3 Stratospheric Rossby waves

3.1 Vertical propagation on a β-plane

[Charney & Drazin, *J. Geophys. Res.*, **66**, p83, 1961; AHL]

We consider the properties of small-amplitude quasigeostrophic waves on a basic state consisting of static stability $S(z) = dT_0/dz + g/c_p = HN^2(z)/R$, zonal wind $U(z)$, potential vorticity $Q(y,z) = y\frac{\partial Q}{\partial y}(z)$ and in which the meridional circulation is zero (to leading order, *i.e.*, it would be zero if the waves were not there).

The linearized QGPV equation is

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right) q' + v'\frac{\partial Q}{\partial y} = \mathcal{X}\,'$$

where

$$\frac{\partial Q}{\partial y} = \beta - \Delta^2 U = \beta - \frac{1}{\rho}\frac{\partial}{\partial z}\left(\frac{\rho f_0^2}{N^2}\frac{\partial U}{\partial z}\right)$$

(since $U = U(z)$). We assume, for now, conservative flow so that $\mathcal{X}\,' = 0$.

$$q' = \Delta^2 \psi' = \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} + \frac{1}{\rho}\frac{\partial}{\partial z}\left(\frac{\rho f_0^2}{N^2}\frac{\partial \psi'}{\partial z}\right)$$

and so

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)\left(\frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} + \frac{1}{\rho}\frac{\partial}{\partial z}\left(\frac{\rho f_0^2}{N^2}\frac{\partial \psi'}{\partial z}\right)\right) + \frac{\partial \psi'}{\partial x}\frac{\partial Q}{\partial y} = 0 \; .$$

We look for solutions of the form

$$\psi' = \mathbf{R}e \; \Psi(z) e^{ik(x-ct)+i\ell y} \; ,$$

where $\Psi(z)$ is to be determined. Then we have

$$(U-c)\left(\frac{1}{\rho}\frac{d}{dz}\left(\frac{\rho f_0^2}{N^2}\frac{d\Psi}{dz}\right) - \left(k^2 + l^2\right)\Psi\right) + \frac{\partial Q}{\partial y}\Psi = 0 \; . \tag{1}$$

In general, (1) is a straightforward differential equation that can readily be solved numerically using standard techniques, given boundary conditions. But doing so does not give much insight. So, to keep things transparent, we focus on the case in which U and N^2 are independent of height (which means $\partial Q/\partial y = \beta$). Then (1) can be written

$$\frac{f_0^2}{N^2}\left(\frac{d^2\Psi}{dz^2} - \frac{1}{H}\frac{d\Psi}{dz}\right) - \left(k^2 + l^2\right)\Psi + (U-c)^{-1}\frac{\partial Q}{\partial y}\Psi = 0 \; . \tag{2}$$

The equation now has constant coefficients, and we can clean it up further by extracting a part of the solution that grows with height as $\exp(z/2H)$; specifically, we look for solutions of the form

$$\Psi(z) = \Psi_0 e^{z/2H} e^{imz}$$

so that the full solution is

$$\psi' = \mathrm{Re}\, \Psi_0 e^{i(kx+ly+mz)-ikct}\ e^{z/2H}\ .$$

Then from (2) we have

$$m^2 = \frac{N^2}{f_0^2}\left(\frac{\partial Q/\partial y}{(U-c)} - k^2 - l^2\right) - \frac{1}{4H^2}\ . \tag{3}$$

Alternatively, we may write this as

$$c - U = -\frac{\partial Q}{\partial y}\left(k^2 + l^2 + \frac{f_0^2}{N^2}m^2 + \frac{f_0^2}{4N^2H^2}\right)^{-1}\ .$$

Eq.(3) is the dispersion relation for three-dimensional Rossby waves of given k, ℓ, c in a stratified atmosphere.

If the $r.h.s$ of (3) is negative, m is pure imaginary, so that Ψ simply grows or decays with height without change of phase. If $m = \pm i\mu$,

$$\Psi = \Psi_0 e^{z/2H} \exp\left(\mp \mu z\right).$$

The kinetic energy density of the wave is

$$\frac{1}{2}\rho(\overline{u'^2} + \overline{v'^2}) = \frac{1}{4}\rho\left(k^2 + l^2\right)|\Psi|^2 = \frac{1}{4}\rho_0\left(k^2 + l^2\right)|\Psi_0|^2 \exp\left(\mp 2\mu z\right)\ .$$

We thus ensure boundedness as $z \to \infty$ by choosing the positive sign for μ.

If the $r.h.s.$ of (3) is positive, the vertical wavenumber m is real and so we have vertically-propagating waves. These may be upward or downward-propagating, depending on the sign of m. The group velocity in the meridional plane is $\mathbf{c}_g = (c_{gy}, c_{gz})$ where

$$c_{gy} = k\frac{\partial c}{\partial \ell} = 2kl\ \Lambda\ ;\ c_{gz} = k\frac{\partial c}{\partial m} = 2\frac{f_\circ^2}{N^2}km\ \Lambda\ ,$$

where

$$\Lambda = \frac{(U-c)^2}{\partial Q/\partial y}.$$

If we define $k > 0$ (which we are free to do without loss of generality), then, if $\partial Q/\partial y > 0$, we see that $m > 0$ implies upward propagation, and $m < 0$ downward propagation.

Since ψ' varies with height as $\psi' \simeq e^{imz}e^{z/2H}$,

$$|\psi'|^2 \simeq e^{z/H} \simeq \rho_0/\rho \ .$$

In the atmosphere, the amplitude of a vertically-propagating wave (m real) increases with height inversely as the square root of pressure. For a wave originating at the surface (1000 hPa) and propagating up to the stratopause (1 hPa), its amplitude would (other things being equal — we shall see later that they are not) increase thirtyfold. This must be the case in the absence of dissipation, as the EP theorem shows; we will address this below.

3.1.1 Wave propagation into the stratosphere and mesosphere

The Rossby wave dispersion relation gives us a lot of insight into the large-scale wave climatology of the middle atmosphere and, in particular, why much of the tropospheric wave activity is not evident there. Waves with $m^2 > 0$ will propagate readily upward and (to the extent that our conservative model is a realistic one) increase in amplitude with height. Those with $m^2 < 0$, on the other hand, will either increase in amplitude with height more slowly or, if $m^2 < -1/4H^2$, decrease with height and will therefore be somewhat weak relative to the propagating waves. We can easily see that m^2 will be positive if the "Charney-Drazin" condition:

$$0 < U - c < U_c \ ,$$

is satisfied where U_c—the "Rossby critical velocity"—is

$$U_c = \frac{\partial Q}{\partial y}\left(k^2 + l^2 + \frac{f_0^2}{4N^2H^2}\right)^{-1} .$$

There is thus a propagation "window" for the mean winds. Note that U_c decreases with increasing κ, so the window becomes narrow for small-scale waves. An example is shown—for constant U (so $\frac{\partial Q}{\partial y} = \beta$)—in Fig 1, using the values $H = 7km$, $N^2 = 4\ 10^{-4}s^{-2}$ and $f_0 = 10^{-4}s^{-1}$ and $\beta = 1.6\ 10^{-11}m^{-1}s^{-1}$.

For a synoptic scale wave with, say, $k = \ell = \pi/(1000km)$ whence $\kappa^2 = 1.96\ 10^{-11}m^{-2}$, the propagation window is tiny (about $1ms^{-1}$ wide); since U in fact varies considerably with height a wave of such scale must be evanescent almost everywhere, if not everywhere. In fact if $U - c \geq 10ms^{-1}$, μ turns out to be about $(1km)^{-1}$ so the wave amplitude decays much more rapidly than $e^{-z/2H}$; clearly we do not expect to see such waves in the middle atmosphere, except just above the tropopause. This, then, explains the absence of tropospherically-forced, synoptic-scale eddies in the stratosphere and mesosphere.

The window becomes wider for larger-scale waves. For the very largest planetary scale wave with 14000km and 6000km half-wavelength in the x and y directions respectively, U_c is about $35ms^{-1}$, which means it is easier for these waves to propagate. Note, however, that $U - c$ must be *positive*; for the quasistationary waves with $c \approx 0$, no waves can propagate

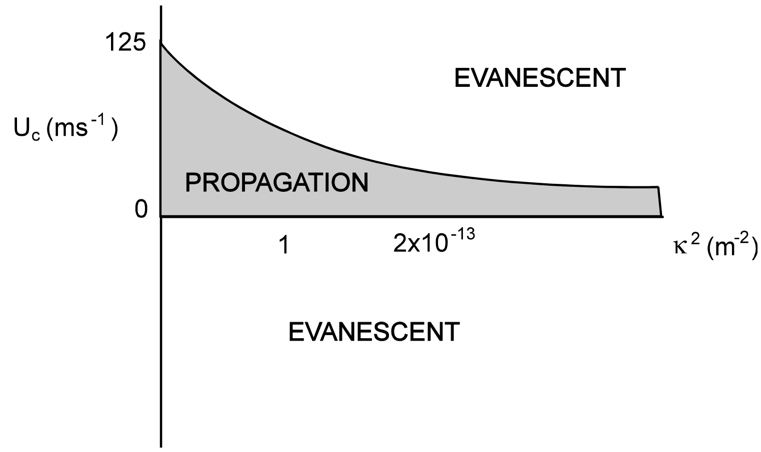


Figure 1: Rossby critical velocity U_c vs. $\kappa^2 = k^2 + l^2$.

through easterlies. Since we saw at the beginning that stratospheric winds are easterly in summer, we thus have an explanation for the absence of large-scale wave activity in summer. During northern hemisphere winter, planetary waves (of only the very large scales) may propagate through the westerlies (or at least through enough of them to reach and sustain large amplitudes)[1]. In the strong westerlies of midwinter, we do not expect to see deep propagation, but we seem to do so; why? Is this theory of waves on a β-plane and unifrom flow inadequate? There is a notable difference between the northern and southern hemispheres, however. While waves are present right through the northern winter (albeit with large amplitude fluctuations) there is a double-peaked structure to the southern hemisphere cycle, with a midwinter minimum. This is not yet fully explained, but it seems likely that the reason for this minimum lies in the fact that the midwinter westerlies are, as we saw, stronger than in the northern hemisphere. This may mean that $U > U_c$ and therefore propagation is forbidden by these arguments, because the westerlies are simply too strong. In the northern hemisphere winter, the observations of deep wave penetration may indicate shortcomings of this theory, or maybe the waves are evanescent, but the decay ($e^{-\mu z}$) is overcome by the density factor $e^{z/2H}$.

3.1.2 Eliassen-Palm fluxes

Now,

$$(F_y, F_z) = \left(\rho \overline{\frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial y}}, \frac{\rho f_0}{N^2} \overline{\frac{\partial \psi'}{\partial x} \frac{\partial \psi'}{\partial z}} \right) .$$

[1] But Charney & Drazin suggested that they can propagate only in spring and fall, when the westerlies are weak enough to permit deep propagation.

Using our expression for ψ' in the propagating regime (real m), and writing $\varphi = kx+ly+mz$ for shorthand, we have

$$\begin{aligned}\frac{\partial\psi'}{\partial x} &= \mathbf{R}e\ ik\Psi_0 e^{z/2H}\ e^{i\varphi}, \\ \frac{\partial\psi'}{\partial y} &= \mathbf{R}e\ i\ell\Psi_0 e^{z/2H}\ e^{i\varphi}\ , \\ \frac{\partial\psi'}{\partial z} &= \mathbf{R}e(im+\frac{1}{2H})\Psi_0 e^{z/2H} e^{i\varphi}\ ,\end{aligned}$$

whence

$$F_y = \tfrac{1}{2}\rho_0\ k\ell|\Psi_0|^2\ ,$$

and

$$F_z = \tfrac{1}{2}\rho_0\ km\frac{f_0^2}{N^2}|\Psi_0|^2\ .$$

[Since F_y is independent of y, $\nabla\cdot\mathbf{F} = \partial F_z/\partial z = 0$ in this case, from the EP theorem. Note that we can interpret the $e^{z/2H}$ growth of amplitude with height as a manifestation of constant EP flux in the presence of decreasing ρ.]

Now, the wave activity density is

$$A = \frac{\rho}{2}\frac{\overline{q'^2}}{\partial Q/\partial y}$$

and

$$q' = -\left(k^2+l^2+\frac{f_0^2}{N^2}(m^2+\frac{1}{4H^2})\right)\mathbf{R}e\ \Psi_0 e^{i\varphi}e^{z/2H}\ ,$$

whence

$$A \cong \frac{\rho_0}{4\Lambda}|\Psi_0|^2\ .$$

Therefore

$$F = (F_y, F_z) = \left(2k\ell\Lambda,\ 2\frac{f_0^2}{N^2}km\Lambda\right)A = (c_{gy}, c_{gz})A\ ;$$

as we noted (but did not prove) before, the EP flux is equal to the wave activity density times group velocity.

For an upward propagating wave with $k>0$ ($c_{gz}>0$, $m>0$), the wave structure in the $(x-z)$ plane is qualitatively as shown in Fig 2.

In reality, the growth of atmospheric waves with height is limited by dissipation, by refraction (the waves spread out in latitude and therefore the wave activity is diluted) and, perhaps most of all, nonlinear effects (as we shall see later). Nevertheless, planetary wave

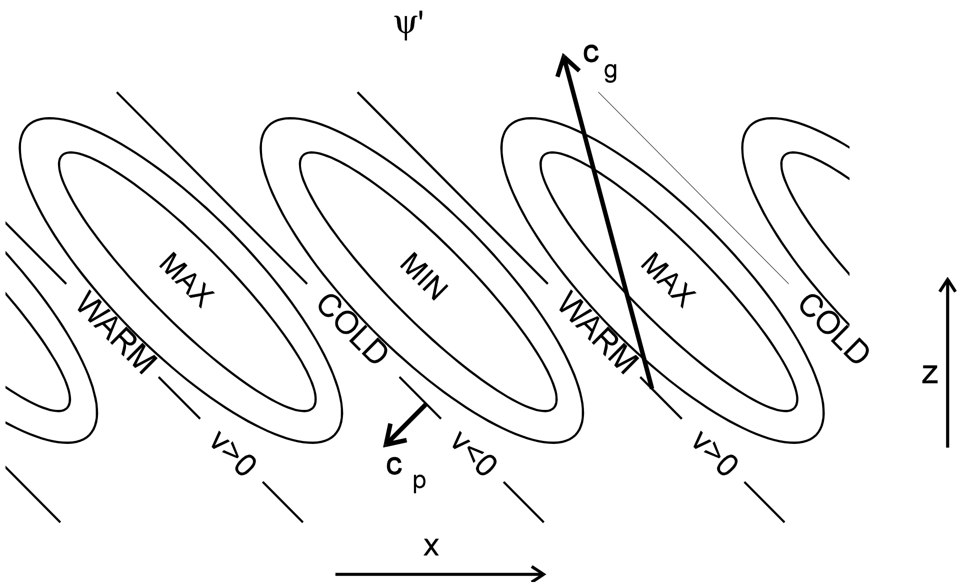


Figure 2: A vertically propagating planetary wave in the northern hemisphere (WARM/COLD are reversed in the S Hem). Note that this should be multiplied by $\exp(z/2H)$ to give the actual result for ψ'.

amplitudes can become large at high levels. Note the westward tilt with height ($km > 0$), and the fact that (for the northern hemisphere) the northward velocity fluctuation is in phase with the temperature fluctuation, *i.e.*, $\overline{v'T'} > 0$; this implies that F_z is positive, which it must of course be for an upward propagating wave. In general the requirement that F_z be upward means that an upward (/downward) propagating wave necessarily has poleward (/equatorward) eddy heat flux.

3.2 Waves on a sphere with nonuniform U

[Matsuno, JAS, **27**, 871-883, 1970]

3.2.1 Basic equations

Basic equations for log-pressure coordinates in spherical geometry (under "primitive equation" assumptions of a thin atmosphere):

$$\begin{aligned}
\frac{Du}{Dt} - \frac{1}{a}\, uv \;\; tan\varphi - 2\Omega v \; sin\varphi &= -\frac{1}{a \; cos\varphi}\frac{\partial \Phi}{\partial \lambda} + \mathcal{G}_\lambda \;, \\
\frac{Dv}{Dt} + \frac{1}{a} u^2 tan\varphi + 2\Omega u \; sin\varphi &= -\frac{1}{a}\frac{\partial \Phi}{\partial \varphi} + \mathcal{G}_\varphi \;, \\
\frac{D\theta}{Dt} &= (\rho\Pi)^{-1}\, \mathcal{J} \;,
\end{aligned}$$

where (λ, φ) are longitude and latitude, a and Ω the Earth radius and rotation rate and where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{u}{a \; cos\varphi}\frac{\partial}{\partial \lambda} + \frac{v}{a}\frac{\partial}{\partial \varphi} + w\frac{\partial}{\partial z} \;.$$

The hydrostatic and continuity equations are as before:

$$\frac{\partial \Phi}{\partial z} = \frac{\kappa \Pi}{H}\tilde{\theta}$$

and

$$\nabla \cdot \rho \mathbf{u} = 0 \;,$$

where, for any vector $\mathbf{U} = (U, V, W)$,

$$\nabla \cdot \mathbf{U} \equiv \frac{1}{a \; cos\varphi}\frac{\partial U}{\partial \lambda} + \frac{1}{a \; cos\varphi}\frac{\partial}{\partial \varphi}(V \cos\varphi) + \frac{\partial W}{\partial z}.$$

3.2.2 Quasigeostrophic equations

We need a set of equations under quasigeostrophic scaling valid over most of the sphere (quasigeostrophic equations are never going to be valid very close to the equator). As usual for QG theory, as a small parameter for the scaling we choose $\epsilon = L/a \ll 1$ where L is the *latitudinal* length scale of the motion. (In fact all trigonometric functions of φ—including the Coriolis parameter—are then slowly-varying on the scale L). We further assume that Rossby number $U/\Omega L \leq O(\epsilon)$ where U is a horizontal velocity scale; that time-dependence is slow, *i.e.*, $(\Omega\tau)^{-1} \leq O(\epsilon)$,where τ is a characteristic time-scale; and that the isentropic slope is small: $|\nabla_h\theta|/|\frac{\partial\theta}{\partial z}| \leq O(\epsilon) \rightarrow |\nabla_h T|/\ S \leq O(\epsilon)$. In fact, this now means that S may be not only a function of z (as usual) but also may be a *slowly-varying* function of horizontal position (*i.e.*, it may vary on the *planetary* scale). Finally, we assume that frictional effects are weak at leading order.

The leading order terms in the horizontal momentum equations then become

$$\begin{aligned} 2\Omega v\ sin\varphi &= \frac{1}{a\ cos\varphi}\frac{\partial\Phi}{\partial\lambda} , \\ 2\Omega u\ sin\varphi &= -\frac{1}{a}\frac{\partial\Phi}{\partial\varphi} , \end{aligned}$$

where Φ is now the departure from the horizontally uniform reference state. It is tempting (and common) to use these equations as definitions of the geostrophic velocities. However this causes problems at next order in ϵ, since then

$$\nabla\cdot\mathbf{u} = -\frac{1}{2\Omega a^2\sin^2\varphi}\frac{\partial\Phi}{\partial\lambda} = -\frac{1}{a}\cot\varphi\ v\ .$$

So the geostrophic wind is not nondivergent, though the divergence is ageostrophic (*i.e.*, it is $O(v/a) \approx \epsilon v/L$). To proceed in this way one has to remember to add a term $-a^{-1}\cot\varphi\ v$ to the continuity equation for the ageostrophic wind—something that is messy and inconvenient (and is not always remembered). An alternative approach is to *insist* at the outset on a definition of geostrophic wind that is exactly nondivergent. One way of doing this is to replace the Coriolis parameter in the geostrophic equations with $2\Omega sin\varphi_0$, where φ_0 is a representative latitude. But then the balance is only valid in the immediate vicinity of φ_0—in which case one might as well use a beta-plane approach.

Another way of defining a nondivergent geostrophic velocity is as follows. In the full equations of motion, we use the identity

$$sin\varphi\frac{\partial}{\partial\varphi}\left(\frac{\Phi}{\sin\ \varphi}\right) = \frac{\partial\Phi}{\partial\varphi} - \Phi\cot\varphi\ ,$$

where we note that the second term is $O(\epsilon)$ smaller than the first [since $\partial\Phi/\partial\varphi \approx \frac{a}{L}\Phi$].

Therefore we may rewrite the equations of motion as

$$
\begin{aligned}
\frac{Du}{Dt} - \frac{1}{a}uv \ \ tan\varphi - 2\Omega v \ sin\varphi &= -\frac{tan\varphi}{a}\frac{\partial}{\partial\lambda}\left(\frac{\Phi}{sin\varphi}\right) + \mathcal{G}_\lambda \ , \\
\frac{Dv}{Dt} + \frac{1}{a}u^2 tan\varphi + 2\Omega u \ sin\varphi &= -\frac{sin\varphi}{a}\frac{\partial}{\partial\varphi}\left(\frac{\Phi}{sin\varphi}\right) - \frac{cot\varphi}{a}\Phi + \mathcal{G}_\varphi \ .
\end{aligned}
$$

Then our geostrophic equations are naturally written

$$
v = \frac{1}{acos\varphi}\frac{\partial\psi}{\partial\lambda} \ ; \ u = -\frac{1}{a}\frac{\partial\psi}{\partial\varphi} \ ,
$$

where

$$
\psi = \frac{\Phi}{2\Omega \ sin\varphi} \ .
$$

We have thus arrived at a definition of an exactly nondivergent geostrophic velocity, for which ψ is the geostrophic streamfunction. The price we pay for this is an extra term in the v-momentum equation (which in my opinion is preferable to having a corresponding term in the continuity equation—it is a matter of convenience and taste). The hydrostatic balance equation is simply

$$
\frac{\partial\psi}{\partial z} = \frac{g}{2\Omega \ \sin\varphi}\frac{\kappa\Pi}{H}\tilde{\theta} \ .
$$

The quasigeostrophic equations (to next order in ϵ) now become

$$
\begin{aligned}
D_g u - \frac{1}{a}uv\tan\varphi - 2\Omega v_a \ \sin\varphi &= \mathcal{G}_\lambda \ , \\
D_g v + \frac{1}{a}u^2\tan\varphi + 2\Omega u_a \sin\varphi &= -\frac{2\Omega\psi}{a}\cos\varphi + \mathcal{G}_\varphi \ , \\
D_g\tilde{\theta} + w_a\frac{d\Theta_0}{dz} &= (\rho\Pi)^{-1}\mathcal{J} \ .
\end{aligned}
$$

where D_g is the time derivative following the geostrophic flow

$$
D_g \equiv \frac{\partial}{\partial t} + \frac{u}{a \ \cos\varphi}\frac{\partial}{\partial\lambda} + \frac{v}{a}\frac{\partial}{\partial\varphi} \ ,
$$

and the ageostrophic continuity equation is

$$
\nabla \cdot \rho\mathbf{u}_a = 0 \ .
$$

3.2.3 QGPV equation

In spherical geometry, vorticity is defined by

$$\zeta = \frac{1}{a\ cos\varphi}\left(\frac{\partial v}{\partial \lambda} - \frac{\partial}{\partial \varphi}(u\ \cos\varphi)\right) = \frac{1}{a^2\ \cos^2\varphi}\frac{\partial^2\psi}{\partial\lambda^2} + \frac{1}{a^2\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial\psi}{\partial\varphi}\right) \ ,$$

and potential vorticity by

$$q = 2\Omega\ \sin\varphi + \Delta\psi \ ,$$

where

$$\Delta \equiv \frac{1}{a^2\cos^2\varphi}\frac{\partial^2}{\partial\lambda^2} + \frac{1}{a^2\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial}{\partial\varphi}\right) + \frac{1}{\rho}\frac{\partial}{\partial z}\left(\rho\frac{4\Omega^2}{N^2}\sin^2\varphi\frac{\partial}{\partial z}\right) \ .$$

We form the vorticity equation by taking the curl of the equations of motion as usual. We use the identities

$$D_g\left(\frac{1}{\cos\ \varphi}\frac{\partial}{\partial\varphi}(u\cos\varphi)\right) = \frac{1}{\cos\ \varphi}\left(\frac{\partial}{\partial\varphi}[\cos\varphi\{D_g u - \frac{uv}{a}\tan\varphi\}] - \frac{\tan\ \varphi}{a}\frac{\partial}{\partial\lambda}(u^2)\right)$$

$$D_g\left(\frac{1}{\cos\ \varphi}\frac{\partial v}{\partial\lambda}\right) = \frac{1}{\cos\ \varphi}\frac{\partial}{\partial\lambda}[D_g v] \ .$$

to obtain the result

$$\begin{aligned} & D_g\zeta + \frac{2\Omega v}{a}\cos\varphi + \frac{2\Omega}{a\ \cos\varphi}\left(\frac{\partial}{\partial\lambda}(u_a\sin\varphi) + \frac{\partial}{\partial\varphi}(v_a\sin\varphi\ \cos\varphi)\right) \\ = & \frac{1}{a\ \cos\varphi}\left(\frac{\partial\mathcal{G}_\varphi}{\partial\lambda} - \frac{\partial}{\partial\varphi}(\mathcal{G}_\lambda\cos\varphi)\right) \end{aligned}$$

Note that the third term on the lhs is not (as in the Cartesian case) the horizontal divergence of $\mathbf{u}_a$, but of $(\mathbf{u}_a\sin\varphi)$—the $\sin\varphi$ comes from the Coriolis term. Therefore this term is

$$\frac{2\Omega\ \sin\varphi}{a\ \cos\varphi}\left(\frac{\partial u_a}{\partial\lambda} + \frac{\partial}{\partial\varphi}(v_a\cos\varphi)\right) + \frac{2\Omega v_a}{a}\cos\varphi = \frac{2\Omega v_a}{a}\cos\varphi - \frac{2\Omega\ \sin\varphi}{\rho}\frac{\partial}{\partial z}(\rho w_a) \ .$$

Therefore the vorticity equation becomes

$$\begin{aligned} & D_g\zeta + \frac{2\Omega v}{a}\cos\varphi + \frac{2\Omega v_a}{a}\cos\varphi - \frac{2\Omega\ \sin\varphi}{\rho}\frac{\partial}{\partial z}(\rho w_a) \\ = & \frac{1}{a\ \cos\varphi}\left(\frac{\partial\mathcal{G}_\varphi}{\partial\lambda} - \frac{\partial}{\partial\varphi}(\mathcal{G}_\lambda\cos\varphi)\right) \end{aligned}$$

The extra term we have gained—the advection of planetary vorticity by the ageostrophic wind—is formally small (being $O(\epsilon)$ smaller than the second term) and so we may neglect it. [This term was, in part, retained by Matsuno (1970) and so his final result is slightly different from ours.] Therefore the vorticity equation becomes

$$D_g(\zeta + 2\Omega sin\varphi) - \frac{2\Omega\ sin\varphi}{p}\frac{\partial}{\partial z}(pw_a) = \frac{1}{a\ cos\varphi}\left(\frac{\partial \mathcal{G}}{\partial \lambda}\varphi - \frac{\partial}{\partial \varphi}(\mathcal{G}_\lambda cos\varphi)\right)\ .$$

Since

$$w_a = \frac{1}{\Theta_{0,z}}\left(\left(\rho\Pi^{-1}\right)\mathcal{J} - D_g\tilde{\theta}\right) \approx \frac{(\rho\Pi)^{-1}\mathcal{J}}{\Theta_{0,z}} - D_g\left(\frac{\tilde{\theta}}{\Theta_{0,z}}\right),$$

we finally have the quasigeostrophic potential vorticity equation

$$D_g q = \mathcal{X}\ ,$$

where the potential vorticity is

$$q = 2\Omega\sin\varphi + \zeta + \frac{2\Omega\sin\varphi}{\rho}\frac{\partial}{\partial z}\left(\rho\frac{\tilde{\theta}}{\Theta_{0,z}}\right) \equiv 2\Omega\ \sin\ \varphi + \Delta\psi\ ,$$

and the nonconservative source of potential vorticity is

$$\mathcal{X} = \frac{1}{a\ \cos\varphi}\left(\frac{\partial \mathcal{G}_\varphi}{\partial \lambda} - \frac{\partial}{\partial \varphi}(\mathcal{G}_\lambda\cos\varphi)\right) + \frac{2\Omega\ \sin\varphi}{\rho}\frac{\partial}{\partial z}(\frac{\mathcal{J}}{\Pi\Theta_{0,z}})\ .$$

3.2.4 Waves on a zonal flow

We now as usual consider the problem of small-amplitude linear waves on a zonal flow $U(\varphi, z)$ $[V = W = 0]$ with uniform static stability N^2 and a corresponding potential vorticity $Q(\varphi, z)$ such that

$$\frac{1}{a}\frac{\partial Q}{\partial \varphi} = \frac{2\ \Omega}{a}\cos\varphi - \Delta U\ ,$$

where we have again assumed that φ is slowly-varying compared to U. The linearized potential vorticity equation is then

$$\left(\frac{\partial}{\partial t} + \omega\frac{\partial}{\partial \lambda}\right)q^{'} + \frac{v'}{a}\frac{\partial Q}{\partial \varphi} = \mathcal{X}\ '\ ,$$

where $\omega = U/(a\ cos\varphi)$ is the angular velocity of the mean flow.

If we look for forced, stationary $(\partial/\partial t = 0)$ waves of zonal wavenumber s $(\partial/\partial\lambda \equiv is)$, then

$$is\omega\ \Delta\psi' + \frac{\partial Q/\partial\varphi}{a^2}\cos\varphi\ is\psi' = \mathcal{X}\ '.$$

If dissipation and internal forcing are zero so that $\mathcal{X}\,' = 0$, then this may be written as

$$\mathcal{O}(\Psi) + \nu_s \Psi = 0 \ ,$$

where

$$\psi' = \mathbf{R}e \ \Psi(\varphi, z) e^{is\lambda} e^{z/2H} \ ,$$

and where the operator $\mathcal{O}$ is

$$\mathcal{O} \equiv \frac{1}{\cos\varphi} \frac{\partial}{\partial\varphi} \left(\cos\varphi \frac{\partial}{\partial\varphi} \right) + h^2 \sin^2\varphi \frac{\partial^2}{\partial z^2}$$

where[2] $h = 2\Omega a/N$. $\mathcal{O}$ has the character of a dimensionless Laplacian in φ and z. Finally, ν_s, the "refractive index square" is

$$\nu_s = \frac{1}{\omega \ \cos\varphi} \frac{\partial Q}{\partial\varphi} - \frac{h^2}{4H^2} \sin^2\varphi - \frac{s^2}{\cos^2\varphi} \ .$$

N.B. ν_s is singular at the "critical line" where $\omega = U/\left(a\cos\varphi\right) = 0$. We can remove the singularity there by adding dissipation (nonzero $\mathcal{X}\,'$) to the problem. In fact (as we shall see) the processes of dissipation near to the critical line are complex and all we can hope to do with a linearized approach is to model them in a crude way. So we may as well do it as simply as possible; this is achieved by assuming Rayleigh friction and Newtonian cooling with equal coefficients α :

$$(\mathcal{G}'_\lambda, \mathcal{G}'_\varphi) = -\alpha(u', v')$$
$$\mathcal{Q}' = -\alpha T' \ ,$$

when we find easily that

$$\mathcal{X}\,' = -\alpha q'.$$

Then all we have to do is replace ω by $\omega - i\alpha/s$ in the wave equation.

If the basic state is slowly-varying in both φ and z, then we will find WKB solutions of the form $\Psi \approx e^{il\varphi} e^{imz}$ where l and m are dimensionless wavenumbers and where

$$\ell^2 + h^2 \sin^2\varphi \ m^2 - \nu_s = 0 \ . \tag{4}$$

So we can have wave-like solutions in φ **and** z (real ℓ and m) only if $\nu_s > 0$. This has some (but only some) parallels with the Charney-Drazin criterion for the β-plane case. [$\nu_s > 0$ gives a criterion for *total wavenumber*, whereas the Charney-Drazin criterion is about *vertical* wavenumber. We'll come back to this.] $\nu_s > 0$ requires $0 < U < U_s$, where

$$U_s = a \frac{\partial Q}{\partial\varphi} \left(\frac{h^2}{4H^2} \sin^2\varphi + \frac{s^2}{\cos^2}\varphi \right)^{-1} \ .$$

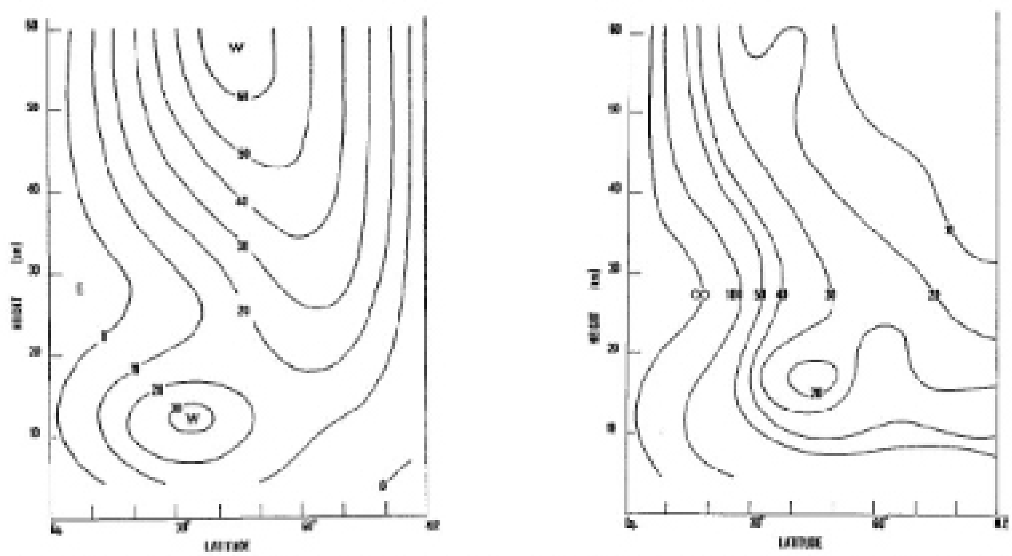


Fig. 1. The model basic state zonal wind distribution (m sec^{-1}) in the winter Northern Hemisphere.

Fig. 3. The refractive index square Q_0 for the $m=0$ wave.

Figure 3: Basic state zonal wind and refractive index from Matsuno's calculation.

Fig. 3 shows ν_0 calculated by Matsuno for the basic state shown. ν_0 becomes very large in the tropics where ω becomes small. Karoly and Hoskins (J.Met.Soc.Japan, **60**, p109, 1982) showed that wave rays (in the direction of the group velocity vector) tend to be refracted toward large ν_s; this is shown in the EP fluxes of Fig. 4 for Matsuno's profile and from Karoly and Hoskins' calculation. Hoskins and Karoly showed that, even if ν_s were positive at all heights, thus apparently allowing upward propagation everywhere, penetration to high levels is limited simply because most of the upward propagating wave activity is refracted away from the vertical, mostly toward the equator (which, in fact, reflects the tendency of planetary waves to propagate along great circles, other things being equal).

"Refractive index diagnostics"—the practice of calculating ν_s and interpreting the result in terms of wave propagation—is a useful tool in assimilating data. One application seems to undermine what was said earlier about the penetration of planetary waves into the southern winter stratosphere. It was argued that this could be prevented when U becomes strong in midwinter. However, ν calculated from southern hemisphere data by Randel (QJRMS, 1988) shows no sign of a major change through the winter. The reason is not really to do with spherical geometry, but with the shear contribution to $\partial Q/\partial y = \beta - \partial^2 U/\partial y^2$. If U is constant in y, $[\partial Q/\partial y]/U = \beta/U \to 0$ as U becomes large, thus preventing propagation. However, if locally $U \approx U_\circ \cos \ell y$, $[\partial Q/\partial y]/U \to \ell^2$ (independent of $U_\circ$) as $U_\circ$ becomes much larger than β/ℓ^2, so that ν loses its sensitivity to U. However, there may be reasons to question the validity of the WKB approximation under such circumstances (since, in a strong narrow jet, Ψ tends to take on the structure of the jet[3], thus violating the slowly-varying assumption).

This latter consideration highlights the contrast between the Matsuno criterion and the Charney-Drazin criterion for vertical propagation. The Matsuno criterion, as already noted, is based on (4)

$$\ell^2 + h^2 \sin^2\varphi\, m^2 - \nu_s = 0$$

which, in order to give a criterion for real *vertical* wavenumber, must be rewritten

$$h^2 \sin^2\varphi\, m^2 = \nu_s - \ell^2 \ ,$$

such that real m requires that $\nu_s > \ell^2$ and ℓ^2 is, *a priori*, unknown. If, for example, the latitudinal wavenumber $\ell = \pi/L$, where L (in radians) is the width of the jet (say, 0.5) then we require $\nu_s > 39$ for vertical propagation. If we look at Fig. 3 as an example, we se that vertical propagation would be excluded from almost the entire stratosphere poleward of about 30°. In fact, if we reformulate the problem as one for m only (assuming we know ℓ) we get

$$m^2 = \left(h^2 \sin^2\varphi\right)^{-1} \left[\frac{1}{\omega\ \cos\varphi}\frac{\partial Q}{\partial\varphi} - \frac{h^2}{4H^2}\sin^2\varphi - \frac{s^2}{\cos^2\varphi} - \ell^2\right]$$

[2]With typical numbers, $\Omega = 7.27 \times 10^{-5}\mathrm{s}^{-1}$, $a = 6370$ km, $N = 0.02\ \mathrm{s}^{-1}$, $h = 46$ km.

[3]Simmons, A. J., *Quart. J. Roy. Meteor. Soc.*, **101**, 185 (1974).

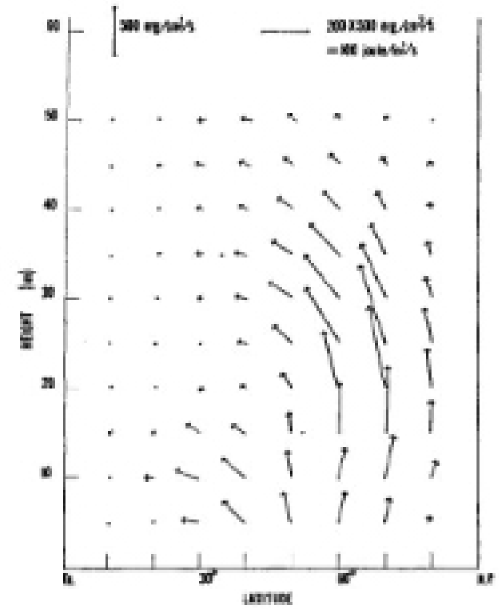


FIG. 13. Computed distribution of energy flow in the meridional plane associated with the $m=1$ wave shown in Fig. 5.

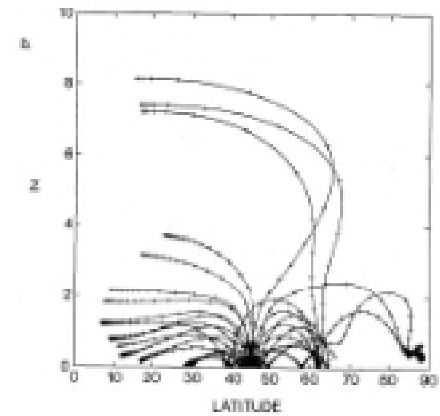


Fig. 6 Rays for zonal wavenumber one in the winter basic state. (a) Rays and phases for a source at 60° and z=0.4. It should be recalled that this wave cannot occur in the region in Fig. 5 where $K_s<1$. (b) Rays and propagation times for a source at 45° and z=0.4.

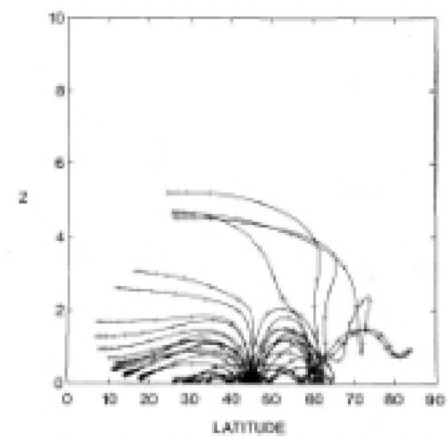


Fig. 7 Rays and propagation times for zonal wavenumber two in the winter basic state from a source at 45° and z=0.4. It should be recalled that this wave cannot occur in the region in Fig. 5 where $K_s<2$.

Figure 4: Arrows parallel to EP flux for Matsuno's solution (left) and wave rays (parallel to c_g) for N Hem winter state (Karoly & Hoskins; right). [Upper frame, $s = 1$; lower frame, $s = 2$.]

which means we get vertical propagation where $0 < U < U_c$, where

$$U_c = \frac{a^{-1}\partial Q/\partial\varphi}{[\Omega^2/N^2H^2 - s^2/(a^2\cos^2\varphi) - \ell^2/a^2]} ,$$

which is obviously identical to the Charney-Drazin result. So it may be very important to remember that ℓ may have a lower limiting value, and hence that deducing propagation from ν_s alone may give a very misleading picture of the potental for vertical propagation. In fact, Harnik and Lindzen (*J. Atmos. Sci.*, **58**, 2872–2894, 2001) diagnosed meridional wavenumber from observations to show examples in which vertical propagation is forbidden in southern winter.

3.2.5 EP flux in spherical geometry

From our linearized potential vorticity equation

$$\left(\frac{\partial}{\partial t} + \omega\frac{\partial}{\partial\lambda}\right) q' + \frac{v'}{a}\frac{\partial Q}{\partial\varphi} = \mathcal{X}' ,$$

it is straightforward, using the same procedures as in the beta-plane case, to obtain the EP relation

$$\frac{\partial\mathcal{A}}{\partial t} + \nabla\cdot\mathbf{F} = \mathcal{D}$$

where the density is

$$\mathcal{A} = \frac{1}{2}\rho a\ \cos\varphi\frac{\overline{q'^2}}{\partial Q/\partial\varphi} ,$$

the nonconservative term is

$$\mathcal{D} = \rho a\ \cos\varphi\frac{\overline{q'\mathcal{X}'}}{\partial Q/\partial\varphi} ,$$

and the EP flux is

$$F = (F_\varphi\ , F_z) = (-\rho\overline{u'v'}\cos\varphi\ ,\ \frac{2\Omega\rho}{S}\ \sin\varphi\ \cos\varphi\ \overline{v'T'}) .$$

The extra factor (as compared with the beta-plane result) of $\cos\varphi$ can be understood as an indication of the fact that, in spherical geometry, the EP flux represents (aside from constant factors) a flux of *angular*, rather than *linear*, momentum.

3.2.6 Observations of stratospheric planetary waves

We have seen that stratospheric planetary waves are large in winter (early and late winter in the southern hemisphere), are dominated by the quasi-stationary waves and grow with

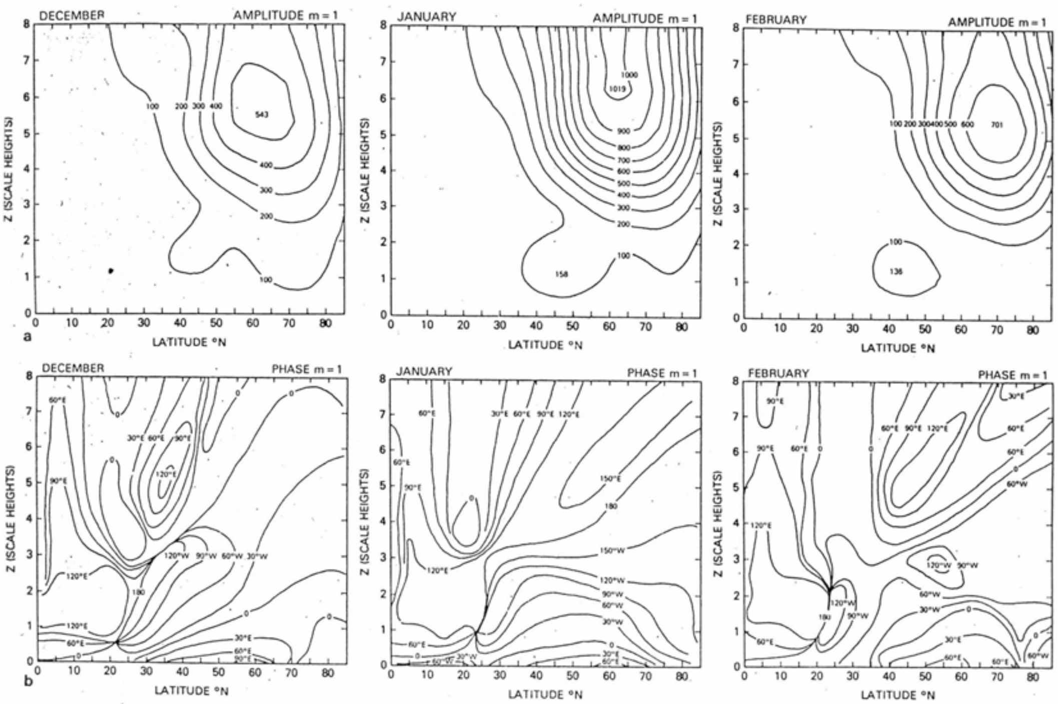


FIG. 3. Northern Hemisphere average December (left), January (middle), and February (right), (a) geopotential height amplitudes (m), and (b) phases for the zonal wavenumber 1 in geopotential for the winters 1978–79 through 1981–82. Phase values indicate the longitude of the geopotential ridge.

Figure 5: Wave 1, N Hem winter.

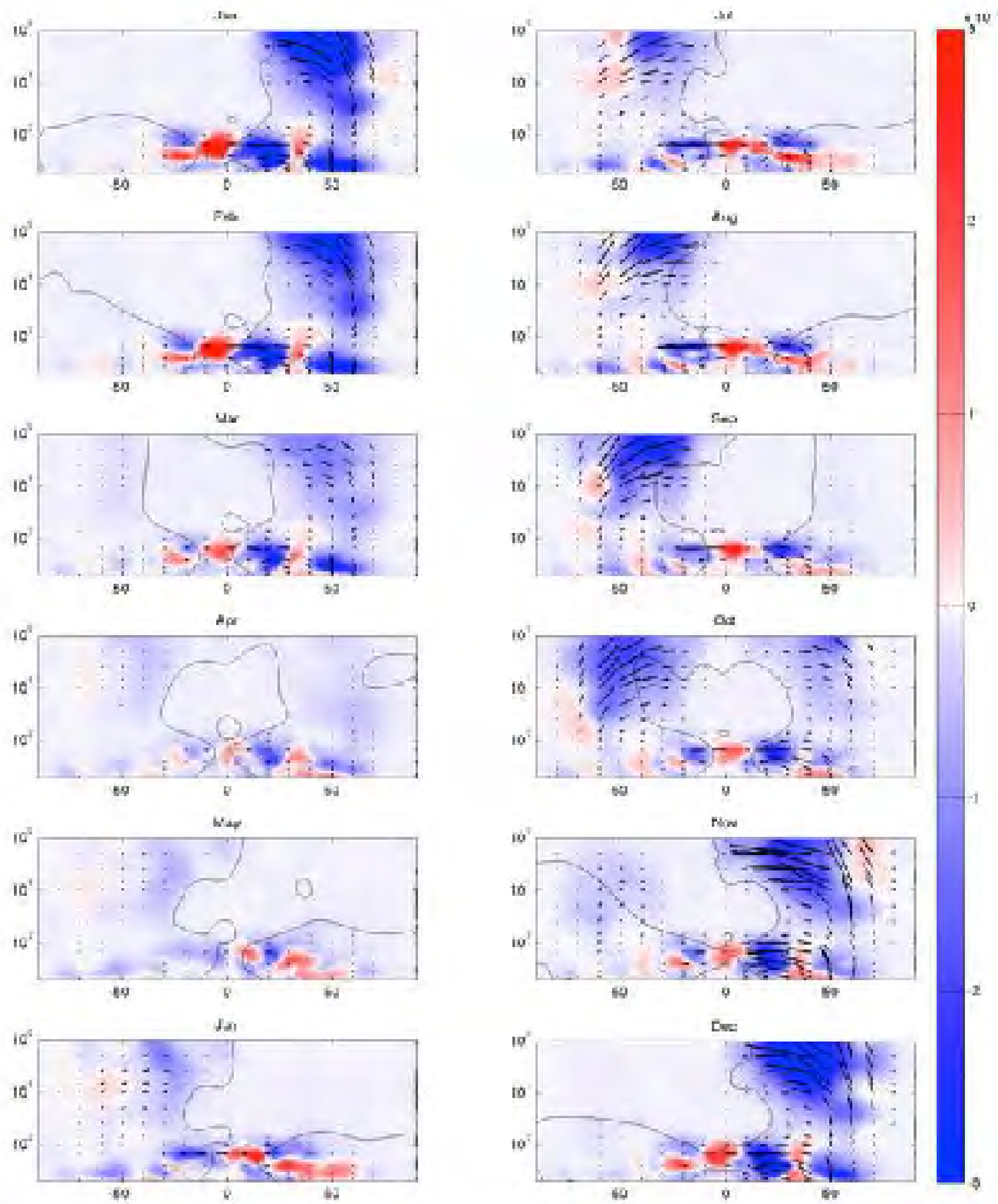


Figure 6: Monthly climatological EP cross-sections for stationary waves. Shading shows the EP flux divergence.

height through the stratosphere. Fig. 5 shows the wave amplitudes from a global climatology for northern winter (when the waves are large in the northern hemisphere) The montly mean EP fluxes and their divergences for stationary waves are shown in Fig 6.together with the separate contributions from the monthly-mean and transient waves (these transient waves are the departures from the monthly-mean wave and include slow variations of the quasi-stationary modes as well as truly transient, travelling, disturbances).

The EP fluxes show the characteristic upward propagation/equatorward refraction we discussed from the theoretical models. (Note that the "transient" eddies are the larger component in the troposphere (see the paper by Edmon, Hoskins and McIntyre, J.A.S., 1980, for a thorough discussion of the tropospheric observations). In the stratosphere, the stationary component that is shown is (marginally) the larger. Upward/equatorward EP flux means poleward heat flux / poleward momentum flux. Qualitatively, the southern hemisphere waves are similar in structure to the northern hemisphere waves (except as regards seasonal cycle, *q.v.*) but of consistently smaller amplitude.

3.2.7 Transient planetary-scale waves

Thus far, we have ignored the nature of the transients in the stratosphere. Some of them are undoubtedly a simple manifestation of the fact that the tropospheric processes driving planetary waves (flow over topography, diabatic heating, nonlinear effects) are not in fact steady themselves. Other components (as we shall see) probably arise from nonlinear effects within the stratosphere. Nevertheless, the structure and frequencies of much of the transient wave field appears to indicate that the atmosphere is "ringing": *i.e.*, in *normal modes*. The atmosphere possesses (in theory) a whole sequence of normal mode frequencies; two of the large-scale ones of these are at 4 days (zonal wave 2) and 10 days (zonal wave 1). Examples are shown in Fig. 7. [See Salby, *Rev. Geophys. Spa. Phys.*, 22, 209, 1984, for a review of normal mode traveling waves].

3.2.8 Seasonal cycle of stratospheric planetary waves

We have noted that the "broad brush" description in terms of strong waves in the winter westerlies and weak waves in summer may be a misinterpretation of the Charney-Drazin criterion. Fig. 8 shows the height amplitude of wave 1 at 1hPa as a function of latitude and time through two years 1979-81. During northern winter, wave amplitudes fluctuate but they areusually large through the winter. In southern winter, by contrast, wave amplitudes are large (though usually less than in N Hem winter) in early winter and spring, but weak in midwinter. (This is typical behavior, though the southern winter of 2002 was more NHem -like.) Why? The simple answer is that the midwinter mean zonal winds are too strong to allow propagation, but this begs the question as to why this is not the case in the NHem. The answer in turn seems to be that the waves in northern winter are normal mode structures

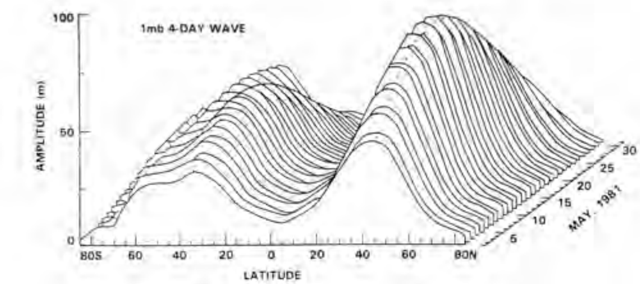


Figure [illegible]. Structure and evolution of the 4-day wave at stratopause level, observed by Tiros-N SSU and NOAA-A HIRS. Corresponds to wavenumber 2 at 1 mb, bandpassed between 3.8-4.5 days over May 1981. Systematic retrogression is simultaneously observed. [After Hirota and Hirooka, 1984.]

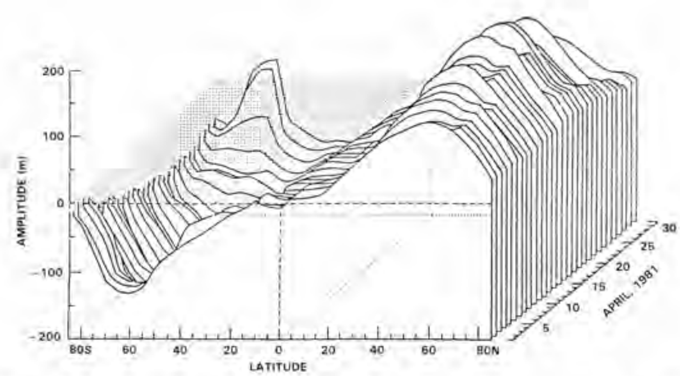


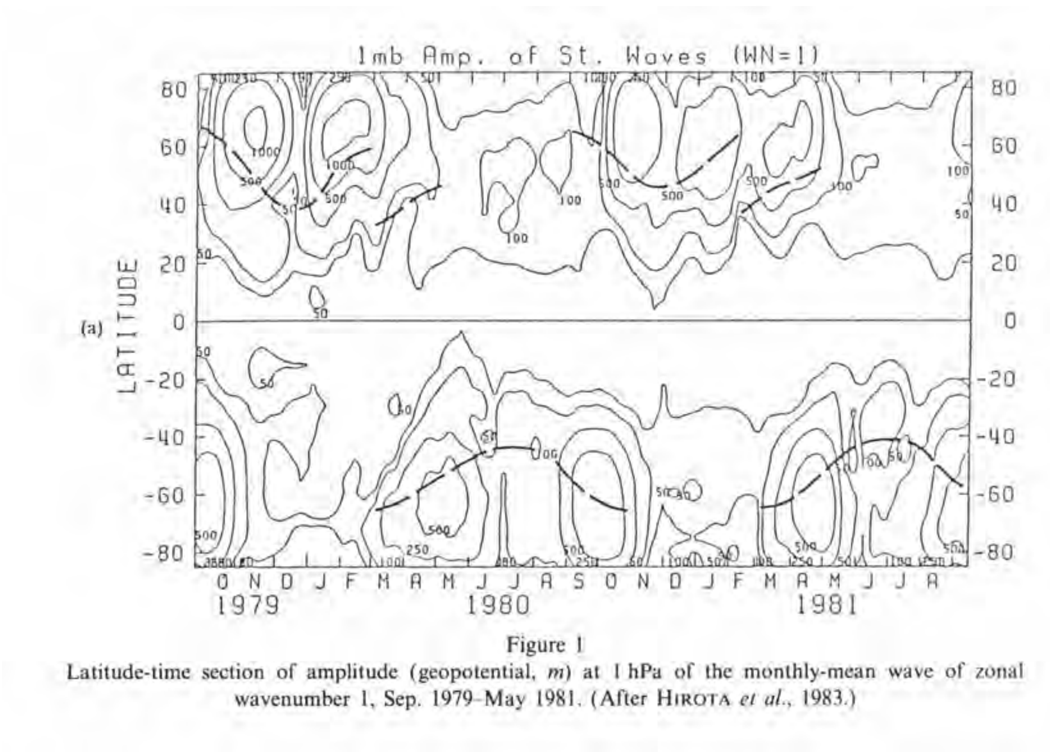
Figure [illegible]. Structure and evolution of the 10-day wave at stratopause level, observed by Tiros-N SSU. Wavenumber 1 geopotential at 1 mb band passed about 9.2 days during April 1981. [Adapted from Hirooka and Hirota, 1985.]

7.jpg

Figure 7: Structures of 2 global normal modes.

wave1

seasonality

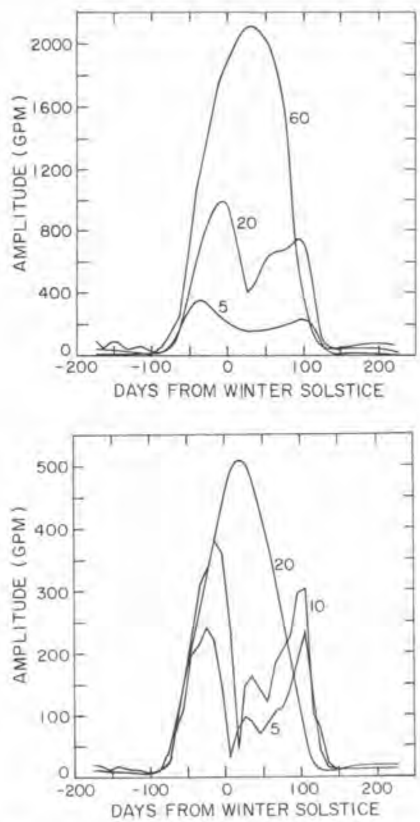


8.jpg

Figure 8: Wave 1 seasonal variations.

strong enough to prevent the winds from getting too strong (by reducing the equator-to-pole temperature gradient). Fig. 9 shows results from a "quasi-linear" β-channel model in which

seasonality



(a)

(b)

Figure 5.
Time series through winter of wave amplitude (m) at $z = 50$ km. (a) Zonal wavenumber 1; (b) zonal wavenumber 2. Labels on curves are forcing amplitudes Φ_0 (in m).

9.*jpg*

Figure 9: Results from quasi-linear 1-D model.

linear waves propagating on a mean flow that is forced by Newtonian relaxation toward a seasonally varying equilibrium are allowed to interact with the background flow through their EP flux divergence. When the wave forcing (in this case, determined by specifying the wave amplitude on the lower boundary) is weak, the system behaves linearly and the waves can propagate into the stratosphere only in the weak westerlies of early and late winter (a S Hem - like situation). With stronger wave forcing, the mean flow in midwinter is reduced enough to allow wave propagation all winter (N Hem - like). So the different seasonalities in the two hemispheres may be a consequence of the different wave forcings.

We should note here that in many years the springtime peak of southern hemisphere wave activity is much larger that the autumn peak. Scott & Haynes (*J. Atmos. Sci.*, **59**, 803-822, 2002) argued that the system passes through resonance in spring — we'll return to this issue later.

3.2.9 Wave dissipation

Another important issue to be addressed concerns the decrease of EP flux with height. This is not evident from the EP flux diagrams (which are inconsistently scaled for clarity) but it is clear from Fig. 5 that wave amplitudes do not grow anything like as $\rho^{-1/2}$, as theory says an undissipated propagating wave should. It is clear, therefore, that the waves are being dissipated rather strongly. We found a simple way to include radiative dissipation in our linear calculations. There is no doubt that this makes a contribution to the real damping of the waves; it is doubtful, however, whether it is strong enough to explain the observed decay with height of wave activity (simple linear models usually give amplitudes too large in the upper stratosphere). There are in fact other dissipative processes acting; consideration of these brings us to the thorny question—which we have glossed over thus far—of the behavior near the "critical line" where $\bar{u} = 0$.

3.3 Rossby wave breaking

3.3.1 Quasi-linear waves near the critical line

We have seen that, under normal circumstances, planetary waves on the sphere propagating up through the stratosphere will be refracted equatorward toward the "critical" region where the winds change sign, the undamped linear problem becomes singular, and where (as well now see) the assumptions on which our linearization was based (*i.e.*, that we can neglect $u'\,[\partial q'/\partial x]$ compared with $\bar{u}\,[\partial q'/\partial x]$) breaks down. What happens in this region determines much of the character of wave transport in the stratosphere.

Consider small-amplitude, adiabatic stationary planetary waves in a shear flow $\bar{u}(y)$ which contains a critical line $\bar{u} = 0$. The linearized quasigeostrophic potential vorticity equation is

$$\left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right) q' + v'\frac{\partial \bar{q}}{\partial y} = 0 \ .$$

Let the northward displacement (in a Lagrangian sense) of an air parcel be η' such that

$$\left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right) \eta' \equiv v' \ .$$

Then, if $q' = 0$ when $\eta' = 0$ (and we can choose to define η' in this way), we have

$$q' = -\eta'\frac{\partial \bar{q}}{\partial y} \ .$$

(We could have got this from a Taylor expansion about the mean position of a q-contour). Therefore

$$\overline{v'q'} = -\overline{v'\eta'}\frac{\partial \bar{q}}{\partial y} = -\frac{\partial}{\partial t}(\frac{1}{2}\overline{\eta'^2})\frac{\partial \bar{q}}{\partial y} \ .$$

This says what we already found in the course of developing the nonacceleration theorem; $\overline{v'q'}$ is zero for a steady, adiabatic wave; what this form makes clear is that it is $\overline{\eta'^2}$—the mean square Lagrangian displacement—that must be steady.

Now, suppose that the wave is steady *in an Eulerian sense*, with geostrophic streamfunction

$$\psi' = \mathbf{R}e \ \Psi(y)e^{ikx} \ ,$$

so that

$$v' = \mathbf{R}e \ ik\Psi \ e^{ikx} \ .$$

Then, assuming steadiness, we have $\bar{u}\partial\eta'/\partial x = v'$, or

$$\eta' = \mathbf{R}e\frac{\Psi}{\overline{u}}e^{ikx} \ ,$$

which is singular where $\bar{u} = 0$. The reasons for this singularity in the steady case are not hard to find. Consider the initial value case where $\eta' = 0$ at $t = 0$. Then if

$$\eta' = \mathbf{R}e \ Y(y,t)e^{ikx} \ ,$$

$$\left(\frac{\partial}{\partial t} + ik\bar{u}\right) Y = ik\Psi \ ,$$

whence

$$Y = \frac{\Psi}{\overline{u}}(1 - e^{-ik\bar{u}t}) \ .$$

For $\bar{u}$ finite, this means that Y just oscillates around the steady solution $Y_0 = \Psi/\overline{u}$ and the dynamics are wavelike, as shown in Fig 10. For $k\bar{u}t \ll 1$, however (and this will be true for an increasingly long time as $\bar{u} \to 0$),

$$Y \approx ikt \ \Psi$$

—so η' just grows linearly with time (*i.e.*, $\partial\eta'/\partial t = v'$). There is therefore (see Fig 10) no oscillation at the critical line; the parcel displacements just increase systematically with time. This is more characteristic of a chaotic flow than a wavelike one and, indeed, is indicative of the presence of closed eddies (see Fig 11, below)—rather than waves on a prevailing westerly flow—near the critical line. Note that q' will behave in the same way as η'.

To the extent that linear theory is valid, the EP flux divergence is

$$\nabla \cdot \mathbf{F} = \rho\overline{v'q'} = -\frac{\rho}{2}\frac{\partial}{\partial t}(\overline{\eta'^2})\frac{\partial \bar{q}}{\partial y} \ .$$

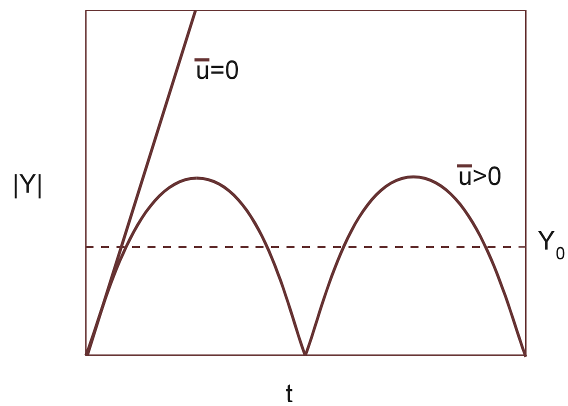


Figure 10: Plot of $|Y|(t)$.

Since we have seen that $\overline{\eta'^2}$ must increase systematically with time as long as linear theory is valid, it follows that we expect

1. At least some absorption of wave activity (convergent EP flux) at the critical line when $\partial\overline{q}/\partial y$ is positive,

2. Reflection when $\partial\overline{q}/\partial y$ is zero (since $\nabla \cdot \mathbf{F} = 0$; there can be no transmission, since stationary Rossby waves do not propagate through easterlies, unless the region of easterlies is narrow enough to permit barrier penetration) and

3. Overreflection (divergent EP flux) when $\partial\overline{q}/\partial y$ is negative.

As we shall see below, the absorption may or may not be sustained at longer times when linear theory breaks down.

3.3.2 The finite amplitude critical layer

Clearly, given enough time, our linear assumption will break down at the critical line (and our assumption of a *steady* linear wave is *never* valid there). In fact, as shown in Fig. 11, we should think not of a critical *line*, where $\overline{u} = 0$, but of a critical *layer*, comprising anticyclones within the "Kelvin cat's eyes" shown in Fig. 11. In the limit of zero wave amplitude, the cat's eyes shrink to the critical line; they become progressively broader as the wave amplitude is increased.

In the vicinity of the critical line $y = 0$, the mean flow is $\overline{u} \simeq \Lambda y$, so the total streamfunction is

$$\psi \simeq -\frac{1}{2}\Lambda y^2 + \Psi(0)\cos kx$$

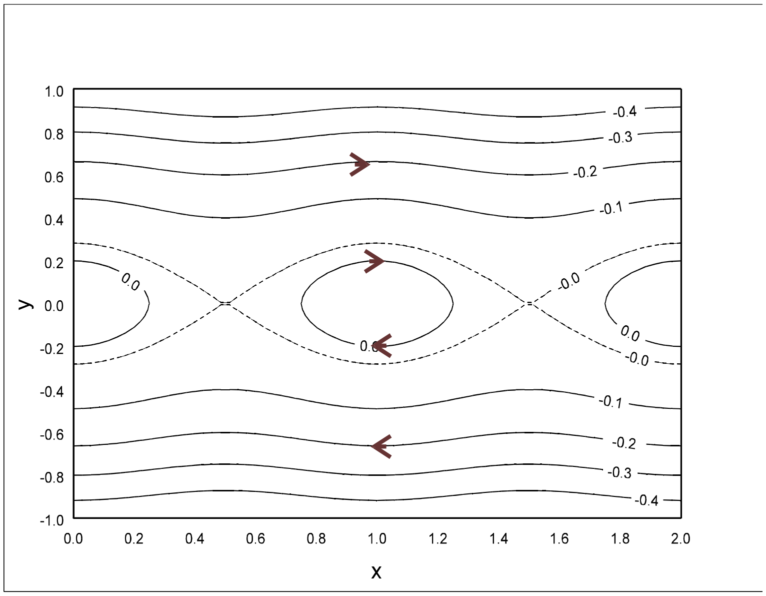


Figure 11: Streamlines in the vicinity of the critical layer. Dashed contour encloses the "Kelvin cat's eyes" region of closed eddies. The undisturbed zero wind line where $\overline{u} = 0$ is at $y = 0$.

where we have chosen the origin of the x-axis to coincide with the maximum of ψ', and where we assume that Ψ varies sufficiently slowly with y that $\Psi(y) \simeq \Psi(0)$ near $y = 0$. Now, the closed anticyclones are maxima of ψ, located where $kx = (2n + 1)\,\pi$. This means that along $y = 0$, the boundaries of the cat's eyes (which are *stagnation points* of the flow) are minima of ψ, the value of which is

$$\psi_{\text{min}} = -\Psi(0) \; .$$

At the opposite phase (the longitude of the center of the closed eddies), $\cos kx = 1$, and so the edge of the cat's eyes—the same streamline—is found where

$$-\frac{1}{2}\Lambda y^2 + \Psi(0) = -\Psi(0)$$

i.e., where

$$y^2 = \frac{4}{\Lambda}\Psi(0) \; .$$

So the width of the cat's eyes increases as the square root of the wave amplitude (as long as the width is sufficiently small).

3.3.3 Absorption or reflection?

We saw that, in the linear case, the critical layer will absorb wave activity if $\partial\overline{q}/\partial y$ is positive there. However, since wave absorption (nonzero $\nabla \cdot \mathbf{F}$) implies nonzero $\partial\overline{q}/\partial t$, through rearrangement of the PV contours within the critical layer, and there is only a finite amount of such rearrangement that can be done, it seems unlikely that absorption can go on forever. To make things simple, suppose that the case shown in Fig. 11 is barotropic, and that the Rossby wave is propagating into the critical layer southward, through the westerlies. If we take some latitude y_N north of the critical layer, the EP flux across that latitude is $F(y_N) = -\left(\overline{u'v'}\right)_N$. We assume that at some latitude y_S, well to the south of the critical layer and deep in the easterlies, $F(y_S) = 0$ (the wave is evanescent there, so the EP flux vanishes). Now, since

$$\overline{v'\zeta'} = -\frac{\partial}{\partial y}\left(\overline{u'v'}\right)$$

it follows that

$$-F(y_N) = -\int_{yS}^{y_N} \overline{v'\zeta'}\, dy \ . \tag{5}$$

Therefore the net flux of wave activity into the critical layer, $-F(y_N)$, is directly related to the integrated PV flux within the critical layer. (Note that, if the wave is steady outside the critical layer, $\nabla \cdot \mathbf{F} = 0$ whence the PV flux is zero there, and so the only contribution comes from within the critical layer.)

Now, consider Fig. 12. At $t = 0$, the PV contours are aligned zonally. During the early, linear phase of the evolution, the PV contours are distorted as shown in (a). Low PV is being advected northward, and high PV southward, whence $\overline{v'\zeta'} < 0$, and so (5) tells us that $F(y_N) < 0$: there is a net flux of wave activity into the critical layer, which is thus absorbing wave activity. After some time (b), however, the advected "tongues" of PV start to become wrapped around the critical layer anticyclones, and thus to move eastward and westward, rather than northward and southward. At this stage then, $\overline{v'\zeta'} \to 0$, and so $F(y_N) \to 0$: there is no net wave activity into the critical layer, which is now becoming perfectly *reflective*. By stage (c), the tongues are being wrapped further around the anticyclones; the high PV air is heading back northward, and the low PV air southward. Thus $\overline{v'\zeta'} > 0$, whence $F(y_N) < 0$; the critical layer is a net emitter of wave activity, and thus is becoming *overreflective*. As time goes on, the PV tongues become wrapped around and around the anticyclones, and the critical layer performs a decaying oscillation between absorption, reflection, and overreflection, as depicted schematically in Fig. 13. The integrated absorption of wave activity—the area under the curve—remains finite as $t \to \infty$, a consequence of the fact that there is only a finite amount of PV rearranging that can be done.

In reality, other effects can change this picture. For one thing, the "tongues" of PV can become unstable. IF we neglect the background flow, then an isolated strip of PV is

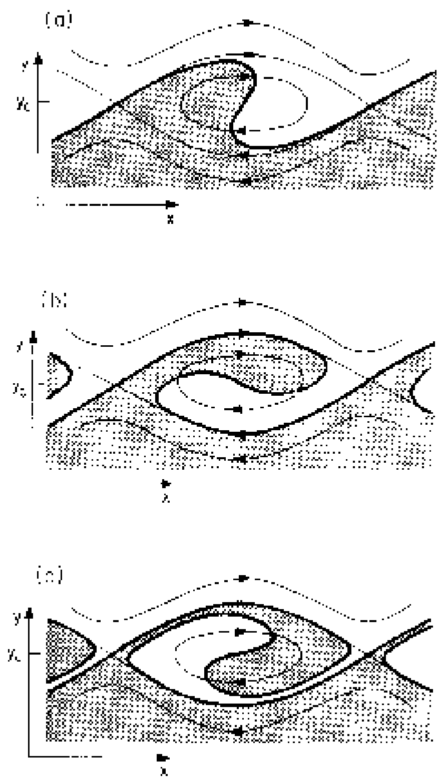


Figure 12: [Fig 5.13 of Andrews et al.]

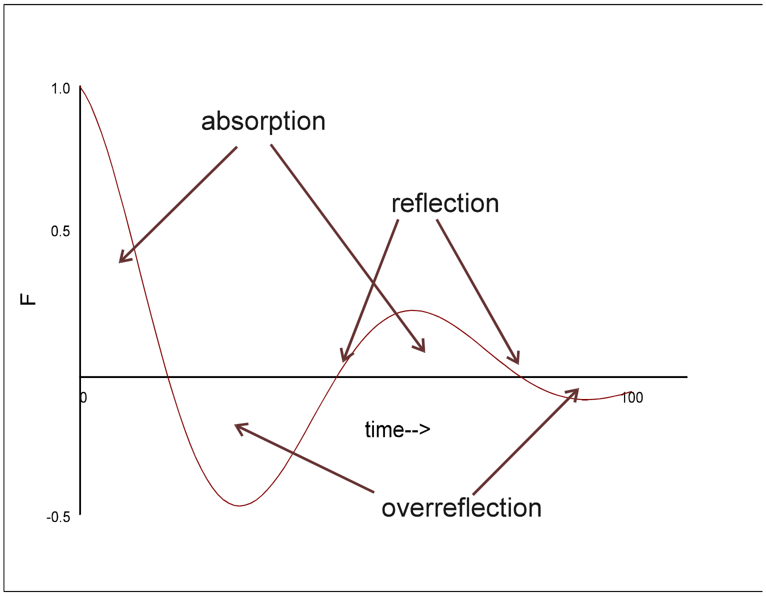


Figure 13: Net flux into the critical layer (schematic).

unstable to barotropic instability (since the PV is an extremum there). Haynes [*J. Fluid Mech.*, **207**, 231-266 (1989)] showed that this tends to eliminate the oscillations in Fig. 13 by disorganizing the roll-up of the PV tongues. In fact, this kind of instability is less common in the stratosphere than one might think,because of shear across the strip that is not present in the case of an isolated strip.

A second potentially modifying factor is dissipation. If, for example, dissipation tends to restore the PV to its original, undisturbed, state, it can hinder or even prevent the roll-up of the PV tongues seen in Fig. 12(c). If such dissipation is strong enough, it could keep the PV distribution looking like Fig 12(a), even in steady state, and thus allow permanent southward PV flux and hence indefinitely sustained absorption.

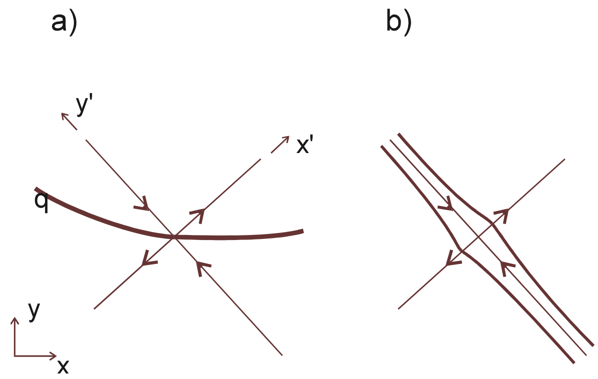
Third, time-dependence (other than simple translation of a steady wave pattern) can also lead to sustained absorption.

3.3.4 The stagnation points

In practice, the stagnation points where $u = 0$ and $v = 0$[4] are important in helping us understand critical layer behavior. All of the above discussion is relevant only if there is some PV structure within the critical layer to be rearranged. A sufficient condition for this

[4]This is for a stationary wave. The key thing is that the flow vanishes in a frame of reference in which the wave is steady so, in general, $u - c = 0$ and $v = 0$ there.

is that there be a nonzero gradient of PV at the stagnation points. Consider Fig. 3.3.4. If



there is a nonzero PV gradient at the stagnation point, we may draw a PV contour through it, as shown in Fig. 3.3.4(a). We place (not shown) two marked particles along the PV contour, one at the stagnation point, the other a short distance **r** away. the two points will separate at a rate

$$d\mathbf{r}/dt = \mathbf{u}(\mathbf{r}) \simeq (\mathbf{r}\cdot\nabla)\,\mathbf{u}\ ,$$

or, in tensor notation

$$\frac{dr_i}{dt} = r_j \partial_j u_i$$

where $\partial_j \equiv (\nabla)_j$. Therefore

$$\frac{dr_i}{dt} = S_{ij} r_j \tag{6}$$

where S is the rate-of-strain tensor

$$S = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} -\psi_{xy} & -\psi_{yy} \\ \psi_{xx} & \psi_{xy} \end{pmatrix}\ .$$

The general solution to (6) is

$$r_i = r_i^{(1)} \exp\left(\sigma^{(1)} t\right) + r_i^{(2)} \exp\left(\sigma^{(2)} t\right)$$

where $\sigma^{(n)}$ are the eigenvalues of S.

In the vicinity of the stagnation point, the flow is hyperbolic: if we rotate coordinates to the principal axes of the strain (x' and y' in Fig. 3.3.4) $\psi = Axy$, say, and then

$$S = \begin{pmatrix} -A & 0 \\ 0 & A \end{pmatrix}\ .$$

So the two solutions ($\sigma = \pm A$) comprise one growing exponentially (along the axis of extension—x' in Fig. 3.3.4), and one decaying exponentially (along the axis of compression y'). [N.B. recall that we have linearized by evaluating S at the stagnation point; after a finite time, the marked particle will be a finite distance away and so will feel a different strain field.]

If an initial PV contour does not coincide with one of the principal axes, it must move. In fact, even if it is aligned with one of the axes, an immediately adjacent contour (and, if the PV gradient is nonzero, there must be such a contour) must move. As shown in Fig. 3.3.4(b), a narrow material strip along the axis of compression will be stretched out along the axis of extension. Therefore, the flow cannot be steady if there is a nonzero gradient of PV at a stagnation point.

If the flow were exactly steady, one would not expect the formation of filaments to go on indefinitely: once the PV contours become aligned with streamlines, no further transport will take place. But one can get sustained transport if the flow is not quite steady. There is now a substantial body of theory called "lobe dynamics"[5] to describe this transport in nondivergent flows. The simplest (and mathematically most precise) examples are for steady flows like we considered in Fig. 11, but with the addition of a small time-periodic component. The introduction of the unsteady component means that the stagnation points are no longer quite relevant; they are replaced by *hyperbolic trajectories*, which trace out the movement of the material hyperbolic points with maximum material stretching (in fact these are the only such trajectories as $t \to \infty$). At any instant in time (see Fig. 14, which shows the streamlines through the hyperbolic points), the flow looks like Fig 7.2. If one puts a cloud of

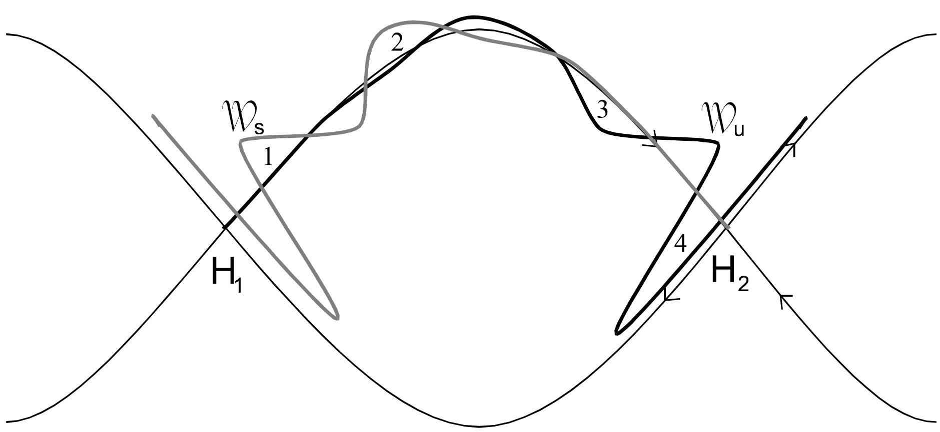


Figure 14: Showing the hyperbolic points H_1, H_2, and the manifolds W_u and W_s.

[5]Mostly done by Steve Wiggins and collaborators; e.g., see Wiggins, *Chaotic transport in dynamical systems*, Springer, NY, 1992; Malhotra and Wiggins, *J. Nonlinear Sci.*, **8**, 401, 1998.

marker particles around H_1, the cloud will become compressed in one direction and stretched out in the other—along something like the axis of extension (in the steady case). As the cloud approaches hyperbolic point H_2[6], it becomes stretched exponentially in the direction across the instantaneous streamline. The *unstable manifold* W_u is shown as a snapshot in time in Fig. 14. Similarly, one can construct the *stable manifold* W_s by putting a cloud of particles near H_2 and integrating backwards in time. Once every period of the flow, the manifolds look exactly the same; this fact, together with the proof (not given here) that the manifolds are material surfaces, allows some simple deductions. The fluid contained in a "lobe" between the two manifolds must remain in a lobe, but not the same one; the lobe advances one wavelength each period. So, lobe 1, initially on the "outer" side of the dividing streamline, becomes successively lobe 2, 3, 4[7]—and thus a filament of outer air entrained into the cat's eye. Note that not all the lobes are shown—there are an infinite number of lobes after 4, each longer and thinner than the preceding one. Thus, the filament is systematically stretched down to ever decreasing scales.

3.3.5 Wave breaking on an isolated vortex

Consider (see Fig. 15) a circular "vortex" patch in shallow water on an f-plane with uniform

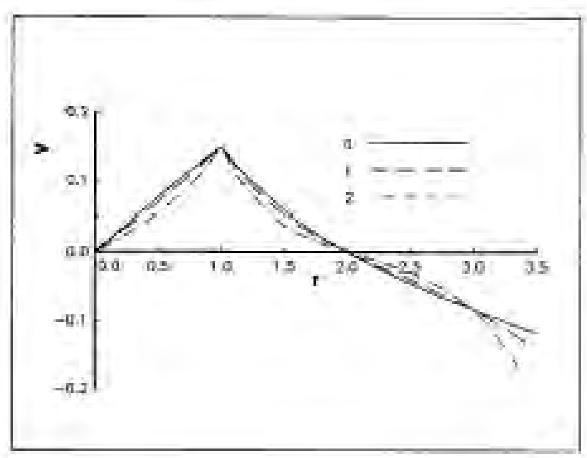


Figure 15: Structure of the basic state vortex (velocity as function of radius). (Polvani & Plumb, JAS, 1992).

cyclonic vorticity $q = Q_i(> f)$ in $r < r_0$ and uniform anticyclonic vorticity $q = Q_o(< f)$

[6] H_1 and H_2 may in fact be the same point, e.g. in an $m = 1$ Rossby wave flow.

[7] Each lobe has the same area, even if my drafting does not look that way!

in $r > r_0$. We will allow this vortex to be disturbed by a wavy topography h; the relevant conservative quasigeostrophic potential vorticity equation is $dq/dt = 0$, where

$$q = f + \nabla^2 \psi - f \frac{h}{D} ,$$

where D is the mean depth. (The flow will be barotropic if D is large enough, if the Rossby radius $(gD)/f >> r_0$.) The interface at the edge of the vortex will be disturbed by the topography, which generates Rossby waves there (since the sharp potential vorticity gradient at the interface will support Rossby waves). Since q is conserved, then for all time $q = Q_i$ everywhere inside the interface, and $q = Q_o$ everywhere outside. Thus, knowledge of the interface position gives knowledge of the entire potential vorticity distribution, which (through PV inversion) determines the entire flow field. In turn, the flow field entirely determines the PV evolution, since q is simply advected. So everything one needs to know about the system is implicit in the location of the single PV contour. (As we shall see, this contour may become highly convoluted.)

Therefore, the system can be integrated using "contour dynamics" [see Dritschel, *J. Comp. Phys.*, 77, 240, 1988], the basis for which is:

1. Given knowledge of the PV contour's location at time t, calculate the velocity at each of many points along the contour. This velocity can be calculated as

$$[u(x), v(x)] = (2\pi)^{-1}(Q_o - Q_i) \int \ln(|x - x'|)[-dx', dy'] + (-\frac{\partial \psi_f}{\partial y}, \frac{\partial \psi_f}{\partial x})$$

   where the integral is around the contour and ψ_f is the streamfunction obtained from inversion of the topographic component of PV:

$$\nabla^2 \psi_f = -f \frac{h}{D} .$$

2. The contour is then advected to a new position at time $t + \delta t$.

3. Go back to (1).

Thus, the computational problem of predicting the flow evolution becomes the one-dimensional one of resolving the contour (vortex edge) to sufficient accuracy, which can be done at very high resolution at modest computational cost.

In the experiments shown below [and described in more detail in Polvani & Plumb, JAS, 49, 462, 1992], values of Q_0 and Q_i were chosen to give a westerly jet of $65ms^{-1}$ at the vortex edge and a zero wind line at $r = 2r_0$ (and r_0 chosen to be 3000km). The wind profile (dimensionless) is shown by the solid line in Fig 15. The topography is wavenumber 1, of the form

$$\frac{h}{D} = H_0(1 - e^{-t/\tau})J_1(1.6r/r_0) \ .$$

The term involving $\tau = 2.5$ days is to ensure a smooth switch-on; the argument of the Bessel function is chosen so that the peak of the main mountain is near the vortex edge.

It turns out that there is a threshold in the response. At subcritical amplitude ($H_0 = 0.15$; not shown here), the mountain distorts the vortex (with stationary and transient Rossby waves) but does not disrupt it. At supercritical amplitude ($H_0 = 0.18$; Fig 16) the wave

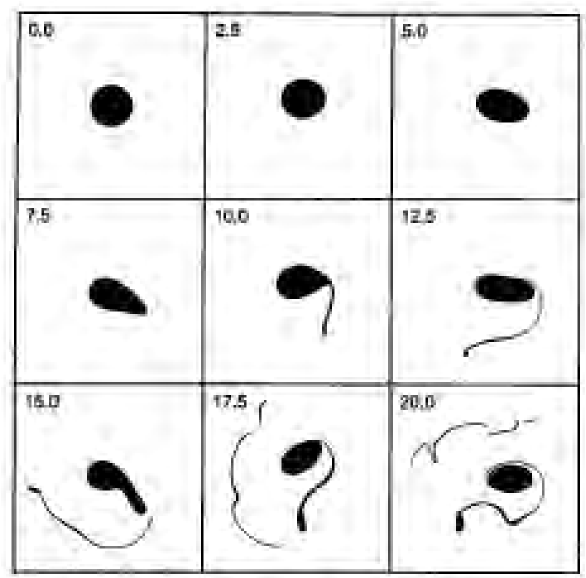


Figure 16: Evolution of the vortex in the contour dynamics model with $H_0 = 0.18$. (Polvani & Plumb, JAS, 1992).

"breaks", and vortex material is ejected at this location at around 10 days. Subsequently, the ejected material is sheared out into a thin filament, and other breaking events occur. (Note the "roll-up" of the tips of the filaments.)

Why the transition? Fig 17 shows the velocity fields at 0, 7.5 and 17.5 days for the just-subcritical case and at 15 days for the just-supercritical. The first frame shows the undisturbed wind field, with the zero wind line (a critical line for stationary waves) at $r = 2r_0$. As the wave grows, the critical line becomes a critical layer, the anticyclone (Kelvin cat's eye) being evident centered near 180^o with a stagnation point (the point of the cat's eye) at about -20^o. This point, which wobbles around in the presence of the transient Rossby wave generated by switching on the topography) moves to about $+20^o$ at 17.5 days. It remains,

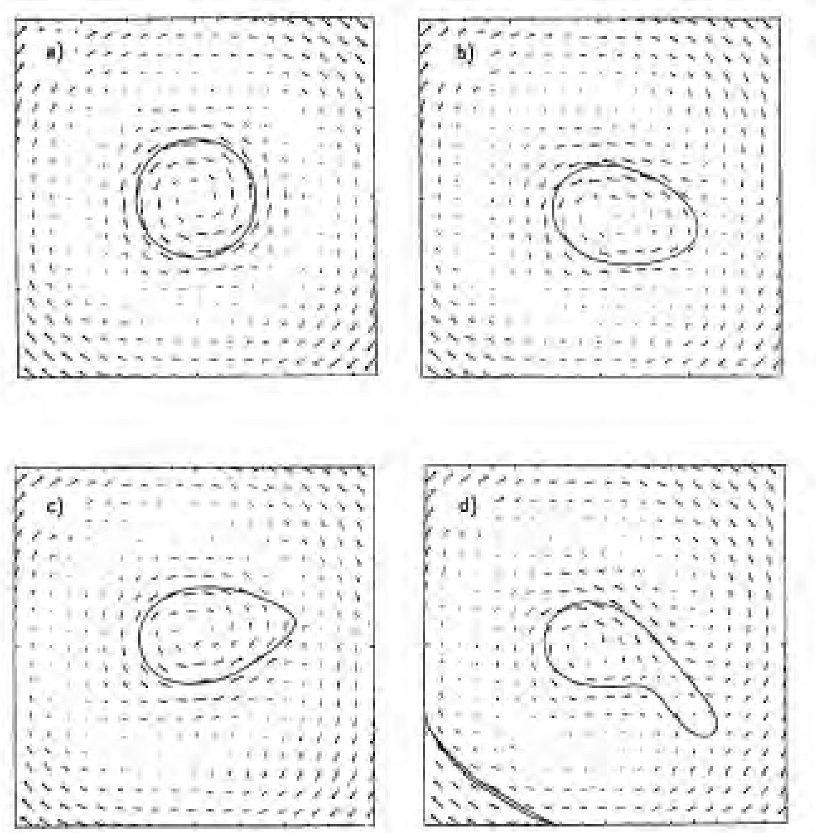


Figure 17: Velocity and contour edge location for $t = 0, 7.5$d, 17.5d of the subcritical case $H_0 = 0.15$ and (d) 15 d for the supercritical case $H_0 = 0.18$ (*cf.* previous figure; Polvani & Plumb, JAS, 1992).

however, outside the vortex; the flow at the vortex edge is still eastward everywhere and the edge is not disrupted. At larger forcing amplitude, however, the stagnation point moves inside the vortex edge, for two reasons: the stagnation point moves inward and the vortex edge in the vicinity of the stagnation point moves outward. When the stagnation point is inside the contour the flow is not systematically eastward along the edge. It seems unlikely that there can be a steady solution when the stagnation point is inside the edge. At this stage, the wave breaks and vortex material is ejected into (entrained by) the anticyclone in the cat's eye.

Of course, the "critical layer" still exists even in subcritical cases, but it does not reach the vortex and therefore does not entrain vortex air. This is shown in Fig 18 in which passive,

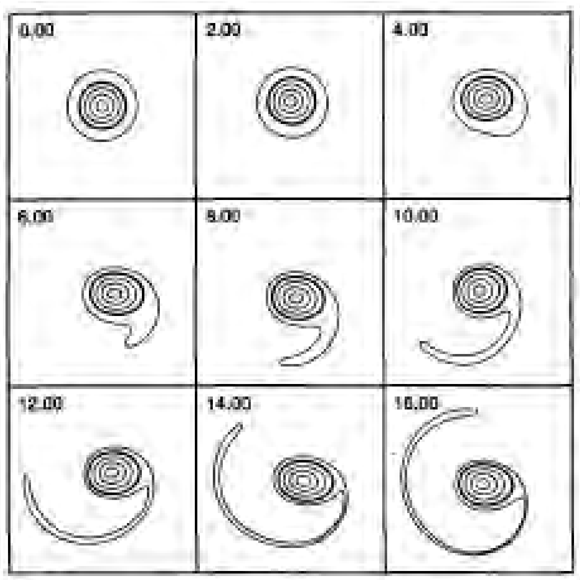


Figure 18: Evolution of the PV contour (heavy) and passive tracer contours (light) for a subcritical case.

marker contours (no PV contrast and therefore no influence on the dynamics) are included in the first case. The single PV contour is the thick one. Even though the wave does not break in the sense that the PV contour is not irreversibly distorted[8], material between the vortex edge and the outer passive contour is entrained into the critical layer. Note how this inevitably leaves behind a sharpened gradient at the vortex edge: this "vortex stripping" is undoubtedly the reason for the creation of the sharp edge to the stratospheric vortices.

Breaking is not always filamentary. Fig. 19 shows the evolution of a more supercritical

[8]As we shall see below, the irreversible distortion of dynamically relevant material surfaces is a useful criterion for "breaking".

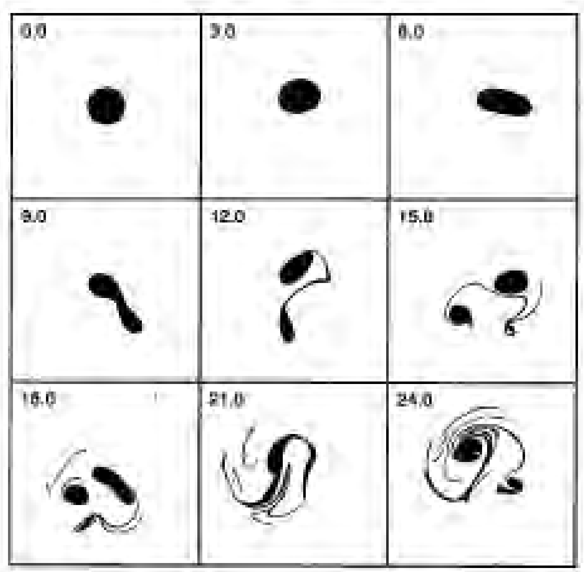


Figure 19: Evolution of the vortex in the contour dynamics model for a more supercritical case with $H_0 = 0.25$.

case ($H_0 = 0.25$), in which a larger "blob" of vortex air is torn off to form a secondary vortex. In general, the stronger the wave disturbance, the bigger the ejected blobs of material.

One important aspect of the above (and other) numerical results is that, although vortex material is ejected outward, there is no corresponding intrusion of outside material into the vortex. Thus, the vortex acts as a "containment vessel" in which interior air is isolated from its surroundings. The reason for this asymmetry is simply that the critical layer (surf zone) is located outside the vortex and so it entrains vortex material outward. In the experiments shown, there is no critical layer inside the vortex. It is in fact possible to contrive a basic wind field so that there is an interior surf zone (though the flow is unlike the usual stratospheric flow) in which case inward breaking occurs.

3.3.6 Rossby wave breaking in the stratosphere

The behaviour of parcel displacements in the stratosphere can be investigated by looking at the evolution of a material tracer. The most obvious choice is potential vorticity, which is a tracer for conservative flow. We could use quasigeostrophic potential vorticity, but it is just as easy to calculate Ertel PV, and thereby avoid imposing quasigeostrophic assumptions. Since both Ertel potential vorticity Q and potential temperature θ are conserved under such circumstances, *i.e.*,

$$\frac{dQ}{dt} = 0 \ ; \ \frac{d\theta}{dt} = 0 \ ,$$

it follows that

$$\left.\frac{dQ}{dt}\right|_{\theta} = 0 \ ,$$

where the derivative is taken on a isentropic surface θ =constant. Therefore isentropic maps of PV will reveal air motions over time scales (10 days or so) on which Q and θ can be assumed conserved.

First, a note about uncertainties inherent in the calculation of $Q = g\zeta_a \frac{\partial\theta}{\partial p}$, where ζ_a is absolute vorticity. For the most part, stratospheric analyses rely on satellite radiance retrievals. There are very few direct wind measurements (and none above 30hPa). So, even though modern analyses are quite sophisticated, in the end they rely on the thermal radiance data (its history as well contemporaneously). The relationship between vorticity iand the data is thus implicitly involves differentiation. In order to derive Q, therefore, it is necessary (implicitly) to take the Laplacian of the geopotential (the vertical integral of temperature). This procedure necessarily amplifies the small scales, which may be susceptible to errors in the data. The bottom line is that one needs to treat potential original data. The saving grace is that the phenomena we will be looking at are planetary in scale and the analyses seem to be capable of producing self-consistent potential vorticity analyses for these scales even in the southern hemisphere. (In fact, they seem to deal with small scales remarkably well, presumably becuase the forecasts hat ar factored into the assimilations do a good job.)

Examples of geopotential and potential vorticity analyses for the middle northern stratosphere in mid-winter are shown in Figs 20 and 21. [These are taken from McIntyre & Palmer, Nature, 1983; see also their longer paper in J.Atmos.Terr.Phys., 1984]. On 17 Jan 1979, the geopotential shows a polar vortex disturbed by waves 1, 2 and 3. This flow is primarily wavelike, in the sense that air parcels will orbit the pole, oscillating about their mean latitudes as they do so. On 27 Jan, wave 1 has grown much larger and a large, separate, cut-off anticyclone appears over the north Pacific (the "Aleutian anticyclone", a common feature of the northern winter stratosphere). Clearly, parcel displacements are no longer all wavelike: a parcel at A will (assuming this flow is steady, which of course it is not, really) orbit the cyclonic polar vortex, but a parcel at B will be "peeled off" around the anticyclone.

This behaviour is evident on the potential vorticity plots. On the 17th, we see high Q marking the vortex, with a region of weak Q gradients outside. On the 27th we can see high$-Q$ vortex air being peeled off around the anticyclone. The sustained equatorward displacements are reminiscent of the behaviour we found in the region of a critical line; in fact, the entire anticyclone (where the total velocity changes sign) is a finite-amplitude analogy to the critical line. Because of the deformation in this part of the flow (where it leaves the vortex on the eastern side of the anticyclone), there is a sustained cascade of Q to small scales as the "tongue" of vortex air is stretched out lengthwise. This kind of

594 ARTICLES NATURE VOL. 305 13 OCTOBER 1983

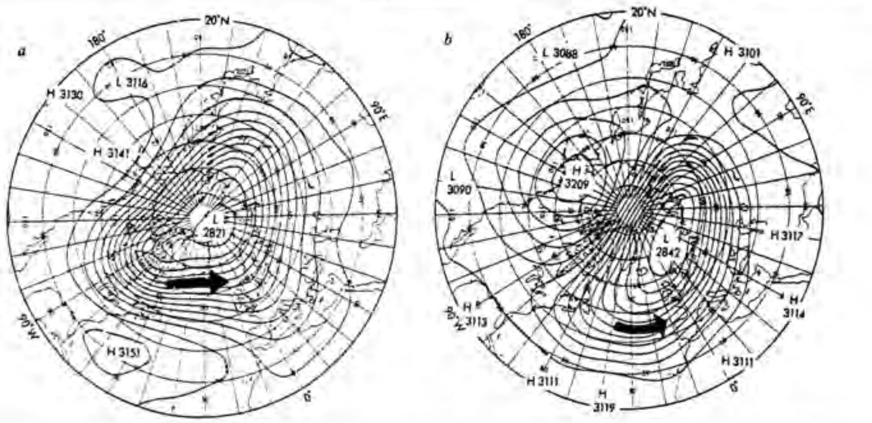


Fig. 1 Geopotential height in decametres (numerically nearly the geometric altitude above sea level) of the 10-mbar isobaric surface on 17 (*a*) and 27 (*b*) January 1979, at 00 h GMT. Contour interval is 24 decametres. Map projection is polar stereographic. The southernmost latitude circle shown is 20° N.

Figure 20: [McIntyre & Palmer, *Nature*, 1983]

NATURE VOL. 305 13 OCTOBER 1983 ARTICLES 595

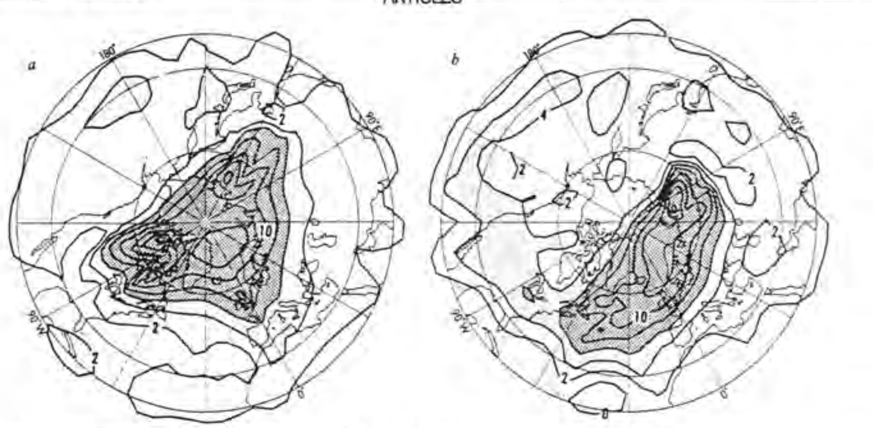


Fig. 2 Coarse-grain estimates of Ertel's potential vorticity Q on the 850 K isentropic surface (near the 10-mbar isobaric surface) on 17 (*a*) and 27 (*b*) January 1979, at 00 h GMT. The southernmost latitude circle shown is 20° N; the others are 30° N and 60° N. Map projection is polar stereographic. For units see equation (5) onwards. Contour interval is 2 units. Values greater than 4 units are lightly shaded, and greater than 6 units heavily shaded.

Figure 21: [McIntyre & Palmer, *Nature*, 1983]

behaviour is common in northern winter and is also evident, if less dramatically, in the southern hemisphere in spring. Overall, the impression is one of mixing of vortex air into midlatitudes; McIntyre & Palmer called this region the "surf zone", by analogy with ocean waves. The surf zone is just the stratospheric manifestation of the nonlinear Rossby critical layer.

As in the numerical experiments, the behaviour seen here can be interpreted as the planetary waves "breaking", just as ocean waves do running up a beach. Ocean waves do this by overturning the water surface in a vertical plane; planetary waves do it by overturning the potential vorticity contours sideways (*i.e.* in a horizontal plane)—and recall that it is stability of these potential vorticity contours that the wave is "riding" on, just as water waves "ride" on the stable water surface. McIntyre and Palmer defined "breaking" as *the irreversible distortion of the dynamically relevant material surfaces* (in this case, the PV contours).

This behavior is broadly similar to that shown in numerical results, but the fine-scale features are not apparent in the analyses. Of course, these satellite-based global analyses cannot resolve fine-scale filaments of the kind calculations predict. Some idea of what the stratospheric (or, for that matter, the tropospheric) potential vorticity structure might look like, if only we could resolve it, can be deduced from Figs 22 and 23. These figures show

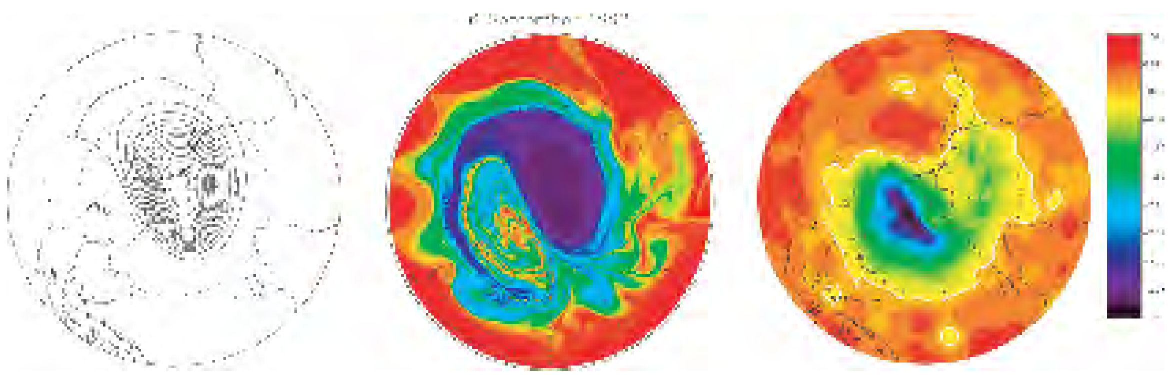


Figure 22: Maps for the southern hemisphere middle stratosphere, 6 September 1992. Left: Geopotential height (km) on the 10 hPa isobaric surface (ECMWF reanalysis data). Center: results of a "reverse domain filling" calculation, in which a tracer whose concentration is initially equal to latitude is advected by winds on the 1100 K isentropic surface (near 5hPa, 35km)for 10 days, ending 6 September (figure courtesy of Darryn Waugh). Right: net diurnally averaged diabatic heating rate (K day^{-1}) on the 10 hPa isobaric surface; white line is the $Q = 0$ contour, and the color scale for this frame is shown at right (data courtesy of Joan Rosenfield). The Greenwich meridian is at the right; the outer circle is the equator.

geopotential height (left frame), the results of high resolution, barotropic, numerical simula-

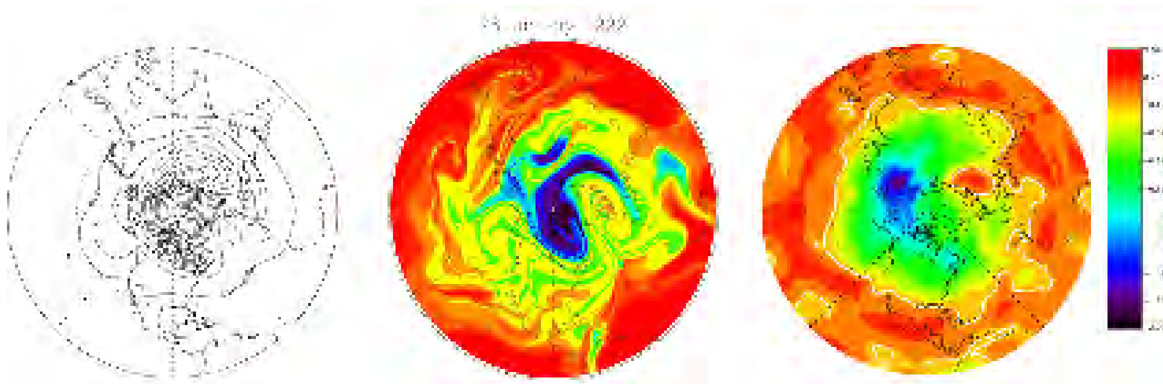


Figure 23: Maps for the northern hemisphere lower stratosphere, 28 January 1992. Left: Geopotential height (km) on the 50 hPa isobaric surface (ECMWF reanalysis data). Center: results of a “reverse domain filling" calculation, in which a tracer whose concentration is initially equal to latitude is advected by winds on the 480 K isentropic surface (near 60hPa, 19km)for 10 days, ending 28 January (figure courtesy of Darryn Waugh). Right: diurnally averaged net diabatic heating rate Q (K day^{-1}) on the 50 hPa isobaric surface; white line is the $Q = 0$ contour, and the color scale for this frame is shown at right (data courtesy of Joan Rosenfield). The Greenwich meridian is at the right; the outer circle is the equator.

tions of an initially smooth tracer field, advected isentropically for 12 days with the analyzed winds (center frame), and calculated diabaic heating rate(right frame). Since PV is quasi-conserved, we would expect it to exhibit similar fine structure to that of the tracer field (but of course this cannot be seen in low-resolution, global, observations). This does not imply that such features are evident in geostrophic streamfunction or geopotential, however; since (for barotropic flow)

$$q = \zeta = f + \nabla^2\psi \ ,$$

then

$$\psi = \nabla^{-2}(q - f) \ .$$

Therefore ψ is a “smoothed out” version of $(q - f)$ and the small scales would not be nearly so evident. Conversely, seeing a ψ (or Φ) field dominated by the largest scales (as it is in the stratosphere) should not delude us into thinking that the flow field is linear and laminar; Figs 20, 21, 22 and 23 show us clearly that it can be chaotic.

In situ aircraft observations show that the lower stratosphere (at least) is indeed rich in fine-scale features that appear to be filament crossings. An example of what is apparently a filament-crossing outside the edge of the Antarctic vortex is shown in Fig. 24.

We saw in numerical results that breaking is usually outward, given reasonable wind profiles. In reality, intrusions do occasionally occur, at least into the Arctic vortex. In fact

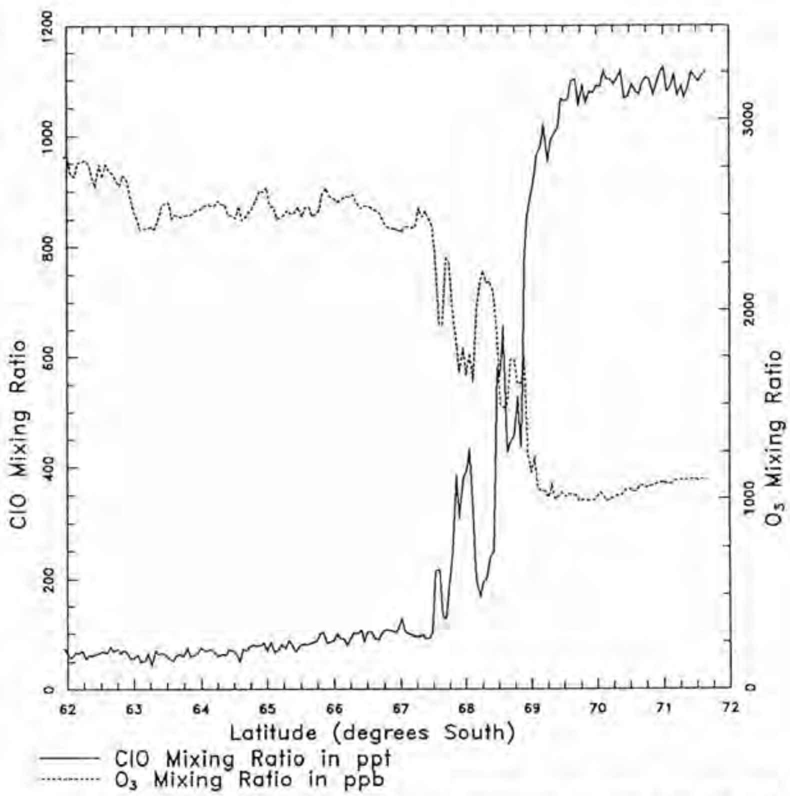


Fig. 14. Simultaneously observed ClO and O_3 obtained on September 16, 1987, by the ER-2, with corrections made for variations in potential temperature. Results shown here correspond to what the aircraft would have observed on a 450 K isentropic surface.

Figure 24: In situ observations of the concentrations of O_3 and ClO along the 450K isentropic surface, 1987 Sep 16. Note the sharp edge to the region of low O_3/high ClO, and the fine-scale (˜50km) features just outside the edge.

an example was shown in Fig. 23 (during an unusually disturbed period in the northern lower stratosphere); a magnified version of the central frame is shown in Fig. 25. Not only

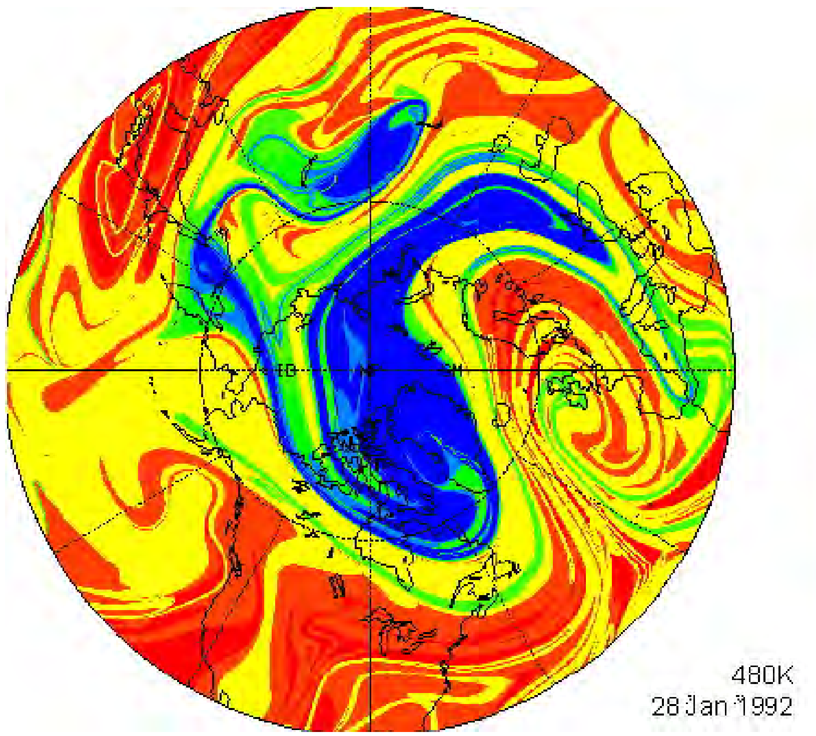


Figure 25: Structure of a conserved tracer advected for 12 days using analyzed winds on the 480K isentropic surface from smooth intial conditions (aligned with analyzed PV). Outer circle is 30^oN.

is vortex air entrained into the midlatitude "surf zone," but a filament of midlatitude air has intruded into the vortex, and has become stretched out within the vortex air by Jan 28. The presence of the intrusion was confirmed in aircraft data on 920124 (made very visible—to airborne lidar observations—by the presence in midlatitude, but not vortex, air of high concentrations of sulfate aerosol following the eruption of Mt Pinatubo the previous year).

Generally, however, intrusions are weaker and much less frequent than outward breaking events. Hence, air within the vortex is usually relatively isolated from midlatitude air during the course of the winter, an important factor in the chemical balances within the vortex.

Most of what we have discussed thus far involves entrainment of vortex air into the surf zone (and, occasionally, vice-versa). However, according to our simple picture in Fig. 11, the critical layer (*i.e.*, the surf zone) should have another boundary at its southern edge (in northern winter). As we shall see later, there is accumulating evidence from tracer data

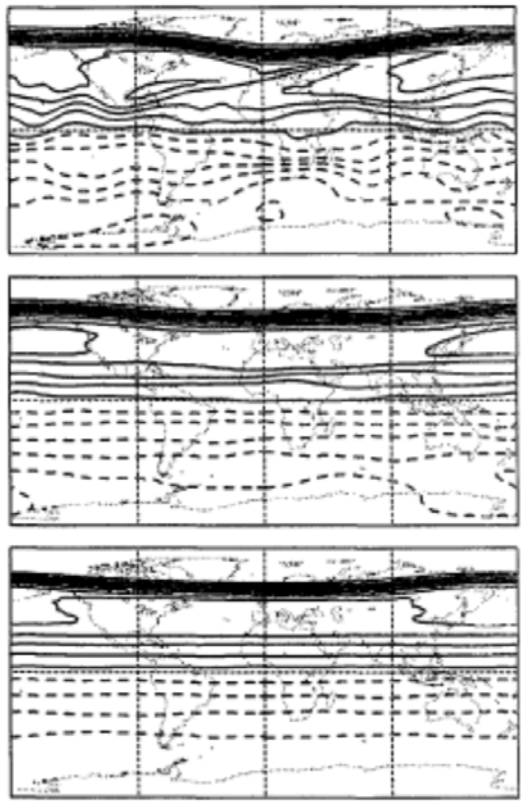


FIG. 13. The PV field at $t = 60$ days for the three profiles of Fig. 12. The [top, middle, bottom] figures correspond to the [solid, dashed–dotted, dotted] initial profiles in Fig. 12; that is, subtropical shear is smallest for the top figure and largest for the bottom one. Note how the stronger subtropical shear confines the surf zone to a narrow latitudinal band. Contour levels are as in Fig. 1.

that such an edge exists in some form in the winter subtropics. Moreover, observations and Lagrangian diagnostic modeling, such as was shown in Fig. 25, do indeed indicate entrainment into the surf zone of tropical air, just like what we have seen with polar air. However, there may be more complexity to this edge than is apparent at first sight. Fig. 3.3.6 shows results from a series of experiments with a shallow water model [Polvani et al., JAS, 52, p1288, 1995]. In this model, Rossby waves were forced by a mountain of zonal wavenumber 1, in a flow somewhat like what is typical of the winter stratosphere, with a polar night jet and tropical easterlies. Fig. 3.3.6 shows the PV distribution after 60d. of integration from 3 experiments with different initial wind profiles. In all three cases, a clear band of strong PV gradients marks the edge of the polar vortex. In the case with strongest tropical easterlies (bottom of Fig. 3.3.6), there is also a well-defined surf zone of weak gradients, which has a clear edge at about 30^{o}N, with a stronger PV gradient, and little activity, south of this. In the top figure, with the weakest tropical easterlies, there is still a band of relatively strong PV gradients in the northern subtropics, but it is less distinct and there is a significant amount of wave activity (and, in fact, wave breaking) going on across the equator and throughout the summer hemisphere. As the Rossby wave forcing is stationary, this is surprising, since we do not expect stationary waves to propagate through the tropical easterlies. Although the forcing is stationary, Rossby waves with nonzero (in fact, strongly westward) phase speeds are produced, mostly as a secondary effect following breaking of the primary, stationary, wave. These transients can (and do) propagate though the easterlies and break there, amongst other things rendering the subtropical surf zone edge indistinct. For the case with strong tropical easterlies, most of this activity breaks on the northern (winter) side of the equator, and leaves a sharp edge to the surf zone.