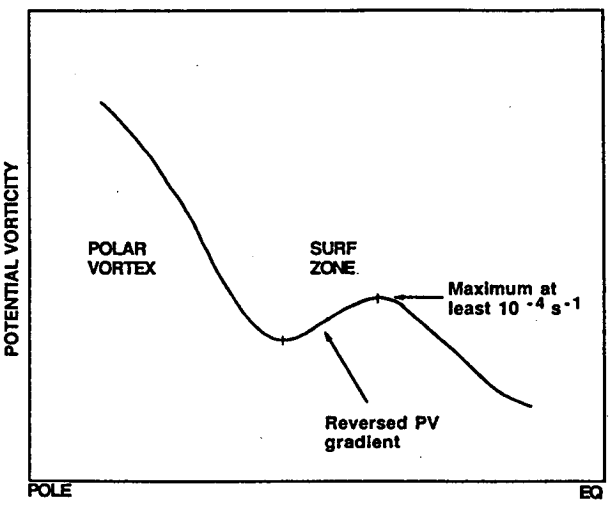


FIG. 2. Schematic illustration of cross section of potential vorticity during a wave breaking event. Maximum value in midlatitude maximum must be at least 10^{-4} s^{-1} to be counted as wave breaking by computational procedure (see text).

Daily maps showing many wave breaking events (not shown) are consistent with the view that the local maxima do represent high PV air streaming off the vortex, typically over North America and to a lesser extent, over Europe. The southward component to the spiral is not so much due to a tendency for the debris to migrate equatorward, but to the displacement of the vortex, which is typically shifted toward Eurasia by the Aleutian high. It is interesting that once the debris reaches the Atlantic region, it does not tend to merge with the main vortex, but tends to remain near 20°N.

The two diagrams in Fig. 3 show clearly that the wave breaking signatures occur frequently during the winter months and that they occur in geographically preferred regions. To the extent that these features are a signature of planetary wave breaking, the ideas of McIntyre and Palmer are confirmed in a statistical sense.

It should be emphasized that wave breaking is not *defined* by any signature. Wave breaking has no unique signature or shape by which one can *always* be sure of recognizing it (McIntyre and Palmer, 1985). We have assumed that reversals of the latitudinal gradient of PV, localized in longitude and latitude, may be regarded as signatures of wave breaking.

Results using PV maps are largely qualitative and few methods have been developed for the quantitative dynamical analysis of PV data. In Fig. 4 two such quantitative PV diagnostics are shown for each October–April winter period. One of these is the area diagnostic of Butchart and Remsberg (1986), developed from a suggestion by McIntyre and Palmer (1983, 1984), which shows the horizontal projection of the area enclosed by isopleths of PV on the 850 K surface. In Fig. 4 the lowest isopleth represents 0.4×10^{-4} s^{-1}. These plots show the development of the vortex/surf zone structure during each winter.

The other diagnostic is designed to show the degree to which each isopleth of PV is involved in the wave breaking process. The diagnostic measures the relative frequency with which a given PV contour is encountered in a negative PV gradient. For each day a sum is made of the number of grid points (weighted by $\cos\phi$) at which a given contour is found in a negative PV gradient. Thus, the diagnostic measures the degree to which each contour is involved in a "wave breaking signature." The diagnostic counts each tongue or blob of high PV air which involves a given isopleth.

The weighting factor of cosine of latitude is used to give equal weighting by area on the sphere. This factor

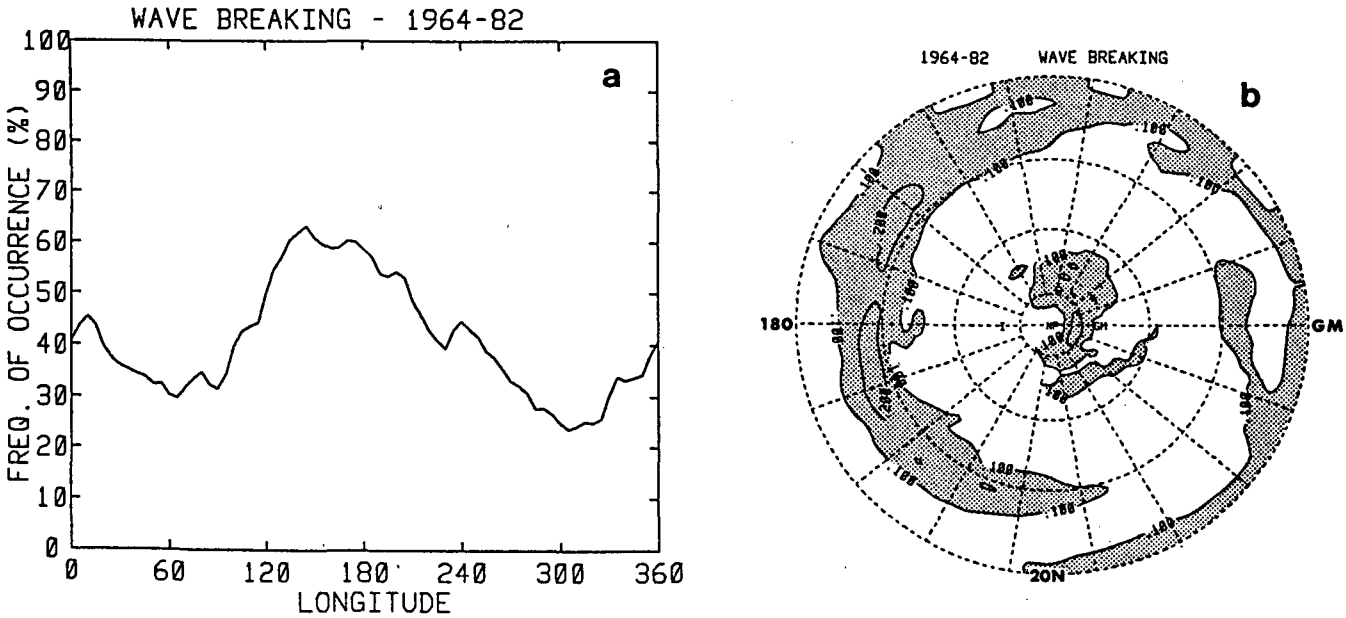


FIG. 3. Climatology of wave breaking at 850 K for 19-year NMC winter (DJF) data. (a) Frequency of occurrence of a wave breaking profile as a function of longitude. See text for a complete explanation. (b) Relative frequency of occurrence (arbitrary units) of a meridional maximum in the potential vorticity. Contour values are dependent on grid spacing. See text.

evenly weights features of the same size, regardless of where they are found on the grid.

If, at a particular longitude, there were two tongues of PV involving the 1.0×10^{-4} s^{-1} contour, then both tongues would be counted, weighted by the cosine of the latitude at which the contour (within a negative PV gradient) is encountered.

If the PV distribution on a given day were entirely less than 1.0×10^{-4} s^{-1} except for a ring of higher PV air between 30° and 45°N, then only the 1.0×10^{-4} s^{-1} contour at 45°N would count since it (and not the contour at 30°N) is involved in the negative PV gradient. To obtain a value to be plotted, the total number of crossings (72) is weighted by the cosine of 45° (0.707) and then divided by the number of longitudes checked (72) for a value of 0.707.

In order to clarify how this diagnostic is calculated using real data, the calculation will be illustrated using the two example days in Fig. 1. As discussed earlier, all values are weighted by the cosine of latitude. For simplicity, that weighting will be neglected in these example calculations. For further simplicity, only the 1.4×10^{-4} s^{-1} contour will be considered.

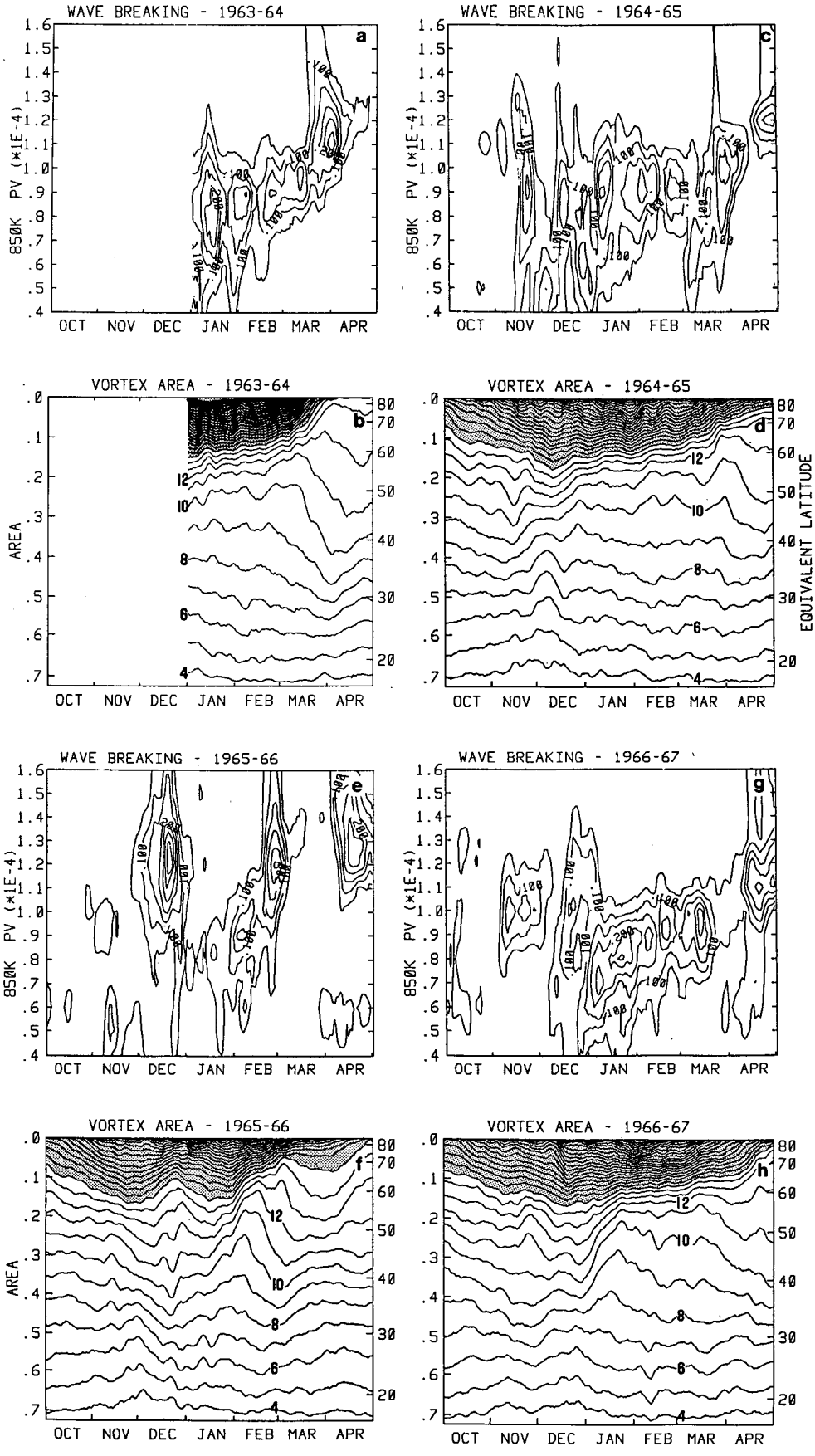
On 1 January 1964, the 1.4×10^{-4} s^{-1} contour was involved in a negative PV gradient in two places, between 110° and 130°W and very slightly between 99° and 100°W. Ignoring the cosine φ weighting, the 10^{-4} s^{-1} contour was involved in a "wave breaking signature" at 21°/360° for a value of 0.058.

On 23 February 1979, the 1.4×10^{-4} s^{-1} contour was involved in a negative PV gradient in four locations: between 11° and 17°W, 63° to 120°W, 124° to 126°W, and 21° to 118°E. This gives a value (again ignoring the cosine 0 weighting) of 162°/360°, or 0.45. If the cosφ weighting were factored in, then these values would be reduced somewhat. It is worth noting that the actual calculation is performed every 5° of longitude, and is therefore less precise than this example. The results shown in Fig. 4 have been smoothed using a five-day running mean.

The wave breaking diagnostic interprets the PV pattern on 23 February 1979 as wave breaking, and could be considered misleading when applied to a single day. However, if one were to view a time history of this diagnostic (Fig. 4ee), supplemented by daily maps of PV such as Fig. 1d, a consistent picture of the wave breaking/vortex erosion process emerges. Often, when the vortex is rapidly displaced from the pole, a "spike" can be seen in the wave breaking diagnostic (e.g., Fig. 4e).

This diagnostic complements the area diagnostic by clearly showing which isopleths of PV are involved in wave breaking. It also shows movement of the vortex away from the pole. If the main vortex is displaced far enough from the pole, as during a Canadian warming, it is counted by the computational procedure. Since this type of episode involves the highest contours and occurs suddenly, (e.g., Dec 1966; Nov–Dec 1979) it can be distinguished from typical wave breaking episodes.

Figure 4 shows the wave breaking diagnostic described above and the area diagnostic, for each winter from 1963/64 to 1981/82. Rather than reviewing the features of each winter, various features will be dis-


WAVE BREAKING - 1963-64
a
WAVE BREAKING - 1964-65
c
850K PV (*1E-4)
VORTEX AREA - 1963-64
b
VORTEX AREA - 1964-65
d
AREA
EQUIVALENT LATITUDE
WAVE BREAKING - 1965-66
e
WAVE BREAKING - 1966-67
g
VORTEX AREA - 1965-66
f
VORTEX AREA - 1966-67
h
OCT NOV DEC JAN FEB MAR APR

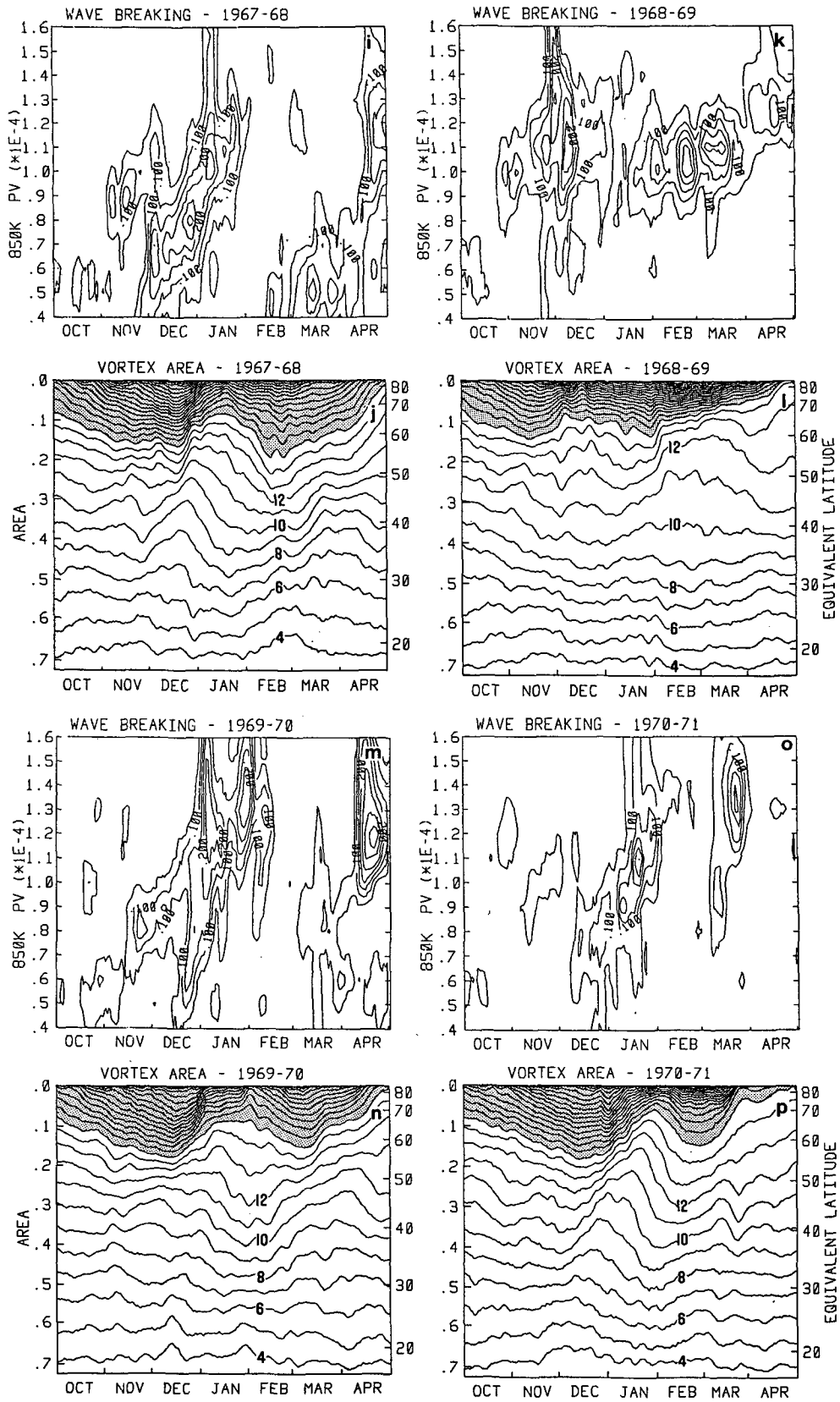


FIG. 4. Wave breaking diagnostic (top) and vortex area (bottom) for nineteen years of NMC data at 850 K. The value of the wave breaking diagnostic is a measure of the degree to which each PV contour (the ordinate) is involved in regions of a negative meridional PV gradient. See text for a complete explanation. The area diagnostic has an ordinate that gives the percent of the hemispheric area enclosed by a given PV contour. Labels are scaled by 10^{-5} s^{-1}. Contour interval is 10^{-5} s^{-1}.

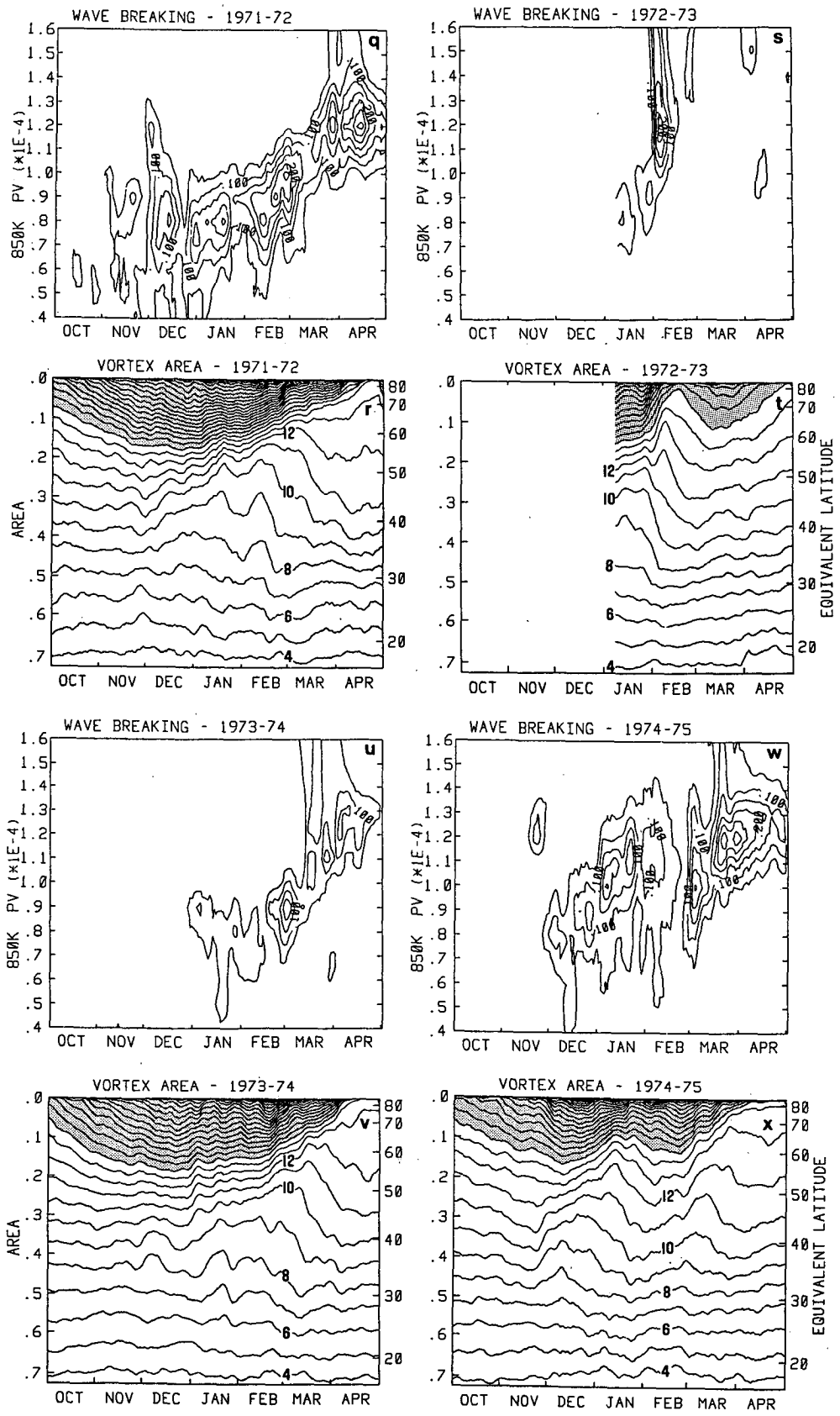


FIG. 4. (*Continued*)

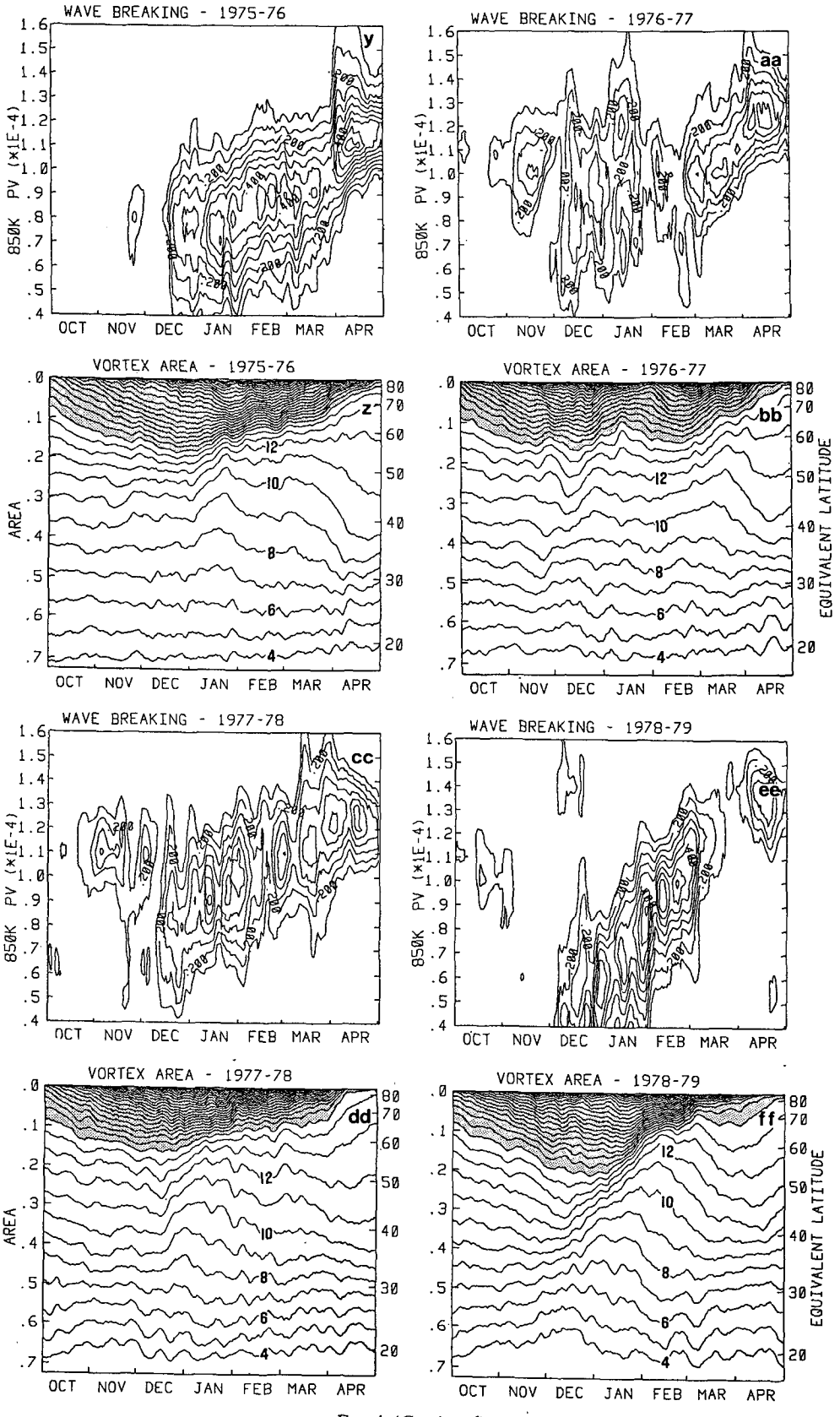


FIG. 4. (*Continued*)

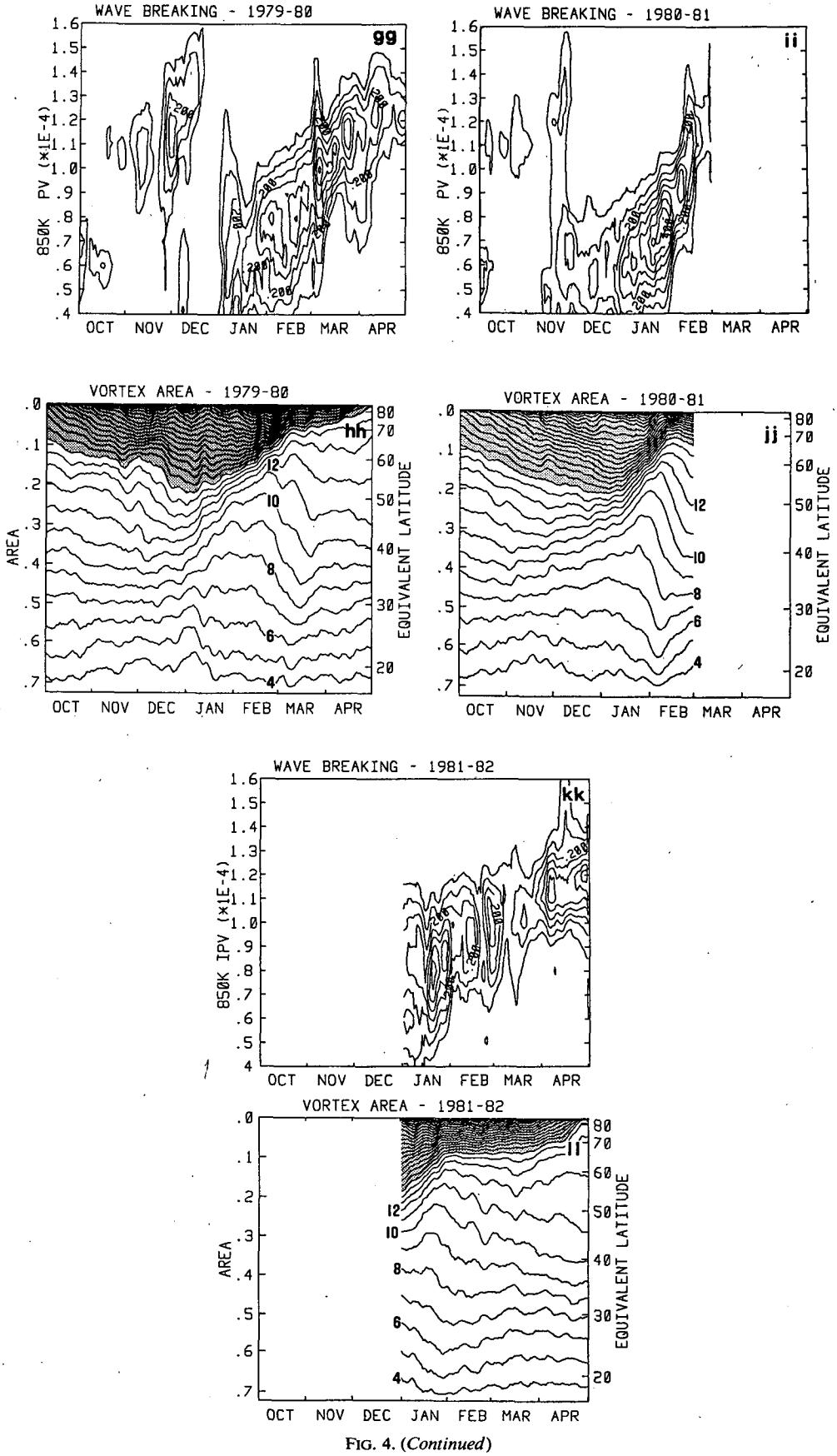


FIG. 4. (*Continued*)

cussed and examples in the data will be offered. Note that the results for 1975/76 to 1981/82 are noisier than the earlier data, because of the increased use of satellite data, which began on 17 December 1975. The minimum contour level for the wave breaking diagnostic is set higher for the years 1975/76 to 1981/82 in order to eliminate spurious noise. (The lowest contour for the earlier years is 0.05, while it is 0.10 for 1975/76, 1979/80, 1980/81 and 1981/82. It is 0.15 for 1976/77, 1977/78 and 1978/79).

The early seasons (October–November) are remarkably similar during each year, as reflected in both diagnostics. There is very little visible wave breaking and the vortex/surf zone structure has not yet developed. Any wave breaking that does manifest itself tends to begin in the latter part of November and is often related to an early Canadian warming (Labitzke, 1981), which originates through the pulsation of the Aleutian anticyclone with a possible reversal of the temperature gradient poleward of 60°N.

The middle and late winters show a high degree of interannual variability, principally due to the presence or absence of sudden warmings and also due to the variability of lesser wave breaking events. Butchart and Remsberg showed that during the 1978/79 winter the vortex/surf zone structure began to be visible at the time of the Canadian warming in December and that the size of the vortex decreased monotonically through the major warming. As the area of the vortex decreased, high PV air was stripped away by the wave breaking process and consequently, the vortex edge was defined by higher and higher values of PV. In the area diagnostic this temporal evolution takes the form of higher and higher values of PV entering the surf zone of low PV gradients.

For comparison with a hypothetical atmosphere with no wave activity, Butchart and Remsberg's Fig. 6 shows the area diagnostic computed for a zonally symmetric model with a radiatively governed annual cycle. In their case the edge of the vortex cannot be defined and late winter looks like a time-reversal of early winter, and quite unlike the disturbed real winters.

Several of the real winters shown in Fig. 4 were remarkably close to this hypothetical case. The winters of 1964/65, 1966/67, 1971/72, 1973/74, 1975/76 and 1977/78 lacked major wave breaking events and tended to exhibit a symmetry in time of the polar vortex area, centered about mid-to-late January, similar to that of the model simulation. For these winters, the wave breaking diagnostic indicates only a moderate level of wave breaking (e.g. 1964/65). In the early winter, the low valued contours tend to be involved in the wave breaking and slowly over the course of the winter, higher value contours enter the process. As can be seen from the wave breaking diagnostic, the time scale is that of the winter season. In these little-disturbed winters usually a final warming takes place during April, after the vortex area has been greatly reduced, largely by radiative effects. The signature of the final warming can be seen as a sudden increase in wave breaking at high PV values, often showing a "spike" as the vortex breaks up or is displaced well away from the pole. This burst tends to be short-lived as the vortex is destroyed completely in a few weeks. The area diagnostic shows a more pronounced vortex/surf zone structure than the model results due to the erosion-sharpening at the edge of the vortex. The final warmings can often be identified by the abrupt end of any vortex/surf zone structure.

The signature of a major warming contrasts sharply with the slow changes associated with the quiescent years. The wave breaking diagnostic for the years 1965/66, 1967/68, 1969/70, 1970/71, 1972/73, 1978/79, 1979/80 and 1980/81 show a very distinctive signature for the sudden warmings. These events are characterized by wave breaking beginning with low PV values and rapidly expanding to include higher values. This process then culminates in a spike as the vortex is displaced or breaks up. The typical time scale for this development is about six weeks, contrasting with a scale of four or five months during quiescent years. The area diagnostic indicates that the 1978/79 winter was not an unusual warming. The process of rapid erosion of the polar vortex accompanied by a sharpening of the vortex edge occurred during each of these warming years.

During each warming, the vortex was first eroded by wave breaking to a large degree. Thus, the observations confirm the hypothesis of McIntyre (1982) that the vortex must be preconditioned by wave breaking (or by radiative effects, in spring) before a major mid-winter warming can occur.

The signature of the Canadian warmings, which involve a modest displacement of the vortex from the pole, can be seen as wave breaking events involving high values of PV and occurring very suddenly during the years 1965/66, 1966/67, 1968/69, 1976/77, 1977/78, 1978/79, 1979/80 and 1980/81. The major difference between the signatures of the Canadian warmings and the major warmings is that the Canadian warmings are not preceded by extensive wave breaking. The vortex is not substantially eroded before the event. Thus, according to the McIntyre preconditioning hypothesis, the upward propagating planetary waves are not sufficiently focused onto the polar cap to produce a major warming.

The winter of 1965/66 provides perhaps the best example of a winter which includes a Canadian, a major and a final warming. October and November showed no significant departures from radiative behavior. During December, visible wave breaking began, centered on approximately the 1.2×10^{-4} s^{-1} contour. Simultaneously, the area enclosed by the higher contours decreased while that enclosed by lower contours increased slightly. The 850 K PV maps for the major warming period (see Fig. 5) confirm that wave breaking

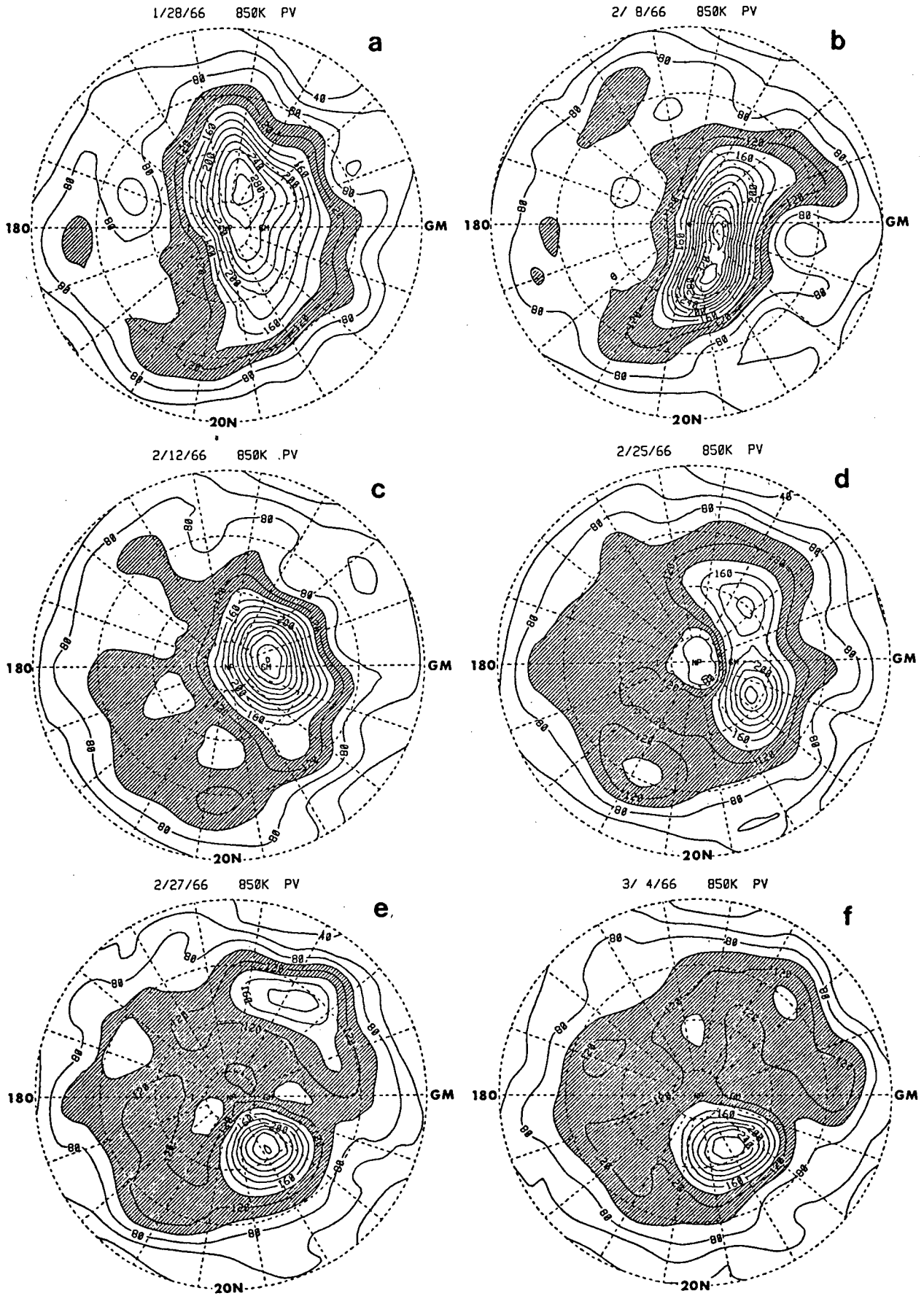


FIG. 5. Representative plots of potential vorticity during the major stratospheric warming of 1966. Diagrams are as in Fig. 1.

occurred associated with an expanding Aleutian high. By late December (not shown) the vortex was displaced over Asia and the wave breaking diagnostic's characteristic spike in Fig. 4e indicates that an area of high PV was encountered away from the pole (or, equivalently, the PV value at the pole was low). By the end of December the vortex had returned to the polar cap and the wave breaking ended rather suddenly. The area of the vortex had been substantially reduced as indicated by the area diagnostic. During January an area of high PV air separated from the main vortex, but the area of the vortex actually increased. At the end of January a series of wave breaking events began which reduced the size of the vortex (Fig. 5). Higher values of PV became involved in the process until the vortex was displaced from the polar cap and split at the end of the month. There was little activity during March and in April the final warming displaced and destroyed the vortex.

Qualitatively, the vortex is always observed to decrease in size prior to every major warming. In order to quantify this relationship, it is necessary to define the edge of the vortex in some objective manner. As suggested by McIntyre and Palmer (1984), the vortex area diagnostic could be used to locate the region of steep PV gradients. This is a logical approach, but as found by Butchart and Remsberg (1986), the vortex is difficult to define at most times. An inspection of the area diagnostic for these nineteen winters indicates that there are frequently periods when the vortex/surf-zone structure is ill-defined, especially in the beginning of the winter. Often the main vortex has sharp PV gradients and the surf zone has weak gradients, but the transition between the two is not precise. Even in McIntyre and Palmer's (1984) conceptual model (their Fig. 12a) there is no precise dividing line between the main vortex and surf zone.

An alternative is to approximate the edge of the vortex by a particular PV contour. Figure 4 indicates that the 1.5×10^{-4} s^{-1} contour always enclosed less than 11 percent of the area of the Northern Hemisphere prior to every major warming. This contour typically encloses more than 11 percent of the area from late October to early March, and only rarely dips below the 11 percent threshold during the December–February winter period.

In addition to every major warming, the 1.5×10^{-4} s^{-1} contour dipped below the 11 percent threshold five other times during these 19 winters. Two of these (December 1968 and December 1975) corresponded to Canadian warmings, while the January 1982 event coincided with a minor warming. The January 1975 and February 1978 events corresponded with substantial distortions of the vortex, which appear to have been close to producing major warmings.

The rate of decrease of the vortex area also appears to be a factor in the timing and strength of the event. Major warmings typically follow sharp decreases in vortex area, while more gradual changes in vortex area (e.g. 1974/75 and 1977/78) tend to precede events which do not constitute major warmings. This is not surprising, since rapid erosion of the vortex is seen to be accompanied by substantial distortions of the vortex, and the persistence of this erosion would lead to a very small vortex area.

5. The 850 K potential vorticity during major warmings

The changes in PV during a major warming have been documented in the literature only for the 1979 warming and preconditioning (McIntyre and Palmer, 1984; Dunkerton and Delisi, 1986). In this section maps of the changes in PV are presented for several representative warmings which occurred during the winters (December–February) of 1963/64 to 1981/82. For each of the warmings six plots are shown to document the changes in the vortex from the initial wave breaking to the post-warming recovery (if any).

The January–February 1966 warming (Fig. 5) involved wave breaking associated with the Aleutian high, which expanded to displace and split the vortex. Between 28 January and 8 February considerable wave breaking sharpened the gradients in the polar vortex and decreased its size. By 12 February the vortex center was displaced somewhat by the Aleutian high and was preconditioned for the warming. The vortex remained in this state until 19 February, when an expanding Aleutian high began to split it. Figure 5d, for 25 February, shows the vortex splitting and the debris from previous wave breaking at approximately 120°W. The vortex proceeded to rapidly split, with the smaller piece drifting toward 60°E and dissipating by 4 March.

The 1966 warming could be considered a typical and a simple warming. The expanding Aleutian high was responsible for the preconditioning and for the splitting of the vortex. The two pieces of the vortex never reattached. There appeared to be about a week during which the vortex was preconditioned before planetary wave amplitudes grew to disrupt the vortex. Labitzke's (1982) Fig. 1 shows that (at 30 mb, 60°N) wave 1 was dominant until about 25 February, when wave 2 grew as the vortex split. This is entirely consistent with observed displacement of the vortex and its subsequent splitting.

The 1968 warming was unusual in that a European anticyclone was associated with most of the preconditioning as well as the breakup of the vortex. During late December (Fig. 6a) an anticyclone centered at about 120°E was displacing the vortex and some high PV air was being stripped off the main vortex. By 27 December (Fig. 6b) an anticyclone at 60°W had begun to intrude on the main vortex and to remove high PV air. This anticyclone proceeded to grow and to position itself at 30°E by 30 December. Simultaneously, a weak Aleutian high was acting to pinch the vortex. During early January these two anticyclones continued to

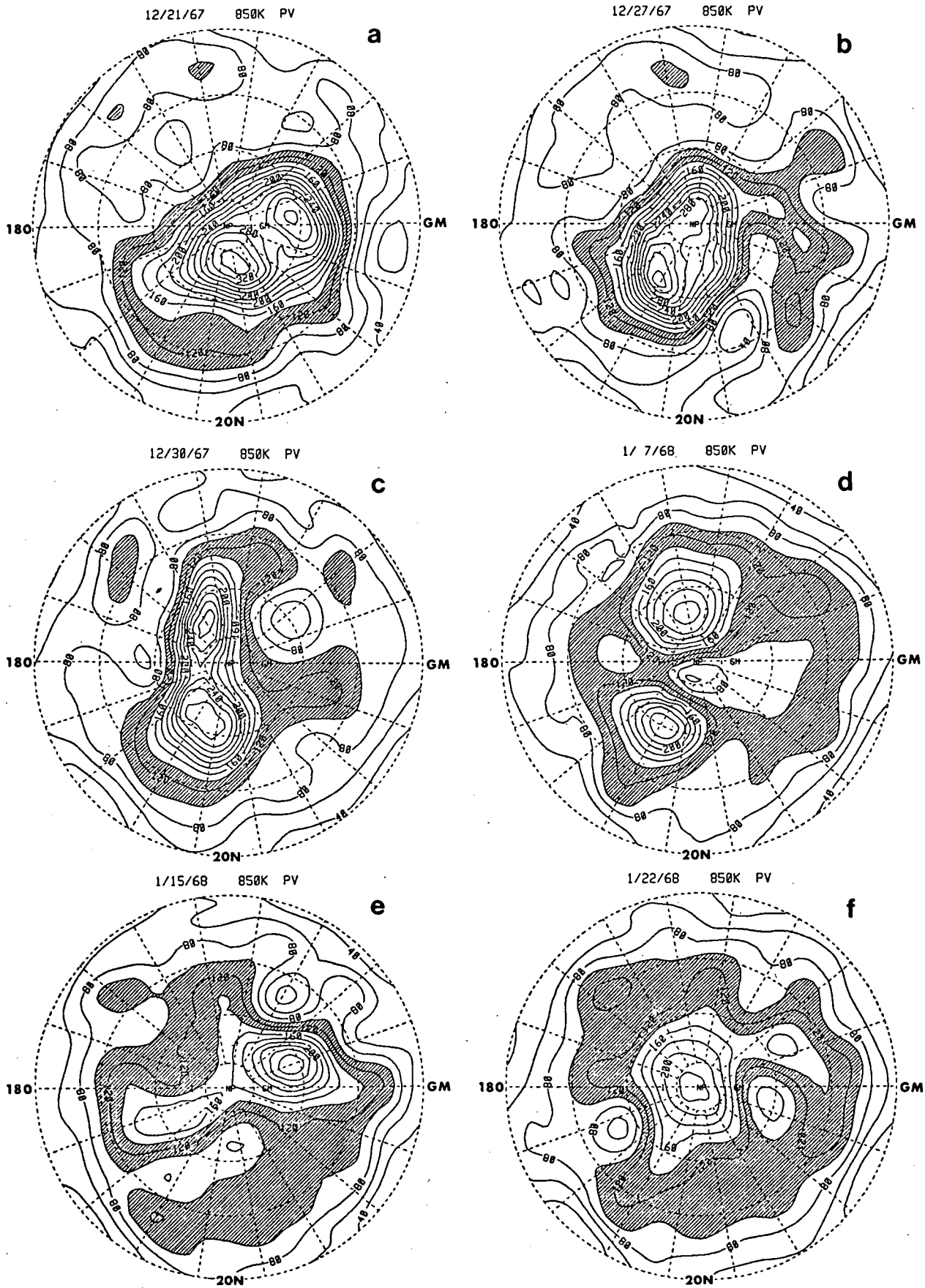


FIG. 6. Representative plots of potential vorticity during the major stratospheric warming of 1967/68.

pinch the main vortex into two centers. This process can be seen in Fig. 1 of Labitzke (1982) as wave 1 is replaced by wave 2 at the end of December. By 7 January the European high was dominant and the warming had peaked. Subsequently, the two parts of the vortex appeared to rejoin on 15 January and the vortex then reformed over the polar cap.

The 1970 warming was interesting in that the vortex never split, and wave 1 was dominant throughout the warming. During late December, the anticyclone at 130°E (Fig. 7a) rapidly grew and intruded on the main vortex to form a very strong Aleutian high associated with which was a very large-scale wave breaking event. The PV debris at 140°E in Fig. 7c appeared to merge with the main vortex during the next two days. This apparent reattachment, together with the continued removal of high PV air from the main vortex resulted in an anticyclone nearly centered at the pole and surrounded by high PV air (Fig. 7d) on 3 January. The long, drawn out vortex broke up into smaller bits and by 10 January (not shown) left just a single slightly elongated main vortex. By 21 January the main vortex had reconsolidated.

A comparison of the area diagnostic (Fig. 4) and the wave breaking diagnostic shows that the winters of 1978/79, 1979/80 and 1980/81 were remarkably similar although the planetary waves involved (Labitzke, 1982, Fig. 1) differed somewhat (on the basis of the 30 mb height field), with wave 1 being dominant for the latter two warmings.

During the 1980 warming (Fig. 8), the vortex was elongated principally in association with the Aleutian high while wave breaking stripped off a considerable fraction of the main vortex. By the end of February, the vortex had partially recovered but was still displaced due to the Aleutian high. The warming consisted simply of the expanding Aleutian high displacing the main vortex as the high moved eastward and poleward. Wave breaking acted to diminish the vortex during the entire period.

6. Conclusions

Although the midstratospheric height field is characterized mainly by zonal wavenumbers 1 and 2, the flow at this level is rich in smaller-scale structure, consideration of which is essential for an understanding of the climatology of the stratosphere. Because the calculation of vorticity involves two spatial differentiations of the height field, it is sensitive to errors in the height field. The main source of data for previous studies has been satellite-derived temperature measurements, which are of sufficient quality for low-resolution calculations of potential vorticity. It is possible, however, to calculate good quality potential vorticity maps at the 850 K level from NMC data as early as 1964. In fact, the early years appear to be of higher quality than the late 1970s, when extensive assimilation of satellite data began. The NMC-based potential vorticity plots show good day-to-day continuity, even for small scale features.

By examining midstratospheric vorticity and potential vorticity maps for the same days, it is clear that outside the polar vortex, vorticity contains almost all the information in potential vorticity, for the features resolved in the coarse vertical resolution NMC data. The maximum in the static stability corresponds fairly well with the vorticity maximum of the polar vortex. It is only in the interior of the polar vortex, where static stability is increased, that potential vorticity differs much from vorticity. Comparisons of vorticity and potential vorticity maps show that the potential vorticity maps have a much more intense looking vortex, but that outside that region the features are essentially the same.

The examination of 19 years of NMC data at the 850 K isentropic surface provides strong qualitative support for the wave breaking/vortex erosion hypothesis of McIntyre and Palmer (1983, 1984). The polar vortex appears to be eroded by breaking planetary-scale waves, which strip off high potential vorticity air and circulate it equatorward, where it slowly dissipates presumably due to diabatic effects. Diabatic effects in the polar region tend to increase the concentration of potential vorticity over the polar cap. The signatures of breaking planetary-scale waves are seen during every winter, and their effects are documented by changes in the area of the polar vortex. There is, however, strong interannual variability in the intensity of wave breaking. In the quiescent years (e.g., Fig. 4h), the area of the polar vortex is seen to follow a pattern more similar to the simulated wave-free winter stratosphere case reported by Butchart and Remsberg (1986). The degree of observed wave breaking and its effects are minimal during these years.

During most winters bursts of planetary-wave activity cause wave breaking and marked decreases in the vortex area. Typically, such wave breaking episodes initially involve low values of potential vorticity, but involve higher values as the vortex area is decreased. The culmination of many such events is typically a displacement and/or breakup of the polar vortex as low potential vorticity air moves over the pole. These events manifest themselves as stratospheric warmings as seen in zonal mean wind and temperature profiles. Thus, a consistent dynamically-based definition of a major sudden warming is a displacement of the polar vortex, in which low PV air moves over the pole.

The wave breaking signature, as defined by the present objective criterion, occurs in preferred geographic locations: North America has the most episodes and Europe has a secondary maximum. The Aleutian and European anticyclones tend to remove high PV air and advect it equatorward and then westward. This results in a "track" encircling the globe, south of 40°N, where blobs of cyclonic PV debris are most likely to be found. The track is not zonally symmetric, but seems to be centered on a vortex displaced by the Aleutian high.

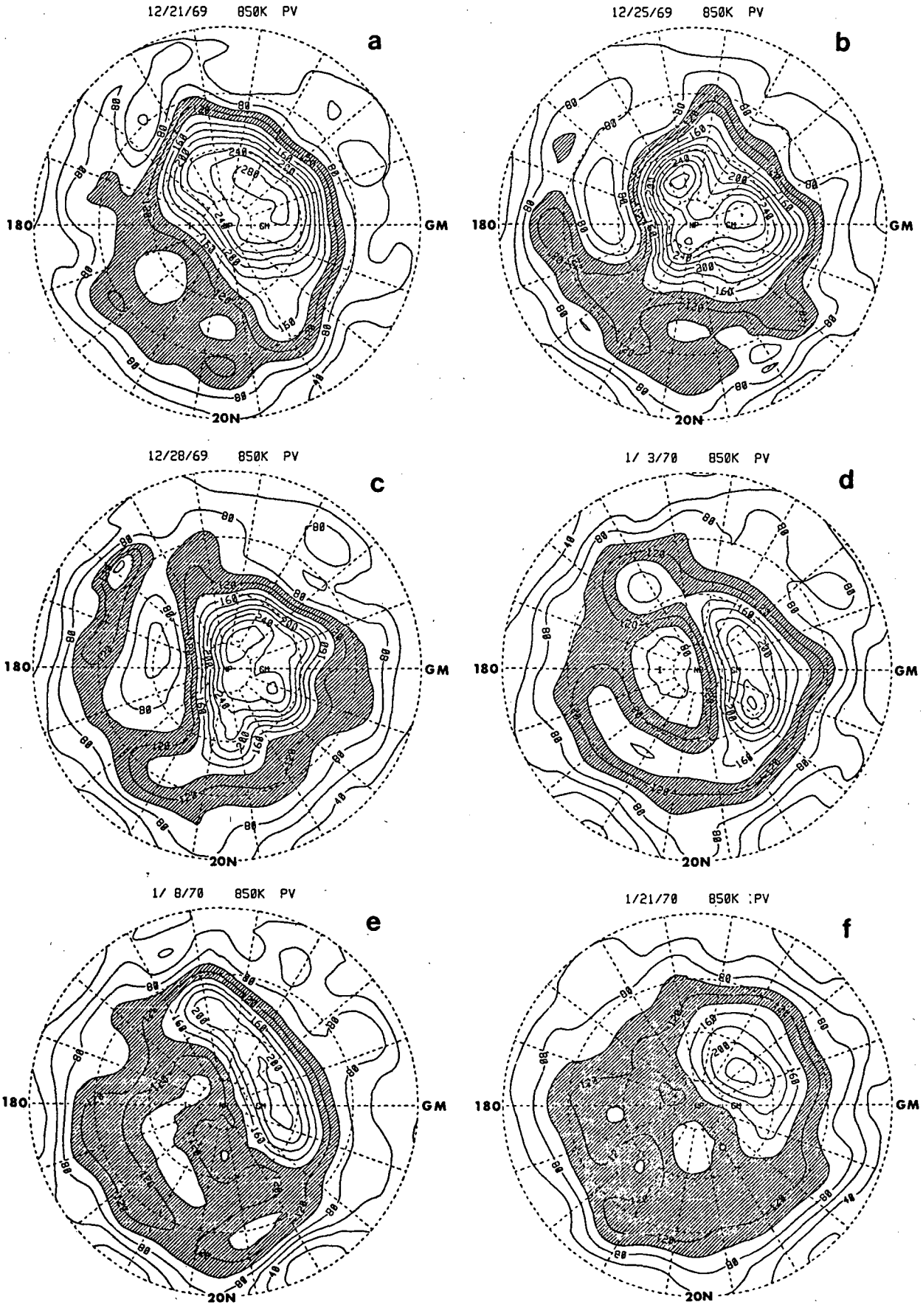


FIG. 7. Representative plots of potential vorticity during the major stratospheric warming of 1969/70.

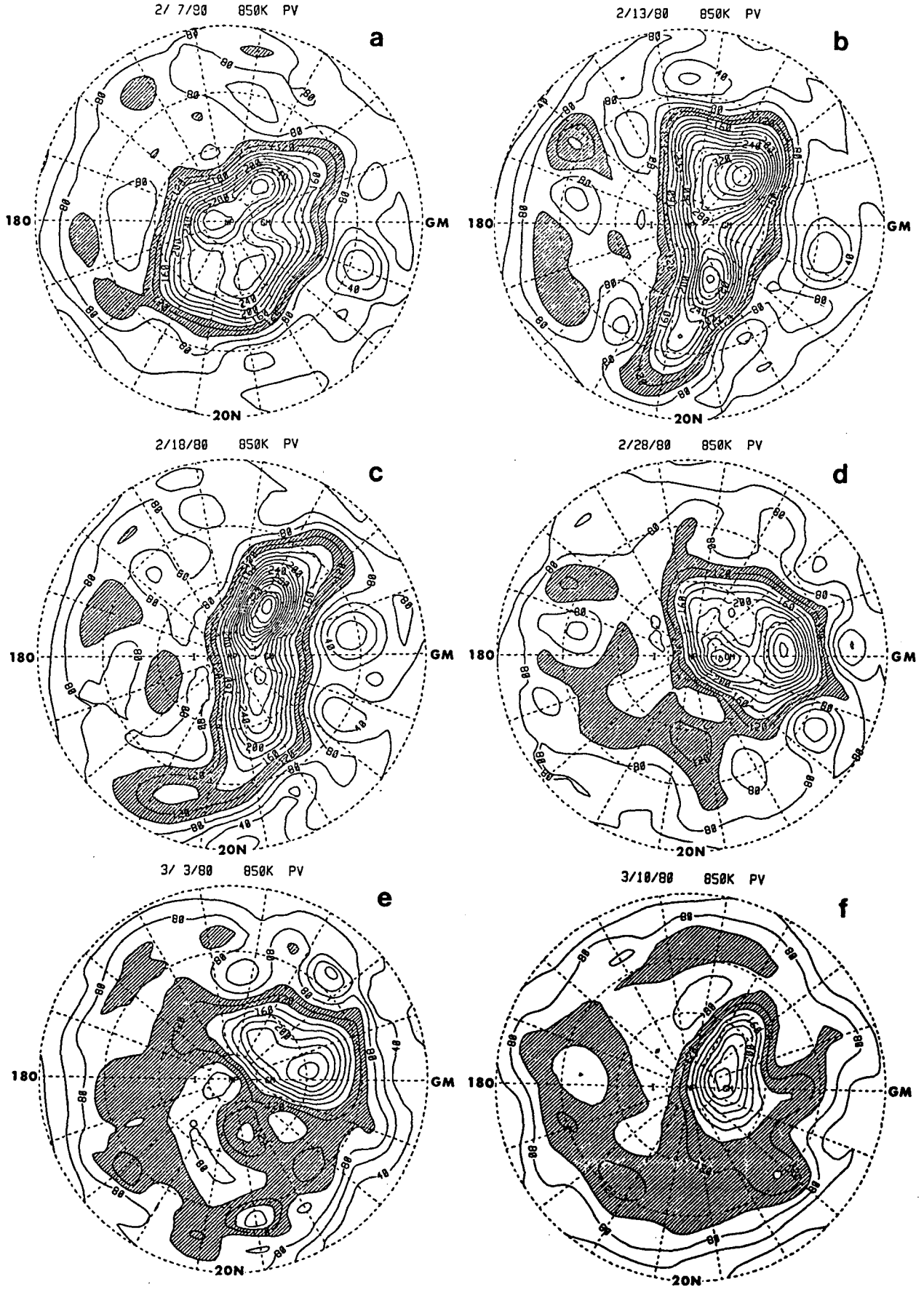


FIG. 8. Representative plots of potential vorticity during the major stratospheric warming of 1980.

Prior to a major warming the main vortex is always preconditioned (as defined by McIntyre, 1982, and McIntyre and Palmer, 1983) by wave breaking so that the area of the vortex is relatively small. Quantitatively, the area enclosed by the 1.5×10^{-4} s^{-1} PV contour always enclosed less than 11 percent of the area of the Northern Hemisphere at the time of every major warming in the 19-year climatology. This is at most a necessary condition for a major warming (adequate planetary wave activity propagating up from a variable troposphere being another), since this condition has been followed also by minor warmings and temporary distortions of the vortex.

Conversely, Canadian warmings often occur when the vortex is large, but none of these warmings, which usually occur in December, was sufficient to be labeled a major warming. In summary, preconditioning does not, of course, guarantee a major warming, but in all cases when the vortex appeared to be preconditioned, a substantial distortion or disruption of the vortex always followed within a few weeks.

The process of preconditioning and the subsequent sudden warmings appear as one continuous event when viewed in terms of wave breaking. The buildup for major sudden warmings begins by involving low-valued PV contours in wave breaking, but then rapidly involve higher value contours as the vortex area is diminished. The entire process typically lasts less than six weeks.

The interannual variability of sudden warmings is considerable, when seen in terms of 850 K PV maps. The Aleutian anticyclone is usually associated with the wave breaking and preconditioning as well as the warming itself. But the European anticyclone is sometimes involved in one or both stages. The major warmings always involve the displacement of the vortex from the pole. Often, the vortex splits and can be represented as wave 2 in the height field. Sometimes the vortex remains largely intact, but is simply displaced in a wave 1 pattern.

At finer scales (smaller than wavenumber 15–20) the process of wave breaking is not well documented observationally, but has been simulated using a high-resolution barotropic model (Juckes and McIntyre, 1987). Without finer resolution, it cannot be determined whether or not "blobs" of potential vorticity are the result of local instabilities which cause them to "roll up."

Acknowledgments. We wish to thank Professor J. M. Wallace for his advice and encouragement, and Professor W. A. Robinson and Dr. T. J. Dunkerton for stimulating conversations. We also wish to thank Prof. M. E. McIntyre and an anonymous reviewer for many helpful comments. This work was supported by the Meteorology Program of the National Science Foundation, NSF Grant ATM83-14111 and by the NASA Upper Atmosphere Research Program, NASA Grant NAGW-662.

REFERENCES

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33,** 1607–1623.

Boville, Byron A., 1987: The validity of the geostrophic approximation in the winter stratosphere and troposphere. *J. Atmos. Sci.*, **44,** 443–457.

Butchart, N., and E. E. Remsberg, 1986: The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface. *J. Atmos. Sci.*, **43,** 1319–1339.

Clough, S. A., N. S. Grahame and A. O'Neill, 1985: Potential vorticity in the stratosphere derived using data from satellites. *Quart. J. R. Met. Soc.*, **111,** 335–358.

Dunkerton, T. J., and D. P. Delisi, 1986: Evolution of potential vorticity in the winter stratosphere of January–February 1979. *J. Geoph. Res.*, **91,** 1199–1208.

——, C.-P. F. Hsu and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38,** 819–843.

Hartmann, D. L., 1977: On potential vorticity and transport in the stratosphere. *J. Atmos. Sci.*, **34,** 968–977.

Hitchman, M. H., C. B. Leovy, J. C. Gille and P. L. Baily, 1987: Quasi-stationary zonally asymmetric circulations in the equatorial lower mesosphere. *J. Atmos. Sci.*, **44,** 2219–2236.

Holton, J. R., 1980: The dynamics of sudden stratospheric warmings. *Annual Review of Earth and Planetary Science,* **8,** Annual Reviews, 169–190.

Hoskins, B. J., M. E. McIntyre and A. W. Robertson, 1985: On the use and significance of isentropic potential-vorticity maps. *Quart. J. R. Meteor. Soc.*, **111,** 877–946. Also **113,** 402–404 (update on references).

Juckes, M. N., and M. E. McIntyre, 1987: A high-resolution one-layer model of breaking planetary waves in the stratosphere. *Nature,* **328,** 590–596.

Kanzawa, H., 1982: Eliassen-Palm flux diagnostics and the effect of the mean wind on planetary wave propagation for an observed sudden stratospheric warming. *J. Meteor. Soc. Japan,* **60,** 1063–1073.

Labitzke, K., 1982: On the interannual variability of the middle stratosphere during northern winters. *J. Meteor. Soc. Japan,* **60,** 124–139.

McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? *J. Meteor. Soc. Japan,* **60,** 37–65.

——, and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature,* **305,** 593–600.

——, and ——, 1984: The 'surf zone' in the stratosphere. *J. Atmos. Terr. Phys.*, **46,** 825–849.

——, and ——, 1985: A note on the general concept of wave breaking for Rossby and gravity waves. *Pure Appl. Geophys.*, **123,** 964–975.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28,** 1479–1494.

Palmer, T. N., 1981: A diagnostic study of wavenumber-2 stratospheric sudden warming in a transformed Eulerian-mean formalism. *J. Atmos. Sci.*, **38,** 844–855.

Plumb, R. A., 1981: Instability of the distorted polar night vortex: A theory of stratospheric warmings. *J. Atmos. Sci.*, **38,** 2514–2531.

Quiroz, R. S., A. J. Miller and R. M. Nagatami, 1975: A comparison of observed and simulated properties of sudden stratospheric warmings. *J. Atmos. Sci.*, **32,** 1723–1736.

Randel, W. J., 1987: The evaluation of winds from geopotential height data in the stratosphere. *J. Atmos. Sci.*, **44,** 3097–3120.

Robinson, W., 1986: The application of the quasi-geostrophic Eliassen-Palm flux to the analysis of stratospheric data. *J. Atmos. Sci.*, **43,** 1017–1023.

Rood, R. B., 1985: A critical analysis of the concept of planetary wave-breaking. *Pure Appl. Geophys.*, **123,** 733–755.

Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112,** 2524–2529.