Chapter 5 | Extratropical Planetary-Scale Circulations

5.1 Introduction

This chapter will discuss observational and theoretical aspects of some of the planetary-scale "climatological" features of the extratropical stratosphere and mesosphere, which vary slowly from month to month during the annual cycle and recur regularly from year to year. Such features include the westerly zonal-mean winds that occur in the winter hemisphere, the easterly zonal-mean winds that occur in summer, and the geographically fixed planetary-scale wave patterns observed primarily in the northern winter.

A natural way to isolate features of this type is to perform time averages for individual calendar months in a record extending over many years. Thus, from 10 years of data, say, one can average the 10 Januaries to obtain a "mean January" field, and so on. The resulting climatological monthly mean flow patterns tend to take a wavy form in the Northern-Hemisphere winter: these patterns can, if desired, be separated into zonal-mean and zonally varying parts, as in Section 3.3. Their zonally varying parts are known as "stationary waves" and can be further separated into zonal Fourier components. The time-dependent departures from the climatological average are often known as "transient eddies." (Note that this definition of "transient" differs from that of Sections 3.6 and 4.1.) Space–time spectral analysis of these transient components sometimes reveals the presence of large-scale, coherent, zonally propagating "traveling waves."

5.2   The Observed Annual Cycle

It was mentioned in Section 1.5 that global observation of stratospheric temperature began in earnest in the late 1960s with the advent of satellite-borne infrared radiometers. Using instruments of this kind, together with conventional radiosonde measurements up to about the 10-mb level in the lower stratosphere, our knowledge of the planetary-scale structure of the middle atmosphere has gradually expanded over the last 20 years. There now exists a data base potentially large enough to construct fairly stable climatologies of temperature, geopotential height, geostrophic wind, and other quantities up to mesopause levels in the extratropics. So far, little detailed analysis of this climatological data set has been performed, but some general remarks about the gross planetary-scale structure of the middle atmosphere can be made.

We first describe the annual cycle of certain monthly mean fields. These can be displayed in several different ways: for example, the mean geopotential fields, $\langle \Phi \rangle$, say (where $\langle \ldots \rangle$ here denotes the climatological monthly average), for each month can be contoured on polar stereographic charts at various log-pressure levels. Alternatively, the monthly mean geopotential field can be expanded in zonal Fourier harmonics up to some zonal wave number S representing the limit of resolution of the data,

$$\langle \Phi \rangle = A_0(\phi, z) + \sum_{s=1}^{S} A_s(\phi, z) \cos[s\lambda + \alpha_s(\phi, z)], \qquad (5.2.1)$$

and the amplitude A_s and phase α_s of each harmonic contoured in the (ϕ, z) plane for each month. We present examples of both of these approaches for various atmospheric fields.

5.2.1   The Zonal-Mean Flow

The most basic monthly mean fields are the zonal means, corresponding to the term A_0 in Eq. (5.2.1). Schematic solstice cross sections of zonal-mean temperature and zonal-mean zonal wind in the middle atmosphere were presented in Figs. 1.3 and 1.4, which were based primarily on Northern-Hemisphere data.

To give some idea of the seasonal cycle and interhemispheric differences, Figs. 5.1 and 5.2 show monthly mean data for January, April, July, and October, from the 1986 COSPAR International Reference Atmosphere (CIRA) compilation, based on approximately 5 years of satellite data. Figure 5.1 shows the zonal-mean temperature for these months and, like Fig. 1.3,
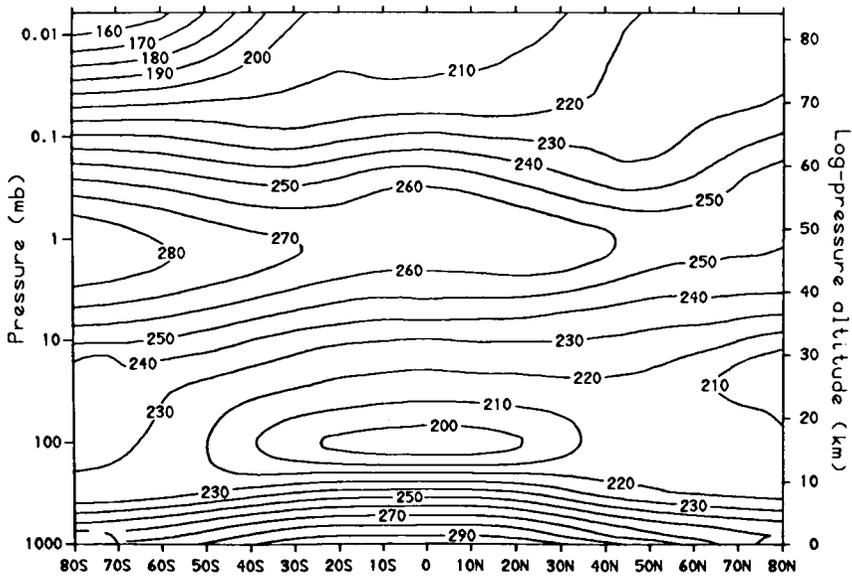
(a)   Zonal mean temperature (K)   January
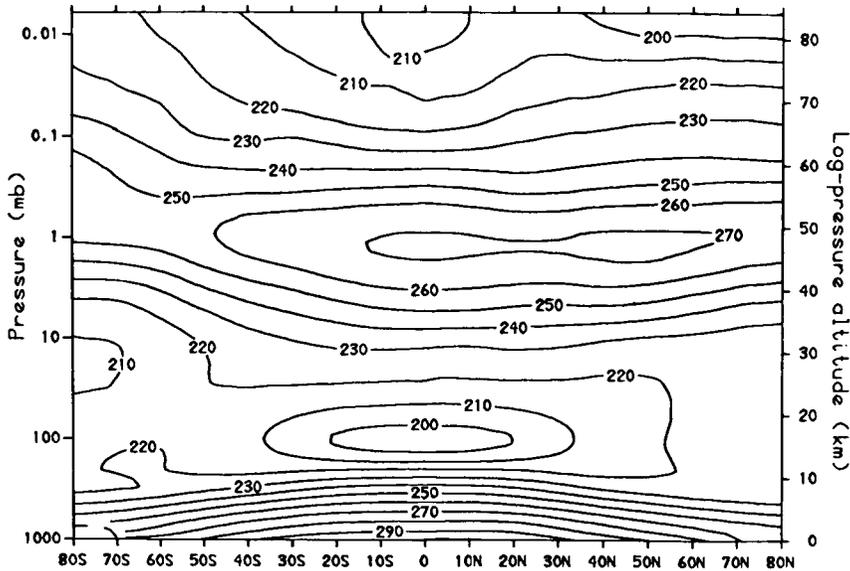
(b)   Zonal mean temperature (K)   April

Fig. 5.1. Monthly and zonally averaged temperature (K) for altitudes up to approximately 90 km, based on about 5 years of data from the *Nimbus* 5 and 6 satellites (January 1973–December 1974 and July 1975–June 1978, respectively) above 30 mb; data supplied by Berlin Free University at 30 mb and Oort's (1983) climatology for 50 mb and below. (a) January, (b)

222
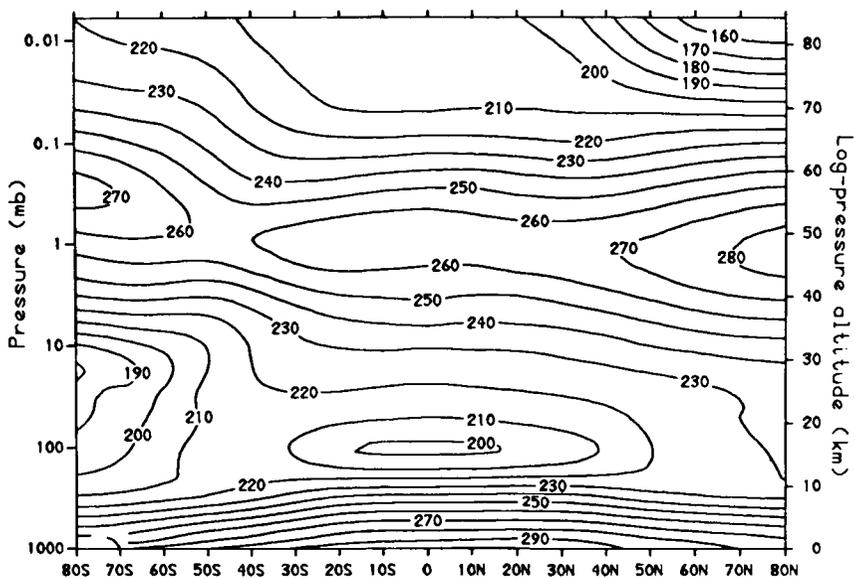
(c) Zonal mean temperature (K) July
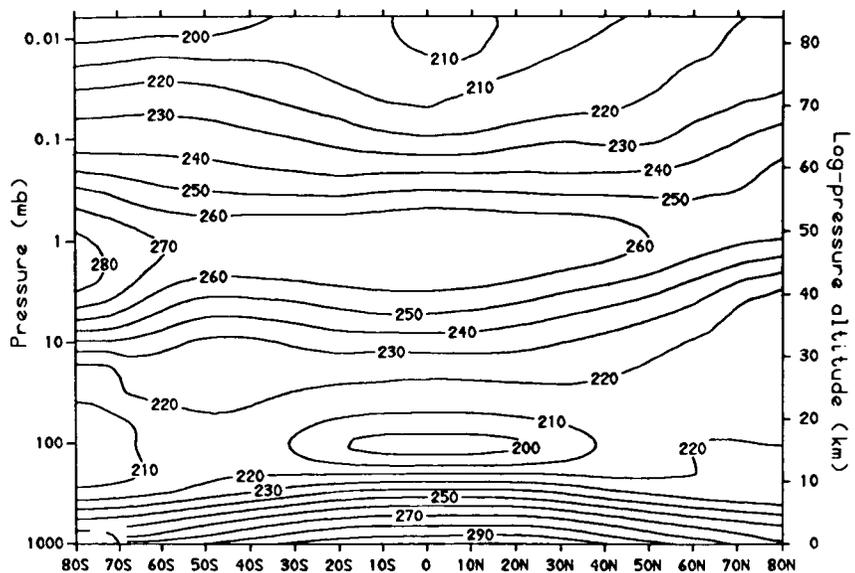
(d) Zonal mean temperature (K) October

April, (c) July, (d) October. From the 1986 CIRA compilation, courtesy of J. J. Barnett and M. Corney, Department of Atmospheric Physics, Oxford University. (See also Barnett and Corney, 1985a.)

(a) Zonal mean wind (m/s)   January

(b) Zonal mean wind (m/s)   April
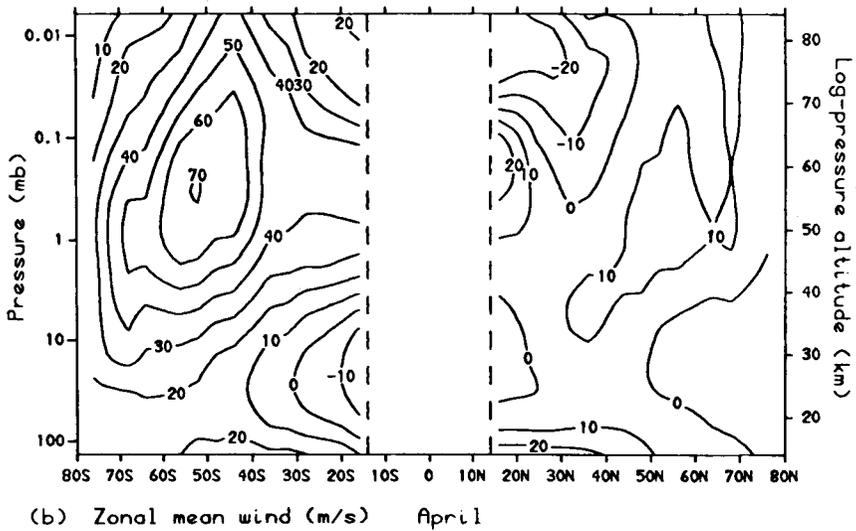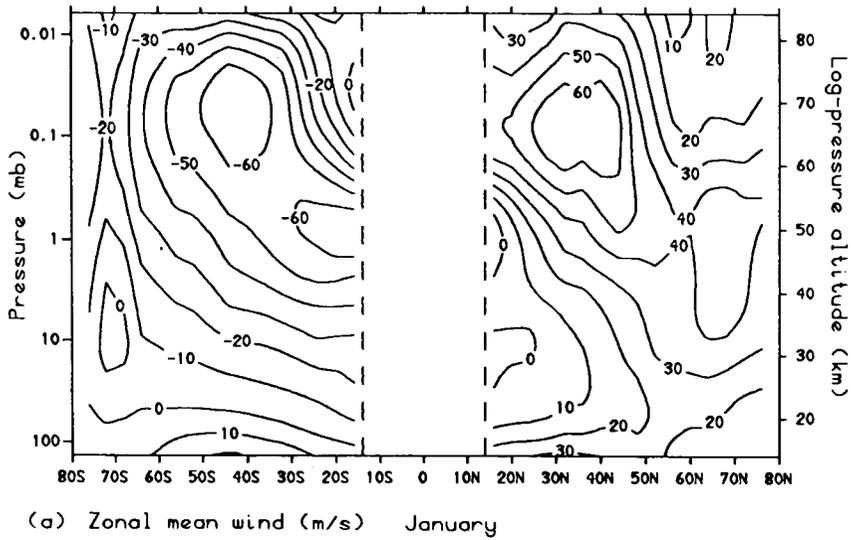
Fig. 5.2. As for Fig. 5.1, but showing geostrophic zonal winds in m s^{-1} (eastward winds positive, westward winds negative), above about 15 km.

(c)   Zonal mean wind (m/s)   July


(d)   Zonal mean wind (m/s)   October
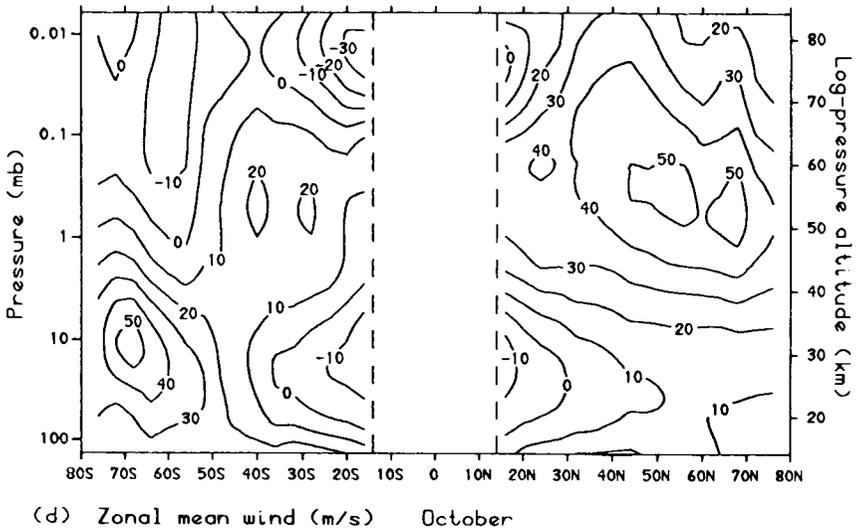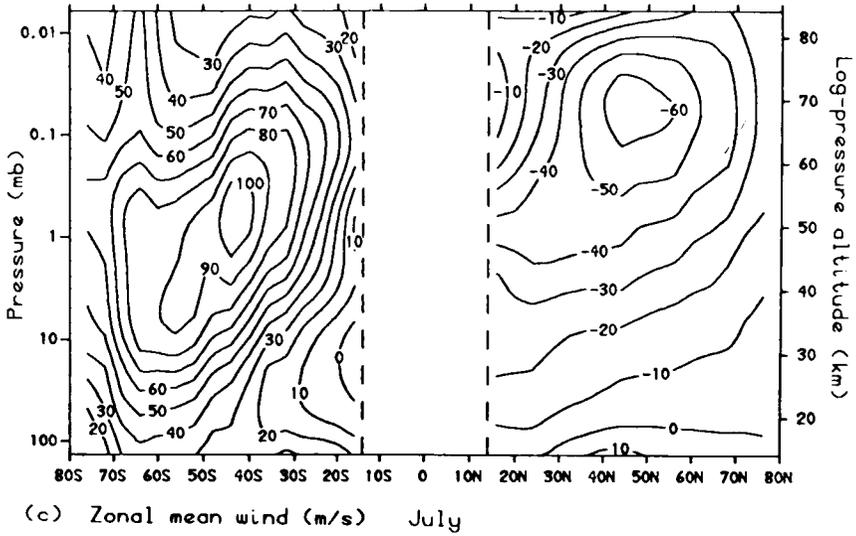
Fig. 5.2 (*continued*)

illustrates the basic features of a cold tropical tropopause, warm stratopause, and cold summer mesopause. The pattern in July is approximately a mirror image of January. However, the polar stratosphere is somewhat colder (and thus nearer to radiative equilibrium) in the southern winter (Fig. 5.1c) than in the northern winter (Fig. 5.1a). The most dramatic interhemispheric asymmetry occurs at the equinoxes; in October the southern polar stratopause (Fig. 5.1d) is much warmer than the northern polar stratopause, or than either polar stratopause in April (Fig. 5.1b); it is even warmer than the equatorial stratopause in April or October.

Figure 5.2 presents zonal-mean geostrophic winds, computed from the temperatures in Fig. 5.1 by integrating the thermal wind equation [Eq. (3.4.1c), but neglecting the term in tan ϕ]

$$f\frac{\partial \bar{u}}{\partial z} = -\frac{R}{aH}\frac{\partial \bar{T}}{\partial \phi} \qquad (5.2.2)$$

in the vertical, using climatological data at 30 mb as a lower boundary condition. The equatorial regions are omitted, partly because of the antici-pated breakdown of Eq. (5.2.2) there. Features similar to those of Fig. 1.4 are apparent; these include the westerly jets in the winter midlatitude mesosphere (stronger in the Southern Hemisphere than the Northern), which extend down to the "polar night jets" in the winter polar stratosphere. Easterlies of somewhat weaker magnitude appear in the summer mesosphere and stratosphere. Away from the subtropics, equinoctial winds tend to be westerly or weakly easterly in both hemispheres. In Chapter 7 we discuss the dynamical reasons for some of these observed features of the climatologi-cal zonal-mean circulation. Interannual variability in the stratosphere is discussed in Section 12.5.

5.2.2 Stationary Waves

We next consider zonally asymmetric aspects of the monthly mean data. The most graphic way of depicting these is by means of maps of the various fields at different levels. For example, Fig. 5.3 shows monthly mean tem-peratures at 10 mb for January, April, July, and October in the form of polar stereographic charts for each hemisphere. (These maps give the *total* field, and include the zonal mean as well as the zonally asymmetric com-ponents.) Some clear differences between the seasons and between the hemispheres are immediately evident: thus, the Northern-Hemisphere win-ter 10-mb temperature field (January) is characterized by strong departures from zonal symmetry, indicating the presence of stationary waves, while the Southern-Hemisphere winter field (July) is more symmetric about the
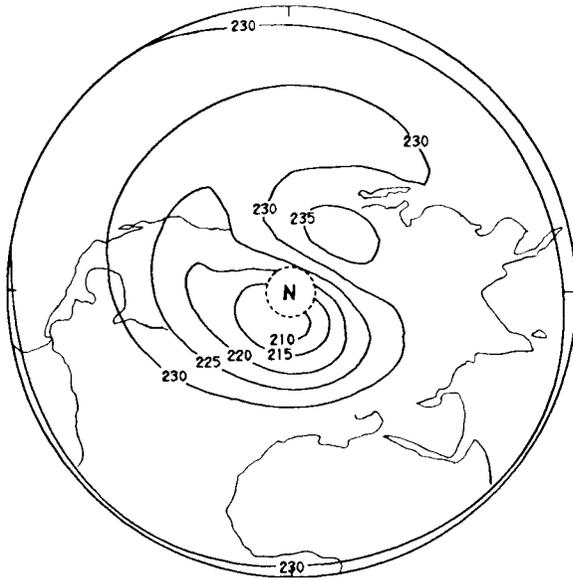
TEMPERATURE (K) AT 10 MB : JANUARY
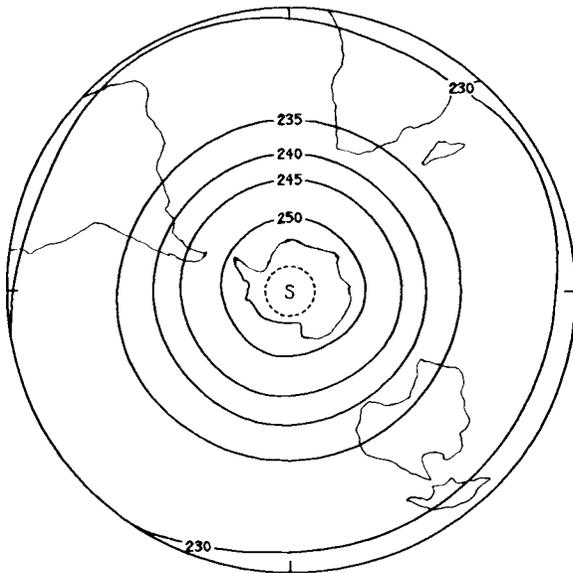
(a)

Fig. 5.3. Polar stereographic maps of monthly averaged temperature (K) at 10 mb (approximately 30 km altitude) for (a) January, (b) April, (c) July, (d) October, for the Northern Hemisphere (above) and the Southern Hemisphere (below). Outer circle, equator; inner circle; 80° latitude. (Courtesy of J. J. Barnett and M. Corney, Department of Atmospheric Physics, Oxford University. See also Barnett and Corney, 1985b.) *Figure continues.*
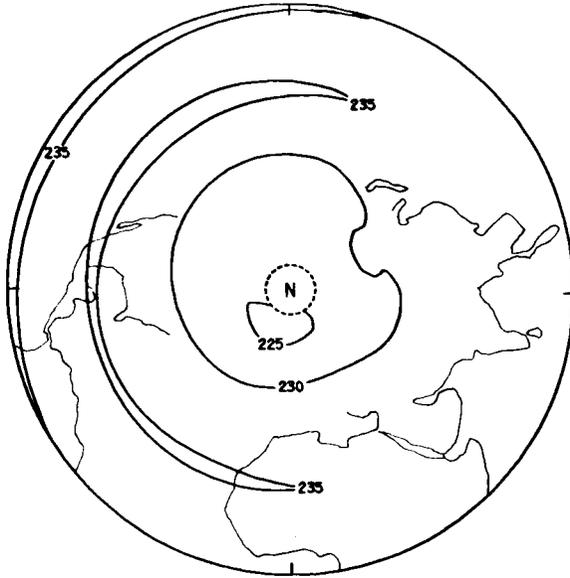
TEMPERATURE (K) AT 10 MB : APRIL
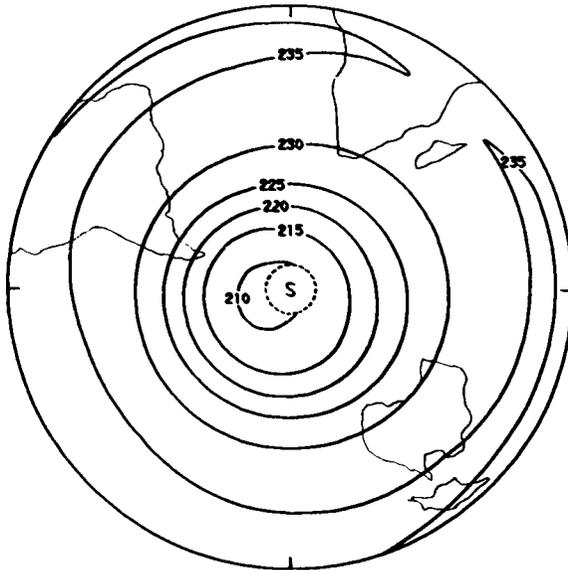
(b)

Fig. 5.3 (*continued*)
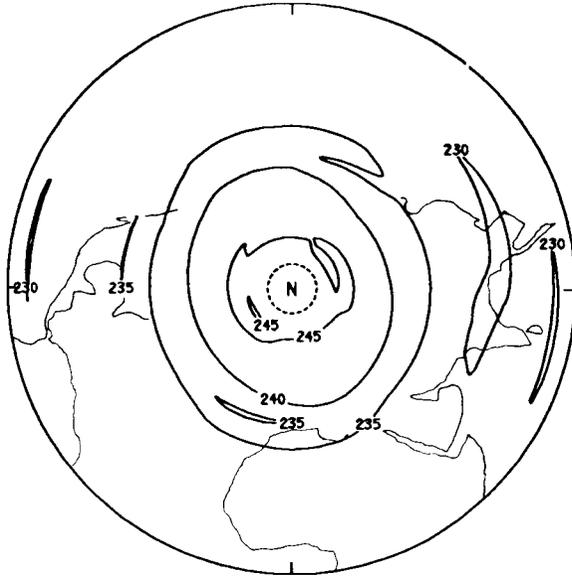
TEMPERATURE (K) AT 10 MB : JULY

(c)
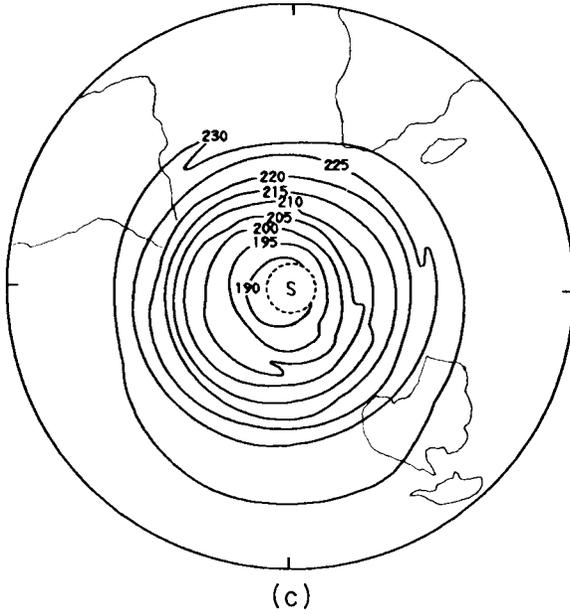
Fig. 5.3 (*figure continues*)
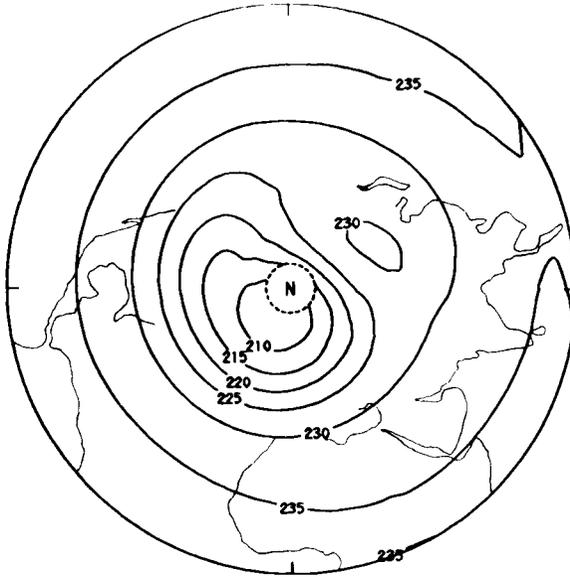
TEMPERATURE (K) AT 10 MB : OCTOBER
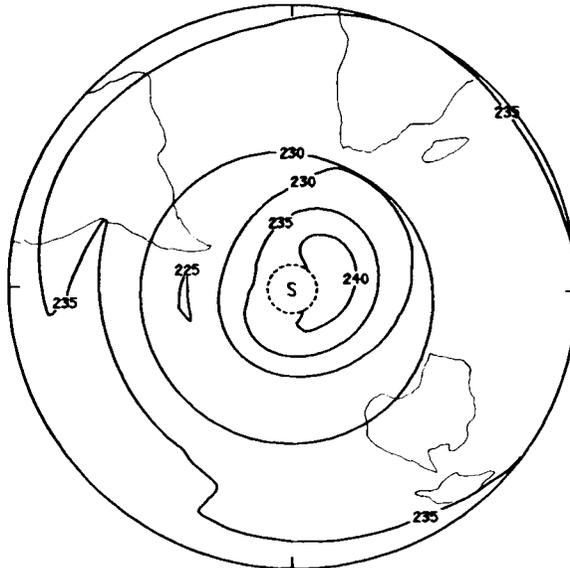
(d)

Fig. 5.3 (*continued*)

pole and exhibits strong latitudinal temperature gradients, as in Fig. 5.1c. Both summer hemispheres are fairly zonally symmetric, with quite weak temperature gradients. The spring equinoctial fields (Northern Hemisphere in April, Southern Hemisphere in October) are close to isothermal, but the autumn equinoxes have noticeable temperature gradients and (in the Northern Hemisphere) significant departures from zonal symmetry.

An alternative method of examining zonal asymmetries is by means of meridional cross sections of amplitude A_s and phase α_s [see Eq. (5.2.1)] for various zonal wave numbers $s \geq 1$. Examples of the geopotential amplitude and (negative) phase for the mean January and July $s = 1$ and 2 components are shown in Fig. 5.4. The summer hemispheres have been omitted from this figure since amplitudes are much smaller there than in the winter hemispheres. It will be seen that winter amplitudes are greater in the Northern than in the Southern Hemisphere. These differences, of course, reflect the differences in asymmetry observed in Fig. 5.3. The lines of constant phase are fairly horizontal in the extratropical lower stratosphere, but tilt equatorward–downward in the upper stratosphere and mesosphere. The longitudes of the ridges of each Fourier component ($-\alpha_1$ for wave number 1 and $-\frac{1}{2}\alpha_2$ and $-\frac{1}{2}\alpha_2 + 180°$ for wave number 2) generally progress westward with increasing height and decreasing latitude. (Note that the phase lines in Fig. 5.4 are labeled with $-\alpha_s$.)

As discussed in Section 5.3 below, there have been many attempts to model the stationary waves appearing, for example, in Fig. 5.3 in terms of the linear theory of forced planetary waves on a zonally symmetric basic state. For this reason another diagnostic that is often applied is the Eliassen–Palm flux vector $\mathbf{F}$ (see Sections 3.5, 3.6, and 4.5.5). A quasi-geostrophic version of $\mathbf{F}$ in spherical geometry is

$$\mathbf{F} = [0, -\rho_0 a(\cos \phi)\overline{v'u'}, \rho_0 a(\cos \phi)\overline{fv'\theta'}/\theta_{0z}) \qquad (5.2.3)$$

[cf. Eqs. (3.5.3) and (3.5.6)], where $f = 2\Omega \sin \phi$. Using the following approximate geostrophic formulas in spherical coordinates,

$$u' = -(fa)^{-1}\Phi'_\phi, \qquad v' = (fa \cos \phi)^{-1}\Phi'_\lambda, \qquad (5.2.4)$$

[cf. Eqs. (3.2.2)–(3.2.4)], together with Eqs. (3.1.3c) and (3.2.13), this reduces to

$$\mathbf{F} = \rho_0(0, \overline{\Phi'_\phi\Phi'_\lambda}/f^2 a, \overline{\Phi'_z\Phi'_\lambda}/N^2), \qquad (5.2.5)$$

where Φ' here represents the departure of the climatological mean geopotential from its zonal mean value [cf. Eq. (4.5.32)]. (This expression, like quasi-geostrophic theory itself, will not be valid near the equator, where f

(a) Geopotential height amplitude : wave number 1

July   January
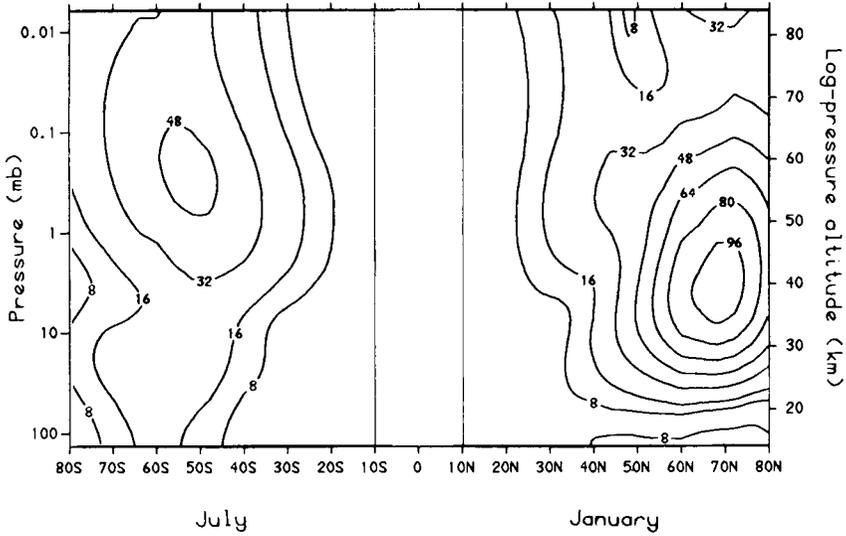
(b) Geopotential height amplitude : wave number 2

July   January
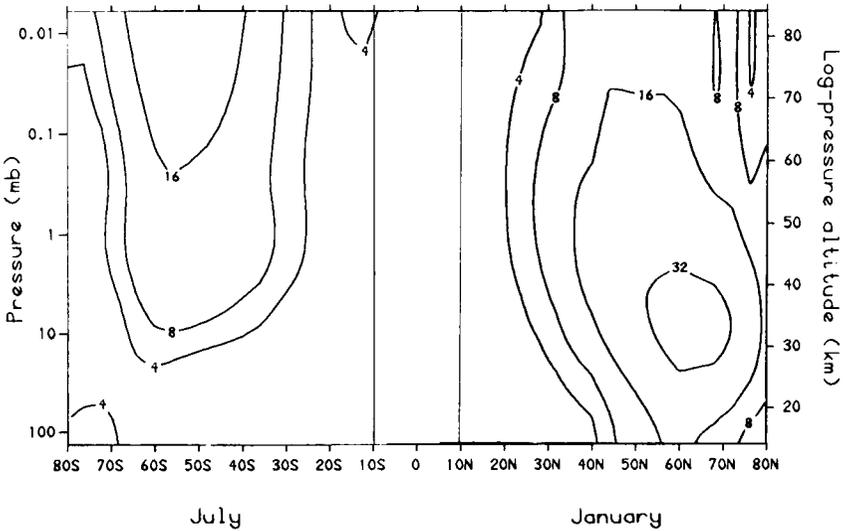
Fig. 5.4. Meridional cross section of the negative phase $-\alpha_s$ (degrees) and amplitude A_s (decameters) of the monthly mean geopotential height $\langle \Phi/g \rangle$ for wave numbers $s = 1$ and 2, from the same data source as Fig. 5.1. Only the winter hemispheres are shown: (a) $s = 1$ amplitude, (b) $s = 2$ amplitude, (c) $s = 1$ phase, (d) $s = 2$ phase.

(c) Geopotential height phase (deg. E)   :   wave number 1

July   January
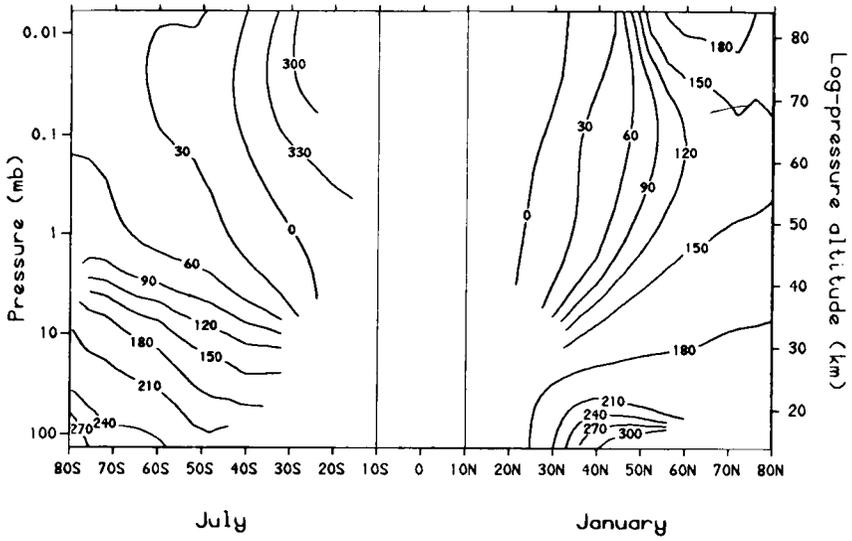
(d) Geopotential height phase (deg. E)   :   wave number 2

July   January
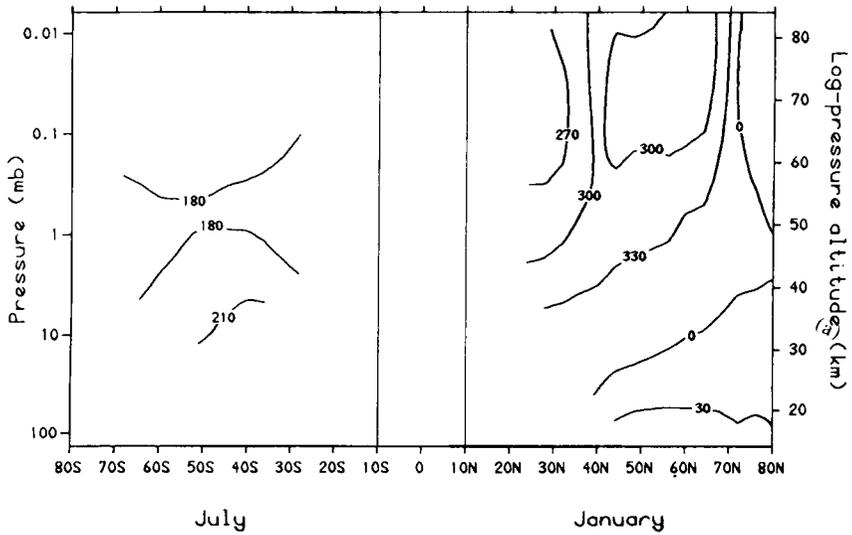
Fig. 5.4 (continued)

is small.) Employing a Fourier decomposition as in Eq. (5.2.1), we have

$$\Phi' = \sum_{s=1}^{S} A_s \cos(s\lambda + \alpha_s)$$

so that Eq. (5.2.5) becomes

$$\mathbf{F} = \frac{1}{2}\rho_0 \sum_{s=1}^{S} sA_s^2 \left(0, \frac{1}{f^2 a}\frac{\partial\alpha_s}{\partial\phi}, \frac{1}{N^2}\frac{\partial\alpha_s}{\partial z}\right), \qquad (5.2.6)$$

showing how the contributions to the ϕ and z components of $\mathbf{F}$ from each zonal harmonic depend on the square of the amplitude and the latitudinal and vertical derivatives of the phase of that harmonic. The philosophy of Section 4.5.5 can then be applied; insofar as the disturbance field Φ' or its individual Fourier components can be regarded as stationary planetary waves embedded in the zonal-mean flow, their propagation in the meridional plane can be described in terms of $\mathbf{F}$ or the contributions to $\mathbf{F}$ from the different Fourier components. (Note that these different components contribute additively to $\mathbf{F}$, since products of terms of differing zonal wavenumber vanish in the zonal mean.) Figure 5.5 shows a meridional cross section in
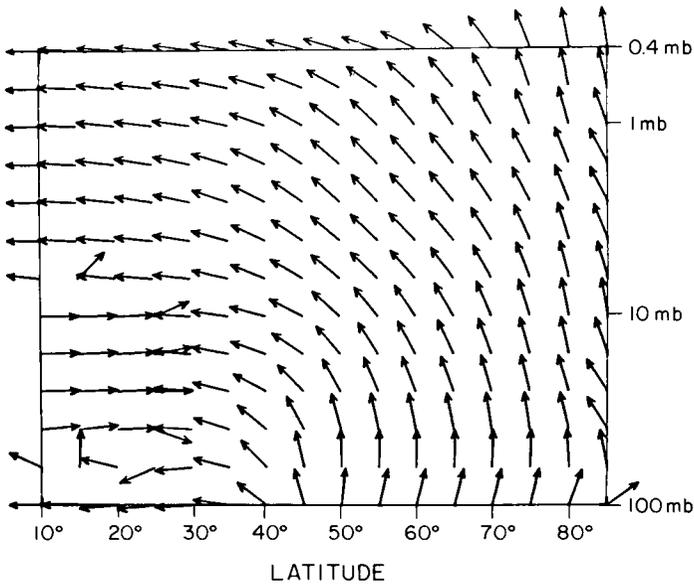
Fig. 5.5. Normalized Eliassen–Palm flux vectors $\mathbf{F}/|\mathbf{F}|$ for January in the Northern Hemisphere stratosphere, based on 4 years of data. The vertical coordinate is z. The vectors are normalized to avoid plotting problems associated with the rapid decrease of $|\mathbf{F}|$ with z. [After Hamilton (1982b). American Meteorological Society.]

which **F** is represented by suitably scaled arrows, which suggest propagation of wave activity from the winter troposphere up into the stratosphere and toward the equator; this is consistent with the ray-tracing calculations of Section 4.5.4 and also with the more detailed numerical models of Sections 5.3 and 11.2. The flux **F** tends to be dominated by $s = 1$ in these cases, since $A_1^2 \gtrsim 2A_2^2$, and the directions of the **F** arrows in Fig. 5.5 tend to be up the $s = 1$ phase gradient in Fig. 5.4c, consistent with Eq. (5.2.6).

5.2.3   Transient Eddies

As mentioned in Section 5.1, the term "transient eddies" (or "transient waves") is often used to describe departures from a time-mean flow such as the zonally asymmetric climatological monthly mean considered above. These departures can take many different forms, and careful space–time filtering, or examination of individual cases of transient development, may be needed to distinguish between the different phenomena. Transient disturbance fields can always be decomposed into zonal Fourier components, including in general a zonal-mean contribution. Whether such a decomposition helps to illuminate the dynamics will depend on circumstances.

The most spectacular transient phenomenon to be observed in the middle atmosphere is the major stratospheric sudden warming; however, this does not occur every year, and its discussion will be postponed to Chapter 6. More ubiquitous large-scale transient features include traveling planetary waves and breaking planetary waves, and these will be mentioned briefly here.

Some observations of traveling planetary waves, such as the 5-day wave and the 2-day wave, were mentioned in Section 4.4, and the simplest theory of the 5-day wave was described there. Detailed space–time spectral analysis of the type used by Mechoso and Hartmann (1982) has revealed a number of coherent, traveling, planetary-scale wave structures in the middle atmosphere. These traveling waves may vary from season to season and between hemispheres; both westward-moving and eastward-moving disturbances are found. Further theoretical attempts to model some of these structures are discussed in Sections 5.4 and 5.5.2.

A phenomenon that has recently been identified from stratospheric satellite data, and that may well turn out to be a common transient process, is the breaking planetary wave. A case study is presented by McIntyre and Palmer (1984, 1985), who use isentropic maps of Ertel's potential vorticity, as well as isobaric charts of geopotential height, from January and February 1979 to discuss the dynamics of the event. The time development is illustrated in Fig. 5.6: on January 26 there is an off-centered cyclonic vortex in the

middle stratosphere, as revealed by the low geopotential at 10 mb over Scandinavia in Fig. 5.6a and the corresponding high potential vorticity on the 850-K isentropic surface (also near 10 mb, or 30 km altitude) in Fig. 5.6b. The latter diagram shows a "tongue" of high potential vorticity emanating from the main vortex and extending westward over North America. A sequence of isentropic potential vorticity maps for the days immediately preceding January 26 suggests that this tongue represents material that has been dragged out of the main cyclonic vortex by the flow associated with the secondary "Aleutian anticyclone" in Fig. 5.6a. This argument is based on the facts that potential temperature θ and Ertel's potential vorticity P are both conserved by air parcels in adiabatic, frictionless flow (see Sections 3.1 and 3.8); in the middle atmosphere they should represent quasi-conservative tracers over periods of a few days. The sugges-
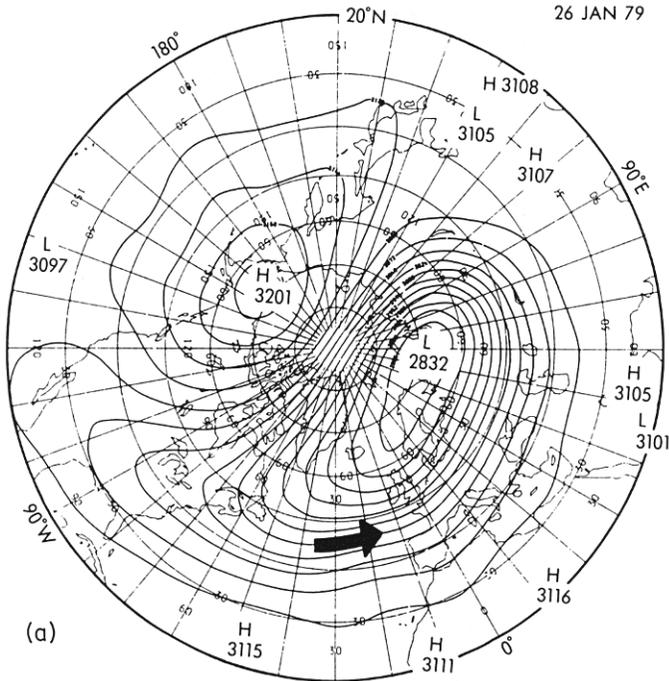
Fig. 5.6. Polar stereographic maps (outer circle: 20°N) of (a) NMC-based analysis of the 10 mb geopotential height (decameters) on January 26, 1979. (b) Coarse-grain estimate of Ertel's potential vorticity divided by gH/p_0 (where $H = 6.5$ km and $p_0 = 1000$ mb) on the 850-K isentropic surface on January 26, 1979, in 10^{-4} K m^{-1} s^{-1}. Values greater than 4 units are lightly shaded, and those greater than 6 units are heavily shaded. The dashed circle shows the position of a local maximum of just under 4 units. (c) As for (b), but for January 27, 1979. [From McIntyre and Palmer (1984), with permission.]
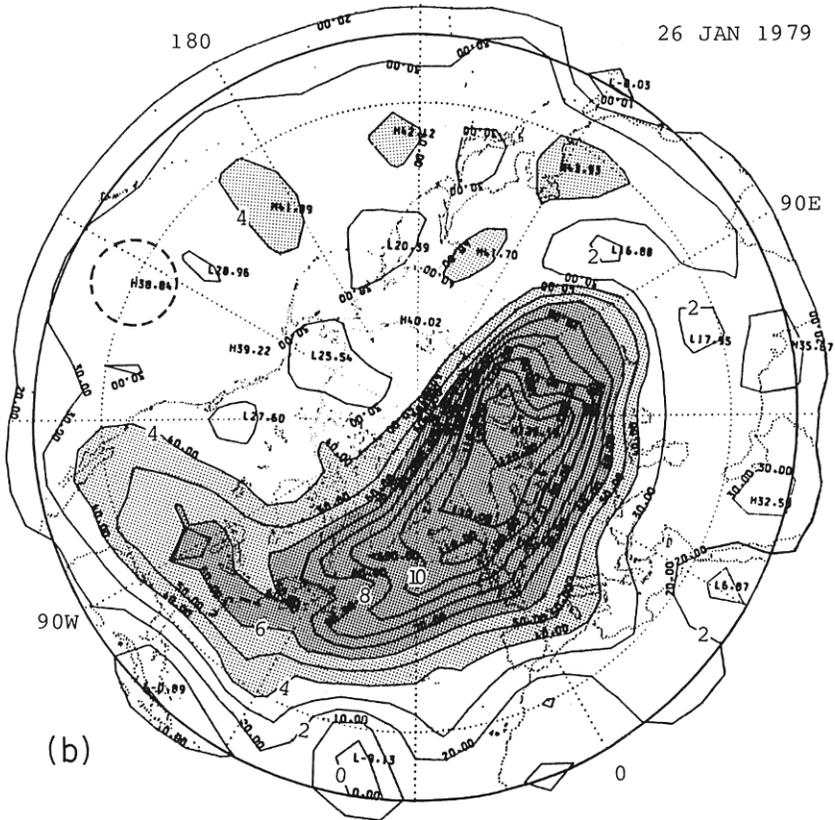
Fig. 5.6 (*figure continues*)

tion is confirmed by calculations of air-parcel trajectories and by independent observations of ozone, which is also quasi-conservative at these altitudes and latitudes in winter (Leovy *et al.*, 1985).

The sequence of events discussed by McIntyre and Palmer is characterized by rapid and irreversible deformation of material contours, as represented for example by the isopleths of P on the $\theta = 850$ K surface (which in small-amplitude wave motions would merely undulate back and forth). For this reason the authors describe the process as "planetary-wave breaking," by analogy with the breaking of ocean waves on a beach (see also the discussion of breaking internal gravity waves in Section 4.6.2). They suggest that isentropic mixing associated with events of this kind may be responsible for eroding the main winter polar vortex, to produce the region of uniform potential vorticity (or "surf zone") that is often observed to surround it in the northern hemisphere. However, Clough *et al.* (1985) find that other
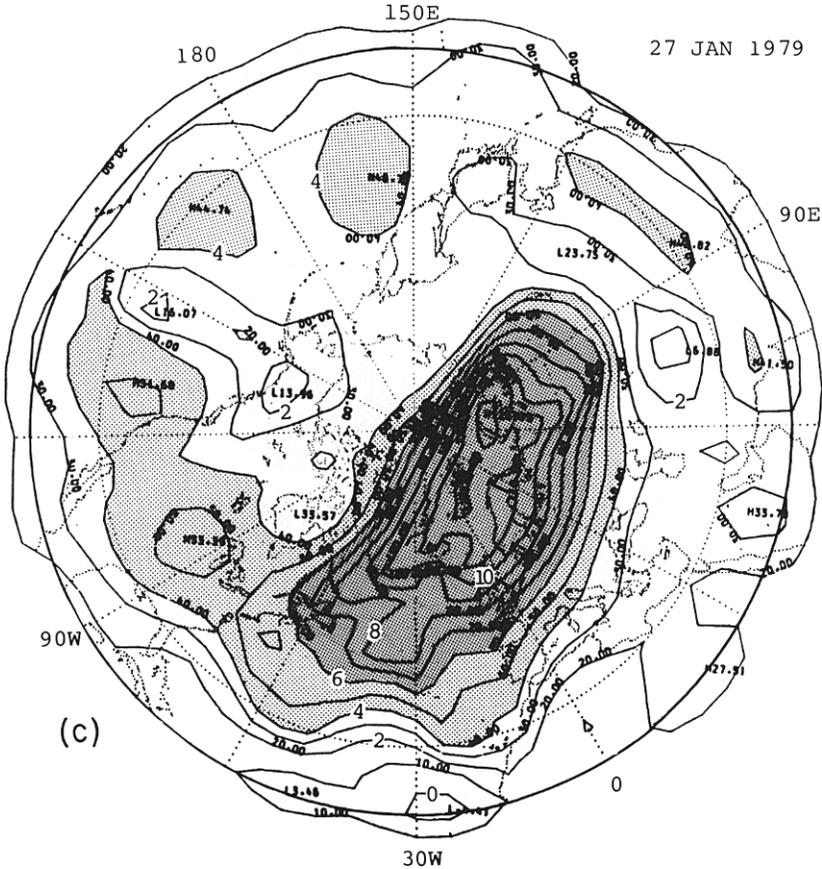
Fig. 5.6 (*continued*)

wave-breaking events, such as that of December 1981, appear to exhibit
much less isentropic mixing than that described by McIntyre and Palmer.
The reasons for these differences in behavior are not yet understood.

It should be noted that on January 27, 1979 the potential vorticity tongue
appears to start breaking up into "blobs": if real, these features could
indicate the presence of barotropic or baroclinic instability (see Section
5.5.2).

We have shown in this section that a variety of transient planetary-scale
eddy phenomena can occur in the middle atmosphere. Through the time-
averaged quadratic eddy flux terms, such transient eddies may exert an
important influence on aspects of the time-mean flow, including the station-
ary eddies. The stationary eddies, in turn, may strongly control the behavior

of the transient eddies. These effects are difficult to interpret and evaluate, since a comprehensive theory for the interaction of transient eddies and a zonally asymmetric time-mean flow (comparable to that of Sections 3.3–3.6, for the interaction of eddies with the zonal-mean flow) is still lacking.

5.3 Detailed Linear Models of Stationary Planetary Waves in the Middle Atmosphere

The modeling of stationary planetary waves in the *troposphere* began with the paper by Charney and Eliassen (1949), who used a simple linear, barotropic beta-plane model. The waves were forced by a uniform eastward basic wind blowing over the surface orography. Later studies (e.g., Smagorinsky, 1953) also included the influence of thermal sources in forcing the waves.

The first major study of stratospheric planetary waves, including stationary waves, was performed by Charney and Drazin (1961), using quasi-geostrophic theory on a beta-plane. Their largely analytical methods were outlined in Section 4.5, and their main results were discussed there.

Detailed quantitative investigation of vertically propagating stationary planetary waves in the stratosphere began with the work of Matsuno (1970). Rather than studying the precise generation mechanism for the waves, he concentrated on the hypothesis that the stationary waves in the Northern Hemisphere winter stratosphere are forced from the troposphere; he therefore imposed 500-mb heights based on observation as a lower boundary condition.

Matsuno's study used a linearized quasi-geostrophic potential vorticity equation in spherical coordinates, modified to include an ageostrophic term (namely, the "isallobaric" contribution to the northward disturbance wind). The equation takes the form

$$\left(\frac{\partial}{\partial t} + \frac{\bar{u}}{a \cos \phi} \frac{\partial}{\partial \lambda}\right) q'_{(\mathrm{M})} + a^{-1} \bar{q}_\phi v' = 0, \qquad (5.3.1)$$

where

$$v' = (fa \cos \phi)^{-1} \Phi'_\lambda, \qquad (5.3.2)$$

$$q'_{(\mathrm{M})} = \frac{1}{fa^2}\left[\frac{\Phi'_{\lambda\lambda}}{\cos^2 \phi} + \frac{f^2}{\cos \phi}\left(\frac{\cos \phi}{f^2}\Phi'_\phi\right)_\phi + \frac{f^2 a^2}{\rho_0}\left(\frac{\rho_0 \Phi'_z}{N^2}\right)_z\right], \quad (5.3.3)$$

and

$$\bar{q}_\phi = 2\Omega \cos \phi - \left[\frac{(\bar{u} \cos \phi)_\phi}{a \cos \phi}\right]_\phi - \frac{a}{\rho_0}\left(\frac{\rho_0 f^2}{N^2}\bar{u}_z\right)_z. \qquad (5.3.4)$$

Here $\bar{u}(\phi, z)$ is the basic zonal flow, Φ' is the geopotential disturbance, and $f = 2\Omega \sin \phi$. The ageostrophic modification ensures energetic consistency and also a generalized Eliassen–Palm theorem of the form of Eq. (3.6.2). On posing a stationary wave solution of zonal wave number s,

$$\Phi' = e^{z/2H} \operatorname{Re} \Psi(\phi, z) e^{is\lambda}, \tag{5.3.5}$$

and taking $N = \text{constant}$, for simplicity, the equation

$$\frac{f^2}{a^2 \cos \phi} \left(\frac{\cos \phi}{f^2} \Psi_\phi \right)_\phi + \frac{f^2}{N^2} \Psi_{zz} + n_s^2 \Psi = 0 \tag{5.3.6}$$

is obtained, where

$$n_s^2 = \frac{\bar{q}_\phi}{a\bar{u}} - \frac{s^2}{a^2 \cos^2 \phi} - \frac{f^2}{4N^2 H^2} . \tag{5.3.7}$$

Equations (5.3.6) and (5.3.7) are the spherical analogs of Eqs. (4.5.27) and (4.5.28) respectively; in particular, n_s^2 is the squared refractive index.

In Section 4.5.4 some cases were mentioned in which the basic flow $\bar{u}$ is simple enough that Eq. (5.3.6), or rather its beta-plane analog [Eq. (4.5.27)], can be solved semianalytically, under certain approximations. Matsuno studied more general flows $\bar{u}(\phi, z)$, representative of the observed zonal-mean wind in the northern winter stratosphere, for which a numerical method of solution was necessary. He considered only the "ultralong" stationary-wave components $s = 1$–3, which by the Charney–Drazin criterion [Eq. (4.5.16)] include those expected to propagate into the winter stratosphere, and for each s forced the model from below by the appropriate zonal Fourier component of the observed monthly mean 500-mb height for January 1967, using a boundary condition of the form of Eq. (3.1.6b). By linearity, the complete solution is the sum of the responses in each Fourier component. A radiation condition was imposed at the top of the model, $z = z_1$, say (near 60 km altitude), by assuming that $\bar{u}$ is independent of z above z_1. This allows separable solutions to be found above z_1, and the upward-propagating solutions can be identified as in Section 4.5.2; these then supply the required upper boundary condition. Only the Northern Hemisphere was considered; the lateral boundary conditions were that $\Psi = 0$ at the pole, to ensure bounded solutions there, and that $\Psi = 0$ at the equator. The latter is somewhat artificial, but can be rationalized by the fact that a critical line (a "zero wind line" $\bar{u} = 0$, for these stationary waves; see Section 5.6) is present near the equator in the chosen basic flow (see Fig. 5.7a). On the equatorward side of this line, n_s^2 is large and negative, and the waves are evanescent there, so that the solution is insensitive to the actual equatorial boundary condition. The singularity at the zero wind line

was removed by replacing $\bar{u}$ in the denominator of the first term on the right of (5.3.7) by $\bar{u} - i\gamma a^{-1} \cos\phi$, where γ is a small damping coefficient.

The basic zonal flow used by Matsuno is shown in Fig. 5.7a; it is reasonably similar to the January monthly mean Northern Hemisphere geostrophic wind of Fig. 5.2a. In Fig. 5.7b is shown the quantity $a^2 n_0^2$, where

$$n_0^2 = \frac{\bar{q}_\phi}{a\bar{u}} - \frac{f^2}{4N^2H^2} = n_s^2 + \frac{s^2}{a^2\cos^2\phi} \cdot \qquad (5.3.8)$$

Plots of the calculated amplitude and phase of Ψ for $s = 1$ and 2 are shown
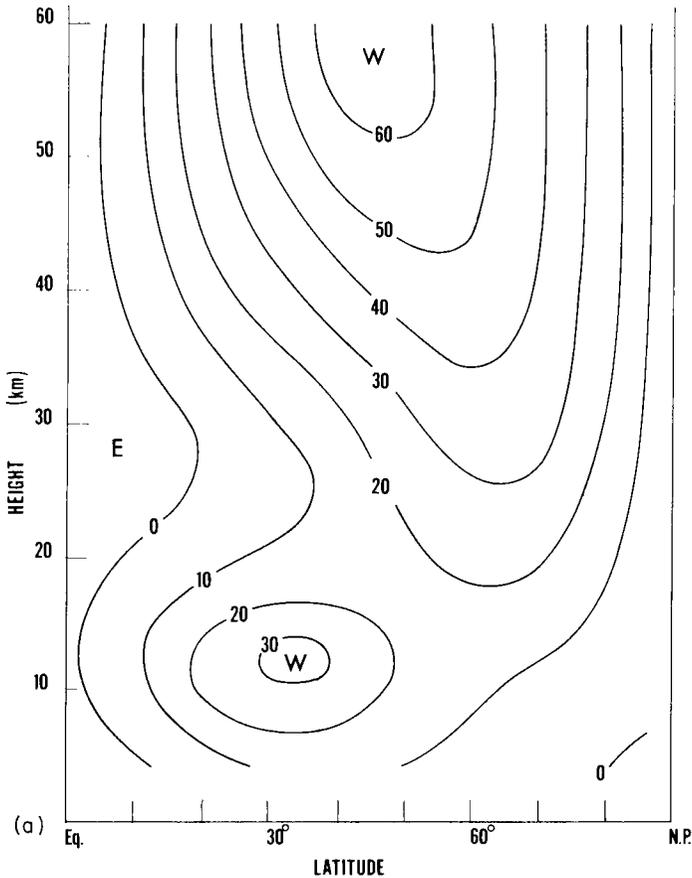
Fig. 5.7. (a) Basic wind distribution $\bar{u}(\phi, z)$ and (b) $a^2 n_0^2$, used by Matsuno in the study of the propagation of stationary planetary waves into the stratosphere. The refractive index squared for zonal wave number s can be obtained from Eq. (5.3.8). [After Matsuno (1970). American Meteorological Society.] *Figure continues.*
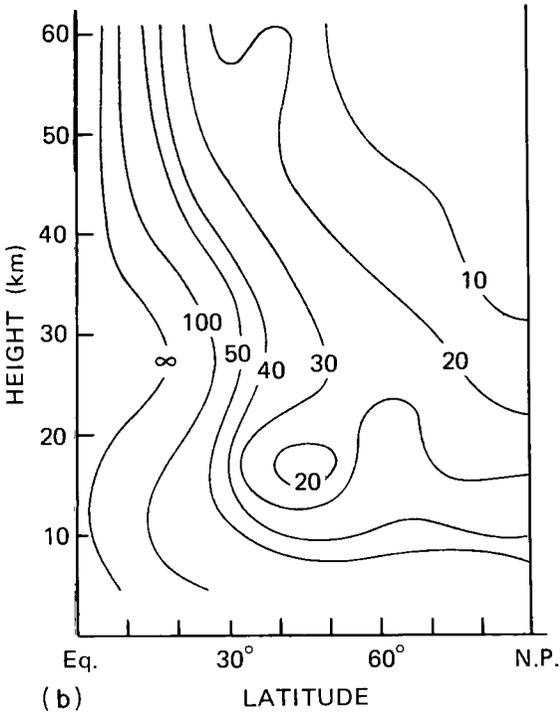
Fig. 5.7 (continued)

in Fig. 5.8; these agree quite well with the observations for January 1967, although the computed wave number 2 geopotential amplitude decays more rapidly with height than the observed.

Another diagnostic used by Matsuno was the wave-energy flux $(0, \rho_0 \overline{v'\Phi'}, \rho_0 \overline{w'\Phi'})$ [cf. Eq. (3.6.3)]. For stationary, conservative, linear waves, this is equal to $\bar{u}\mathbf{F}$ (Eliassen and Palm, 1961), where $\mathbf{F}$ is the EP flux. The pattern of arrows representing the wave-energy flux vectors in Matsuno's calculations shows the characteristic upward and then equatorward orientation noted in the observed stationary-wave EP fluxes in Fig. 5.5, as well as in the ray-tracing calculations of Fig. 4.14a.

Several other authors have used linear quasi-geostrophic numerical models similar to Matsuno's; for example, Schoeberl and Geller (1977) investigated the propagation of stationary planetary waves up to 100 km in several different mean wind structures, with Rayleigh friction and Newtonian cooling present. They found considerable sensitivity of the geopotential amplitude of the waves to the strength of the polar night jet and the magnitude of the Newtonian cooling. They interpreted their results
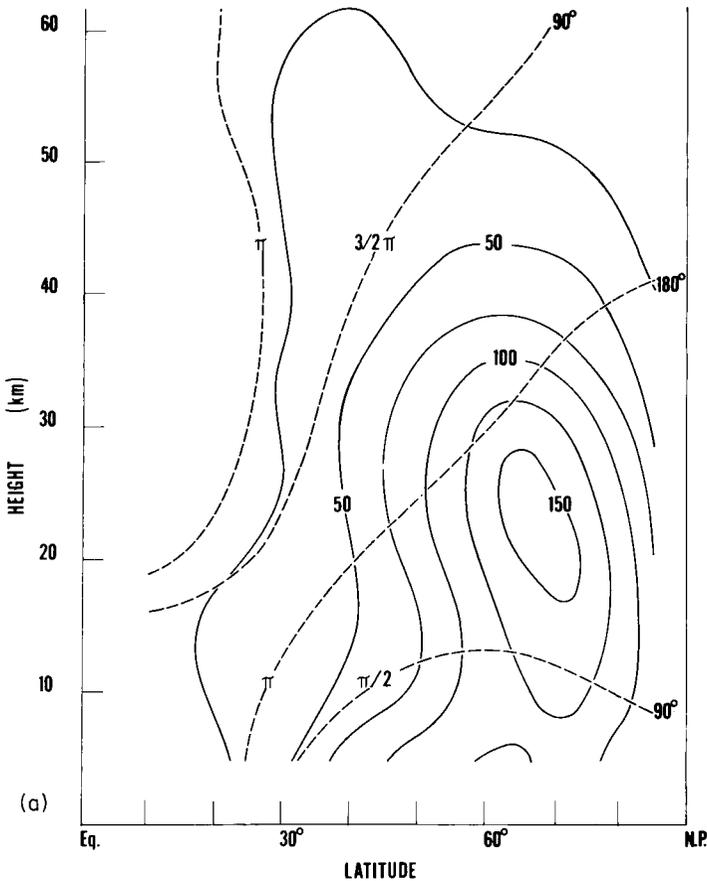
Fig. 5.8. Computed amplitude $|\Psi|$ (in meters, times a factor 1.43: solid lines) and phase of Ψ (longitude of ridges: dashed lines) for (a) $s = 1$, (b) $s = 2$, forced by the observed mean 500-mb height for January 1967. [After Matsuno (1970). American Meteorological Society.] *Figure continues.*

in terms of the vertical propagation of the gravest horizontal modes $\Theta_n(\phi)$ in an expansion of the solution Ψ of Eq. (5.3.6) in the form

$$\Psi = \sum_n \Psi_n(z)\Theta_n(\phi);$$

the Θ_n approximate the planetary-wave Hough functions discussed in Section 4.2.

The most comprehensive linear calculation to date of the winter-mean stationary planetary waves throughout the Northern-Hemisphere strato-sphere is that of Lin (1982). He used a hemispheric primitive-equation model
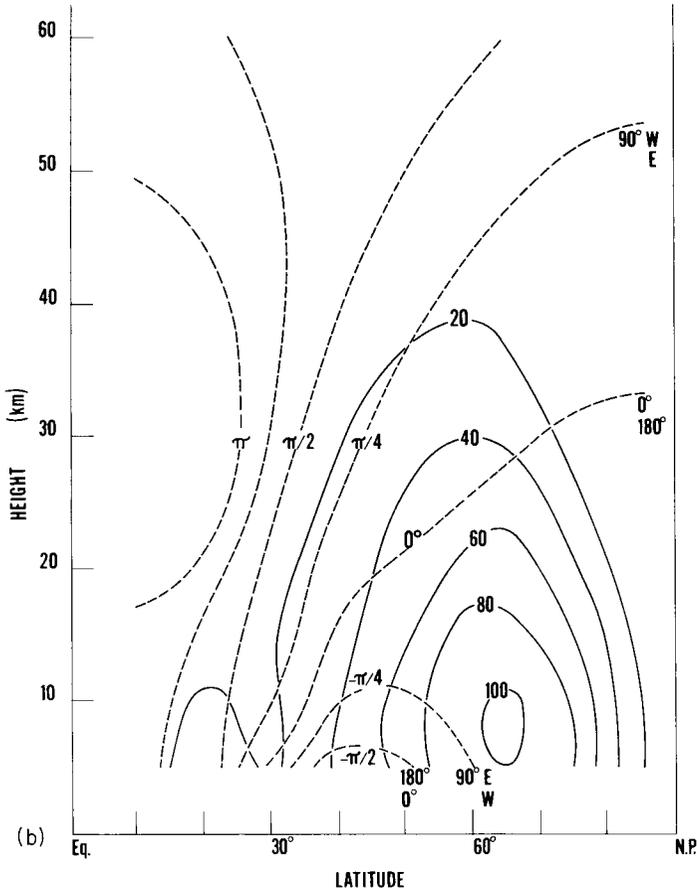
Fig. 5.8 (*continued*)

of the atmosphere from the ground up to 60 km, solving a set of equations equivalent to Eqs. (3.4.2), with various representative forms of $\bar{u}(\phi, z)$. Dissipation in the forms of Rayleigh friction and Newtonian cooling were incorporated, with specified thermal forcing, and topographic forcing included in the lower boundary condition.

Lin's results confirm those of Matsuno and of Schoeberl and Geller in most respects; in particular, they show that the vertical propagation of the stationary waves into the stratosphere is sensitive to the zonal-mean wind structure, and especially its latitudinal curvature and the latitude of the polar night jet. They also indicate that only the $s = 1$ and 2 components can propagate into the stratosphere, in qualitative agreement with the Charney and Drazin criterion. With topographic, but not thermal, forcing,

the solution simulates quite well the observed winter-mean stationary waves in the troposphere and stratosphere. However, inclusion of diabatic heating degrades the results somewhat. The reason for this is not entirely clear, but the good simulation of tropospheric waves when topographic forcing alone is used in this linear model may be partly fortuitous. This is because experiments using *nonlinear* tropospheric general circulation models (see, e.g., Held, 1983) indicate that topographic and thermal forcing produce roughly comparable contributions to the tropospheric stationary waves.

Linear models of the type discussed here do not include the time-averaged nonlinear effects of the transient eddies on the stationary waves. Recent studies suggest that these effects may be important in a number of tropospheric phenomena; the same may also be true in the stratosphere and mesosphere, although no studies have yet investigated this possibility in detail. However, general circulation models of the middle atmosphere do perform nonlinear simulations of the stationary planetary waves, and these are discussed in Section 11.2.

5.4 Detailed Linear Models of Free Traveling Planetary Waves in the Atmosphere

In Section 4.4 we presented a simple linear theory of the "5-day wave," the most prominent observed free traveling Rossby wave or global normal mode. This theory took the basic zonal flow $\bar{u}$ to be zero. In the present section we outline the theoretical methods that can be used to search for free traveling waves in more realistic zonal-mean wind structures $\bar{u}(\phi, z)$.

The mathematical problem amounts *in principle* to seeking eigensolutions of the form

$$\Phi' = \text{Re}[\hat{\Phi}(\phi, z)e^{i(s\lambda - \omega t)}] \qquad (5.4.1)$$

to the linearized primitive equations of Eqs. (3.4.2) on the sphere, subject to an upper boundary condition of decaying wave-energy density (as in Section 4.4) and to the lower boundary condition [Eq. (3.1.6a)]. The linearized version of the latter can be applied at $z = 0$, as well as $z^* = 0$, when topography is absent $(h = 0)$; it takes the form

$$\Phi'_t + \frac{\bar{u}}{a \cos \phi}\Phi'_\lambda + \frac{v'}{a}\bar{\Phi}_\phi + w'\bar{\Phi}_z = 0 \qquad \text{at} \quad z = 0. \qquad (5.4.2)$$

In practice, the easiest way of finding these solutions is to adopt the method of Geisler and Dickinson (1976) and add a forcing term

$$\text{Re}[\hat{W}(\phi)e^{i(s\lambda - \omega t)}]$$

to the right of Eq. (5.4.2), corresponding to g times an assumed nonzero vertical velocity imposed at $z^* = 0$. The integer zonal wave number s and (real) frequency ω are varied until a large, quasi-resonant response occurs in the model atmosphere. (The restriction to real ω means that only stable modes are considered: unstable modes, with complex ω, are discussed in Section 5.5.) An infinite, truly resonant response would correspond to a mode that is a true free mode of the unforced problem. In practice, weak damping in the form of Newtonian cooling and Rayleigh friction is usually included in Eqs. (3.4.2), and this ensures that the response remains finite. Moreover, some of the quasi-resonant modes may in fact be weakly propagating (or "leaky"), rather than evanescent, as $z \to \infty$, and a radiation condition may be required.[1] It is reasonable to suppose that those theoretical modes with the largest response to the imposed forcing may be good candidates for representing free traveling modes in the atmosphere, which are perhaps excited by random or other forcing effects.

Geisler and Dickinson used their method to search for theoretical waves of zonal wave number 1 with periods close to 5 days in zonal-mean winds that are representative of the middle atmosphere at solstice. They found that the period and low-level structure of the quasi-resonant mode are not very sensitive to the zonal wind configuration, and resemble those of the simple solution of Section 4.4 quite closely. In the upper stratosphere and mesosphere, however, the amplitude and phase of the model 5-day wave become strongly asymmetric about the equator, with relatively large geopotential amplitude in the summer mesosphere. Geisler and Dickinson also included Newtonian cooling with a 10-day relaxation time: the main effect of this was to halve the summer mesosphere maximum.

Another free traveling mode to receive much attention is the 2-day wave (Fig. 4.11), which may perhaps be identified with the gravest antisymmetric (in w') mode for $s = 3$ in the absence of mean winds. Here the inclusion of a mean flow $\bar{u}(\phi, z)$ in numerical calculations (e.g., by Salby, 1981a,b) leads to larger amplitudes in the summer hemisphere than in the winter hemisphere, in agreement with observation. The observed 16-day wave can likewise be identified with the second symmetric $s = 1$ mode; this mode is also fairly sensitive to the details of the basic wind field. Some observations of the 16-day wave in the upper stratosphere are presented in Fig. 5.9, and some theoretical calculations are shown in Fig. 5.10. Note that both show roughly equatorially symmetric amplitude structure at equinox, but much

---

[1] One can also regard the (real) quasi-resonant frequencies as approximations to the true (complex) eigenfrequencies of the dissipative problem. From this viewpoint, the distinction between the free modes considered here and the unstable modes of Section 5.5 may become blurred.
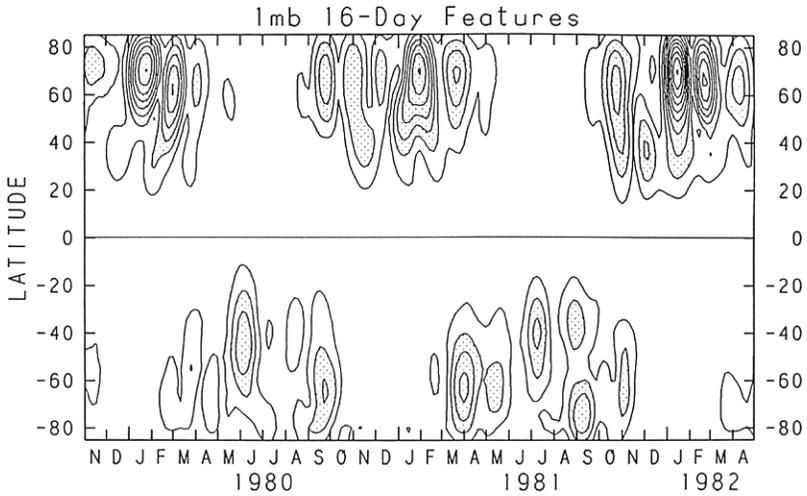
Fig. 5.9. The observed "16-day wave," as revealed by a latitude–time section of the geopotential height amplitude of $s = 1$ westward-traveling waves at 1 mb, band-passed to include periods between 12 and 24 days. (The results are insensitive to the precise bandwidth.) Values are averaged over 10 days. The contour interval is 50 m, and stippling indicates values greater than 100 m. [After Hirooka and Hirota (1985). American Meteorological Society.]
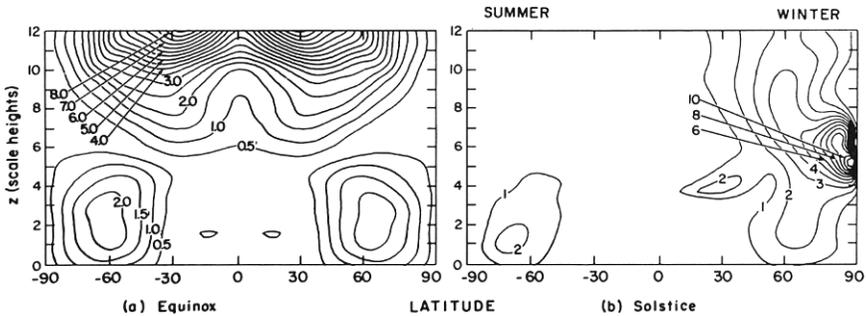
Fig. 5.10. Geopotential height amplitude (arbitrary units) of the second symmetric mode for $s = 1$ in idealized background wind $\bar{u}(\phi, z)$ and temperature $\bar{T}(\phi, z)$ fields for (a) equinox (period = 16.4 days) and (b) solstice (period = 15.7 days). [After Salby (1981b). American Meteorological Society.]

larger amplitudes in the winter hemisphere than in the summer hemisphere at solstice.

5.5 Barotropic and Baroclinic Instability

5.5.1 Necessary Conditions for Instability of Zonally Symmetric Basic Flows

The linear theoretical models of planetary-scale disturbances in the middle atmosphere that have been considered so far in this book have all been for *stable* disturbances; that is, their amplitudes do not increase indefinitely with time. These disturbances have typically been of the form of Eq. (5.4.1), with real frequency ω, or a sum of terms of this kind. We now briefly examine a large class of disturbances whose amplitudes, as predicted by linear theory, grow without limit. These modes thus represent *unstable* disturbances to the basic flow and, if they are of the form of Eq. (5.4.1), their frequencies ω are complex, with positive imaginary parts.

Two types of large-scale instability that may be important in the middle atmosphere are barotropic and baroclinic instability; these are both described by quasi-geostrophic theory. Barotropic instability depends on large horizontal curvature of the basic flow profile, while baroclinic instability depends, roughly speaking, on vertical curvature. Combined barotropic-baroclinic instability can occur in a basic flow that varies both horizontally and vertically.

A more precise statement of some of the necessary conditions for barotropic or baroclinic instability is provided by the theorem of Charney and Stern (1962). This states that, under appropriate boundary conditions, a *necessary* condition for instability of a basic zonal flow $\bar{u}(y, z)$ on a beta-plane to conservative quasi-geostrophic disturbances is that the basic northward quasigeostrophic potential vorticity gradient $\bar{q}_y \equiv \beta - \bar{u}_{yy} - \rho_0^{-1}(\rho_0 \varepsilon \bar{u}_z)_z$ must change sign somewhere in the flow domain. We give a simple proof of this theorem that, unlike most stability proofs, does not restrict the disturbances to the "normal mode" form of Eq. (5.4.1).

We suppose that the flow is bounded by vertical walls at $y = 0, L$, and by a rigid lower boundary $z^* = 0$; within the linear theory considered here, the latter can be replaced by $z = 0$. In the conservative case, the quasi-geostrophic generalized Eliassen–Palm theorem [Eq. (3.6.5)] can be rewritten using Eq. (3.6.10) for A, instead of Eq. (3.6.6). If the resulting equation is integrated over y and z we obtain

$$\frac{\partial}{\partial t} \int_0^\infty \int_0^L \frac{1}{2} \rho_0 \bar{q}_y \overline{\eta'^2} \, dy \, dz = \int_0^L \rho_0 f_0 \overline{v'\theta'} / \theta_{0z}|_{z=0} \, dy \qquad (5.5.1)$$

to second order in amplitude, using the sidewall boundary conditions $v' = 0$ at $y = 0, L$ and assuming that $\rho_0 \overline{v'\theta'}/\theta_{0z} \to 0$ as $z \to \infty$; here η' represents the northward parcel displacement, defined for example in Eq. (3.6.8).

The linearized lower boundary condition [Eq. (5.4.2)] can be used to simplify the right hand side of Eq. (5.5.1). In the quasigeostrophic beta-plane case, this boundary condition becomes

$$\bar{D}\psi' + v'\bar{\psi}_y + w'f_0^{-1}\bar{\Phi}_z = 0 \qquad \text{at} \quad z = 0, \tag{5.5.2}$$

where $\bar{D} = \partial/\partial t + \bar{u}\,\partial/\partial x$ and $\psi = f_0^{-1}[\Phi - \Phi_0(z)]$ by Eq. (3.2.4). The linearized quasi-geostrophic disturbance potential temperature equation is

$$\bar{D}\theta' + v'\bar{\theta}_y + w'\theta_{0z} = 0 \tag{5.5.3}$$

[see Eqs. (3.2.9d) and (3.4.2e)]. Now w' is not the geometric disturbance vertical velocity and does not vanish at $z = 0$; however, it can be eliminated between Eqs. (5.5.2) and (5.5.3). Equations (3.2.3), (3.2.5′), (3.2.13), (3.5.5d), and (3.6.8) then give

$$\bar{D}\left\{\frac{R}{f_0 H}\theta' - \eta'(\bar{u}_z - B\bar{u}) - B\psi'\right\} = 0 \qquad \text{at} \quad z = 0, \tag{5.5.4}$$

where

$$B \equiv N^2 H/R\bar{T} = (T_s/\bar{T})N^2/g.$$

Integrating Eq. (5.5.4), given suitable initial conditions, we obtain

$$\theta' = f_0 H R^{-1}[\eta'(\bar{u}_z - B\bar{u}) + B\psi'] \qquad \text{at} \quad z = 0,$$

and hence

$$f_0\overline{v'\theta'}/\theta_{0z} = \varepsilon(\bar{u}_z - B\bar{u})\overline{v'\eta'} = \varepsilon(\bar{u}_z - B\bar{u})(\tfrac{1}{2}\overline{\eta'^2})_t, \qquad \text{at} \quad z = 0, \tag{5.5.5}$$

by Eqs. (3.2.13), (3.2.16), and (3.6.8). Substitution of Eq. (5.5.5) into Eq. (5.5.1) yields

$$\frac{\partial}{\partial t}\int_0^\infty\int_0^L \frac{1}{2}\rho_0\bar{q}_y\overline{\eta'^2}\,dy\,dz - \frac{\partial}{\partial t}\int_0^L \frac{1}{2}\rho_0\varepsilon(\bar{u}_z - B\bar{u})\overline{\eta'^2}\bigg|_{z=0}\,dy = 0, \tag{5.5.6}$$

or, more compactly,

$$\frac{\partial}{\partial t}\int_{0_-}^\infty\int_0^L \frac{1}{2}\rho_0\tilde{q}_y\overline{\eta'^2}\,dy\,dz = 0, \tag{5.5.7}$$

where

$$\tilde{q}_y \equiv \bar{q}_y - \varepsilon\,\delta(z)(\bar{u}_z - B\bar{u})|_{z=0}, \tag{5.5.8}$$

$\delta(z)$ is the Dirac delta function, and the z integration extends from just below $z = 0$. The term $B\bar{u}|_{z=0}$ in Eq. (5.5.8) is known as a "non-Doppler" term. It is not invariant under a Galilean transformation $\bar{u} \to \bar{u} - u_0$, although it is often negligible compared to $\bar{u}_z$ at $z = 0$.

If $\tilde{q}_y$ is positive everywhere, the integral in Eq. (5.5.7) is positive definite and can be taken as a global measure of disturbance amplitude (see the end of Section 3.6). Equation (5.5.7) states that this quantity is constant in time; in particular it does not grow, and if it is initially small it will remain small. This accords with the usual properties of a globally *stable* disturbance. A similar result holds if $\tilde{q}_y$ is negative everywhere.

On the other hand, these considerations do not apply if $\tilde{q}_y$ takes both positive and negative values. The parcel displacement η' may then perhaps grow indefinitely with time, indicating an *unstable* disturbance, at least while linear theory remains valid. The change of sign of $\tilde{q}_y$ can be due to the interior potential vorticity gradient $\bar{q}_y$ changing sign, or to the boundary term somewhere having the opposite sign to the interior $\bar{q}_y$. An important example of the latter case often occurs when vertical shear—and thus a horizontal potential temperature gradient, by the thermal wind equation, Eq. (3.5.5d)—is present at $z = 0$. It should be recalled that $\bar{q}_y$ can be related to the northward gradient of Ertel's potential vorticity on an isentropic surface, by Eq. (3.8.10).

A stronger condition, due essentially to Fjørtoft (1950), states that if there exists a constant u_0 such that $(\bar{u} - u_0)\tilde{q}_y < 0$ for all y and z, then the flow is stable to conservative quasi-geostrophic disturbances. This can be proved by a method similar to that given above for the Charney–Stern theorem, but using the quasi-geostrophic version of the wave-energy equation, Eq. (3.6.3) in addition to the generalized Eliassen–Palm theorem. The Charney–Stern theorem follows as a special case by choosing $u_0 > \max(\bar{u})$ if $\tilde{q}_y > 0$ everywhere and $u_0 < \min(\bar{u})$ if $\tilde{q}_y < 0$ everywhere.

It should be noted that these theorems only give *sufficient* conditions for *stability* or, conversely, *necessary* conditions for *instability*. They cannot by themselves tell us that a particular flow is definitely unstable; an explicit search for unstable modes will normally be necessary. The simplest cases occur when $\bar{u}$ and $\bar{q}_y$ depend only on y (leading to the possibility of barotropic instability) or when they depend only on z (leading to the possibility of baroclinic instability). Numerous calculations of unstable disturbances to such flows have been made.

5.5.2. Barotropic and Baroclinic Instability Calculations for Representative Middle Atmosphere States

In this section we describe some calculations of unstable modes that have been suggested as possible explanations for some of the observed traveling-wave structures mentioned in Sections 4.4 and 5.2.3, or at least for the initiation of such structures.

The first example is that of the localized eastward-moving "warm pool" documented by Prata (1984) in the Southern Hemisphere winter upper stratosphere and shown in Fig. 4.12. Hartmann (1983) has shown that, in the region where this phenomenon occurs, the spherical analog of $\bar{q}_y$ tends to change sign, owing to strong meridional curvature of the zonal-mean zonal wind on the poleward flank of the stratospheric jet (Fig. 5.11). He performed linear instability calculations for this basic flow, and found barotropically unstable disturbances of zonal wave numbers 1 and 2, which move eastward with periods of 3–4 days and 1.5–2 days, respectively, and which have e-folding times of a few days and geopotential amplitude maxima near 70°S. He suggested that the localized disturbances observed by Prata may represent a "phase-locking" of these wave-number 1 and 2 modes. However, linear theory is only able to describe the early, small-amplitude evolution of disturbances of this kind, and a full understanding will probably need to await nonlinear calculations, in which the unstable disturbances are allowed to grow to (and perhaps equilibrate at) finite amplitude.

The Southern Hemisphere zonal-mean wind structure shown in Fig. 5.11 also has a reversed potential vorticity gradient on the equatorward flank of the jet, and Hartmann (1983) showed that instabilities are associated with this also. These bear some resemblance to observed modes that move slowly
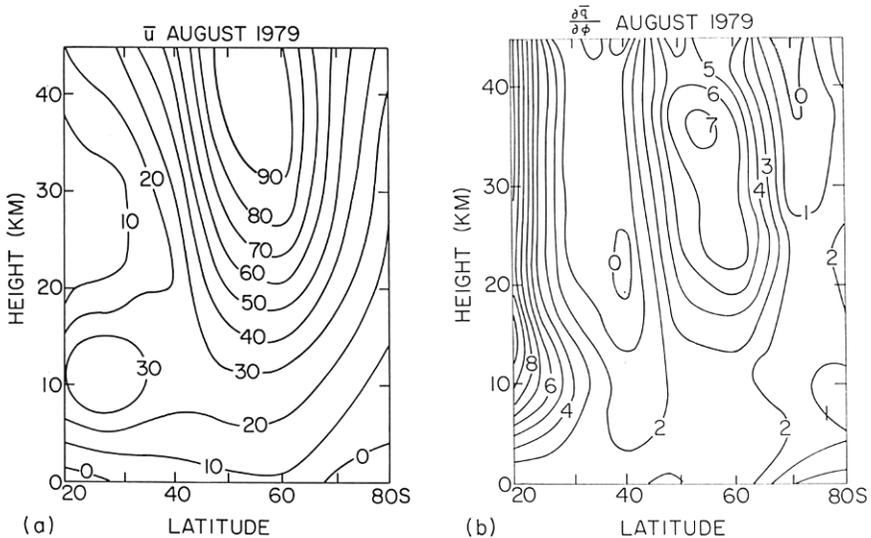

Fig. 5.11. (a) Basic wind distribution $\bar{u}(\phi, z)$ and (b) quasi-geostrophic potential vorticity gradient $\partial\bar{q}/\partial\phi$ (divided by Ω; regions of negative $\partial\bar{q}/\partial\phi$ stippled) for the month of August 1979 in the southern hemisphere. [After Hartmann (1983). American Meteorological Society.]

eastward in the Southern Hemisphere. However, Hartmann (1985) points out that the theoretical modes have a poleward EP flux ($-\rho_0\overline{u'v'} < 0$), while the observed modes have an equatorward EP flux.

Plumb (1983) examined the baroclinic instability of profiles $\bar{u}(z)$, which broadly capture the vertical variation of the zonal wind in the Southern Hemisphere winter mesosphere. He found that a baroclinically unstable mode of zonal wave number 3 and 2-day period can develop when the westerly upper mesospheric shear exceeds $6 \text{ m s}^{-1} \text{ km}^{-1}$. He suggested that this mechanism may elucidate certain aspects of the observed 2-day wave that are not explained by the "free normal mode" theory of Salby (1981a) (see Section 5.4). Further observations and, again, a nonlinear treatment may be needed to resolve the question; it could be that the instability process describes the generation of the 2-day wave and Salby's calculations, despite being linear, describe some aspects of a finite-amplitude equilibrated state.

Plumb's example, like those of Hartmann (1983), depends on a change of sign of $\bar{q}_y$ within the atmosphere to violate the Charney–Stern stability criterion. Other studies have investigated possible baroclinic instabilities in the troposphere, stratosphere, and mesosphere that depend on temperature gradients (or, equivalently, vertical shear) at the ground, so that the delta-function term in Eq. (5.5.8) is opposite in sign to $\bar{q}_y$ in the interior. The resulting unstable modes for typical winter flows tend to fall into two broad classes: the so-called "Charney modes," which decay monotonically with height and are largely confined to the troposphere and lower stratosphere, and the "Green modes," which oscillate somewhat with height and radiate into the stratosphere (Hartmann, 1979; Straus, 1981). The observational study by Mechoso and Hartmann (1982) identified eastward-traveling waves in both troposphere and stratosphere of the Southern Hemisphere, but with little coherence between lower and upper levels. They suggested that the tropospheric disturbances might be identified with Charney modes and the stratospheric disturbances with Green modes. Yet again, this hypothesis deserves a nonlinear investigation, since McIntyre and Weissman (1978) point out that radiating instabilities, which invariably have small growth rates, are in general unlikely to describe far-field behavior accurately. The reason is that, at finite amplitude, the modes that radiate most readily into the stratosphere from an unstable region in the troposphere are likely to be those freely propagating waves whose phase speeds and wavelengths match those of nonlinear disturbances resulting from instability in the troposphere.

Barotropic and baroclinic instability may also play important roles on a smaller scale in the middle atmosphere—for example, in leading to the apparent break-up of tongues of potential vorticity into "blobs," as mentioned at the end of Section 5.2.3 and depicted in Figs. 5.6b,c.

5.6 Planetary-Wave Critical Layers

It was mentioned in Section 4.5 that the theory of linear, steady, conserva-tive Rossby waves breaks down at "critical surfaces," on which the basic zonal wind $\bar{u}(\phi, z)$ matches the zonal phase speed c. This is because the factor $(\bar{u} - c)^{-1}$ in the wave equation [Eq. (4.5.9)] or, more generally, in the refractive index squared [Eq. (4.5.28)], becomes infinite there. The infinity is removed by the inclusion of further physical effects, namely, wave transience, dissipation, and nonlinearity (the effects that violate nonacceler-ation conditions), one or more of which must become important in a region called the "critical layer," which surrounds the critical surface. The detailed dynamical behavior within the critical layer depends on the relative impor-tance of these different effects, and this behavior in turn can have a crucial influence on the wave structure far from the critical surface. An understand-ing of critical-layer dynamics is thus likely to be of great importance for the study of planetary-wave propagation in the presence of critical surfaces, and much theoretical research is currently being devoted to the phenomenon.

A convenient idealized model of critical layer behavior uses beta-plane geometry and takes the flow to be barotropic (i.e., independent of z); in this case the critical surface reduces to a "critical line." We suppose that a steady zonal-mean flow $\bar{u}(y)$ of the form sketched in Fig. 5.12 supports
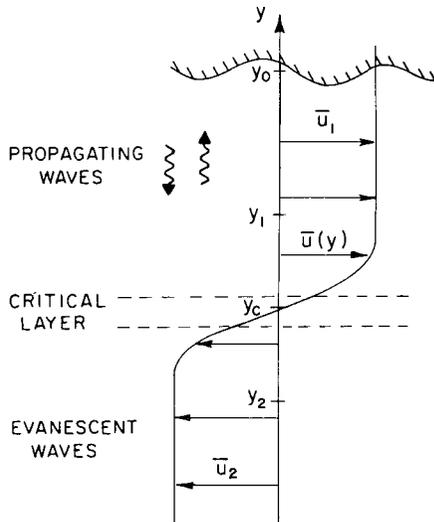

Fig. 5.12. Schèmatic diagram of a basic zonal shear flow $\bar{u}(y)$ containing Rossby waves of zero zonal phase speed forced by a corrugated northern boundary near $y = y_0$. See text for details.

small-amplitude Rossby waves of zonal wave number $k(>0)$ and zero zonal phase speed $(c = 0)$. For the purposes of the model, these waves can be regarded as generated, say, by the flow past a northern boundary near $y = y_0$ containing stationary sinusoidal corrugations. This configuration mimics aspects of the quasi-horizontal equatorward propagation of stationary planetary waves in the mid- and low-latitude stratosphere (Fig. 5.5). It is assumed that $\bar{u}$ is eastward and constant $(\bar{u} = \bar{u}_1 > 0)$ in the northern region $y_1 < y < y_0$; if the waves are essentially steady, linear, and conservative there, it can easily be shown by the methods of Section 4.5 that the disturbance stream function may be written in the form

$$\psi' = C \, \text{Re}[e^{ikx}(e^{-ily} + Re^{ily})] \qquad \text{for} \quad y_1 < y < y_0 \qquad (5.6.1)$$

in that region. Here C is a real constant, R is a complex constant, and $l = +(\beta \bar{u}_1^{-1} - k^2)^{1/2}$, which is real provided that $\bar{u}_1$ is chosen to be less than βk^{-2}. It is easy to verify that the northward Eliassen–Palm flux component [see Eq. (3.5.6)] is given by

$$F^{(y)} \equiv -\rho_0 \overline{v'u'} \equiv \rho_0 \overline{\psi'_x \psi'_y}$$

$$= -\tfrac{1}{2}\rho_0 C^2 kl(1 - |R|^2) \qquad \text{for} \quad y_1 < y < y_0. \qquad (5.6.2)$$

Equation (5.6.1) thus represents the superposition of a southward-propagating wave of streamfunction amplitude C and EP flux $-\tfrac{1}{2}\rho_0 C^2 kl$ and a northward-propagating wave of amplitude CR and EP flux $\tfrac{1}{2}\rho_0 C^2 |R|^2 kl$, which is "reflected" by the shear layer south of y_1.

In the region south of y_2 it is assumed that $\bar{u}$ is westward and constant $(\bar{u} = \bar{u}_2 < 0)$ and so

$$\psi' \propto e^{ikx + \Lambda y} \qquad \text{for} \quad y < y_2, \qquad (5.6.3)$$

where $\Lambda = (\beta |\bar{u}_2|^{-1} + k^2)^{1/2}$. This represents a disturbance that is evanescent with decreasing latitude; it can be verified that

$$F^{(y)} \equiv 0 \qquad \text{for} \quad y < y_2. \qquad (5.6.4)$$

Thus far, the theory is simple; the difficulties arise when the wave solutions are sought in the intervening "shear zone" between y_1 and y_2, which includes the critical line, $y = y_c$ say, where $\bar{u} = c \equiv 0$. The mathematical theory describing the dynamics of the waves in the shear zone, and in the critical layer in particular, is quite complicated, and we only quote the main results here. For simplicity, we suppose that the curvature of the $\bar{u}(y)$ profile is so small that partial reflections of the waves from outside the critical layer can be neglected. We concentrate on the ways in which the details of the flow in the critical layer affect the "reflection coefficient" R, and thus the amplitude of the reflected waves in $y > y_1$, far from the critical line.

It has long been known that if the waves are steady, linear, and dissipative in the critical layer (as for example in Matsuno's model of Section 5.3), the layer is perfectly absorbing under the hypotheses mentioned above, with $R = 0$. A similar result was shown by Dickinson (1970) and Warn and Warn (1976) to hold at moderately large times under the long-wave approximation $k(y_1 - y_2) \ll 1$ (specifically, for $1 \ll t^* \ll \alpha^{-1/2}$, where $t^* = k\bar{u}_1 t$ and α is a small dimensionless wave-amplitude parameter) for linear, conservative, transient waves that are "switched on" at some initial instant $t = 0$ and maintained thereafter. However, at larger times ($t^* \sim \alpha^{-1/2}$), nonlinear effects must become important in the critical layer [which then has a thickness $\sim \alpha^{1/2}(y_1 - y_2)$]. Stewartson (1978) and Warn and Warn (1978) showed that the conservative, nonlinear critical layer then oscillates between partial absorption ($|R|^2 < 1$), reflection ($|R|^2 = 1$), and over-reflection ($|R|^2 > 1$), tending to a state of perfect reflection ($|R|^2 = 1$) at still larger times ($t^* \gg \alpha^{-1/2}$). It has recently been shown that this flow is, in fact, barotropically unstable (Killworth and McIntyre, 1985; Haynes, 1985); while, at first sight, the resulting mixing in the critical layer might be expected to introduce enhanced dissipative effects, and thus to bring about absorption, Killworth and McIntyre demonstrate that absorption does not necessarily occur: under fairly general conditions the critical layer remains a perfect reflector in a time-integrated sense.

A partial physical description of the reflection process is as follows. In the Stewartson–Warn–Warn model the flow within the critical layer takes the form of "Kelvin's cats'-eyes." As shown schematically in Fig. 5.13, this flow tends to wrap up the contours of absolute vorticity $\zeta \equiv f - u_y + v_x$ [to which the quasi-geostrophic potential vorticity q reduces in this barotropic model: see Eq. (3.2.15)]. The barotropic analog of Eq. (3.5.10) is

$$\overline{v'\zeta'} = \rho_0^{-1} \frac{dF^{(y)}}{dy};$$

integrating this over the shear layer and using Eqs. (5.6.2) and (5.6.4) we obtain

$$\int_{y_1}^{y_2} \overline{v'\zeta'}\, dy = -\tfrac{1}{2}C^2 kl(1 - |R|^2). \tag{5.6.5}$$

Since nonacceleration conditions hold outside the critical layer, $\overline{v'\zeta'} \equiv 0$ there, and the left of Eq. (5.6.5) can be replaced by

$$\int_{\text{critical layer}} \overline{v'\zeta'}\, dy. \tag{5.6.6}$$

Examination of the sequence of diagrams in Fig. 5.13 shows that Eq. (5.6.6) oscillates between negative and positive values of continually decreasing
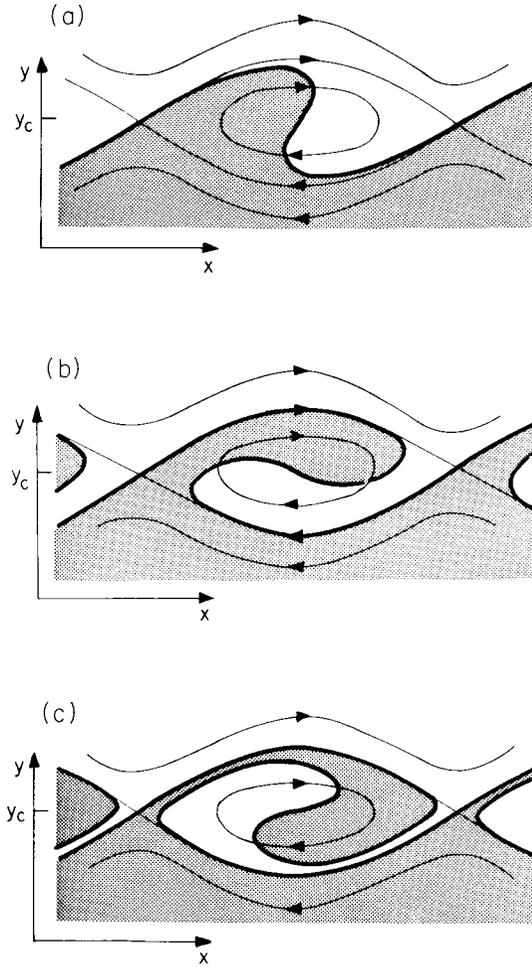
Fig. 5.13. Time-dependent analytical solution for a barotropic Rossby-wave nonlinear critical layer at non-dimensional times (a) $\alpha^{1/2}t^* = 2$, (b) $\alpha^{1/2}t^* = 4$, (c) $\alpha^{1/2}t^* = 6$ (Stewartson, 1978; Warn and Warn, 1978). The flow is periodic in x, and the y scale is greatly exaggerated: the initial critical line was at $y = y_c$. The thin lines indicate streamlines, and the "Kelvin's cats'-eyes" are the lens-shaped regions of closed streamlines; their width is of order $\alpha^{1/2}(y_1 - y_2)$. The thick line shows the successive positions of the material absolute vorticity contour, $\zeta = \zeta_c$ say, that initially lay along $y = y_c$. Thus $\zeta < \zeta_c$ in the stippled regions and $\zeta > \zeta_c$ in the unstippled regions. For this model, $\bar{v} \equiv 0$, so $\overline{v'\zeta'} = \overline{v'\zeta} = \overline{v'(\zeta - \zeta_c)} = \overline{v(\zeta - \zeta_c)}$. In (a) it can be seen that most of the stippled region within the cats'-eyes has $v > 0$, and most of the unstippled region has $v < 0$. Thus $\overline{v'\zeta'} = \overline{v(\zeta - \zeta_c)} < 0$ throughout most of the critical layer, Eq. (5.6.6) is negative, and $|R| < 1$ by Eq. (5.6.5), indicating partial absorption. Similar arguments show that in (b), $\overline{v'\zeta'} \approx 0$ and $|R| \approx 1$, indicating near-perfect reflection, and in (c) $\overline{v'\zeta'} > 0$ and $|R| > 1$, indicating overreflection. (Courtesy of P. H. Haynes.)

magnitude as the velocity field wraps up the vorticity contours more and more. Thus, by Eq. (5.6.5), $|R|^2$ oscillates between values less than and greater than 1, approaching 1 (perfect reflection) at large times. The picture is complicated by the fact that the Stewartson–Warn–Warn flow goes unstable, leading to complex small-scale features in the velocity and vorticity fields. Nevertheless, Killworth and McIntyre's general result shows that $|R|^2$ cannot depart systematically from 1 at large times in the model, even when this happens.

While great strides have recently been made in understanding these models of critical layers, the assumptions made in the theory (for example, that the waves are small-amplitude disturbances to an initially zonal flow) are of course still rather poor idealizations of actual atmospheric situations. Thus, the implications of the theory for the modeling of planetary waves and for the interpretation of atmospheric observations are not yet clear. It is, however, of interest that the time-dependent nonlinear theories predict behavior within the critical layer that bears some resemblance to the breaking-wave structures observed by McIntyre and Palmer (1983, 1984) in isentropic potential vorticity maps of the stratosphere (see Section 5.2.3). (The z-independent vertical absolute vorticity component in the barotropic theory plays the role of the potential vorticity in the more general case.) In particular, the theoretical flows exhibit the same kind of rapid and irreversible deformation of material contours (Fig. 5.13) as is suggested by the stratospheric maps of McIntyre and Palmer (Figs. 5.6b,c), and for this reason these authors propose that nonlinear critical-layer theory may model certain aspects of the observed breaking planetary wave events. [Rapid, irreversible deformation of material contours can also occur, for example, when vortices interact nonlinearly (see, e.g., Dritschel, 1986). Such flows, too, may turn out to provide useful idealizations of features of breaking planetary waves.]

References

5.2. Recent compilations of stratospheric monthly mean climatological data include those of Hamilton (1982a,b), Geller *et al.* (1983), and Wu *et al.* (1984). The most comprehensive collection is to be found in the 1986 COSPAR International Reference Atmosphere [Labitzke *et al.* (1985)]. Recent reviews of observed stationary and transient eddies in the *troposphere* are those of Wallace (1983) and Holopainen (1983), respectively.

5.3. The fundamentals of the theory and modeling of *tropospheric* stationary planetary waves are discussed by Held (1983). Developments of the Charney and Drazin semianalytical theory of stationary planetary waves in the stratosphere are presented by Dickinson (1968) and Simmons (1974), for example. Discussions of

the self-consistency of Eqs. (5.3.1)–(5.3.4) are given by Matsuno (1970) and Palmer (1982). The paper by Alpert *et al.* (1983) gives a detailed study of the sensitivity of a linear quasi-geostrophic stationary planetary-wave model to differing representations of the topographic and thermal forcing. Jacqmin and Lindzen (1985) present calculations of stationary planetary waves in the troposphere and lower stratosphere derived from a high-resolution, linear, primitive-equation model. The time-averaged nonlinear effects of transient eddies on stationary waves in the troposphere are considered by Hoskins (1983).

5.5. The proof given in Section 5.5.1 is due essentially to Bretherton (1966a), who used ideas originally developed by Taylor (1915). "Non-Doppler" effects are considered for example by White (1982). The proof of Fjørtoft's (1950) result using methods similar to those of this section is given by Blumen (1978). Examples of barotropically and baroclinically unstable modes, and thorough discussions of thei dynamics, are given by Pedlosky (1979) and Gill (1982), among others.

5.6. Killworth and McIntyre (1985) review the current understanding of the theory of planetary-wave critical layers and present some important new results.