

*And beyond it, the deep blue air, that shows
Nothing, and is nowhere, and is endless.
Philip Larkin, High Windows, 1974.*

CHAPTER 17

The Stratosphere

THE STRATOSPHERE IS THE REGION OF THE ATMOSPHERE above the troposphere and below the mesosphere; thus, it extends from the tropopause at a height of about 8–15 km, or a pressure of around 200–300 hPa, to the stratopause at about 50 km or about 1 hPa (see Fig. 15.24 on page 574). The *middle atmosphere* is the somewhat larger region that also includes the mesosphere, and so that extends up to the mesopause at about 90 km or 2×10^{-3} hPa, but we won't consider the mesosphere here. Our goal in this chapter is to provide an introduction to the dynamics giving rise to the structure and variability of the stratosphere.¹

The outline of this chapter is roughly as follows. We begin with a rather descriptive overview of the stratosphere as a whole. Then, starting in Section 17.2, we discuss the Rossby and gravity waves that in many ways serve to drive the circulation. We come back to the circulation itself in Section 17.4, focusing mainly on the generation of zonal flows and the meridional residual overturning circulation. We round out the chapter with discussions of two striking examples of stratospheric variability, namely the quasi-biennial oscillation in Section 17.6, and extratropical variability and sudden warmings in Section 17.7, with these terms to be defined in the sections ahead.

17.1 A DESCRIPTIVE OVERVIEW

In the troposphere the stratification is determined by dynamical processes — largely by convection at low latitudes and additionally by baroclinic instability at high latitudes — and the tropopause is the height to which the dynamical activity reaches, as discussed in Chapter 15. In contrast, in the stratosphere the temperature is determined to a much greater degree by radiative processes and the dynamics are, compared to those in the tropopause, slow. Over much of the stratosphere the temperature actually increases with height, and this is due to a layer of ozone that absorbs solar radiation in the mid-stratosphere between about 20 and 30 km. If there were no ozone we would certainly have a tropopause and a stratosphere, but the temperature in the stratosphere would increase much less with height than it in fact does.

The radiative-equilibrium temperature for January is illustrated in Fig. 17.1. This temperature is that which would putatively ensue without any stratospheric fluid motion, although we take the distribution of absorbers (such as ozone) to be those present in the actual, moving, atmosphere, and the calculation involves a linearization around the observed temperature.² There is quite a strong lateral gradient in the winter hemisphere and a weaker reversed temperature in the summer hemisphere, and in fact the part of the stratosphere with the highest radiative equilibrium

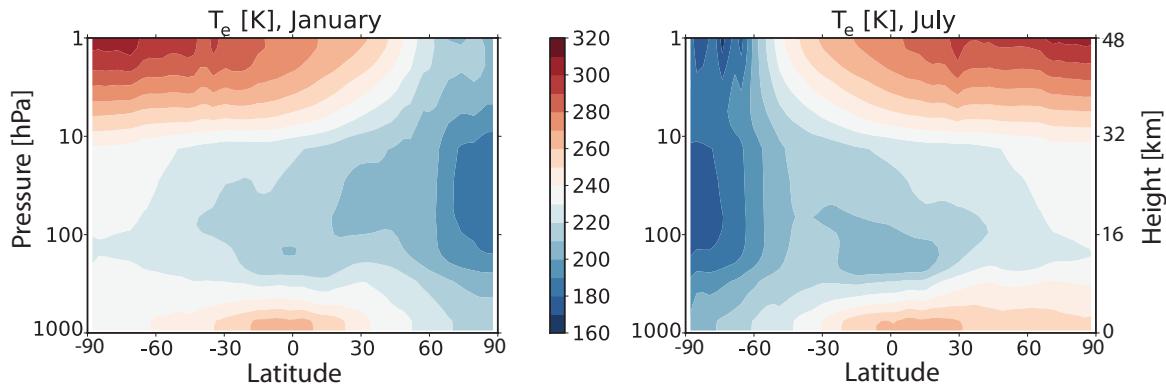


Fig. 17.1 The zonally averaged radiative-equilibrium temperature in January and July; that is, the temperature that would nominally arise in the absence of fluid motion in the stratosphere, but with the actual distribution of radiative absorbers. Above about 50 km the equilibrium temperature generally diminishes with height. The ordinate is pressure, and the height values are approximate.

temperature is the upper-stratosphere summer pole, at around 1 hPa. The actual observed zonally averaged temperature and zonal-wind structure are plotted in Fig. 17.2. From these figures we infer the following:

- The stratosphere is very stably stratified, with a typical lapse rate corresponding to $N \approx 2 \times 10^{-2}$ s, about twice that of the troposphere on average. This is in part due to the absorption of solar radiation by ozone between 20 and 50 km.
- In the summer the solar absorption at high latitudes leads to a reversed temperature gradient (warmer pole than equator) and, by thermal wind balance, a negative vertical shear of the zonal wind. The temperature distribution is not far from the radiative equilibrium distribution, and over much of the summer stratosphere the mean winds are negative (westward).
- In winter high latitudes receive very little solar radiation and there is a strong meridional temperature gradient and consequently a strong vertical shear in the zonal wind. Nevertheless, this temperature gradient is significantly weaker than the radiative equilibrium temperature gradient, implying a poleward heat transfer by the fluid motions.

How do the dynamics of the stratosphere differ from the troposphere? One way is that there is little, if any, baroclinic instability in the stratosphere — for various reasons. Suppose we first think of the stratosphere in isolation. There is no clear reversal of the potential vorticity gradient and no real opportunity for counter-propagating edge waves or Rossby waves to interact in the stratosphere, and hence stratosphere alone may simply be baroclinically stable. If the stratosphere were baroclinically unstable the instability would be much weaker, because of its higher stratification. A typical value of the static stability in the stratosphere is $N \approx 2 \times 10^{-2}$ s⁻¹, and using a height scale of 20 km gives a value of the deformation radius NH/f of about 4000 km, as opposed to the canonical value of 1000 km in the troposphere. (The stratospheric estimate is very approximate because the scale height, H_s , is much less than 20 km, and a relevant deformation radius is then $N\sqrt{HH_s}/f$. But on the other hand one could also take $H > 20$ km, so 4000 km may be a fair estimate.) Thus, even with the same horizontal temperature gradient as the troposphere, a typical instability scale (of the stratosphere in isolation) would be large, perhaps at wavenumber 2 rather than wavenumber 8. The stratospheric growth rate would then be much less than in the troposphere: the Eady growth rate is given by $\sigma_E \equiv 0.31\Lambda H/L_d = 0.31U/L_d$, where Λ is the shear, giving the growth rate that is several times smaller than its tropospheric counterpart. Of course, if baroclinic instability has a modal form then the instability has the same horizontal scale and grows at the same rate in

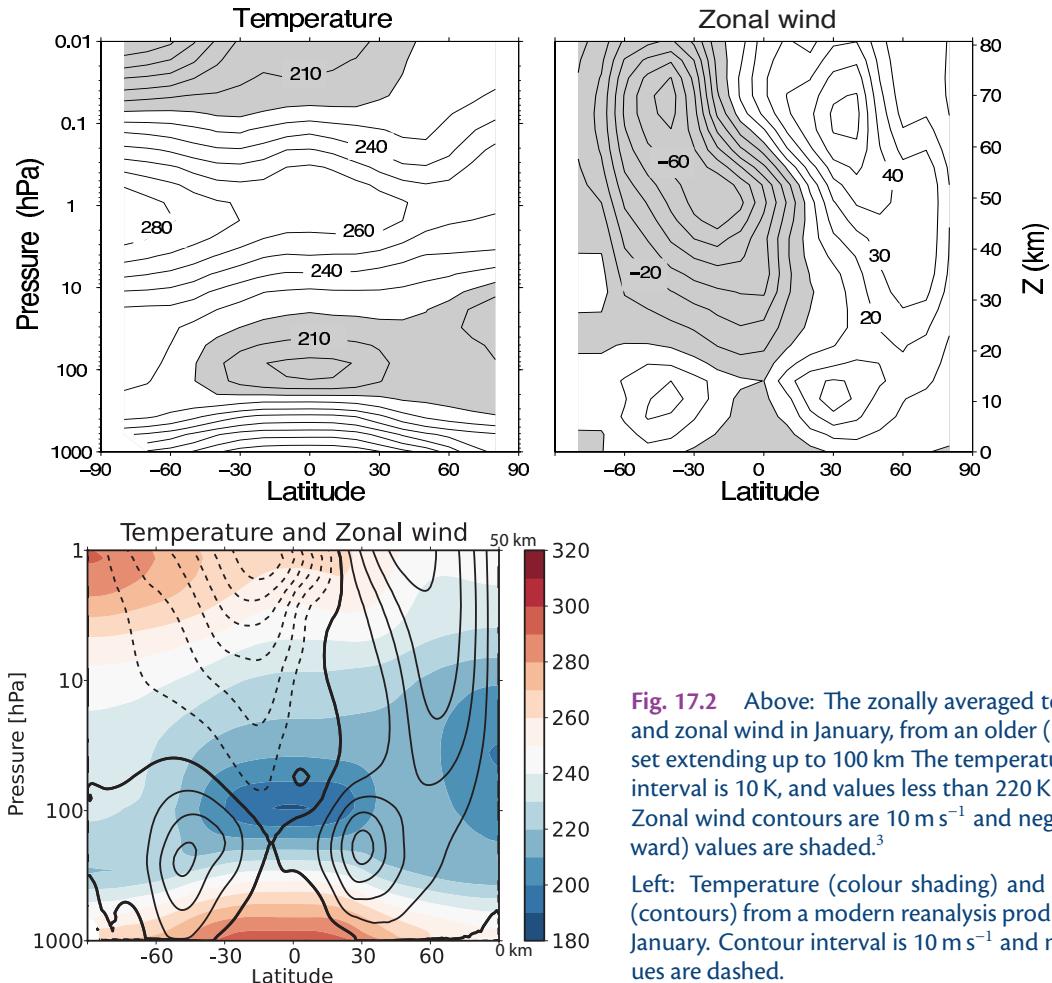


Fig. 17.2 Above: The zonally averaged temperature and zonal wind in January, from an older (1980s) data set extending up to 100 km. The temperature contour interval is 10 K, and values less than 220 K are shaded. Zonal wind contours are 10 m s⁻¹ and negative (westward) values are shaded.³

Left: Temperature (colour shading) and zonal wind (contours) from a modern reanalysis product, also for January. Contour interval is 10 m s⁻¹ and negative values are dashed.

the stratosphere as the tropospheric one — it is the same mode! But in this case the higher lapse rate suppresses the amplitude of the stratospheric instability, as shown in Fig. 9.21.

For all these reasons, baroclinic instability is not the main process leading to a circulation in the stratosphere — the main process is the propagation and subsequent breaking and dissipation of gravity and Rossby waves from the troposphere to the stratosphere. This breaking will produce an acceleration of the zonal flow, and/or a meridional overturning circulation, and a good fraction of this chapter will be devoted to describing that process. But first we'll provide a little more description about the circulation itself, and it is convenient to divide that into two parts:

- (i) a quasi-horizontal circulation;
- (ii) a meridional overturning circulation (MOC) that is most usefully described as a residual circulation (the residual meridional circulation, or RMOC) using the TEM formalism.

17.1.1 The Quasi-Horizontal Circulation

In the extra-tropics the stratification is high and the Rossby number small and, at least to the extent that the scales of motion are not truly hemispheric the circulation is well described by the quasi-geostrophic equations. Now, not only does any stratospheric baroclinic instability tend to occur on a large scale, but so does any wave activity that arises from the propagation of Rossby waves up from

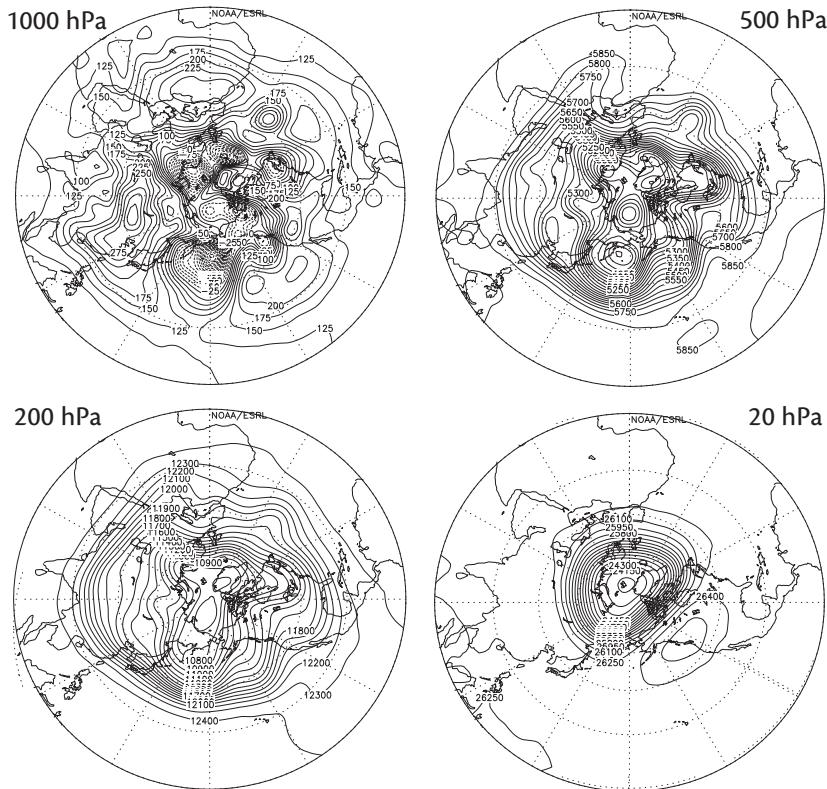


Fig. 17.3 The geopotential height on 1 February, 2000, at various levels in the atmosphere — 1000 and 500 hPa are in the troposphere, 200 hPa is around the tropopause and 20 hPa is in the mid-stratosphere, at about 30 km. Note the general increase in the scale of the variations of the geopotential with height.

the troposphere. This is because of Charney–Drazin filtering, summarized in Fig. 16.6: the smaller the wavelength the smaller is the range of zonal winds through which the waves can propagate. If the wind is too high the waves encounter a turning surface, whereas if the wind is too low they encounter a critical layer. Thus, we would expect that the general horizontal scale of motion is larger in the stratosphere than in the troposphere, and this is borne out by inspection of Fig. 17.3, which shows geopotential height at various levels. The complex patterns of the lower and mid-troposphere are well filtered, and in mid-stratosphere the pattern is dominated by wavenumbers 1 and 2. Indeed it seems from the figure (which is typical) that much of the motion is concentrated around a *polar vortex*.

Looking at geopotential (which roughly corresponds to a streamfunction) gives a somewhat misleading impression of the lack of activity away from the poles. Here, because diabatic effects occur on a rather longer time scale than advective processes, the flow may be characterized by the advection of potential vorticity on more slowly evolving isentropic surfaces, as illustrated in Figs. 17.4 and 17.5. Both the potential vorticity and the tracer are evocative of two-dimensional turbulence. We see Rossby waves breaking and vortices stretched into filaments and tendrils, the features of an enstrophy cascade. We also perceive some idea of the spectral non-locality of the enstrophy transfer — a single large vortex overturns and breaks and there is little sense of a spectrally-local cascade of enstrophy to dissipative scales. For this reason, the mid-latitude region is sometimes known as the *surf zone*. It is precisely this wave breaking that gives rise to the enstrophy flux to

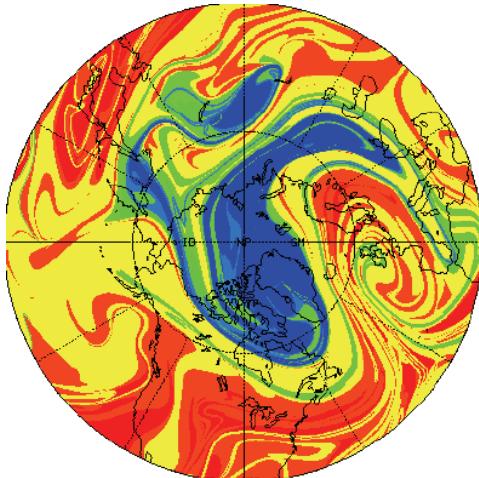


Fig. 17.4 The tracer distribution in the northern hemisphere lower stratosphere on 28 January 1992. The tracer was initialized on 16 January by setting it equal to the potential vorticity field calculated from an observational analysis, and then advected for 12 days by the observed wind fields.⁴

small scales and its dissipation, and which in turn gives rise to the overturning circulation that we discuss below.

The surf-zone does not usually extend to the pole, and in winter dense cold air over the pole forms itself into a cyclonic vortex, apparent in both Fig. 17.3 and 17.5. Although the vortex is ultimately the result of diabatic forcing, and has a preferred location, the tendency of quasi-two-dimensional flow to organize itself into vortices (as we see in Figs. 9.6 and 11.8) contributes to its coherence and isolation from the rest of the hemisphere. The boundary of the vortex, as measured by the value of the potential vorticity or of the tracer, is quite sharp with the value of PV often jumping by a factor of 2 or so, and the vortex is quite persistent — in fact it is a near-permanent feature of the winter hemisphere. Within the vortex potential vorticity tends to homogenize, and once formed the main communication that the vortex has with the surf zone is via occasional wave breaking at its boundary. It is interesting that, although the potential vorticity gradient is strong at the edge of the vortex, the exchange of properties is weak, implying a failure of notions of diffusion, or at least diffusion with a constant value of diffusivity; the edge of the vortex is a *mixing barrier*. We saw this property before, in our discussion of potential vorticity staircases in Section 12.1.3.

Stable as it is, the polar vortex is nevertheless sometimes disrupted by wave activity from below; this tends to occur when the wave activity itself is quite strong, and when the mean conditions are such as to steer that wave activity polewards. Occasionally, this activity is sufficiently strong so as to cause the vortex to break down, or to split into two smaller vortices, and so allow warm mid-latitude air to reach polar latitudes — an event known as a *stratospheric sudden warming*, and one such is illustrated in Fig. 17.22. We come back to the mechanism of such warmings later in the chapter.

17.1.2 The Overturning Circulation

That there is a meridional overturning circulation in the stratosphere was inferred by A. Brewer and G. Dobson based on observations of water vapour and chemical transport, and it is now often called the *Brewer–Dobson circulation*.⁷ Brewer and Dobson both inferred the circulation on the basis of tracer transports, rather than performing an Eulerian average of velocity measurements (which would have been impossible then, and is still difficult now). Thus, the circulation they inferred was, in modern parlance, a residual circulation and some modern observations of this circulation are shown in Fig. 17.6. The figure actually shows the observed thickness-weighted circulation, which is almost equivalent to the residual circulation (Section 10.3.3), and which represents both the Eulerian mean and eddy-contributed components. We see a single, equator-to-pole cell

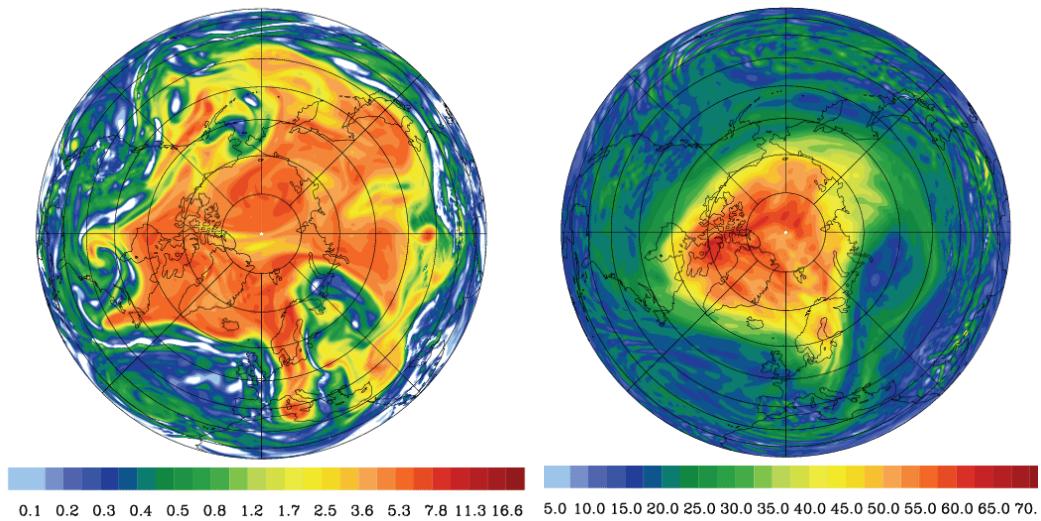


Fig. 17.5 The potential vorticity on two isentropic surfaces, the 310 K surface (left) and the 475 K surface (right), on 19 January, 2005. The shaded bar is in PV units. The 310 K surface is mainly in the troposphere (see Fig. 15.16) where baroclinic instability is abundant. The 475 K surface is at about 20 km altitude, and on it we see a polar stratospheric vortex with a fairly sharp boundary where the PV gradient is high, and a mid-latitude region of smaller-scale features and wave breaking.⁵

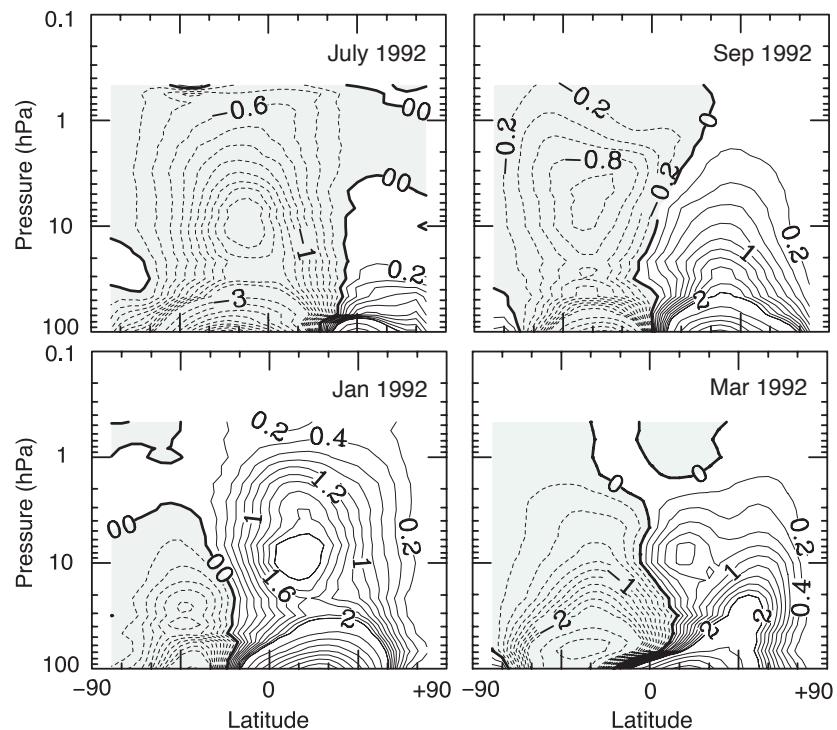


Fig. 17.6 The observed thickness-weighted (residual) streamfunction in the stratosphere, in Sverdrups (10^9 kg s^{-1}). The circulation is clockwise where the contours are solid.

The circulation is stronger in the winter hemispheres, whereas the equinoctial circulations (September, March) are more inter-hemispherically symmetric.⁶

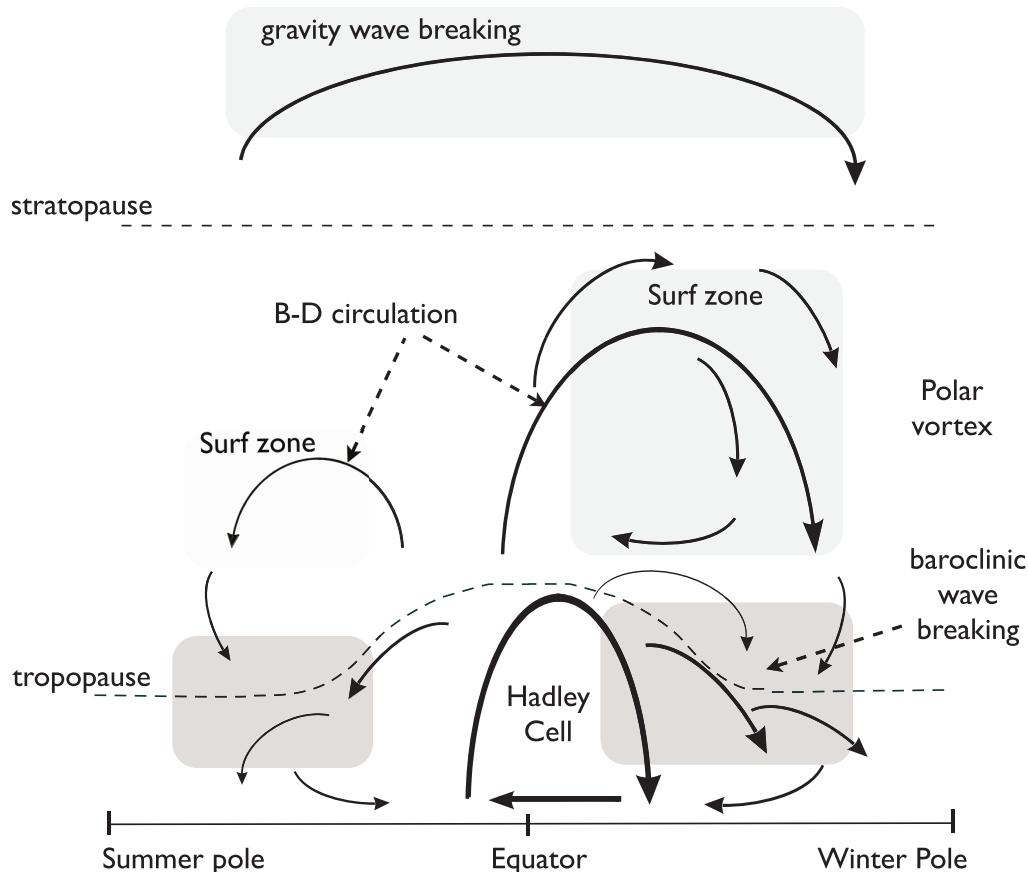


Fig. 17.7 A sketch of the residual mean meridional circulation of the atmosphere. The solid arrows indicate the residual circulation (B-D for Brewer–Dobson) and the shaded areas the main regions of wave breaking (i.e., enstrophy dissipation) associated with the circulation. In the surf zone the breaking is mainly that of planetary Rossby waves, and in the troposphere and lower stratosphere the breaking is that of baroclinic eddies. The surf zone and residual flow are much weaker in the summer hemisphere. Only in the Hadley Cell does the residual circulation consist mainly of the Eulerian mean; elsewhere the eddy component dominates.

in each hemisphere, stronger in the winter hemisphere where it goes high into the stratosphere. There is also a distinct lower branch to the circulation, present in all seasons although strongest in winter, that is confined to the lower stratosphere and is in some ways a vertical extension of (the residual circulation of) the tropospheric Ferrel Cell. Not all the upper circulation is ventilated by the troposphere — some of it recirculates within the stratosphere. This circulation and some of the associated dynamics are illustrated schematically in Fig. 17.7,⁸ and three regions may usefully be delineated: (i) a tropical region; (ii) a mid-latitude region; (iii) the polar vortex. The tropical region is relatively quiescent, an area of upward motion where air is drawn up from the troposphere. In mid-latitudes the residual flow is generally directed poleward before sinking at high latitudes. In winter the extreme cold leads to the formation of the polar vortex, a strong cyclonic vortex that appears quite isolated from mid-latitudes although, especially in the Northern Hemisphere, it is not always centred over the pole.

Let us now turn to the dynamics of the circulation, and since this is in large measure dependent on the waves that exist in the stratosphere we first discuss them.

17.2 WAVES IN THE STRATOSPHERE

Both gravity waves and Rossby waves are important in the stratosphere and we have already discussed aspects of both. Our goal now is to see how they affect the stratosphere, first in mid-latitudes and then in equatorial regions, where Rossby waves and gravity waves are intertwined.

17.2.1 Linear Equations of Motion

Because we are dealing explicitly with compressible atmosphere we will use the ideal gas equations with log-pressure coordinates, as discussed in Section 2.6.3. (Nevertheless we will often find that, to a decent approximation, the equations reduce to the Boussinesq form, with compressibility having only a small effect.) We will restrict attention to the hydrostatic case, thereby limiting ourselves to relatively large scales. On a β -plane the equations of motion in log-pressure coordinates, linearized about a resting state, may be written as

$$\frac{\partial u}{\partial t} - fv = -\frac{\partial \Phi}{\partial x}, \quad \frac{\partial v}{\partial t} + fu = -\frac{\partial \Phi}{\partial y}, \quad (17.1a,b)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_R} \frac{\partial(\rho_R w)}{\partial z} = 0, \quad \frac{\partial}{\partial t} \frac{\partial \Phi}{\partial z} + w N_*^2 = 0. \quad (17.1c,d)$$

The equations are, respectively, the two horizontal momentum equations, the mass continuity equation and the thermodynamic equation. The notation is as in Section 2.6.3 except we use z , not Z , and w , not ω ; thus, $z = -H \ln(p/p_R)$ where p_R is a constant reference pressure and H is a reference height, and $w = Dz/Dt$. As usual u and v are the horizontal velocities and Φ is the geopotential. The density profile ρ_R is an exponential, $\rho_R(z) = \rho_0 \exp(-z/H)$, where we may take $\rho_0 = 1$, and N_*^2 is a reference stratification parameter similar to but not exactly the same as the buoyancy frequency; we will take it to be constant and drop the subscript *. The Coriolis parameter f varies as $f = f_0 + \beta y$; when we consider equatorial waves we will take $f_0 = 0$, and when we consider gravity waves in mid-latitudes we will take $\beta = 0$.

As in Section 16.5 it is convenient to extract that part of the solution that grows exponentially with height, and so seek wave solutions of the form

$$[u, v, w, \Phi] = [\tilde{u}(y), \tilde{v}(y), \tilde{w}(y), \tilde{\Phi}(y)] e^{z/2H} e^{i(kx+mz-\omega t)}. \quad (17.2)$$

We cannot assume a simple harmonic form in the y -direction because the equations of motion have coefficients (i.e., f) that depend on y . Substituting (17.2) into (17.1) yields

$$-i\omega\tilde{u} - f\tilde{v} = -ik\tilde{\Phi}, \quad -i\omega\tilde{v} + f\tilde{u} = -\frac{\partial \tilde{\Phi}}{\partial y}, \quad (17.3a)$$

$$ik\tilde{u} + \frac{\partial \tilde{v}}{\partial y} + i \left(m + \frac{i}{2H} \right) \tilde{w} = 0, \quad -i\omega \left(\frac{1}{2H} + im \right) \tilde{\Phi} + \tilde{w} N_*^2 = 0. \quad (17.3b)$$

Perhaps surprisingly, in many situations we can ignore the factor $1/2H$ in this system. Many observed stratospheric waves have a vertical wavelength, λ , that is of order a few kilometres and usually less than 10 km. Also, $T_0 = 240$ K then $H \approx 7$ km. The $1/2H$ factor is small when $m \gg 1/2H$ or $4\pi H/\lambda \gg 1$. In fact, it is the square of this ratio that needs to be large, and this is true for all but the deepest stratospheric waves. Compressibility remains in the system because, using (17.2), all the perturbation variables grow exponentially with height, albeit slowly.

In the sections that follow we look at some of the waves supported by this system. The analysis is more complicated for equatorial regions, where gravity and Rossby waves are intertwined, so we begin with the mid-latitudes.

17.2.2 Waves in Mid-Latitudes

In mid-latitudes there is a good frequency separation between Rossby waves and gravity waves so they can be treated separately, and since we have already treated both in Chapters 7 and 16 our discussion focuses on their stratospheric relevance.

Rossby waves

If we neglect the factor of $1/H^2$ the x - and z -components of the group velocity are

$$c_g^x = \frac{(k^2 - l^2 - f_0^2 m^2/N^2)\beta}{(k^2 + l^2 + f_0^2 m^2/N^2)^2}, \quad c_g^z = \frac{2km\beta f_0^2/N^2}{(k^2 + l^2 + f_0^2 m^2/N^2)^2}. \quad (17.4a,b)$$

Since $k < 0$ for Rossby waves then, in order for the waves to be upwardly propagating, (17.4b) requires that $m < 0$. Thus, the lines of constant phase tilt westward with height. The ratio of the vertical to the horizontal components of group velocity is not, unlike the case with gravity waves, a simple function of the wavenumbers and it is not possible to determine whether c_g^z is positive or negative without knowing the value of l , the meridional wavenumber. To obtain a typical value of the vertical group velocity in the atmosphere we may take $k^{-1} = 1000$ km, $m^{-1} = 10$ km, $f_0/N = 10^{-2}$, $\beta = 10^{-11}$ m $^{-1}$ s $^{-1}$, giving $c_g^z \sim 0.1$ m s $^{-1} \approx 10$ km/day.

Because Rossby waves grow in amplitude as they ascend the linearity assumption will eventually fail and the waves may break and dissipate, and in doing so they will deposit momentum. They may also break and/or dissipate if they encounter a critical line. This deposition is responsible for the production of the stratospheric meridional overturning circulation that we discuss later.

Gravity waves

Gravity waves may also propagate up into the stratosphere from the troposphere. If we take f to be a constant, f_0 , then, except for the factor of ρ_R the set (17.1) is the same as the f -plane hydrostatic Boussinesq equations, namely (7.144) on page 280. It is a straightforward matter to show that the dispersion relation is

$$\omega^2 = \frac{f^2 m'^2 + (k^2 + l^2)N^2}{m'^2} = f^2 + \frac{N^2(k^2 + l^2)}{m'^2}, \quad (17.5)$$

where $m'^2 = m^2 + 1/4H^2$. As noted above the factor of $1/4H^2$ is often small and we shall ignore it. If we suppose that the horizontal component of the wave vector is aligned with the x -axis (i.e. $l = 0$) then the group velocity components are

$$c_g^x = \frac{N^2 k}{\omega m} = \frac{N^2}{\omega m} \cos \vartheta, \quad c_g^z = -\frac{N^2(k^2 + l^2)m}{\omega m^4} = \frac{-N^2}{\omega m} \cos^2 \vartheta, \quad (17.6a,b)$$

where $\cos^2 \vartheta = k^2/(k^2 + m^2) \approx k^2/m^2 \ll 1$. The above expressions are most easily derived directly from (17.5) but are also the hydrostatic limit of the full expression (7.140). The directional aspects of these expressions are the same as those given for the non-rotating case in (7.74) with $\sin \vartheta = 1$, consistent with the hydrostatic limit — indeed we can obtain (17.6) from (7.74) by setting $\cos \lambda = 1$, $\sin \vartheta = 1$, $\omega = N \cos \vartheta$ and $\kappa = m$. Thus, the relation of the group velocity to the phase speed is much the same as for the gravity waves considered in Chapter 7, and in particular we have $c_g^z/c_g^x = -k/m$. If the waves are generated in the troposphere then c_g^z must be positive and so m must be negative. In mid-latitudes there is however no requirement that the horizontal propagation be in any particular direction. The distributions of velocity, pressure and temperature are illustrated in the two panels of Fig. 17.8 for waves with a positive and negative horizontal wavenumber and a negative vertical wavenumber.

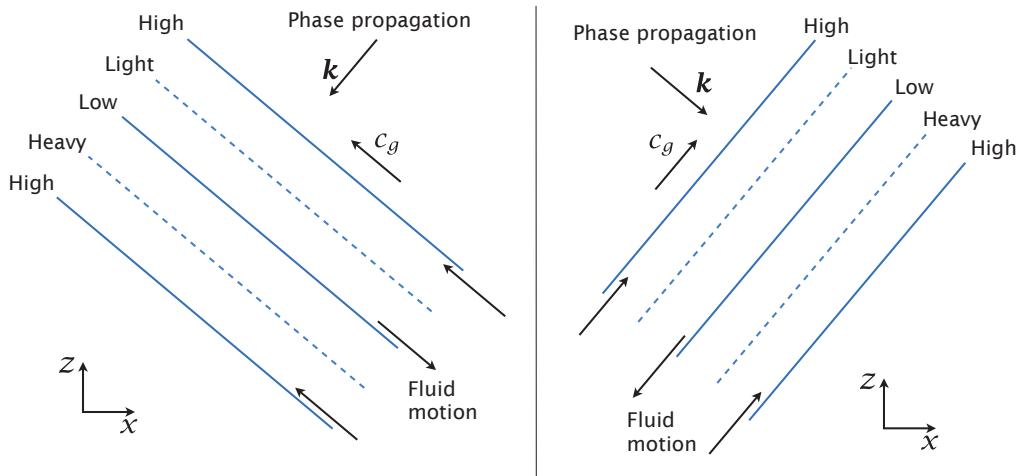


Fig. 17.8 Phase relationships for two examples of upwardly propagating gravity waves. The sketch on the left shows waves propagating to the left, with $k < 0$, and the one on the right shows waves with $k > 0$. The solid and dashed lines are contours of constant phase: ‘high’ and ‘low’ refer to pressure and ‘light’ and ‘heavy’ refer to density and correspond to warm and cold, respectively. In both sketches m is negative and the group velocity is directed upward and phase propagates downward. For hydrostatic flow the phase lines would be nearly horizontal. The figure may be compared with Fig. 7.3.

17.2.3 Waves in the Equatorial Stratosphere

Equatorial waves in the stratosphere — Kelvin waves and Rossby waves — turn out to be particularly important in generating stratospheric variability. It transpires that there can be vertically propagating waves with both eastward and westward phase speeds, even at relatively low frequencies, and this difference in phase speed has important consequences. We first look at Kelvin waves, for these provide a gentle introduction via a special treatment, and follow this by a more general treatment that includes Rossby and gravity waves.

Kelvin waves

We obtain the Kelvin wave solution by setting $\tilde{v} = 0$ everywhere in (17.3), whence, after eliminating \tilde{w} , (17.3) straightforwardly becomes

$$\omega\tilde{u} = k\tilde{\Phi}, \quad f\tilde{u} = -\frac{\partial\tilde{\Phi}}{\partial y}, \quad \omega\left(m^2 + \frac{1}{4H^2}\right)\tilde{\Phi} - N^2k\tilde{u} = 0. \quad (17.7a,b,c)$$

Equations (17.7a,b) give $\omega\partial\tilde{u}/\partial y + kf\tilde{u} = 0$, which upon integration and with $f = \beta y$ yields

$$\tilde{u}(y) = \tilde{u}_0 e^{-\beta y^2/2c_p}, \quad (17.8)$$

where $c_p = \omega/k$ and \tilde{u}_0 is the value of \tilde{u} at the equator. The exponential fall-off is familiar from our earlier studies of Kelvin waves in Chapter 8 and requires that $c_p > 0$, meaning that the phase speed of the waves is eastward. Also, since $\omega > 0$ by convention, the x -wavenumber is positive (i.e., $k > 0$). The dispersion relation for Kelvin waves follows easily from (17.7a,c) and is

$$\omega^2 = \frac{N^2 k^2}{m^2 + 1/4H^2}. \quad (17.9)$$

Aside from the factor of $1/4H^2$, which in any case is often small compared to m^2 , (17.9) is essentially the same as the dispersion relation for hydrostatic gravity waves, namely (7.60) on page 261. The zonal and vertical components of the group velocity are

$$c_g^x = \frac{N}{(m^2 + 1/4H^2)^{1/2}}, \quad c_g^z = \frac{\partial \omega}{\partial m} = \frac{-Nkm}{(m^2 + 1/4H^2)^{3/2}}. \quad (17.10)$$

Now, for upwardly propagating waves (and so for waves that emanate from the troposphere) we require $c_g^z > 0$ and therefore (because $k > 0$) $m < 0$. The combined conditions of $k > 0$ and $m < 0$ mean that the phase lines tilt eastward with height, as in the right panel of Fig. 17.8. Finally, note that the frequency of Kelvin waves, unlike inertia-gravity waves, is uninfluenced by rotation and thus, as seen in Fig. 8.6, can extend over a broad range.

A more general treatment of equatorial waves

For simplicity let us assume that the scale height H is very large compared to the vertical wavelengths of interest; that is, $m^2 \gg 1/H^2$ and $\exp(-z/H) = 1$. Eqs. (17.1b) and (17.1c) then combine to give

$$\frac{\partial}{\partial t} \frac{\partial^2 \Phi}{\partial z^2} - N^2 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0. \quad (17.11)$$

If we assume a vertical structure of the form $\Phi(x, y, z, t) = \tilde{\Phi}(x, y, t) \exp(imz)$, and similarly for u and v , then we obtain

$$\frac{\partial \tilde{\Phi}}{\partial t} + \frac{N^2}{m^2} \left(\frac{\partial \tilde{u}}{\partial x} + \frac{\partial \tilde{v}}{\partial y} \right) = 0, \quad (17.12a)$$

with corresponding momentum equations

$$\frac{\partial \tilde{u}}{\partial t} - f\tilde{v} = -\frac{\partial \tilde{\Phi}}{\partial x}, \quad \frac{\partial \tilde{v}}{\partial t} + f\tilde{u} = -\frac{\partial \tilde{\Phi}}{\partial y}. \quad (17.12b,c)$$

Evidently, (17.12) are isomorphic to the linear shallow water equations (8.20) on page 304 with the replacement $c^2 = N^2/m^2$, where in (8.20) $c^2 \equiv gH_e$ where H_e is the equivalent depth. (Note that c is not necessarily the phase speed in this problem; we will denote that as c_p .) The correspondence between the continuously-stratified and shallow water equations is a general one, as we found in Section 3.4. All of the machinery following (8.20) can now be applied to (17.12), with Rossby-gravity waves and Kelvin waves emerging in the same way (and the Kelvin waves identified above emerge as a special case); thus, in what follows we draw directly from Section 8.2. We will use m to denote the vertical wavenumber and n to denote the order of the Hermite function, akin to a meridional wavenumber.

Rossby-gravity waves

The dispersion relation that emerges from (17.12) is, by analogy with (8.37b), (8.53) and (8.63),

$$\omega^2 - c^2 k^2 - \beta \frac{kc^2}{\omega} = (2n+1)\beta c, \quad n > 0, \quad (17.13a)$$

$$\omega^2 - wkc - \beta c = 0, \quad n = 0, \quad (17.13b)$$

$$\omega = ck, \quad n = -1, \quad (17.13c)$$

where $c = N/m$. The case with $n = 0$ is the Yanai wave (a Rossby-gravity wave) and the cases with $n \geq 1$ are planetary waves or gravity waves. The ' $n = -1$ ' case is the Kelvin wave, and all cases are illustrated in Fig. 8.6.

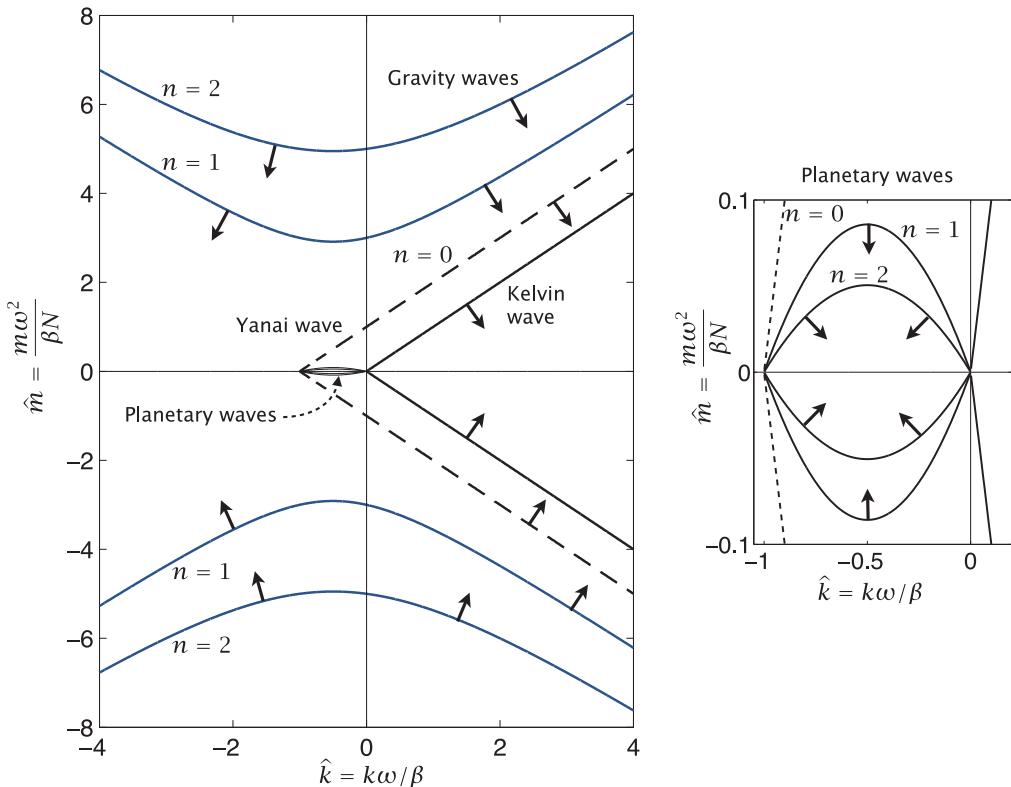


Fig. 17.9 Dispersion curves plotted in $\hat{m}-\hat{k}$ space, using (17.15). Shown are gravity waves, the Yanai wave and the Kelvin wave for positive and negative \hat{m} , with the plot at the right showing a magnification of the region near the origin. The arrows in the figure indicate the group velocity, which, being the gradient of the frequency, is perpendicular to the curves. Upward propagating waves occur for negative m . Compare with Fig. 8.6.

For waves whose origin is in the troposphere we may think of the frequency as being given and (17.13) then provides a condition on the vertical wavenumber. This in turn suggests that we nondimensionalize the wavenumbers by defining

$$\hat{m} = \frac{\omega^2}{\beta N} m = \frac{\omega^2}{\beta c}, \quad \hat{k} = \frac{\omega}{\beta} k, \quad (17.14)$$

with the hats denoting nondimensional variables. Equations (17.13) become

$$\hat{m}^2 - (2n+1)\hat{m} - \hat{k}^2 - \hat{k} = 0, \quad n > 0, \quad (17.15a)$$

$$\hat{m} - \hat{k} - 1 = 0, \quad n = 0 \quad (17.15b)$$

$$\hat{m} = \hat{k}, \quad n = -1. \quad (17.15c)$$

These equations define a set of curves in $\hat{m}-\hat{k}$ space that are similar to the curves in $\omega-k$ space defined by (17.13), although (17.15) are lower order. Equation (17.15a) has the solution

$$\hat{m} = \left(n + \frac{1}{2}\right) \pm \left[\left(\hat{k} + \frac{1}{2}\right)^2 + n(n+1)\right]^{1/2}, \quad (17.16)$$

and the complete set of curves is plotted in Fig. 17.9. The curves at the top and bottom are gravity waves, corresponding to the positive sign in (17.16), and the planetary waves are the curves just to

the left of the origin, corresponding to the negative sign. The $n = 0$ curve (the Yanai wave) and the $n = -1$ curve (the Kelvin wave) are labelled.

We can infer the group velocity from the figure by noting that, since β and N are constant, and using (17.14), the curves are contours of constant frequency. The group velocity is the gradient of frequency in wavenumber space and so is at right angles to these curves and directed toward a higher frequency, and is marked by short arrows. For waves propagating up from the troposphere the group velocity must be upward, and therefore have negative m , as can be seen from (17.6) and (17.10).

17.3 WAVE MOMENTUM TRANSPORT AND DEPOSITION

Steady waves by themselves don't affect the mean flow. Rather, they affect the mean flow when they are generated or dissipated, and typically they are generated in the troposphere and dissipated in the stratosphere. Let's look into that, beginning with Rossby waves.

17.3.1 Rossby Waves

The vertical transport of zonal momentum by eddies is given by $\overline{w'u'}$, where an overbar denotes a zonal average. However, directly evaluating this expression from the quasi-geostrophic equations is not particularly simple because w is not a first order variable — it results from the divergence of the ageostrophic horizontal velocities.⁹ In fact, under quasi-geostrophic scaling we neglect the vertical eddy flux divergences, but nevertheless the eddy fluxes may certainly make themselves felt aloft, among other things by generating a form stress that acts to transfer momentum vertically and generating a meridional overturning circulation, as discussed in Section 10.4.3.

To proceed we will use the transformed Eulerian mean (TEM) framework of Section 10.3, for which the inviscid and adiabatic zonally-averaged momentum and thermodynamic equations are

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \overline{v'q'}, \quad \frac{\partial \overline{b}}{\partial t} + N^2 \overline{w}^* = 0, \quad (17.17a,b)$$

where potential vorticity flux is related to the EP flux, \mathcal{F} , by

$$\overline{v'q'} = \nabla \cdot \mathcal{F}, \quad \mathcal{F} = -\overline{u'v'} \mathbf{j} + \frac{f_0}{N^2} \overline{v'b'} \mathbf{k}, \quad (17.18)$$

and now for simplicity we use the Boussinesq equations with b as buoyancy. If the eddy fluxes are due to the presence of waves that satisfy a dispersion relation then, as shown in Section 10.2.2, the EP flux is related to the group velocity by

$$\mathcal{F} = (\mathcal{F}^y, \mathcal{F}^z) = c_g \mathcal{P}. \quad (17.19)$$

Combining the above equations we obtain

$$\frac{\partial \overline{u}}{\partial t} - f_0 \overline{v}^* = \frac{\partial}{\partial y} (c_g^y \mathcal{P}) + \frac{\partial}{\partial z} (c_g^z \mathcal{P}) = \nabla \cdot \mathcal{F}. \quad (17.20)$$

Now, repring (10.29a), the wave activity, specifically the pseudomomentum, \mathcal{P} , satisfies a conservation law of the form

$$\frac{\partial \mathcal{P}}{\partial t} + \nabla \cdot \mathcal{F} = \mathcal{D}, \quad (17.21)$$

where $\mathcal{P} = \overline{q'^2}/2\beta$, which is a positive quantity, and \mathcal{D} represents dissipation. If $\mathcal{D} = 0$ and the waves are steady then $\nabla \cdot \mathcal{F} = 0$ so that the left-hand side of (17.20) is zero. (This is the non-acceleration result of Section 10.4.2.) Evidently in order to get a zonal flow acceleration or an MOC we need to invoke some dissipative or time-dependent processes.

Consider the case in which Rossby waves propagate up from the troposphere, with $c_g^z > 0$. Suppose that there is some dissipation in the system (and/or that Rossby waves break as they ascend) and that wave activity \mathcal{P} falls with height, and suppose further that we are in a statistically steady state. In this case $\nabla \cdot \mathbf{F} < 0$ and from (17.17a) this will produce a mean flow deceleration (i.e., a westward acceleration) and/or a polewards residual flow. The balance between these two possibilities is discussed later, but one may intuit that close to the equator the more likely outcome is a zonal acceleration rather than a meridional circulation. Why is there a preferred sense of acceleration when the waves break? Ultimately it is because of the beta effect which distinguishes east from west, and the pseudomomentum \mathcal{P} is proportional to beta. The beta effect leads to a particular orientation of the phase of Rossby waves and this carries westward momentum away from the source region. When the waves break, be it in the tropospheric subtropics or in the stratosphere, a westward momentum is deposited.

17.3.2 Gravity and Kelvin waves

Now consider the vertical momentum transport in gravity waves. If these are uninfluenced by rotation, or if they reside on the f -plane, then there is no preferred horizontal direction of propagation. Kelvin waves, on the other hand, propagate their phase eastward only. The upward transport of momentum from waves originating in the troposphere can occur only for waves with a positive group velocity, and thus for either of the examples illustrated in Fig. 17.8. Those waves that propagate phase westward (i.e., have $k < 0$) have $\overline{u'w'} < 0$, and those that propagate phase eastward, such as equatorial Kelvin waves, have $\overline{u'w'} > 0$. The contribution to the zonal flow acceleration by the wave transport is given by

$$\frac{\partial \bar{u}}{\partial t} = -\frac{\partial}{\partial z} \overline{u'w'} + \text{other terms.} \quad (17.22)$$

and if the amplitude of the waves stays constant with height then no mean flow acceleration is induced. However, if the amplitude diminishes with height, because of dissipative processes, then the waves that have a westward (eastward) phase propagation will cause the zonal flow to accelerate westward (eastward). Thus *the dissipation of Kelvin waves as they propagate vertically will cause an eastward acceleration of the zonal flow*.

17.3.3 ♦ Processes of Wave Attenuation

We now explicitly consider the dissipation of waves and the associated momentum deposition as gravity waves propagate vertically. (The reader may wish to skim Section 16.3 before proceeding, for there we consider similar but algebraically simpler problems.) To keep the algebra manageable we will consider the propagation of two-dimensional (x - z) gravity waves in a Boussinesq fluid uninfluenced by rotation. The momentum and buoyancy equations, linearized about a zonal flow $U(z)$ and constant stratification N^2 , are

$$\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} = -\frac{\partial \phi}{\partial x}, \quad \frac{\partial w}{\partial t} + U \frac{\partial w}{\partial x} = -\frac{\partial \phi}{\partial z} + b, \quad (17.23a)$$

$$\frac{\partial b}{\partial t} + U \frac{\partial b}{\partial x} + w N^2 = -\alpha b. \quad (17.23b)$$

We include a damping term, $-\alpha b$, where α is a constant, in the buoyancy equation but neglect viscous effects in the momentum equation. If we cross-differentiate the momentum equation and use the mass continuity equation ($\partial u / \partial x + \partial w / \partial z = 0$) we obtain the linear vorticity equation

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} \right) \nabla^2 \psi + \frac{\partial \psi}{\partial x} \frac{d^2 U}{dz^2} = \frac{\partial b}{\partial x}, \quad (17.23c)$$

where ψ is such that $w = \partial\psi/\partial x$ and $u = -\partial\psi/\partial z$. We seek solutions of (17.23) in the form

$$[\psi, b] = [\tilde{\psi}(z), \tilde{b}(z)] e^{ik(x-ct)}, \quad (17.24)$$

and a little algebra reveals

$$i(-kc + Uk) \left(-k^2 \psi + \frac{d^2 \tilde{\psi}}{dz^2} \right) + ik\psi \frac{d^2 U}{dz^2} = ik\tilde{b}, \quad (17.25a)$$

$$i(-kc + Uk)\tilde{b} + ikN^2 \tilde{\psi} = -\alpha\tilde{b}. \quad (17.25b)$$

These last two equations combine to give

$$\frac{d^2 \tilde{\psi}}{dz^2} + m^2(z) \tilde{\psi} = 0, \quad m^2(z) = \left[\frac{N^2 [1 + i\alpha/k(U - c)]}{(U - c)^2 + \alpha^2/k^2} - k^2 - \frac{d^2 U/dz^2}{(U - c)} \right]. \quad (17.26)$$

This is an equation for the vertical structure of the streamfunction. If U and N^2 were constant in (17.26) then m would be constant and its real part would be the vertical wavenumber. There is an imaginary component to m which can be expected to cause the solution to decay in the vertical. In most circumstances the decay is slow but in the neighbourhood of a critical line where $c = U$ then the decay will be rapid. The wave can be expected to deposit momentum and accelerate or decelerate the mean flow, depending on the direction of the phase propagation of the wave. When $\alpha = 0$ there is no dissipation and no deposition (except possibly at the critical line itself). Although the equation seems complicated, we can proceed to a solution if we make some reasonable simplifying assumptions:

- (i) We consider (as is realistic) low aspect ratio flows ($L_z/L_x \ll 1$) so that the factor of k^2 is small compared to m^2 and may be neglected.
- (ii) We assume that the variation of the mean flow occurs on a long vertical scale compared to m , and so neglect the term in d^2U/dz^2 . This assumption also allows us to use WKB methods.
- (iii) We assume dissipation is small, and in particular that $\alpha/k(U - c) \ll 1$.

With these approximations (17.26b) becomes

$$m(z) \approx \left[\frac{N^2 [1 + i\alpha/k(U - c)]}{(U - c)^2 + \alpha^2/k^2} \right]^{1/2} \approx \frac{N}{U - c} \left[1 + \frac{i\alpha}{2k(U - c)} \right], \quad (17.27)$$

and we can proceed with a WKB solution.

WKB solution and momentum flux

The WKB solution to (17.26) is

$$\tilde{\psi}(z) = Am^{-1/2} \exp \left(\pm i \int^z m dz' \right), \quad (17.28)$$

where A is a constant. The wave momentum flux, F , associated with the wave is

$$F_k(z) = \overline{u'w'} = -ik \left(\tilde{\psi} \frac{\partial \tilde{\psi}^*}{\partial z} - \tilde{\psi}^* \frac{\partial \tilde{\psi}}{\partial z} \right), \quad (17.29)$$

where the overbar denotes a zonal average and the right-hand side is always real, and the subscript on F indicates we are considering the effects of a single wave of wavenumber k .

Now, the fact that m varies only slowly with z (specifically $m^2 \gg |dm/dz|$) means that when we take the vertical derivative of $\tilde{\psi}$ we can ignore the derivative of the vertical derivative of the amplitude, $Am^{-1/2}$. Given this, and using (17.28) and (17.27) in (17.29) we obtain

$$F_k(z) = F_0 \exp \left(i \int_0^z (m - m^*) dz' \right) = F_0 \exp \left(\int_0^z \frac{-N\alpha}{k(U - c)^2} dz' \right). \quad (17.30)$$

where F_0 is the value of the flux at $z = 0$ and we have chosen the sign in the exponent to be appropriate for upwardly propagating gravity waves. The integrand in the right-most expression is the attenuation rate of the wave and, referring to (7.80) on page 266, it can be written as

$$\text{Attenuation rate} = \frac{\alpha}{k(U - c)^2/N} = \frac{\text{Dissipation rate}}{\text{Vertical group velocity}}. \quad (17.31)$$

As $U - c$ diminishes the group velocity falls, giving the dissipative processes more time to act. The result of (17.31) is a general one; we found an almost identical result when looking at the absorption of Rossby waves in Section 16.3 — see (16.49). (The dissipation rate in expressions like (16.49) and (17.31) is that of wave activity, which in the gravity wave case here equals the thermal dissipation rate.)

Effect on the mean flow

If a wave propagating upward is attenuated there will be a divergence in the eddy momentum flux associated with that wave. In particular, if α in (17.31) is non-zero then momentum flux deposition will increase rapidly as a critical layer is approached, $\partial F_k / \partial z$ will be non-zero and the zonal mean flow will be accelerated or decelerated. For definiteness, consider a wave propagating upward with a positive phase speed (so $m < 0$ and $k > 0$). From (17.30) F_k diminishes with height and the mean flow is accelerated eastward. Similarly, absorption of a wave of negative phase speed leads to a negative, or westward mean-flow acceleration. It is not inconceivable to imagine that the wave deposition will affect the mean flow to an extent that the position of the deposition is significantly altered, leading to interesting dynamical behaviour. Indeed this is precisely what happens in the quasi-biennial oscillation of the equatorial stratosphere. But before we discuss variability let us discuss the maintenance of the mean state.

17.4 PHENOMENOLOGY OF THE RESIDUAL OVERTURNING CIRCULATION

We now return to a discussion of the general circulation of the stratosphere and in particular the maintenance of the residual meridional overturning circulation (RMOC), or the Brewer–Dobson circulation. We expect the circulation to be a consequence of waves coming up from the troposphere and breaking, with both tropospheric baroclinic instability and flow over thermal and topographic zonal asymmetries being sources of wave activity. We may naturally ask such questions as whether wavebreaking can give rise to a circulation of the right strength and the right sense, whether it will give the correct seasonal variability, and what determines vertical extent of the circulation. We begin with some elementary theory and phenomenology, for we will find that it will explain a number of the main features of the RMOC, and the advanced or confident reader may skip ahead to Section 17.5.

17.4.1 Wave Breaking and Residual Flow

The equations of motion governing the mean fields are the zonally averaged momentum and thermodynamic equations, which with quasi-geostrophic scaling and in residual form (Section 10.3.1) may be written as

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \nabla \cdot \mathbf{F} + \bar{F}, \quad \frac{\partial \bar{\theta}}{\partial t} + \frac{\partial \bar{\theta}}{\partial z} \bar{w}^* = \bar{J}, \quad (17.32a,b)$$

where \bar{F} represents frictional effects and \bar{J} represents heating, and on the β -plane the residual velocities are related to the Eulerian means by

$$\bar{v}^* = \bar{v} - \frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{\overline{v' \theta'}}{\overline{\partial_z \theta}} \right), \quad \bar{w}^* = \bar{w} + \frac{\partial}{\partial y} \left(\frac{\overline{v' \theta'}}{\overline{\partial_z \theta}} \right). \quad (17.33)$$

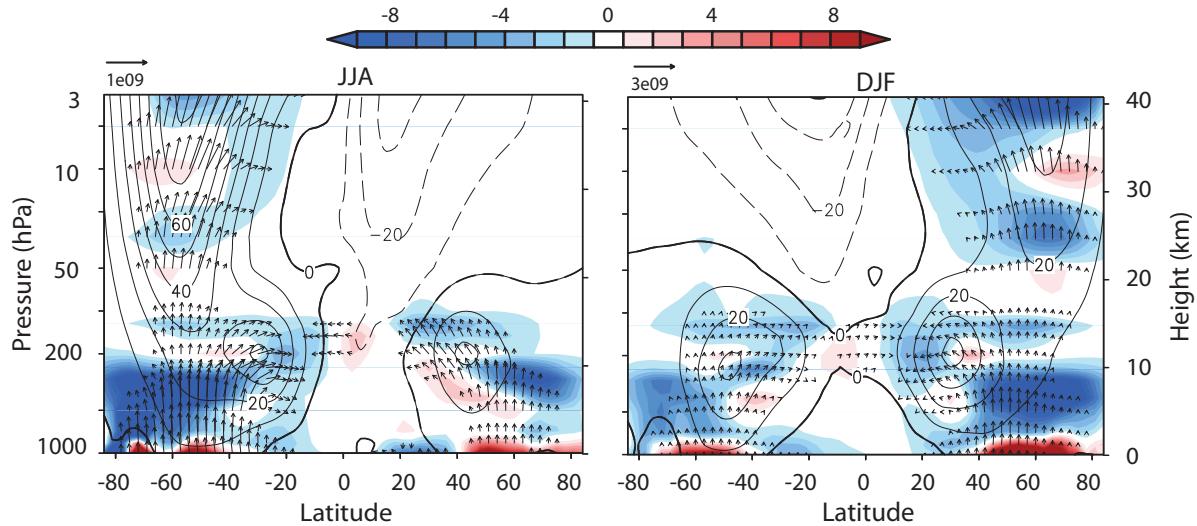


Fig. 17.10 The Eliassen–Palm (EP) flux vectors (arrows), the EP flux divergence (shading) and the zonally averaged zonal wind (contours) for northern hemisphere summer and winter. The tropospheric EP fluxes are of similar magnitudes in the summer and winter hemispheres, but are almost zero in the summer in the stratosphere. Note also strong convergence at high latitudes in the stratospheric winter hemisphere, leading to poleward residual flow and/or zonal flow acceleration.¹⁰

The vector \mathcal{F} is the Eliassen–Palm flux, and this is related to the meridional flux of potential vorticity by $\nabla \cdot \mathcal{F} = v' q'$. The wave activity (or pseudomomentum) obeys the Eliassen–Palm relation

$$\frac{\partial \mathcal{P}}{\partial t} + \nabla \cdot \mathcal{F} = \mathcal{D}, \quad (17.34)$$

where \mathcal{P} is the pseudomomentum, \mathcal{F} is its flux and \mathcal{D} is its dissipation.

From the autumn to the spring, the zonal wind in the stratosphere is generally receptive to planetary-scale Rossby waves propagating up from the troposphere (Fig. 16.6), although at high latitudes in winter there may be a period when the eastward zonal winds are too strong for waves to propagate. If these waves break in the stratosphere then there will be an enstrophy flux to small scales and dissipation. In a statistically-steady state and with small frictional effects the dominant balance in the zonal momentum equation (17.32a) is

$$-f_0 \bar{v}^* \approx \overline{v' q'}, \quad (17.35)$$

where \bar{v}^* is the residual velocity and the potential vorticity flux on the right-hand side is induced by the Rossby wave breaking. In dissipative regions the zonally averaged potential vorticity flux will tend to be down its mean gradient and, if the potential vorticity gradient is polewards (largely because of the β -effect), the residual velocity will be positive if f_0 is positive. That is, the residual flow will be *polewards*, in both hemispheres, and the mechanism giving rise to this is called the ‘Rossby wave pump’. Put another way, Rossby waves propagating up from the troposphere break and deposit westward momentum in the stratosphere, and in the mean this wave drag is largely balanced by the Coriolis force on the polewards residual meridional circulation.

This meridional circulation is weakest in summer mainly because linear Rossby waves cannot propagate upward through the westward mean winds, as illustrated in Fig. 17.10. It is quite striking how the EP vectors avoid the region of westward winds in the summer hemisphere, even though the level of wave activity at low elevations is relatively similar in the summer and winter hemispheres

(look between 10 and 15 km in the figure). We can interpret this by noting that for nearly plane waves the EP flux obeys the group velocity property, meaning that $\mathbf{F} = c_g \mathcal{P}$; however, as discussed in Section 16.5, if the mean winds are westward the waves evanesce and do not propagate, and thus almost the entire summer hemisphere is shielded from upwardly propagating waves, leaving it in a near-radiative equilibrium state. In the other seasons, the EP flux is able to propagate into the stratosphere and a circulation is generated. This acts to weaken the pole-equator temperature gradient, as we see by inspection of the thermodynamic equation: if the heating is represented by a simple relaxation to a radiative equilibrium state, θ_E , then in a steady state we have

$$N^2 \bar{w}^* = \frac{\theta_E - \theta}{\tau}. \quad (17.36)$$

Poleward flow in mid-latitudes must be supplied by rising air at low latitudes, and sinking air at high latitudes. Thus, from autumn to spring, at low latitudes we have $\theta < \theta_E$ and at high latitudes $\theta > \theta_E$.

Although cause and effect can be very difficult to disentangle in fluid dynamical problems, and the ultimate cause of nearly all fluid motions in the atmosphere is the differential heating from the Sun, it is important to realize that the meridional overturning in the stratosphere is not a direct response to differential solar heating: note that the most intense solar heating is over the summer pole, yet here there is little or no ascent. Rather, the circulation is more usefully thought of as a response to potential vorticity fluxes which in turn are determined by the upward propagation of Rossby waves from the troposphere combined with the poleward gradient of potential vorticity in the stratosphere. It is salutary to note that without motion we have $\theta = \theta_E$, so there is no net heating at all — the heating is a consequence of the wave forcing.

17.5 ♦ DYNAMICS OF THE RESIDUAL OVERTURNING CIRCULATION

We now discuss the dynamics of the residual meridional overturning circulation, the RMOC, in rather more detail than in the previous section and with a little repetition of important matters.¹¹ The dynamics of the RMOC can be usefully couched as a quasi-linear problem. That is, the governing equations can be written with the linear terms on the left-hand side and the nonlinear terms as forcing terms on the right-hand side. This cannot be regarded as a full solution, but if the right-hand sides can be determined, if only approximately, by independent means then the equations can be solved fairly straightforwardly and the structure of the RMOC so determined. Such a procedure is likely to be more successful in the stratosphere than in the troposphere; in the latter, the nonlinear terms are a truly essential part of the solution and cannot properly be separated from the linear dynamics (although we might choose to separate the terms to *diagnose* what forces the RMOC). In the stratosphere the nonlinear terms represent the effects of waves and wave breaking on the mean flow. These waves — both gravity waves and Rossby waves — often have their origins in the troposphere, and although the propagation and breaking of the waves does depend strongly on the background flow the basic features of the forcing of the RMOC can still usefully be considered independently of the RMOC itself.

17.5.1 Equations of Motion

Away from the equator the Rossby number is small and the equations governing the large scale flow are in good geostrophic balance. The equations of motion governing the mean fields are the zonally averaged momentum and thermodynamic equations, along with the thermal wind equation and the mass continuity equations. We write the equations in their full form in spherical coordinates using the ideal gas equations in log-pressure coordinates, since both sphericity and compressibility are important, using the TEM formalism, giving

$$\frac{\partial \bar{u}}{\partial t} - f \bar{v}^* = \mathcal{G} + \mathcal{D} = \nabla \cdot \mathbf{F} - \gamma \bar{u}, \quad f \frac{\partial \bar{u}}{\partial z} + \frac{R}{aH} \frac{\partial \bar{T}}{\partial \vartheta} = 0, \quad (17.37a,b)$$

$$\frac{\partial \bar{T}}{\partial t} + \bar{w}^* \mathcal{S} = Q_s + Q_l = \mu(T_R - \bar{T}), \quad \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (\bar{v}^* \cos \vartheta) + \frac{1}{\rho_R} \frac{\partial}{\partial z} (\rho_R \bar{w}^*) = 0. \quad (17.37c,d)$$

These equations are, respectively, the zonal momentum equation, the thermal wind equation, the thermodynamic equation and the mass continuity equation, with an overbar denoting a zonal average. The vertical coordinate, z is log pressure and $S = H_\rho N^2/R$ where R is the gas constant and H_ρ is the scale height used to define z ; thus, z has dimensions of height and $\rho_R = \exp(-z/H_\rho)$. The velocity components \bar{v}^* and \bar{w}^* are the residual, or transformed Eulerian mean, meridional and vertical velocities. The equations have a very similar form if written in height coordinates using the anelastic approximation (Section 2.5); in that case, ρ_R is a reference profile of density and the thermodynamic equation is written using potential temperature or buoyancy as the thermodynamic variable, the factor H_ρ/R no longer appears, and z really is physical height.

The right-hand side of (17.37a) represents wave forcing and friction: $\mathcal{G} = \nabla \cdot \mathbf{F}$ is the divergence of the Eliassen–Palm flux and \mathcal{D} is the frictional force, which we take to be a simple linear drag. Such a drag is a little arbitrary but its form greatly simplifies the ensuing analysis. On the right-hand side of the thermodynamic equation Q_s and Q_l represent the forcing due to solar and long wave radiation; we take $Q_l = -\mu \bar{T}$ where μ is a constant thermal damping rate, and we may write $Q_s = \mu T_r(\vartheta, z)$ where T_r is a radiative equilibrium temperature, assumed known. There are no fluid-dynamic wave-forcing terms in the TEM form of the thermodynamic equation. Typically, the momentum dissipation is small and $\gamma \ll \mu$, and indeed we may take $\gamma = 0$ without much loss of realism, except close to the ground.

17.5.2 An Equation for the RMOC

If the right-hand sides are known, (17.37) constitutes a closed set of equations for the response of the temperature and the three components of the velocity to an applied wave force \mathcal{G} and solar heating Q_s . Although nominally the equations have two time derivatives, the zonal wind and temperature are related through the thermal wind relation and the equations are balanced and no gravity waves are present. Our interest here is in the RMOC, and it is possible to derive a single equation for either \bar{v}^* and \bar{w}^* , and we will focus on \bar{w}^* .

The procedure is similar to that used in Section 14.5.3. Essentially, we differentiate (17.37a) with respect to z and (17.37c) with respect to ϑ and then use (17.37c) to eliminate the time derivatives. We then use the mass continuity equation to obtain a single equation in \bar{w}^* . We are particularly interested in the dependence of the RMOC on the spatial structure and time dependence of \mathcal{G} and Q_s , and to this end it is instructive to consider the case in which the time dependence is harmonic; that is, $\mathcal{G} = \tilde{\mathcal{G}} e^{i\omega t}$, $Q_s = \tilde{Q}_s e^{i\omega t}$ and $\tilde{w} = \bar{w}^* e^{i\omega t}$. After a little algebra, we obtain

$$\begin{aligned} & \frac{\partial}{\partial z} \left[\frac{1}{\rho_0} \frac{\partial(\rho_0 \tilde{w})}{\partial z} \right] + \left(\frac{i\omega + \gamma}{i\omega + \mu} \right) \frac{N^2}{4\Omega^2 a^2 \cos \vartheta} \frac{\partial}{\partial \vartheta} \left[\frac{\cos \vartheta}{\sin^2 \vartheta} \frac{\partial \tilde{w}}{\partial \vartheta} \right] \\ &= \frac{1}{2\Omega a \cos \vartheta} \frac{\partial}{\partial \vartheta} \left[\frac{\cos \vartheta}{\sin \vartheta} \frac{\partial \tilde{\mathcal{G}}}{\partial z} \right] + \left(\frac{i\omega + \gamma}{i\omega + \mu} \right) \frac{R}{4H\Omega^2 a^2 \cos \vartheta} \frac{\partial}{\partial \vartheta} \left[\frac{\cos \vartheta}{\sin^2 \vartheta} \frac{\partial \tilde{Q}_s}{\partial \vartheta} \right]. \end{aligned} \quad (17.38a)$$

The equation is quite a handful, but it is useful to realize that, schematically and without all the metric factors, it is of the form

$$\frac{\partial^2 \tilde{w}}{\partial z^2} + A \frac{N^2}{f^2} \frac{\partial^2 \tilde{w}}{\partial y^2} \sim \frac{1}{f} \frac{\partial}{\partial y} \frac{\partial \tilde{\mathcal{G}}}{\partial z} + \frac{A}{f^2} \frac{\partial^2 \tilde{Q}_s}{\partial y^2}, \quad (17.38b)$$

where

$$A = \frac{i\omega + \gamma}{i\omega + \mu}.$$

Equation (17.38b) is similar to (10.63), since with the addition of diabatic terms and a slight change in notation (10.63) is

$$f_0^2 \frac{\partial^2 \tilde{\psi}}{\partial z^2} + AN^2 \frac{\partial^2 \tilde{\psi}}{\partial y^2} = f_0 \frac{\partial \tilde{G}}{\partial z} + A \frac{\partial \tilde{Q}_s}{\partial y}, \quad (17.38c)$$

where $\tilde{\psi} = \psi^* e^{i\omega t}$ is the amplitude of the residual streamfunction of the overturning circulation. Since $\bar{w}^* = \partial \psi^* / \partial y$, (17.38b) is almost the y -derivative of (17.38c). (In (10.63) we took $\gamma = \mu$ so that $A = 1$.) Much of the physical interpretation in what follows comes from (17.38b) and (17.38c), although we will allow f to vary spatially — that is, we use f and not f_0 .

With quasi-geostrophic scaling the wave forcing term in (17.38) is

$$\mathcal{G} = \overline{v' q'} = -\frac{\partial}{\partial y} \overline{u' v'} + \frac{\partial}{\partial z} \left(\frac{f_0}{N^2} \overline{v' b'} \right) = \nabla \cdot \mathbf{F}, \quad \text{where } \mathbf{F} = -\overline{u' v'} \mathbf{j} + \frac{f_0}{N^2} \overline{v' b'} \mathbf{k}, \quad (17.39)$$

and \mathbf{F} is the Eliassen–Palm flux.

17.5.3 The Nature of the Response

The operator on the left-hand side of (17.38) is elliptic, similar to a Poisson equation. Thus, the response will be much less localized than the forcing itself, a concept that is familiar from potential vorticity inversion. To understand the equation better it is useful to take a heuristic look at some special cases, as follows. The cases are not all ‘orthogonal’ to each other — thus, for example, the low-latitude limit could be either low frequency or high frequency.

(i) The aspect ratio of the response

From (17.38c) the natural aspect ratio of the response, α_r say, is given by

$$\alpha_r = \frac{H_r}{L_r} = \frac{1}{A^{1/2}} \frac{f}{N} = \left(\frac{i\omega + \mu}{i\omega + \gamma} \right)^{1/2} \frac{f}{N}, \quad (17.40)$$

where H_R and L_R are the vertical and horizontal scales of the response. If the thermal and mechanical dissipation are zero, or have the same time scale, then $A = 1$, but more generally the presence of dissipation can alter the aspect ratio considerably. Also, the thermal dissipation can be expected to be much stronger than the mechanical dissipation, meaning that $\mu \gg \gamma$. The high- and low-frequency limits then have somewhat similar behaviour.

(ii) The high-frequency limit

In this case the thermal and mechanical damping are negligible and $A = (i\omega + \gamma)/(i\omega + \mu) \approx 1$. Since μ typically varies between $1/(20 \text{ days})$ in the lower stratosphere and $1/(5 \text{ days})$ in the upper stratosphere, and $1/\gamma$ is an even longer time, phenomena of order a few days fall into this category. Sudden stratospheric warmings are one example, although since the timescale of warmings is of order days thermal effects are not wholly negligible.

Using (17.40) we see that the aspect ratio of the response is simply of the order of Prandtl’s ratio. That is, $\alpha_r = H/L \sim f/N$ and since $f \sim 10^{-4} \text{ s}^{-1}$ and $N \sim 10^{-2} \text{ s}^{-1}$ or larger, the response to rapid forcing is typically quite shallow. Of course, although the Prandtl ratio is small it is a natural scaling of vertical to horizontal scales in atmospheric dynamics and shallowness should be interpreted in that context. Still, as we approach the equator the response shallows still further, although at the equator itself (17.38) ceases to be valid. A quantitative analysis of the right-hand side of (17.38a) further suggests that both waves (the \mathcal{G} term) and solar forcing act to drive the overturning circulation.

(iii) The low-frequency limit

In this case the frequency is less than the thermal damping rate; that is, $\omega \ll \mu$, with one

obvious example being the annual cycle. If as is realistic, $\gamma \ll \mu$ then $|A|$ becomes small. The effect of the solar heating thus also becomes small in (17.38c). The response generally deepens, with the ratio of the vertical to the horizontal scales being given by

$$\alpha_r = \frac{H_r}{L_r} \approx \left(\frac{1}{A^{1/2}} \right) \frac{f}{N} = \left(\frac{\mu}{i\omega + \gamma} \right)^{1/2} \frac{f}{N} \gg \frac{f}{N}, \quad (17.41)$$

if $\mu \gg \gamma$. In the steady-state limit the solar heating is balanced by the thermal relaxation and the right-hand side of (17.37c) nearly vanishes. We discuss this limit in more detail below, in the section on downward control.

(iv) Deep and shallow forces

A force may be regarded as deep or shallow depending on whether its aspect ratio (vertical to horizontal, α_F say) is greater or less than f/N , and it turns out that deep force tends to give rise to an acceleration of the zonal wind (a non-zero $\partial \bar{u}/\partial t$) whereas a shallow force tends to give rise to a meridional circulation. To see this we will consider the simplified forms of the momentum equation

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \mathcal{G}, \quad (17.42)$$

along with the MOC equation, (17.38c) which, if f_0 and N are both constant, is a Poisson equation with a right-hand side equal to $\partial \mathcal{G}/\partial z$. Solutions can be obtained by Fourier series methods with terms having the form

$$\tilde{\psi} = \Psi \cosh k_z z \sin k_y y. \quad (17.43)$$

Let us suppose the forcing is also of this form, so that by a deep forcing we mean that $k_y/k_z \gg f_0/N$ and shallow means $k_y/k_z \ll f_0/N$. We'll also suppose that $A = \mathcal{O}(1)$. From (17.38c) we see that

$$(f_0^2 k_z^2 - N^2 k_y^2) \tilde{\psi} \sim f_0 k_z G, \quad (17.44)$$

where G is the forcing amplitude. From this we can informally infer the form of the solution for deep and shallow forcing, as follows:

- *Deep forcing:* The dominant balance in (17.44) is between the second term on the left-hand side and the right-hand side and therefore $\tilde{\psi} \sim f_0 k_z G / (N^2 k_y^2)$ and

$$\tilde{w} \sim \frac{f_0 k_z G}{N^2 k_y^2}, \quad \tilde{v} \sim \frac{f_0 k_z^2 G}{N^2 k_y^2}. \quad (17.45)$$

Now look at the momentum equation (17.42). The ratio of the Coriolis force to the forcing on the right-hand side is given by

$$\frac{|f_0 v|}{|\mathcal{G}|} \sim \frac{f_0^2 k_z^2}{N^2 k_y^2} \ll 1, \quad (17.46)$$

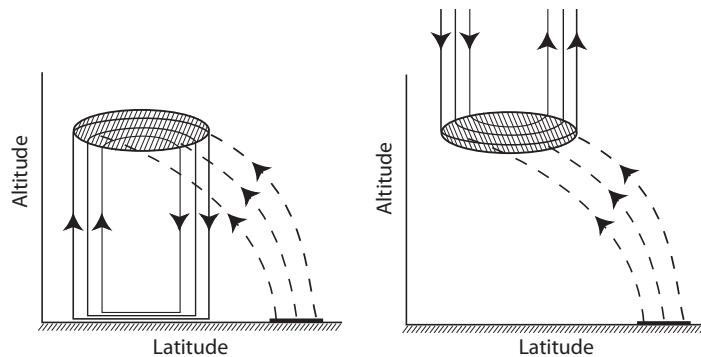
where the inequality follows by definition of what is deep. The dominant balance in the momentum equation must then be between the wave forcing and the acceleration.

- *Shallow forcing:* The dominant balance in (17.44) is between the first term on the left-hand side and the right-hand side and therefore $\tilde{\psi} \sim G / (f_0 k_z)$ and

$$\tilde{w} \sim \frac{k_y G}{f_0 k_z}, \quad \tilde{v} \sim \frac{G}{f_0}. \quad (17.47)$$

The ratio of the Coriolis term to the forcing term in the momentum equation is now $\mathcal{O}(1)$, and therefore the response to a shallow forcing appears in the meridional circulation rather than as an acceleration.

Fig. 17.11 Downward control. Left panel: wave activity propagates upward (dashed lines) from the troposphere, breaking and depositing zonal momentum in the shaded region. This induces an overturning circulation (solid lines) connecting the wave-breaking region with a bottom frictional boundary layer. Right panel: putative ‘upward control’, requiring friction above the wave breaking region for a steady response.



The underlying reason for the two different responses arises from the need to satisfy the thermal wind equation. A shallow force means that a shear is generated, and hence via thermal wind the temperature must change, but this can only be accomplished if there is an RMOC to affect temperature. If the force is deep then the response can and will be in the form of an acceleration. Close to the equator all forces are essentially deep because f is small.

(v) Deep and shallow heating

A similar analysis may be applied to a heating field, but given the intuition we have just developed about the response to a mechanical forcing there is no need to go through the details (although these are straightforward). A shallow heating (perhaps more usefully thought of as a broad heating) can and will produce a direct response in the temperature field itself. However, if the heating is deep (or latitudinally confined) then any direct response in the temperature field would have to be associated with a response in the zonal wind. Instead, a latitudinally confined heating tends to produce a response in the meridional circulation.

(vi) The low-latitude limit

At low-latitudes most forces become deep because f is small. More precisely, the criterion for deepness, that $H_f/L_f \gg N/f$ where H_f and L_f are the vertical and horizontal scales of the forcing, becomes easier to satisfy. Thus, wavebreaking at low-latitudes is more likely to induce an acceleration than a similar wavebreaking in mid-latitudes, which will tend to induce an overturning circulation. By the same token, a heating source at low-latitudes has a greater tendency to induce an overturning circulation than a similar heat source in mid-latitudes. The above statements are rules of thumb and not necessarily quantitative.

17.5.4 The Steady-state Limit and Downward Control

Let us now consider in a little more detail the steady-state response in which $\omega/\mu \rightarrow 0$, and we also take $\gamma = 0$, so that there is no momentum forcing. Although (17.38) of course still holds, it is also useful to look directly at the momentum equation and thermodynamic equations. The momentum equation, (17.37a) reduces to a balance between the Coriolis force and the wave driving, namely

$$-f\bar{v}^* = \mathcal{G}, \quad (17.48)$$

and the thermodynamic equation becomes

$$\bar{w}^*\mathcal{S} = Q_s + Q_l = \mu(T_R - \bar{T}). \quad (17.49)$$

The thermodynamic equation gives us very little information about the vertical velocity because the right-hand side contains the unknown temperature, \bar{T} ; rather, we can glean much of the information we want from (17.48).

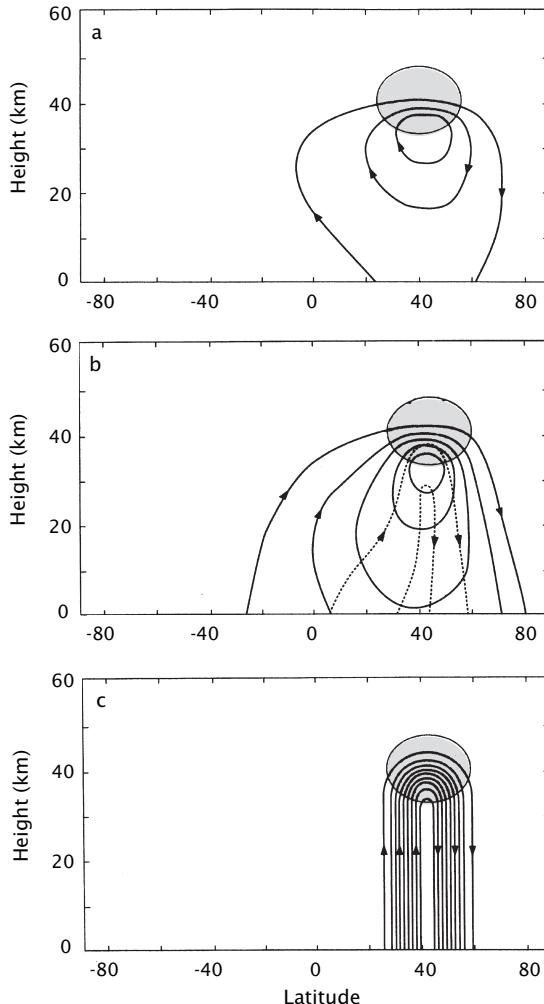


Fig. 17.12 The numerically-computed response of the meridional overturning circulation to a longitudinally symmetric westward force, with the frequency of the forcing decreasing from top to bottom.¹² Contours are streamlines of the residual circulation, with the same uniform interval in all panels, and the shading denotes the forcing region.

(a) Response to high-frequency forcing, $\omega/\mu \gg 1$, $\omega \gg \gamma$. The response is adiabatic and weakly spreads into the opposite hemisphere.

(b) A lower frequency case with $\omega/\mu = 0.34$, corresponding to an annual cycle and a 20-day thermal relaxation timescale. The solid and dashed lines show the response that is in phase and out of phase with the forcing, respectively.

(c) Steady state response, $\omega/\mu \ll 1$. The circulation increases in magnitude and narrows as the frequency decreases, and in panel (c) it is given using the downward control expression (17.51a).

If we differentiate (17.48) with respect to y (or ϑ) and use the mass continuity equation we obtain

$$\frac{1}{\rho_0} \frac{\partial \rho_0 w}{\partial z} = \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} \left(\frac{\mathcal{G} \cos \vartheta}{f} \right). \quad (17.50)$$

This is a first-order partial differential equation for the vertical velocity, and we can obtain the vertical velocity itself by a vertical integration, using a single boundary condition. If we require that vertical velocity stays finite at $z \rightarrow \infty$ and that $\rho_0 w = 0$, then we obtain

$$\bar{w}^*(z) = \frac{1}{a \rho_0(z) \cos \vartheta} \frac{\partial}{\partial \vartheta} \int_z^\infty \left(\frac{\rho_0(z') \mathcal{G}(\vartheta, z') \cos \vartheta}{f} \right) dz'. \quad (17.51a)$$

The Cartesian quasi-geostrophic version of this equation is just

$$\bar{w}^*(z) = -\frac{1}{\rho_R} \frac{\partial}{\partial y} \int_z^\infty \rho_R(z') \frac{\mathcal{G}(\vartheta, z')}{f_0} dz'. \quad (17.51b)$$

Equation (17.51) implies that, in the steady-state limit, the vertical velocity at a given height is determined by the wave forcing *above* that height. The physical situation is illustrated in the left

panel of Fig. 17.11. Here, a wave source in the troposphere propagates upward and breaks in the middle atmosphere depositing momentum. This induces a meridional circulation as illustrated in the left panel, with a response *below* the momentum source. The numerically-computed response to an imposed force is illustrated in Fig. 17.12, showing how the response changes depending on the time-scales of the forcing and damping, with the steady-state response illustrated in the bottom panel.

The above derivation may seem a little disingenuous, for surely we might just as well have assumed $w = 0$ at $z = 0$, leading (in the quasi-geostrophic case) to

$$\bar{w}^*(z) = \frac{1}{\rho_R} \frac{\partial}{\partial y} \int_0^z \rho_R \frac{\mathcal{G} + \mathcal{D}}{f_0} dz'. \quad (17.52)$$

This might appear to give ‘upward control’, as illustrated in the right panel of Fig. 17.11. However, if we are integrating from the ground up then frictional effects are important near the surface and must be included, as represented by the frictional term \mathcal{D} in (17.52). Furthermore, mass conservation demands that

$$\int_0^\infty \rho_R \bar{v}^* dz = 0, \quad \text{implying} \quad \int_0^\infty \rho_R (\mathcal{G} + \mathcal{D}) dz = 0, \quad (17.53a,b)$$

using (17.37a) for steady state conditions. Thus, above the level of the momentum source \mathcal{G} , (17.52) also in fact implies that the vertical velocity is zero, because \mathcal{G} and \mathcal{D} have cancelling effects. Thus, the location of the frictional boundary layer where the momentum is removed is one way to distinguish up from down. Equation (17.53b) tells us that the frictional boundary layer at the bottom must adjust to remove the same amount of momentum that is deposited by wave breaking higher up, if there is to be a steady state. If there were a momentum sink above the momentum deposition region there would be no justification for downward control, for we would have to include that frictional term in (17.51). However, it is hard to envision how such a sink could exist without violating angular momentum conservation. If there were no frictional sink at the ground the disturbance would initially propagate down, but on reaching the ground would then propagate up.

From the point of view of the diagnostic equation for the meridional overturning circulation, in the steady state limit we have $A = 0$, and the solar forcing on the right-hand side and the y -derivative on the left-hand side of (17.38) both vanish, and the equation for the RMOC becomes

$$\frac{\partial}{\partial z} \left[\frac{1}{\rho_0} \frac{\partial(\rho_0 \bar{w})}{\partial z} \right] = \frac{1}{2\Omega a \cos \theta} \frac{\partial}{\partial \theta} \left[\frac{\cos \theta}{\sin \theta} \frac{\partial \tilde{\mathcal{G}}}{\partial z} \right], \quad (17.54a)$$

or, in the quasi-geostrophic limit,

$$f_0 \frac{\partial^2 \tilde{\psi}}{\partial z^2} = \frac{\partial \tilde{\mathcal{G}}}{\partial z}. \quad (17.54b)$$

That is to say, in a steady state the solar forcing provides no input to the meridional circulation! This may seem a little counter-intuitive, but if there is no wave forcing and $\mathcal{G} = 0$ then the vertical velocity is zero and the temperature adjusts to the radiative equilibrium temperature, so that the diabatic forcing is zero.

The temperature field

Given that in a statistically steady state the vertical velocity field is determined by the wave forcing, the temperature field can be determined diagnostically from the thermodynamic equation. Thus, using (17.37c) with no time-dependence and a vertical velocity given by (17.51), we obtain in the quasi-geostrophic case

$$\bar{T}(z) - T_r(z) = \frac{\mathcal{S}}{\mu \rho_R} \frac{\partial}{\partial y} \int_z^\infty \rho_R \frac{\mathcal{G}(\theta, z')}{f_0} dz', \quad (17.55)$$

Parameter	Background	Rossby-gravity waves	Kelvin waves
Static stability, N	$2.2 \times 10^{-2} \text{ s}^{-1}$		
Beta at equator, β	$2.3 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$		
Coriolis parameter, f , at 5°	$1.27 \times 10^{-5} \text{ s}^{-1}$		
Wave period		4–5 days	10–20 days
Zonal wavelength		10,000 km	20,000–40,000 km
Zonal wavenumber dimensionally		4	1–2
Meridional scale		$6.3 \times 10^{-7} \text{ m}^{-1}$	$1.6\text{--}3.2 \times 10^{-7} \text{ m}^{-1}$
Vertical wavelength		1,200 km	1,500 km
Phase speed, relative to ground		4–8 km $20\text{--}25 \text{ m s}^{-1}$ (westward)	6–10 km 25 m s^{-1} (eastward)
Amplitudes:			
zonal velocity		$2\text{--}3 \text{ m s}^{-1}$	$4\text{--}8 \text{ m s}^{-1}$
meridional velocity		$2\text{--}3 \text{ m s}^{-1}$	0
vertical velocity		$1\text{--}2 \text{ mm s}^{-1}$	$1\text{--}2 \text{ mm s}^{-1}$
temperature		1 K	2–3 K
geopotential height		4 m	30 m
F_0 , wave forcing at 17 km, see (17.59).		$3\text{--}6 \times 10^{-3} \text{ m}^2 \text{ s}^{-2}$	$4\text{--}10 \times 10^{-3} \text{ m}^2 \text{ s}^{-2}$
Thermal damping rate, α		$0.5\text{--}1.5 \times 10^{-6} \text{ s}^{-1}$	$0.5\text{--}1.5 \times 10^{-6} \text{ s}^{-1}$

Table 17.1 Typical, approximate, values of parameters appropriate for waves and background flow in the equatorial lower stratosphere.¹³

with a similar but more complicated expression in the full case. The temperature field at a given height is determined purely by the momentum forcing, being given by the meridional gradient of the zonal force *above* that height.

† An oceanic comparison

It is instructive to compare downward control with the Stommel problem in oceanography (Section 19.1.1). In TEM (residual) form, the approximate zonally averaged zonal momentum equation may be written, as in (17.32a), as

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \mathcal{F} + \mathcal{D}, \quad (17.56)$$

The steady version of (17.56) and the equation for the streamfunction for the *horizontal* flow, ψ , in the ocean, (19.6), may thus respectively be written

$$f_0 \frac{\partial \psi^*}{\partial z} = \mathcal{F} + \mathcal{D}, \quad \beta \frac{\partial \psi}{\partial x} = \mathcal{F}_w - r \nabla^2 \psi, \quad (17.57a,b)$$

where \mathcal{F}_w represents the wind forcing at the ocean surface, and the second term on the right-hand side of (17.57b) represents friction. In (17.57a), $\bar{v}^* = -\partial \psi^* / \partial z$ and in (17.57b), $v = \partial \psi / \partial x$. The two equations have a formal similarity — is there more?

In the ocean interior, the frictional term is negligible, and in solving the resulting first-order equation ($\beta \partial \psi / \partial x = \mathcal{F}_w$) we may apply the boundary condition of $\psi = 0$ only at one meridional boundary. The natural choice is to choose the eastern boundary for this, and then invoke frictional processes to bring ψ to zero on the west. It is a natural choice because Rossby waves propagate westward ('westward control', as in Fig. 19.14); thus, the boundary current (e.g., the Gulf Stream) is on

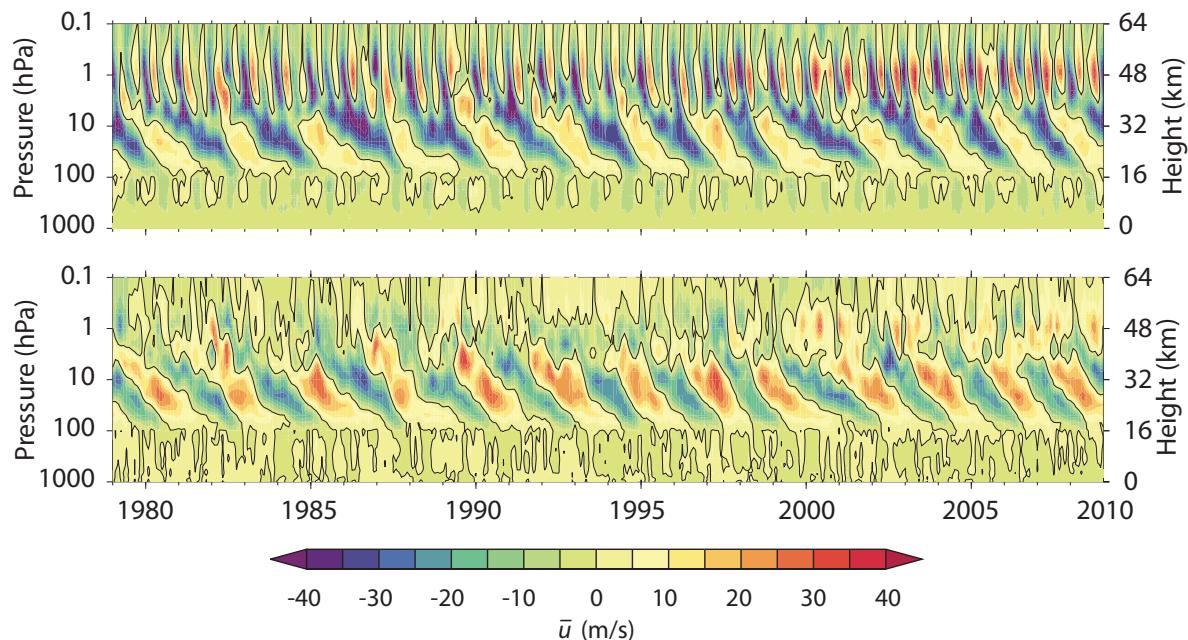


Fig. 17.13 Time-height sections of the observed zonal-mean zonal wind for 1979–2010, averaged from 5° S to 5° N. The contour is the zero wind line, and in the bottom panel the seasonal cycle is removed. The quasi-biennial oscillation, or QBO, is clearly visible between 20 and 40 km, or about 60 and 3 hPa.¹⁵

the west *because* the wind's influence is carried westward by Rossby waves, not vice versa. The consequence is that westward control is an enormously robust effect that pervades almost every aspect of large-scale physical oceanography. In the atmospheric case there is no similar mechanism that demands that the influence of the momentum source be propagated *only* downward. However, even without friction there is an up-down asymmetry in the atmospheric case because of (a) thermal damping, and (b) density variations. If we imagine a case with no boundaries at either top or bottom then the forcing creates meridional cells that propagate both up and down, growing (it turns out) like $t^{1/2}$. However, the solution also indicates that the upward propagating cell would eventually disappear, essentially to satisfy a boundary condition of boundedness at positive infinity (where density vanishes), leaving only a downward influence.¹⁴ In the case with boundaries a finite distance from the source, the final steady state does depend on the location of the frictional layers, and in particular a steady solution with only a downward influence results because the frictional boundary layer is at the bottom, not vice versa.

As we mentioned, the mechanism of downward control is related to that which gives rise to the Ferrel Cell in the troposphere, and that is certainly a strong and robust effect. Whether the downward control effect following wavebreaking *in the stratosphere* is strong enough to influence circulation in the troposphere, or the structure of the tropopause, remains an open question.

17.6 THE QUASI-BIENNIAL OSCILLATION

17.6.1 A Brief Review of the Observations

The *quasi-biennial oscillation*, or QBO, is a nearly periodic reversal of the zonal wind in the equatorial stratosphere, as illustrated in Fig. 17.13 and Fig. 17.14. It is the most dominant variability of that region, and the following lists some of the main features of the phenomenon:¹⁶

- The zonal winds in the equatorial region between about 5 and 100 hPa (about 40 and 18 km)

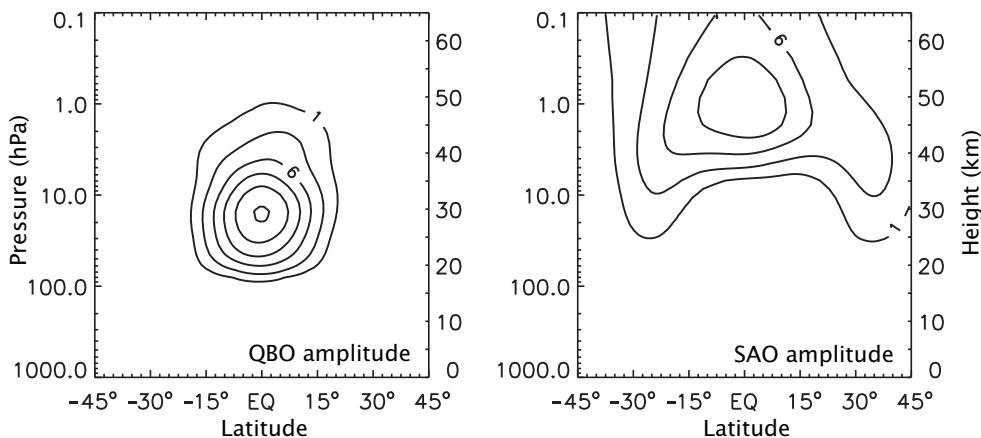


Fig. 17.14 The amplitudes (the root-mean square of the zonally-averaged wind of the oscillations after temporal filtering, with contours at 1, 3, 6, 9, 12 and 15 m s⁻¹) of the QBO and the SAO. The QBO evidently extends roughly from 15° S to 15° N and from 100 hPa to 1 hPa (about 17 km to 50 km). The SAO is broader, higher, weaker and faster.¹⁷

alternate between being eastward and westward with an average period of about 28 months, with the period varying between 22 and 34 months.

- The QBO is almost latitudinally symmetric about the equator. The amplitude is approximately Gaussian with a half width of about 12°.
- The phenomenon is approximately zonal symmetric; that is, the longitudinal variation is small.
- The maximum amplitude of the oscillation is about 30 m s⁻¹ (± 15 m s⁻¹) at about 20 hPa on the equator. The westward winds are slightly stronger than the eastward winds, after removing the annual cycle.
- The wind pattern descends at about 1 km per month with little loss of amplitude until it reaches 100 hPa, and the cycle begins again. (This does not mean that information propagates downward, as we discuss later.)
- The QBO is mildly synchronized to the annual cycle, with transitions between eastward and westward flow having a tendency to occur more commonly in March–June than in the other months.

See also the summary on page 655. Although we will not discuss it here, another oscillatory phenomenon occurs above the QBO known as the semi-annual oscillation, or SAO. The SAO is an oscillation in the zonal wind with an approximate period of six months (it should perhaps be called the quasi semi-annual oscillation) and it occurs between 1 hPa and 0.1 hPa and extends from about 30° N to 30° S (Fig. 17.14).

17.6.2 A Qualitative Discussion of Mechanisms

Candidate mechanisms

We first note that the QBO *must* involve zonally asymmetric motions. Without such asymmetric or eddying motions there can be no maximum of angular momentum within the fluid interior, and therefore no eastward winds at the equator, as explained in Sections 13.5.1 and 14.2.8. Given this, let us ponder for a moment what *might* be the mechanism of the QBO. One might suppose that horizontally propagating planetary waves would be a likely mechanism, for the transport of

momentum by Rossby waves is known to be an important mechanism for the maintenance of jets in mid-latitudes. However, the descent of the wind pattern with no loss of amplitude cannot be easily explained by such a mechanism.¹⁸ Other proposed mechanisms have involved interactions with the annual cycle or its harmonics (natural enough given the period of the QBO) or have invoked external forcing or some nonlinear feedback. However, no candidate mechanism was able to explain all the features noted above, until a mechanism involving the vertical propagation and absorption of gravity waves was proposed, as we now describe.

A gravity wave mechanism

The mechanism for the QBO that is now generally accepted involves the upward propagation and absorption of gravity waves and their effect on zonal flow.¹⁹ We first describe the basic mechanism rather roughly and qualitatively.

A broad spectrum of gravity waves, with phase speeds in both eastward and westward directions, is generated in the upper equatorial troposphere by deep convection and various other instabilities. The waves will in general have a component of the group velocity that is directed upward, and if these waves are dissipated via mechanical or thermal damping then they will force mean flow accelerations (steady, non-dissipative waves cannot force a mean flow acceleration). A critical level, where the phase speed of the waves equals the speed of the mean flow (i.e., where $c = \bar{u}$) is one place where wave absorption and mean-flow acceleration will be particularly effective, because as a wave approaches a critical level it slows, giving more time for dissipation to act. However, it is not necessary for there to be an actual critical level; indeed, waves approaching a critical level will often be largely dissipated before reaching it.

Let us suppose that initially there is a westward shear (that is $\partial\bar{u}/\partial z > 0$) and that there are upward propagating gravity waves with positive phase speed c . These will be very efficiently absorbed as they approach the critical level, depositing momentum and causing the mean flow to accelerate. As pictured in Fig. 17.15, this causes the critical level to descend and hence the subsequent absorption of gravity waves and acceleration of the zonal flow will be at a lower level. The wind anomaly thus descends, and so on. A similar effect will still occur even if there is no critical level, provided there is enough dissipation for the gravity wave to be absorbed somewhere, causing the mean flow to accelerate. Even if this acceleration is insufficient to induce a critical level, the difference between the wave speed and the zonal fluid speed will be reduced (i.e., $c - \bar{u}$ diminishes) and gravity wave absorption is enhanced. Gravity waves are thus absorbed at a lower level than previously and the anomaly in the zonal wind descends, as before.

Eventually, in the above models, the wind anomaly descends to the level of the gravity wave source. Depending on the strength of the dissipation one might imagine that dissipative processes could then wipe out the wind anomaly completely and the whole process would start all over again, or perhaps a low level westward wind anomaly would persist, so redefining the mean flow. However, in either case the zonal wind anomaly would not change sign (i.e., become westward), as is observed in the real QBO. For that to occur we may invoke a second wave in conjunction with an instability, as we now explain.

A two-wave model

Suppose now there are two upward propagating gravity waves with speeds $+c$ and $-c$ (where c itself is positive), each of which will slowly be dissipated as it propagates, with the dissipation enhanced for smaller values of $|\bar{u} - c|$ or $|\bar{u} + c|$, respectively. (There will of course be very large dissipation if there is a true critical level.) Suppose that the mean flow has no shear, then simply by symmetry that state can persist, with the eastward and westward waves being dissipated equally as they ascend with no zonal flow generation. However, that symmetric state is unstable; to see this suppose that there is a small eastward perturbation to the zonal wind, as illustrated in the left panel of Fig. 17.15. The eastward propagating wave will then be preferentially dissipated, because $\bar{u} - c$ is smaller for it than for the westward wave. The eastward anomaly in the zonal wind will therefore grow and

Essentials of the QBO

What is the Quasi-Biennial Oscillation?

- The QBO is a quasi-periodic reversal of the zonal-mean zonal winds between about 20 km and 45 km altitude and 15° S to 15° N, with an irregular period of about 28 months. It is the dominant pattern of variability in the equatorial stratosphere and it is the clearest manifestation of a non-directly forced nearly-periodic phenomenon in the atmosphere.
- The eastward and westward zonal winds appear to propagate downward at about 1 km per month, reversing at the end of each half cycle.
- The half-amplitude of the zonal wind cycle is about 15 m s^{-1} , with the westward winds being slightly stronger.

What is the mechanism?

- The oscillation is caused by the upward propagation and absorption of Kelvin waves and Rossby-gravity waves at the equator. If a wave has an eastward phase speed then, on absorption, it will cause the mean zonal flow to accelerate eastward. Furthermore, the absorption is strongest near a critical layer, where the mean zonal wind speed equals the phase speed. An upward propagating Kelvin wave thus causes the mean flow aloft to accelerate eastward, and then the maximum in eastward winds to move downward and eventually to be dissipated. An upward propagating Rossby wave then generates a westward zonal wind anomaly aloft, which similarly propagates down. In this way the zonal wind oscillates between positive and negative values, as illustrated in Fig. 17.16.
- The waves are generally considered to be primarily excited by moist convection in the upper tropical troposphere.
- The period is determined by a combination of parameters involving the wave and mean flow. In the simplest model of two upwardly propagating gravity waves the period is given by

$$P = \frac{Akc^3}{\alpha N_0 F_0} \quad (\text{QBO.1})$$

where A is a nondimensional number weakly dependent on viscosity, and the other parameters, properties of the waves and mean flow, are defined in the text. The period is not proportional to the period of the waves; rather it is inversely proportional to their strength, F_0 , because stronger waves cause more mean flow acceleration and a faster descent of the pattern.

Why is the phenomenon equatorially confined?

- The mechanism requires there to be upwardly propagating waves with very different phase speeds in order that the mean flow can oscillate between the two values. In equatorial regions such a forcing can be provided by Rossby waves and Kelvin waves.
- In mid-latitudes the tropospheric flow is largely balanced and it is primarily long Rossby waves that reach the stratosphere with a spectrum of phase speeds. If and when they break they would provide only a westward acceleration. Furthermore, in mid- and high latitudes an imposed force tends to induce a mean meridional circulation, not a mean flow acceleration (Section 17.5).

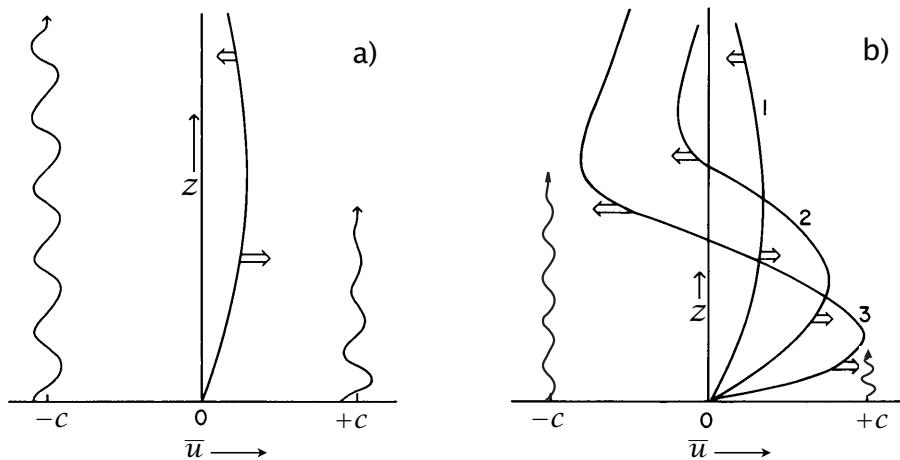


Fig. 17.15 Schema of the initial instability leading to the QBO.²⁰ The solid lines show the zonal flow, the wavy arrows indicate the gravity-wave penetration from below and the double arrows indicate wave-induced acceleration. Initially, as in the left panel, a small eastward perturbation is added to a stationary mean flow. The eastward moving wave is preferentially absorbed and the perturbation is amplified and then descends, with the right panel showing the zonal flow at the successive times indicated, with the gravity wave penetration illustrated at $t = 3$. After the flow develops an eastward component the westward wave penetrates higher before being absorbed, inducing a westward flow aloft that then itself descends, making the eastward anomaly thinner. Subsequent stages and the development of a periodic oscillation are illustrated in Fig. 17.16.

descend, just as described above. The upward propagation of the eastward wave is then limited, but the westward wave is unconstrained so it reaches higher levels before eventually being absorbed, providing a westward acceleration to the zonal flow, as illustrated in the right panel of Fig. 17.15.

As the the westward anomaly descends it squeezes the eastward anomaly which becomes thinner and thinner. Dissipative processes then become more efficient and can erode the eastward anomaly completely, with the flow becoming entirely westward, as illustrated in panel (b) of Fig. 17.16. A high level eastward anomaly is then created (panel (c) of Fig. 17.16), descending and squeezing the westward anomaly, and a mirror image of the first stage takes place. The entire cycle repeats itself and an oscillation is born, with the period of the oscillation being determined by the strength of the gravity waves and the rate of dissipation: stronger gravity waves lead to a faster acceleration of the mean flow and so a greater rate of descent and so a shorter period. Finally, note that the waves need not have speeds symmetric on either side of zero, $+c$ and $-c$. Suppose, for example, the wave speeds were both positive, a and b say. The mean flow could accelerate to an average value of $(a + b)/2$, with the flow then oscillating between a and b in a fashion similar to the symmetric case.

17.6.3 A Quantitative Model of the QBO

We now consider the above wave–mean-flow interaction model a little more quantitatively, and our first goal will be to obtain equations of motion for the interaction. To this end we will parameterize the vertical propagation and absorption of gravity waves by simple expressions resulting from gravity wave theory described in Section 17.2.3. The absorption leads to a zonal flow acceleration, which in turn affects the wave absorption, and so on.

Let us consider a semi-infinite (no top), non-rotating, stratified fluid subject to a standing wave

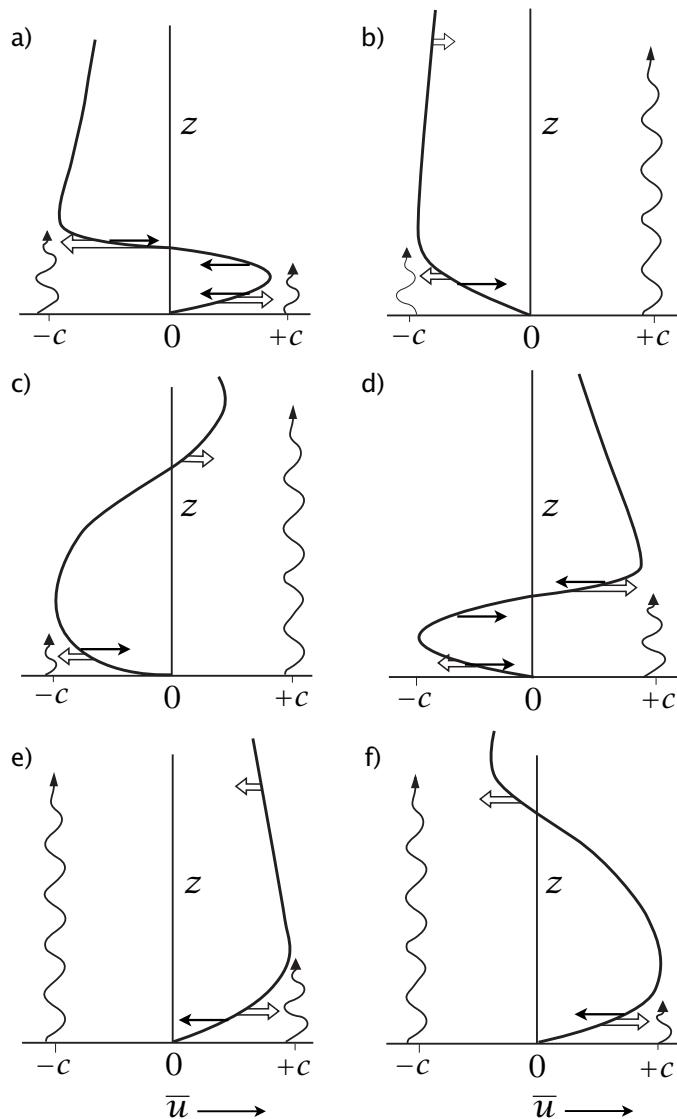


Fig. 17.16 Schema of the evolution of the QBO with gravity-wave forcing from below, following an initial perturbation illustrated in Fig. 17.15.²¹ The solid lines show the mean flow and the wavy lines indicate the propagation of gravity waves. Horizontal double arrows indicate wave forcing and single arrows indicate viscous relaxation.

The panels are at successive times, with the top four panels showing a half cycle, and panels (d), (e) and (f) are mirror images of (a), (b) and (c). Wave-induced acceleration of the mean flow occurs preferentially near critical levels where $\bar{u} = c$.

forcing at the lower boundary. Specifically, the waves are of the form

$$w = \text{Re } \tilde{w}_1(z) e^{ik(x-ct)} + \tilde{w}_2(z) e^{ik(x+ct)}. \quad (17.58)$$

The waves have a dispersion relation as discussed in Section 17.2.3, a positive (upward) group velocity, and we will take $\tilde{w}_1 = \tilde{w}_2$. If there is a source of gravity waves such as convection there is no difficulty in exciting waves with either an eastward or westward phase speed: a Kelvin wave has a purely eastward phase speed ($c_p > 0$), a Rossby-gravity wave has a westward phase speed, and gravity waves completely uninfluenced by rotation can have a phase speed in either direction. The Kelvin and Rossby-gravity waves, probably the most important waves for the QBO, typically have zonal wavenumbers 1–4, and so zonal wavelengths greater than 10 000 km, and periods of 3 days or longer.

As the waves propagate up they are dissipated, primarily by thermal rather than viscous dissipation, and their amplitude diminishes in the vertical and consequently they deposit momentum into the mean flow. From the WKB calculation of Section 17.3.3 the wave momentum flux,

$F_k(z) = \overline{u'w'}$, of a given upward-propagating wave is of the form

$$\bar{F}_k(z) = \bar{F}_k(0) \exp \left[- \int_0^z g_k(z') dz' \right], \quad (17.59)$$

where the subscript k indicates the zonal wave number and the attenuation rate, $g_k(z)$, for a given upward-propagating internal wave is given by

$$g_k(z) = \frac{\text{damping rate}}{\text{vertical group velocity}} = \frac{\alpha}{k(\bar{u} - c)^2/N}, \quad (17.60)$$

where c is the phase speed of the waves. The mean flow, $\bar{u}(z, t)$, is influenced by many such waves and so evolves according to

$$\frac{\partial \bar{u}}{\partial t} = - \sum_k \frac{\partial \bar{F}_k}{\partial z} + \nu \frac{\partial^2 \bar{u}}{\partial z^2}. \quad (17.61)$$

In writing (17.61) we include a dissipative term but neglect terms representing advection by the mean flow (such as $w \partial \bar{u} / \partial z$) and the Coriolis force $f v$. Equation (17.61), with (17.59) and (17.60), is a closed partial differential equation in a single unknown for the mean flow. If the forcing consists of two waves, one with a phase speed c that is positive and one with a negative phase speed, the model produces behaviour that is quite similar to that of the QBO, as we see shortly.

Direction of influence

Although both the observations and the schematic solutions illustrated in Fig. 17.16 suggest that influence is somehow propagating downward, this is in fact not the case when the gravity waves are propagating upward. From (17.59) and (17.61) the wave-driven acceleration of the mean flow, A_w say, is given by

$$A_w = - \frac{\partial \bar{F}}{\partial z} = +\bar{F}(0)g(z) \exp \left[\int_0^z g(z') dz' \right] = +g(z)\bar{F}(z). \quad (17.62)$$

That is, the acceleration is a function only of the profile of g in the region from 0 to z , that is the region through which the wave has propagated. Furthermore, attenuation rate $g(z)$ is itself, from (17.60), a function only of the local value of $\bar{u}(z)$ and not of the derivatives of \bar{u} . Thus, the wave forcing at some level z is a function only of the profile of $\bar{u}(z')$ for $z' < z$ and independent of the profile at higher altitudes. In other words, and in so far as the diffusivity term in (17.61) is negligible, there is no downward propagation of influence of the mean flow and the mean flow evolution is independent of what takes place above. The physical origin of this result is simply that waves are propagating upward and are absorbed by the mean profile as they ascend. If there were a source of waves at very high altitude, or if waves were reflected within the fluid (in which case the first-order WKB approximation is incomplete) then there could be a downward propagation of influence.

17.6.4 Scaling and Numerical Solutions

Scaling the equations — nondimensionalizing in an intelligent way — not only makes numerical integration easier but also indicates what the natural height and time scales are for the problem. Important external parameters that determine the problem are the stratification N (which has units of inverse time, T^{-1}), the damping rate α (also units of inverse time) and the strength of the wave forcing, \bar{F} (units of $(L/T)^2$). A natural horizontal scale is the inverse of the wavenumber k .

Denoting nondimensional quantities with a hat, let

$$\hat{F} = \bar{F}/F_0, \quad \hat{N} = N/N_0, \quad (17.63)$$

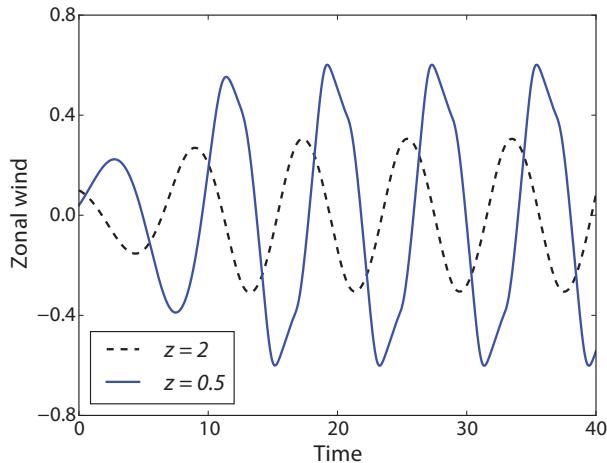


Fig. 17.17 The evolution of the mean zonal wind at two different levels in a numerical solution of (17.67). All the variables are nondimensional, with the only parameter in the problem being viscosity, and here $\hat{v} = 0.15$. A small perturbation is added to \hat{u} , and relatively quickly the solution becomes periodic.

where $F_0 = \bar{F}(0)$ and N_0 is a typical value of N . If N were uniform then we would simply choose $N_0 = N$ whence $\bar{N} = 1$. To obtain sensible nondimensional quantities we note that the attenuation rate, g , has dimensions of inverse height, so using (17.60) we choose a scaling height H as

$$H = \frac{kc^2}{\alpha N_0}. \quad (17.64)$$

This might suggest a time scaling of $T = kc/\alpha N$. However, because we have a forced problem, the form of (17.61) indicates that we choose

$$T = \frac{cH}{F_0} = \frac{kc^3}{\alpha N_0 F_0}, \quad (17.65)$$

with a velocity scaling of $U = F_0 T / H = c$ (note that this is not an advective scaling). The non-dimensional coefficient of viscosity and thermal damping coefficients are then

$$\hat{v} = v \frac{T}{H^2} = v \frac{\alpha N_0}{F_0 kc}, \quad \hat{\alpha} = \alpha T = \frac{kc^3}{N_0 F_0}. \quad (17.66)$$

The nondimensional equation for the mean flow evolution is then, for a single wave,

$$\frac{\partial \hat{u}}{\partial \hat{t}} = -\frac{\partial \hat{F}}{\partial \hat{z}} + \hat{v} \frac{\partial^2 \hat{u}}{\partial \hat{z}^2}, \quad \text{where} \quad \hat{F}(z) = \exp \left[- \int_0^z \frac{1}{(\hat{u} - 1)^2} dz' \right], \quad (17.67a,b)$$

with the hats indicating nondimensional quantities. The great simplification that (17.67) offers over (17.59)–(17.61) is that in (17.67) there are no parameters, save for the viscosity, and so the time and vertical scales of the problem are laid bare. In particular, if viscosity is small the only significant timescale in the system is (17.65) and the period of the oscillation must be proportional to that, and the vertical scale of the oscillation must be given by (17.64). Evidently the period of the oscillation is inversely proportion to the strength of the waves, but the vertical extent and the amplitude of the oscillation are both independent of the wave strength.

A numerical solution

Equation (17.67) may readily be numerically integrated²² and solutions are illustrated in Figs. 17.17, 17.18 and 17.19. The simulations show many of the qualitative features of the observed QBO, including the decay of the pattern with height and its apparent downward propagation. The simulations

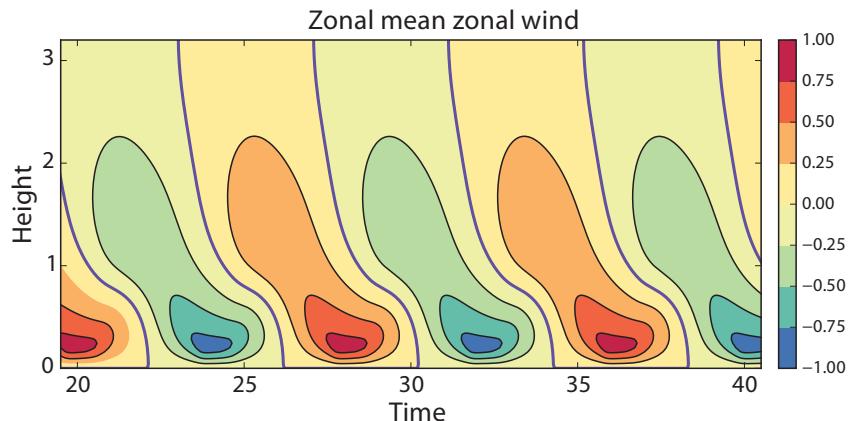


Fig. 17.18 Time height section of nondimensional zonal wind in a numerical solution of (17.67), showing the last 20 time units of the same integration as Fig. 17.17. The zero contour is thicker.

are dependent on the viscosity \hat{v} in that, if $\hat{v} = 0$, the jet at the bottom of the domain cannot be dissipated and the system in fact evolves to a steady state. On the other hand, if the viscosity is large then the jets become too broad, and the boundary layer near $z = 0$ that is evident in Fig. 17.18 is thicker. Still, if the viscosity is small but non-zero then over the bulk of the cycle it plays little role and the oscillation period is only weakly dependent on its value. Thus, for example, in Fig. 17.16 viscosity is needed to wipe out the low level westward jet between panels (d) and (e), but has little role in the rest of the half cycle and so only a small effect on the period. If viscosity is unimportant then the only timescale in the problem is that given by (17.65) and the period is proportional to it, and from numerical integrations we find that it is given by

$$P = A \frac{kc^3}{\alpha N_0 F_0} \quad (17.68)$$

where $A \approx 8$. Note that the period of the QBO is not directly dependent on the period of oscillation of the waves themselves, but is dependent on their strength.

17.6.5 The Roles of Rossby and Kelvin Waves

The waves that propagate into the stratosphere in equatorial regions are of two main types, Kelvin waves and Rossby waves. Kelvin waves are a form of gravity wave but have only an eastward phase propagation, whereas Rossby waves are balanced waves with a westward phase propagation. The theoretical development paralleling Section 17.6.3 is naturally more complex, in part because the problem is now, in principle, a three-dimensional one. However, it is much simplified if we consider motions at the equator and if we take note that, in general, the attenuation rate of a wave is equal to its damping rate divided by its group velocity, as in (17.60). The corresponding attenuation rates for Kelvin and Rossby waves are then given by

$$\text{Kelvin wave: } g_K(z) = \frac{\alpha}{k_K(\bar{u} - c_K)^2/N}, \quad (17.69a)$$

$$\text{Rossby wave: } g_R(z) = \frac{\alpha}{k_R(\bar{u} - c_R)^2/N} \left(\frac{\beta}{k_R^2(\bar{u} - c_R)} - 1 \right). \quad (17.69b)$$

The Kelvin wave attenuation rate is just the same as that for a non-rotating gravity wave, although the wave speed, c_K , is strictly positive. The Rossby wave attenuation rate (whose derivation requires a little work) involves the equatorial beta parameter and a negative phase speed, c_R . The full problem is defined by (17.61) and (17.59), now with $g(z)$ given by (17.69). It is evident that the problem is no longer east–west symmetric, but the essential structure of the problem remains.

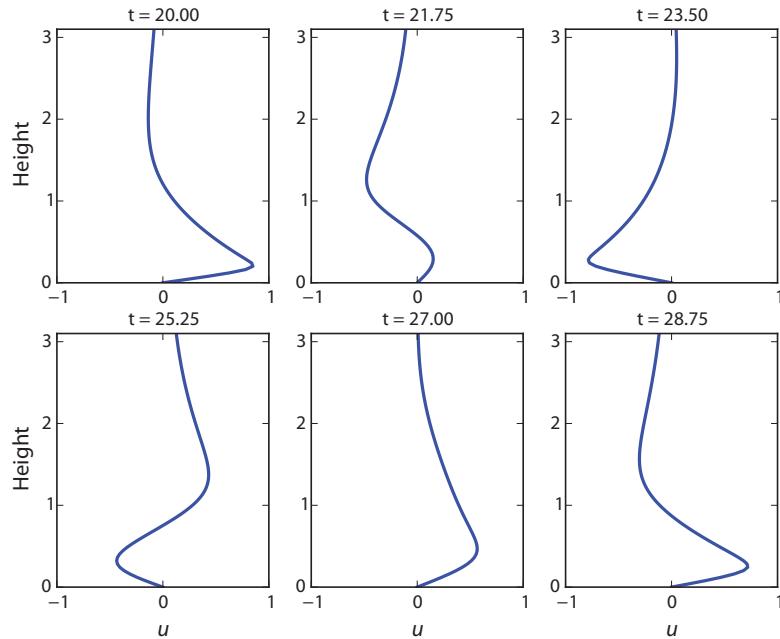


Fig. 17.19 Profiles of zonal wind at the times indicated from the numerical solution of (17.67), shown in Fig. 17.18.

Rossby wave absorption is enhanced near a critical layer where $\bar{u} = c_R$, and Kelvin wave absorption still occurs near $\bar{u} = c_K$, so we expect to see an oscillation contained between these two values. Further, just as in the gravity wave problem, influence propagates upward with the waves.

The equation set (17.61), (17.59) and (17.69) may be numerically integrated and solutions are illustrated in Fig. 17.20. It is the factor g_R that is responsible for the westward acceleration of the mean flow, for ‘dragging’ \bar{u} toward the value c_R , which is negative. But g_R is zero when $\bar{u} - c_R = \beta/k_R^2$ which, for our numerical simulation, occurs when $\bar{u} = 1$. Thus, when \bar{u} is close to its eastward (Kelvin-wave induced) peak the westward acceleration is small. Another source of east–west asymmetry is the likelihood that Rossby waves and Kelvin waves may have different amplitudes. If Kelvin waves were stronger, for example, then the eastward acceleration would be stronger than the westward and that part of the cycle would be faster.

17.6.6 General Discussion

The above sections have described a couple of relatively simple models that seem to capture the essence of the QBO. The simple model using both Rossby waves and Kelvin waves is not noticeably more realistic in its predictions; rather, it is attractive because Rossby waves and Kelvin waves are observed in the equatorial stratosphere and may be more realistic in its assumptions. The observed east–west asymmetry in the observed QBO is not obviously caused by the differences between Rossby waves and Kelvin waves; other possibilities include the effects of a mean circulation and possible differences in the strength of the eastward and westward forcing.

The model with two non-rotating gravity waves is attractive because it allows a more complete analysis of its properties. In particular, the upward propagation of waves leading to a downward propagation of the zonal wind pattern, and the factors determining the period of the oscillation, are made transparent. The period of the problem is given by (17.68). Using Table 17.1 as a guide, let us take the following dimensional values of the parameters: $k = 2 \times 10^{-7} \text{ m}^{-1}$, $\alpha = 1 \times 10^{-6} \text{ s}^{-1}$, $c = 25 \text{ m s}^{-1}$, $N_0 = 2.2 \times 10^{-2} \text{ s}^{-1}$, $F_0 = 5 \times 10^{10} \text{ m}^2 \text{ s}^{-2}$. We obtain a timescale of $T = (kc^3/\alpha N_0 F_0) \approx 160 \text{ days}$ or about 5 months and so, using (17.68), a period of 40 months. Obviously there is considerable uncertainty in the parameters chosen and it would not be difficult

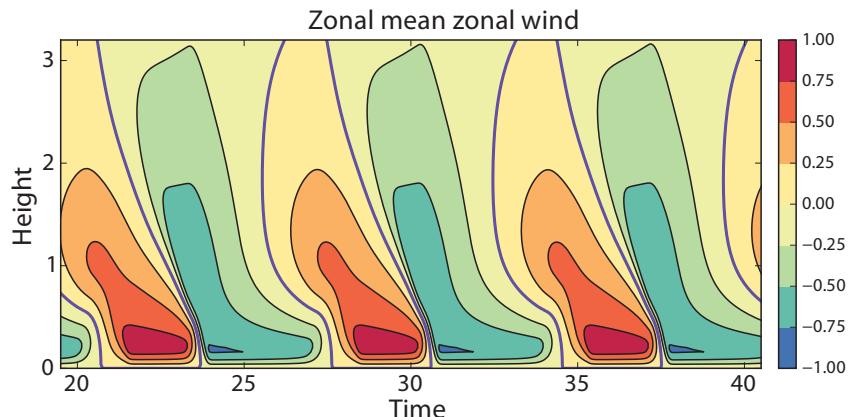


Fig. 17.20 Time height section of nondimensional zonal wind in a numerical solution similar to that of Fig. 17.18, but now using (17.69) and so with both Rossby and Kelvin waves. The zero contour is thicker.

to choose a set of parameters giving the observed value of about 26 months — or for that matter, to choose a set that gave a still longer period.

The vertical scale of the oscillation is given by (17.64) and with the above set of parameters we obtain $H = (kc^2/\alpha N_0) \approx 6\text{ km}$. From the numerical simulations we see that the vertical penetration of the phenomena is 2 or 3 times this, so about 15 km. This value again is reasonably close to the observed value of, from Fig. 17.13, about 20 km, but again we should be wary of too close an agreement, especially using a one-dimensional quasi-Boussinesq model. It is interesting that the period and vertical extent of the observed oscillation vary only a little, implying only a little interannual variability in the forcing strength and other parameters of the problem.

The actual waves themselves are primarily generated by convection in the tropical troposphere, they then propagate up into the stratosphere. It is difficult to numerically simulate a QBO with an explicit representation of gravity waves because of the large range of scales involved in the problem, although three-dimensional simulations with parameterized gravity waves have been quite successful. Still, a striking demonstration of the mechanism above came from laboratory experiments, using an annulus of stratified water subject to a standing wave forced by a flexible lower boundary. Given a strong enough forcing an oscillating mean flow was generated whose structure was found to be in very good agreement with the two-wave theory. Thus, at the very least, the mechanism does describe a real physical phenomenon.²³

There are a great many aspects of the QBO that we have not discussed, including its latitudinal structure, the effects of a mean circulation, and three-dimensional numerical simulations, and a few references that may serve as an introduction to these topics are given here.²⁴ The QBO is also not completely regular, as we see from Fig. 17.13, and one somewhat unusual example occurred in early in 2016 (Fig. 17.21). The QBO normally shows a fairly steady downward propagation of the westerly phase, but in January 2016 the westerly winds in the lower stratosphere switched back to easterlies after only about 6 months, the shortest period of westerlies above 20 km in a record going back to 1953. This behaviour may have been caused by anomalous horizontal propagation of Eliassen–Palm fluxes from mid-latitudes because of the absence of a subtropical critical line, although we cannot be definitive. Still, it is the regularity of the QBO rather than the occasional anomaly that is most striking.

Finally, to make a personal remark, the QBO is both a curiosity and a triumph. The former because its relationship to and influence on tropospheric circulation, and the climate and weather that affect humankind, is not obvious to the casual or even expert observer; it does not have the impact of an El Niño event or a cold winter, for example. Yet, excepting directly forced oscillations like the diurnal and seasonal cycle, it is the clearest example of a nearly periodic phenomenon in the atmosphere and its simple and beautiful explanation must rank as a major achievement in geophysical fluid dynamics.

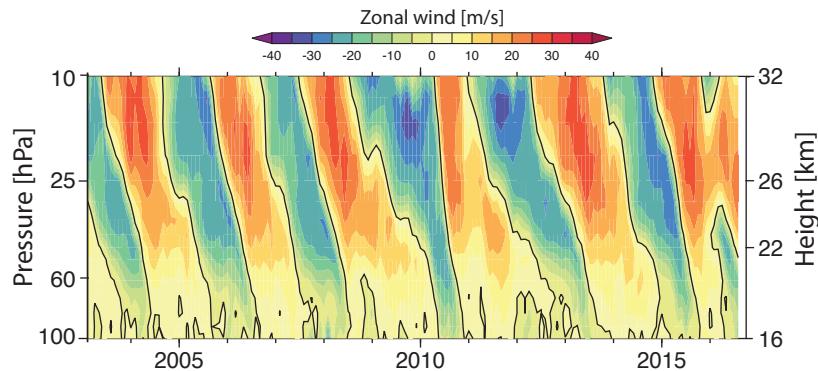


Fig. 17.21 Equatorial zonal wind, similar to Fig. 17.13, but now with cycles of the QBO from 2003 to 2016 and showing the interruption of the eastward winds below about 25 km in early 2016.²⁵

17.7 VARIABILITY AND EXTRA-TROPICAL WAVE–MEAN-FLOW INTERACTION

As noted in Section 17.1, the stratosphere has only mild, if any, baroclinic instability and because of its high stratification the amplitude of a baroclinic mode reaching up from the troposphere thus tends to decay rapidly in the stratosphere, as in Fig. 9.21. If there were to be a baroclinic instability confined to the stratosphere it would be at large scales, perhaps at wavenumbers 1, 2 or 3, as opposed to wavenumber 8 or so in the troposphere. But this is not to say there is no variability in the stratosphere, with the variability arising in two main ways:

- (i) From waves propagating up from the troposphere, with the stratospheric variability arising from the variability of the troposphere.
- (ii) From oscillatory or even chaotic flow arising from the interaction, within the stratosphere, of large-scale planetary waves with themselves and with the mean flow. The forcing may still come from the troposphere but, even when this is steady, intra-stratospheric interactions may give rise to unsteadiness.

In either case the variability tends to be relatively slow (compared to the troposphere) and at a large scale — waves from the troposphere undergo Charney–Drazin filtering and tend to occur at wavenumbers 1 and 2, and as noted any baroclinic instability is also at large scale. It is therefore useful to think of the variability as a wave–mean-flow problem rather than as a problem in geostrophic turbulence.

17.7.1 Upward Propagating Disturbances and Sudden Warmings

Consider planetary waves that are excited in the troposphere and propagate upward, as described in Chapter 16, with this occurring predominantly in winter when the tropospheric forcing is strongest. The wave activity obeys

$$\frac{\partial \mathcal{P}}{\partial t} + \nabla \cdot \mathbf{F} = D, \quad (17.70)$$

where $\mathcal{P} = \overline{q'^2}/2\bar{q}_y$ and $\nabla \cdot \mathbf{F} = \overline{v'q'}$. Thus, if dissipation is small and $\partial \mathcal{P}/\partial t$ is positive, $\overline{v'q'}$ is negative. Now consider the zonal momentum equation in quasi-geostrophic TEM form, namely

$$\frac{\partial \bar{u}}{\partial t} - f_0 \bar{v}^* = \overline{v'q'}. \quad (17.71)$$

Rossby waves propagating into the stratosphere or, by the same token, dissipating Rossby waves in a statistically steady state, thus induce a deceleration (i.e., a westward tendency) of the zonal mean flow and/or a poleward meridional flow, and only for a very deep forcing is the residual circulation response negligible. Also, the zonal-wind response to a wave forcing will tend to be of larger scale

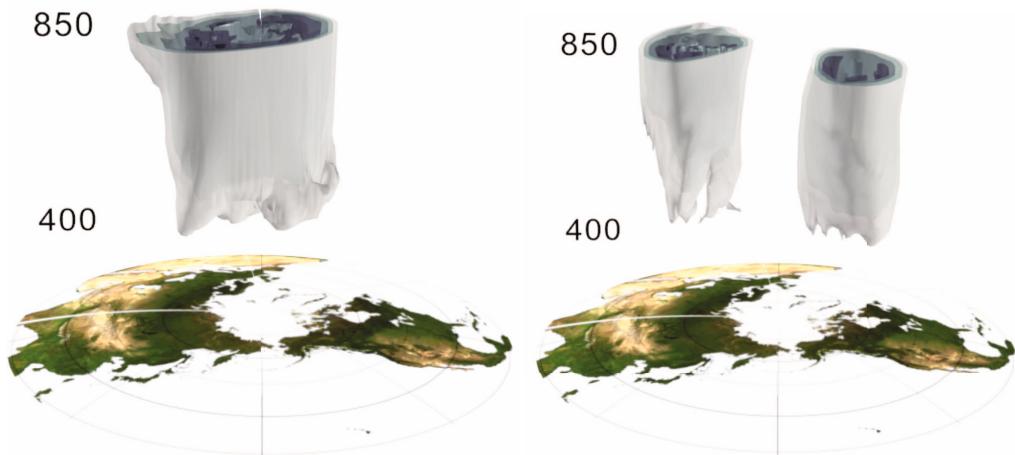


Fig. 17.22 The edge of the stratospheric polar vortex in December 1984. Plotted is the 35 PVU isosurface of $Q^* = Q(\theta/\theta_0)^{-4.5}$, where Q is Ertel PV and $\theta_0 = 475\text{K}$. The vertical coordinate is potential temperature. Like Q , Q^* is materially conserved in adiabatic flow, and roughly compensates for the change in density with height that affects the Ertel PV. The left panel (14 December) shows the vortex in a fairly usual state, and the right panel (30 December) shows a split vortex following a stratospheric sudden warming.²⁶

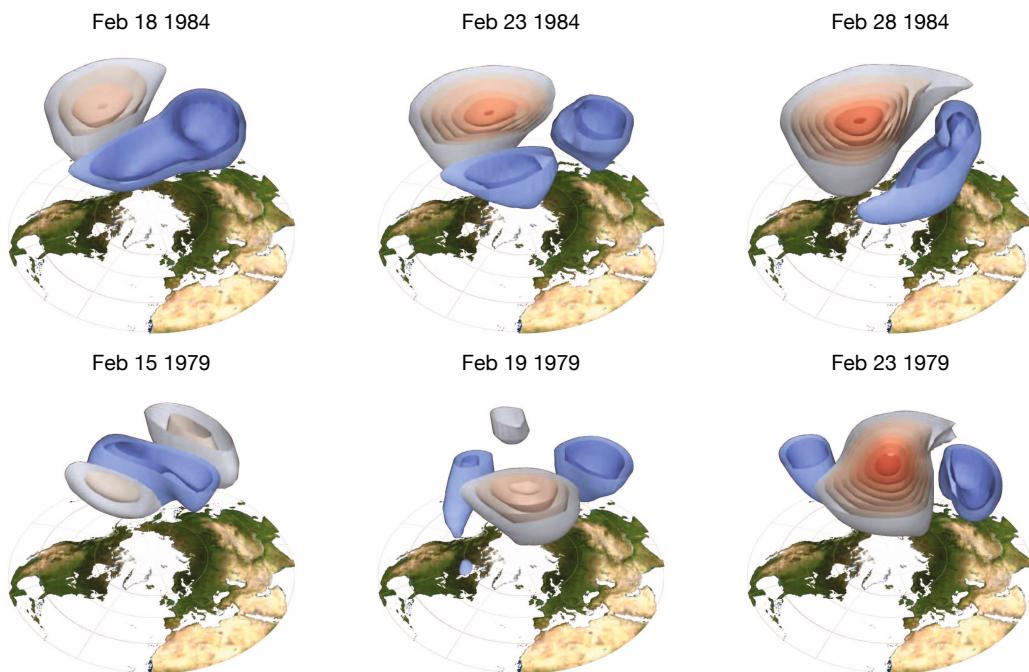


Fig. 17.23 Time sequence of two stratospheric warmings, with the top row showing a displacement of the initial (blue) polar vortex, and the bottom showing a split, with the dates marked. Contours are anomalous geopotential height larger (smaller) than 4km (-4km), between 200 and 10 hPa, spaced at 2km intervals. In both cases the initial polar vortex is cold (blue) with an anomalously warm end state (red).²⁷

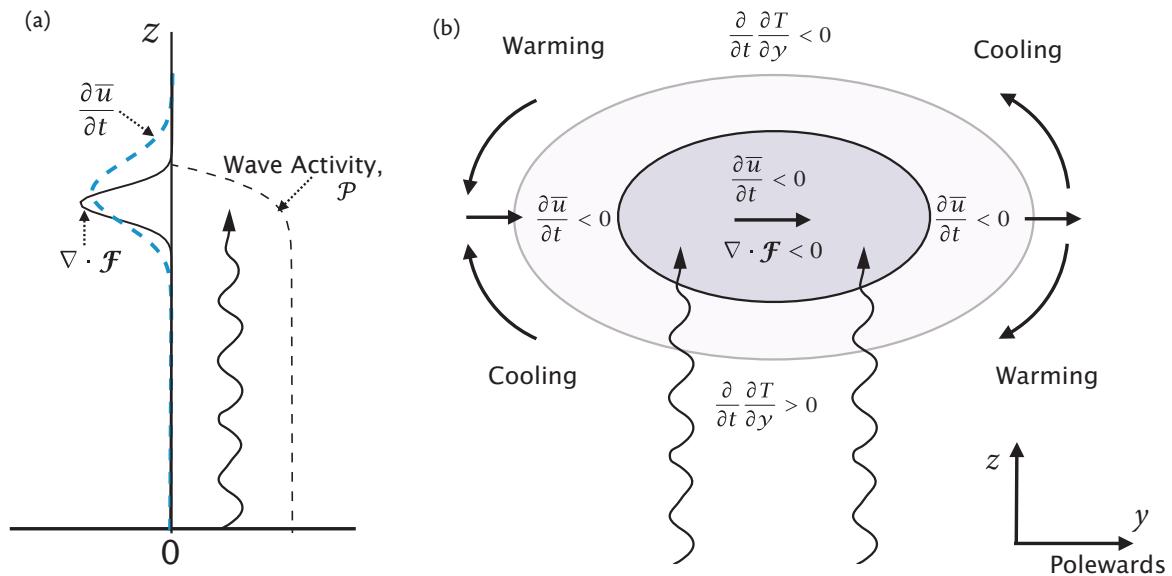


Fig. 17.24 The wave–mean-flow dynamics of a stratospheric warming. (a) Upward propagating Rossby waves (wavy lines) break in the stratosphere, and the wave activity (dashed line) diminishes. The EP flux divergence is negative ($\nabla \cdot \mathcal{F} < 0$), inducing a westward acceleration ($\partial \bar{u} / \partial t < 0$) over a broader region because of the ellipticity in (17.72). (b) The circulation and temperature response induced. The negative EP flux divergence is shaded dark, inducing westward acceleration over the broader region (light shading). Assuming there is no acceleration far away from the wavebreaking, the temperature response can be inferred from thermal wind, with a warming at the lower poleward end of the breaking and an induced residual circulation as shown by the arrows.

than the forcing itself, as can be seen from (10.87),

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z} \left(\frac{f_0^2}{N^2} \frac{\partial}{\partial z} \right) \right] \frac{\partial \bar{u}}{\partial t} = \frac{\partial^2}{\partial y^2} \bar{v}' q'. \quad (17.72)$$

The elliptic nature of the operator acting on $\partial \bar{u} / \partial t$ produces a response on a larger scale than the right-hand side.

Suppose, then, that Rossby waves propagate upward from the troposphere and break in the stratosphere. The mean eastward flow (sometimes called the polar night jet) will be weakened, so allowing more waves to propagate up, since strong eastward flow inhibits propagation (Fig. 16.6). If the process continues the winds will eventually reverse, forming a critical layer (as described in Section 16.6.1) where $\bar{u} = 0$. This completely inhibits further upward propagation and wave breaking is intensified, inducing a westward flow at the level to which the propagation reaches. There is a rapid changeover to westward flow and the critical layer descends. This sequence has an obvious similarity with the westward acceleration phase of the QBO, but in the extra-tropics there is no eastward counterpart as there are no Kelvin waves, and thus no oscillation. Rather, the eastward winds of the polar night jet are gradually restored by radiative effects.

A reduced (or reversed) vertical shear is, by thermal wind, associated with a reduced (or reversed) meridional temperature gradient, so that the polar night jet is replaced by a warmer westward flow. Put simply, the deposition of westward wave momentum leads to a warming of the high-latitude stratosphere. Such an event can at times be strong enough to split asunder the cold polar vortex, as illustrated in Figs. 17.22 and 17.23, and when it does the event is known as a *sudden stratospheric warming*.²⁸

The interaction of the waves and mean-flow is sketched in Fig. 17.24. In the left panel we see a wave propagating up from the troposphere and breaking, with wave activity then falling. The EP flux is negative in the breaking region causing a deceleration of the zonal flow over a somewhat broader region because of the elliptic operator in (17.72). The temperature response, shown in the right panel, can be inferred from thermal wind balance, noting that above the breaking region $\partial_t(\partial u/\partial z) > 0$ so that $\partial_t(\partial T/\partial y) < 0$, and oppositely for below. The direction of the residual circulation then follows by noting that adiabatic warming (cooling) results from descent (ascent). The residual circulation may also be inferred from (17.38b) or (17.38c).

Numerical simulations

To illustrate the above mechanism in a more realistic setting we show some results from a primitive equation simulation that mimics the broad features of the observations quite well. The advantage over showing the observations is that a great many events are simulated and full diagnostics can be obtained.²⁹ To obtain the results, the model (which had a well-resolved stratosphere) was integrated for many decades, during which time many sudden warmings occurred. Composites of these events are shown in Fig. 17.25.

The results show an anomalous upward flux of wave activity (EP flux) that, on dissipating in the stratosphere, induces a westward acceleration of the zonal flow, a weakening of the polar vortex and a warming of the polar regions, extending equatorward as far as 60°. It seems to be the condition of the stratosphere in filtering, or not, upward propagating waves that determines whether or not a warming occurs, rather than anomalous bursts originating in the troposphere. Vortex dynamics also play a role: a vortex, once formed, is rather stable and has a natural tendency to persist rather than break up; in two-dimensional turbulence vortices tend to merge and not split. This stability prevents warmings from occurring too frequently, since the wave activity must be strong enough to overcome the elastic properties of the vortex edge.

17.7.2 ♦ Wave–Mean-Flow Interaction and Stratospheric Variability

Stratospheric variability need not arise solely from waves propagating up from the tropopause, and we can illustrate this with a simple numerical model of wave–mean-flow interaction, similar to those discussed in Section 10.1.3. Specifically the model consists of the following quasi-geostrophic ideal-gas equations.³⁰ The zonally-averaged fields obey

$$\frac{\partial \bar{q}}{\partial t} = \bar{F} - \frac{\partial}{\partial y} \bar{v}' q', \quad (17.73a)$$

where

$$\bar{q}(y, z, t) - \beta y = \left[\frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{f_0^2}{N^2} \frac{\partial \Psi}{\partial z} \right) + \frac{\partial^2 \Psi}{\partial y^2} \right], \quad \left(\bar{u}, \frac{R}{H f_0} \bar{T} \right) = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial z} \right). \quad (17.73b,c)$$

The eddies obey

$$\frac{\partial q'}{\partial t} + \bar{u} \frac{\partial q'}{\partial x} + v' \frac{\partial \bar{q}}{\partial y} = F', \quad (17.74a)$$

where

$$q'(x, y, z, t) = \nabla^2 \psi' + \frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{f_0^2}{N^2} \frac{\partial \psi'}{\partial x} \right), \quad (u', v') = \left(-\frac{\partial \psi'}{\partial y}, \frac{\partial \psi'}{\partial x} \right) \quad (17.74b)$$

and

$$\frac{\partial \bar{q}}{\partial y} = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{f_0^2}{N^2} \frac{\partial \bar{u}}{\partial z} \right). \quad (17.74c)$$

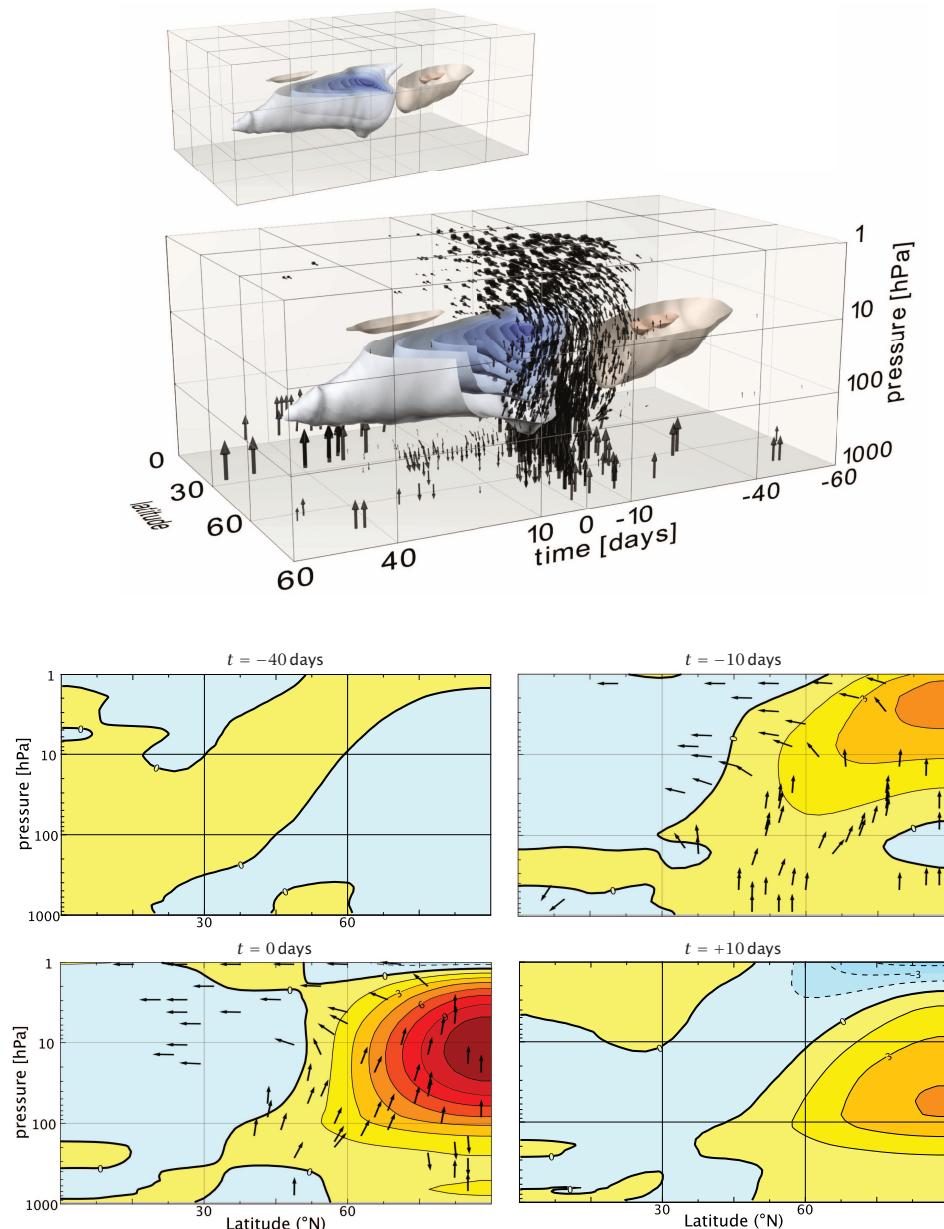


Fig. 17.25 The evolution of a stratospheric warming obtained with a general circulation model, compositing several events together. The top plots show anomalous zonal mean zonal wind, with surface intervals of 2 m/s, with time running from right to left. The bottom plots show time slices of anomalous temperature in colour contours (interval 1.5 K, red is warm), with times relative to the peak warming. In both plots the black arrows show anomalous EP flux.

There is a strengthening and slight northward propagation of the polar vortex prior to the onset, and a strong weakening during and after the onset. At forty days prior to the onset (upper left panel in lower set of plots), anomalies are very small. Ten days before onset (upper right), warming has appeared in the polar upper stratosphere, and anomalously strong EP fluxes appear throughout the atmosphere. At the onset (lower left), the warming and the EP flux anomalies are strongest. Ten days later (lower right), the temperature anomalies are weaker and confined to the lower stratosphere, and the EP fluxes are very weak.³¹

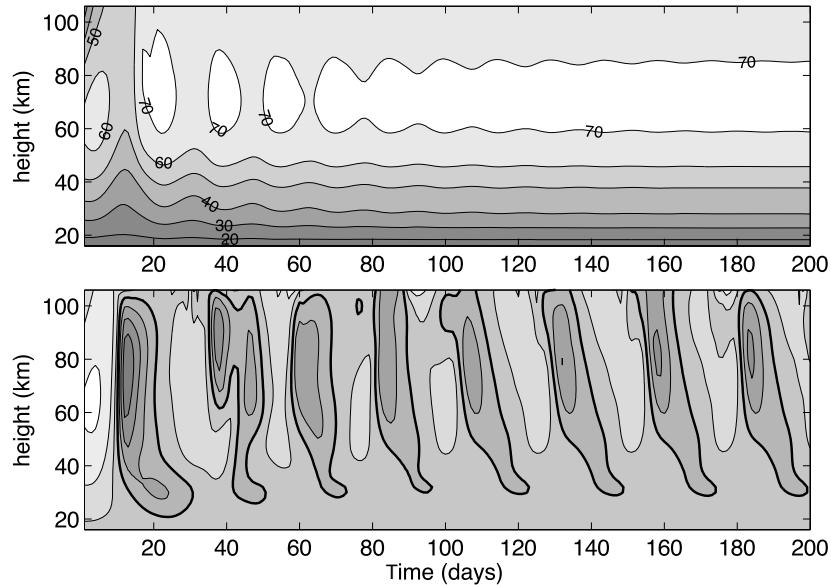


Fig. 17.26 Evolution of the zonal mean zonal wind in a case with steady wave forcing with a value of 200 in the top panel and 300 in the bottom panel. In the bottom panel the contours are every 20 m s^{-1} , positive values have lighter shades and the zero contour is heavy.

The notation follows our usual conventions, with $\rho_R(z)$ being a reference density profile and \bar{F} and F' the forcing/dissipation terms for the mean flow and eddies, respectively. It is the domain, the boundary conditions and forcing that distinguish the model and make it representative of the stratosphere, as we now discuss.

The model domain is a channel nominally centred at 60° and of width 60° , extending upward to about 100 km. The forcing on the zonal flow is a relaxation back to a specified radiative equilibrium temperature field (or equivalently a thermal wind field). In the results shown below this is independent of time and corresponds to a constant shear of 1 m s^{-1} per kilometre, or a temperature difference of about 15 K across the domain, with a relaxation timescale that varies from 20 days at 20 km to 4 days at 50 km. There is also a weak linear drag on the mean flow. The eddies are forced by imposing a constant perturbation at the lower boundary, with wavenumber 2 in the simulations shown. There is a radiative damping on the eddies ensuring that the eddies are mostly damped before reaching the top of the domain. The vertical variations are represented using finite differencing, whereas in the horizontal both the mean flow and the eddies are expanded in a Fourier series with only a very small number of terms retained. Thus, we write

$$[\bar{q}, \Psi] = [Q_0(z, t), \Psi_0(z, t)] \cos ly, \quad [q', \psi'] = \operatorname{Re} [q_0(z, t), \psi_0(z, t)] \sin ly \exp(ikx), \quad (17.75)$$

and after some manipulation we can obtain evolution equations for Q_0 and q_0 with diagnostic equations for Ψ_0 and ψ_0 . The quadratic terms in the equations of motion create higher order terms that are projected back onto the retained terms. (Aside from the severe horizontal truncation the numerical method used to find results is not a key aspect of the model.)

Some numerical results and interpretation

Results of two numerical integrations are shown in Fig. 17.26 and Fig. 17.27. In one integration the geopotential forcing at the lower boundary has an amplitude of 200 m, whereas in the other it has an amplitude of 300 m. In the first case the flow evolves into an absolute steady state, whereas in the second the mean flow and the waves oscillate with a period of about 25 days, with the mean flow actually becoming negative over half the cycle. The streamfunction in the unsteady case is tilting into the mean shear (the right panel of Fig. 17.27), evocative of baroclinic instability. The oscillations are in some way redolent of stratospheric warmings. The climatological eastward winds transition

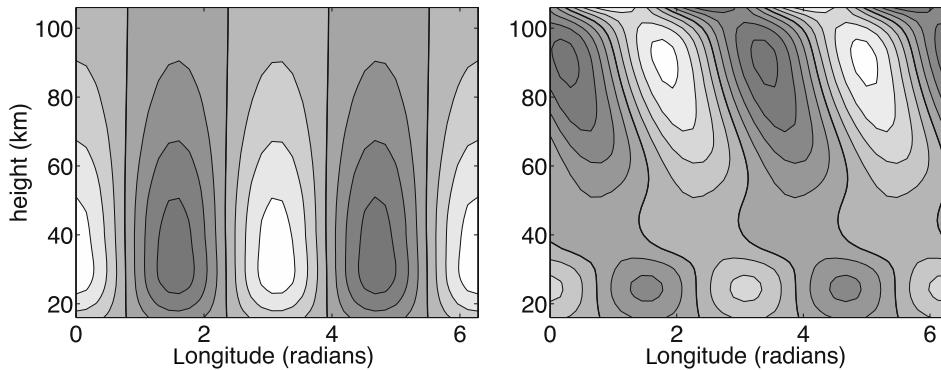


Fig. 17.27 Snapshots of the wave streamfunction in the steady case (forcing of 200) in the left panel and in the oscillatory case (forcing of 300) in the right panel. Zero contour is heavy.

rather quickly to westward winds (darker shading in Fig. 17.26), with the westward winds then descending and with a slower recovery back to climatology. There are two points to be made:

- (i) Interactions that are internal to the stratosphere can give rise to oscillatory motion.
- (ii) The source of energy for the waves may in part arise from a baroclinic instability and in part from tropospheric forcing.

To summarize, stratospheric warmings are not solely a response to tropospheric forcing; the internal dynamics of the stratosphere, and stratosphere-troposphere interactions, can also play a role, albeit a secondary one.

Notes

- 1 To read more about the middle atmosphere see, for example, the review by Hamilton (1998), the collection of articles in *Journal of the Meteorological Society of Japan*, vol. 80, no. 4B, 2002, the book by Andrews *et al.* (1987) and the review by Haynes (2005). I am also grateful to Nili Harnik and Peter Haynes for many comments on this chapter.
- 2 The calculation is described in Jucker *et al.* (2013).
- 3 Upper figure courtesy of J. Wilson of GFDL, using data from Fleming *et al.* (1988). Lower figure uses data from the ECMWF ERA-interim reanalysis.
- 4 Courtesy of D. Waugh.
- 5 Courtesy of A. Dörnbrack.
- 6 Adapted from Eluszkiewicz *et al.* (1997).
- 7 Brewer (1949) and Dobson (1956). Brewer deduced upward motion into the stratosphere at low latitudes based on the water vapour distribution, while Dobson deduced a poleward transport within the stratosphere based on the ozone distribution — the circulation takes ozone from the low latitudes toward the poles. Although originally the Brewer–Dobson circulation was taken to mean the chemical transport circulation, it is now usually taken to mean the residual (i.e., thickness weighted) overturning circulation. The two may differ if there is a mixing of chemicals without a mixing of mass, and the chemical transport may differ among chemicals.
- 8 See also Plumb (2002), which motivated this figure.
- 9 It turns out that $\overline{u'w'} > 0$ for upward propagating Rossby waves and thus, if the waves were to be dissipated, an eastward acceleration would seemingly be implied. In fact it is the form stress that is the most important aspect of vertical momentum transport in such waves, and when the waves are dissipated a westward acceleration ensues.

- 10 Blanca Ayarzagüena Porras kindly made this figure using an ERA-interim reanalysis.
- 11 Drawing from Garcia (1987) and Haynes (2005).
- 12 Adapted from Holton *et al.* (1995).
- 13 Values are taken from Wallace (1973), Plumb (1984) and Andrews *et al.* (1987).
- 14 The solution with no boundaries is not obvious from the description given here and the reader should consult the solution given in Haynes *et al.* (1991), as well as exploring the analogy further.
- 15 I am grateful to Verena Schenzer who kindly made this plot using winds from the ERA-interim reanalysis. See also Gray (2010).
- 16 A broad overview of QBO is provided by Baldwin *et al.* (2001), with updates and additions by Gray (2010), and a more theoretical review is given by Plumb (1984). The term quasi-biennial oscillation seems to have been coined by Angell & Korshover (1964), although the discovery of the QBO is generally credited to R. J. Reed and R. A. Ebdon, independently and at about the same time (Ebdon 1960, Veryard & Ebdon 1961, Reed 1960, Reed *et al.* 1961).
- 17 Figure adapted from Gray (2010), who used the method of Pascoe *et al.* (2005).
- 18 Wallace & Holton (1968), Baldwin *et al.* (2001).
- 19 The first theory of the QBO along these lines was put forward by Lindzen & Holton (1968) and Holton & Lindzen (1972), with clarifications and simplifications by Plumb (1977), and it is these models that we draw from. Prior to Lindzen & Holton's work, Booker & Bretherton (1967) had shown how the momentum deposition by gravity waves can be enhanced near critical lines and this was a key theoretical advance. A host of papers elaborating on the basic mechanism have since appeared, discussing such things as the particular type of gravity waves involved, the role of the Coriolis force and the meridional confinement of the QBO, the possible influence of the solar cycle and El Niño, the impact on tracer transport and so on.
- 20 Adapted from Plumb (1984).
- 21 Adapted from Plumb (1984).
- 22 Code is available to the reader.
- 23 Plumb & McEwan (1978). Regarding atmospheric relevance, and as with some other scientific theories of complex phenomena, it is hard to be absolutely certain that the theory is correct, for the sceptic can always point to observational disagreements or say that another theory might be the correct one. Sometimes the analogue of deciphering a complex communication may be apt: if an encrypted signal is deciphered to reveal a meaningful message, it may seem perverse to ask whether the deciphering is unique, and whether some other message might have emerged from a different decryption.
- 24 Examples of additional theoretical development are Dunkerton (1982, 1997), Boyd (1978) and Plumb & Bell (1982). Simulations of a QBO in an atmospheric GCM have been achieved by Takahashi (1996), Hamilton *et al.* (2001), Scaife *et al.* (2002), Giorgi et al. (2002) and others. The possible effects of the QBO on the extra-tropical circulation are discussed by Holton & Tan (1980, 1982), Jones *et al.* (1998), Kushner (2010), Labitzke *et al.* (2006), Randel *et al.* (1999), Scott & Haynes (1998), Dunkerton *et al.* (1988) and others. The anomalous QBO of 2015–2016 is documented by Newman *et al.* (2016) and Osprey *et al.* (2016).
- 25 This plot was kindly made by Varena Schenzer using data from the Singapore radiosonde. For a description of QBO datasets see <http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/>.
- 26 Figure kindly prepared by M. Jucker using ERA-interim reanalysis. See Lait (1994) for a discussion of the alternative PV.
- 27 Data from ECMWF ERA-Interim, visualization with the software 'pv-atmos', described in Jucker (2014).
- 28 The model described here was proposed by Matsuno (1971) and although nonlinear effects (and, to a lesser degree, non-geostrophic effects) play a quantitative role, Matsuno's model remains the

foundation of our understanding. An early review is that of Schoeberl (1978) and there have been numerous studies since. To name but a few, Dunkerton *et al.* (1981) and Palmer (1981) explored the phenomenon from a TEM perspective, Limpasuvan *et al.* (2000) and Charlton & Polvani (2007) provide a comprehensive view of the observations of warmings using reanalysis datasets, Charlton *et al.* (2007) look at various simulations with GCMS, and Gray *et al.* (2001) look at external influences on the timing of warmings.

- 29 These simulations were kindly performed by Martin Jucker. See also Jucker *et al.* (2014) and Jucker (2016).
- 30 This model was introduced by Holton & Mass (1976) and the numerical results we show use a code adapted from one by J. Holton. Plumb (1981), Yoden (1987, 1990) and others have explored the model further with a view to better understanding the parameters for which steady, oscillatory or chaotic motion was present. Christiansen (1999, 2000), Scaife & James (2000), Scott & Haynes (2000), Sjoberg & Birner (2012), Jucker *et al.* (2014) and others have explored related behaviour using various types of models and observations, for example with the primitive equations and/or including eddy-eddy interactions and with different boundary conditions.
- 31 Figure created using the software 'pv-atmos', described in Jucker (2014).

