The General Circulation of the Atmosphere

Isaac M. Held and GFD/2000 Fellows

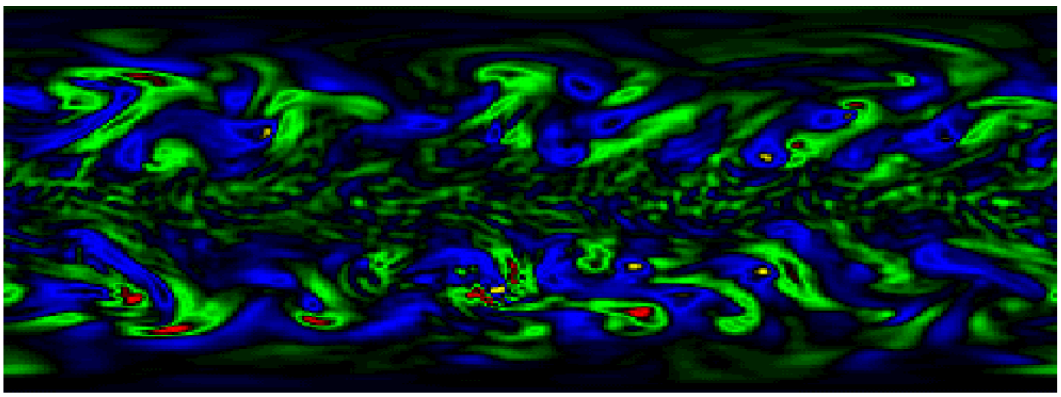

Figure 1: Mid-tropospheric vertical motion in an idealized dry atmospheric model with a zonally symmetric climate, forced as described in [7]. The entire sphere is shown. Note the wave-like structures in midlatitudes (with a NE/SW tilt in the Northern subtropics and the opposite tilt in the Southern subtropics) and the more turbulent, convective, character of the tropics. The model is spectral with T106 resolution. Green → upward; red → strongly upward; blue → downward; yellow → strongly downward.

1 Surface winds and vorticity mixing

Our goal in these lectures will be to develop a qualitative understanding of the general circulation of the atmosphere. As computer power increases, the problem of the general circulation is more and more often tackled with complex numerical simulations. Yet attempts at comprehensive simulations in isolation rarely produce satisfactory understanding. The challenge for theorists in the future will be to combine idealized models and complex, more realistic simulations in such a way as to produce a deeper understanding of the atmospheric circulation.

Studies of the general circulation focus on large scale structures of the atmospheric climate. These include the seasonally varying mean state and the statistics of eddies, on various space and time scales, that influence this mean state. Surveys of existing meteorological data define most of these large-scale structures rather well. The latitude-height distribution of the zonally averaged zonal winds is an important example (Figure 2) on which we will focus much of our attention.

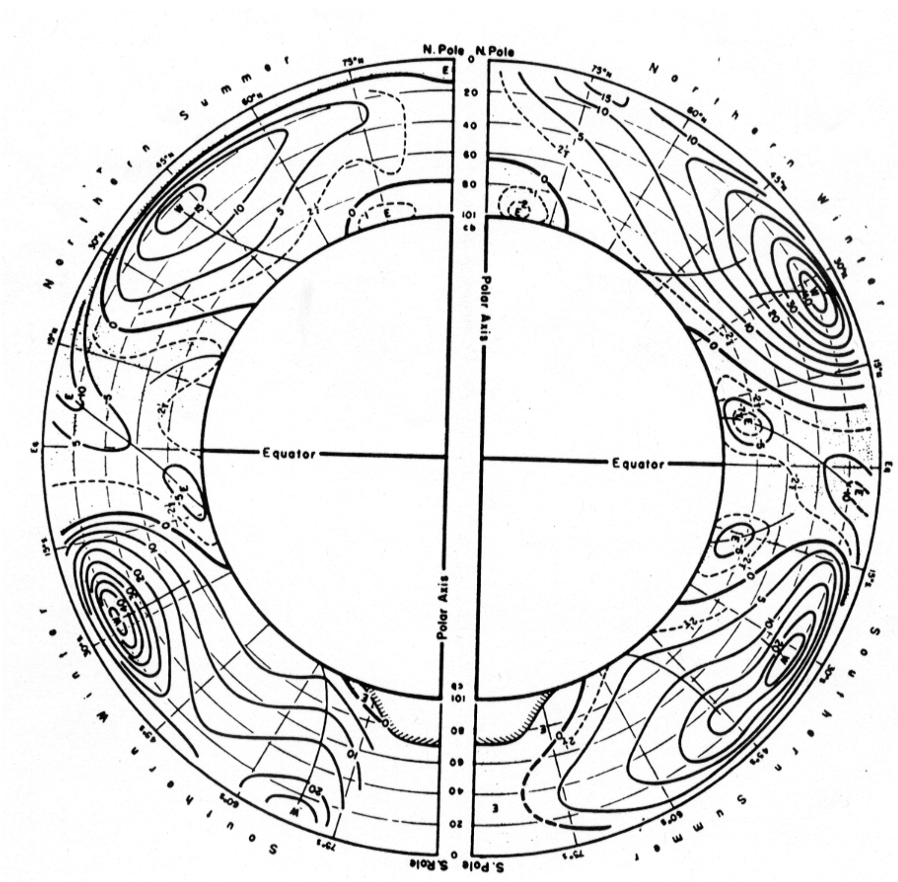

Figure 2: Zonally averaged zonal wind as a function of pressure and latitude for the summer and winter seasons in both Northern and Southern hemispheres, from [11]. The contour interval is 5 m/s. The vertical axis is marked every 200mb.

If the Earth's surface were zonally symmetric (independent of longitude), the atmospheric climate would be zonally symmetric as well. While there are substantial asymmetries in the surface, of course, due to mountains and the land-sea distribution (and the ocean circulation), the basic structure of the zonally averaged atmospheric climate does not depend strongly on these details. The similarity between the zonal mean flow in the Northern and Southern Hemispheres, despite the very different lower boundary structures, is good evidence of this. (The tendency for the upper tropospheric flow in the Southern winter to split into a subtropical and a subpolar jet in the zonal mean is the most distinctive interhemispheric difference in Figure 2.) There is a lot of interesting theory that helps us understand the deviations from zonal symmetry of the climate, but this topic will not be addressed here. Rather, we will focus on a hypothetical Earth with a symmetric lower boundary and a symmetric climate.

There are many starting points that one could choose in a discussion of the atmospheric climate. I have chosen the transport of angular momentum. (The term "angular momentum" in these lectures always refers to the component of the angular momentum vector in the direction of the Earth's rotation.) At first, we will focus on the *horizontal* redistribution of angular momentum by the extratropical circulation, which is intimately tied to the maintenance of the zonal mean surface wind distribution. Then we will move on to the vertical redistribution of angular momentum, which controls the vertical shear of the zonal wind in the extratropics, or, equivalently, the north-south temperature gradient. The angular momentum budget also provides the key ingredient in understanding the fundamental dynamical distinction between the tropical and extratropical atmospheres. We will touch on the interaction of moist convection and large-scale dynamics only in the discussion of the Hadley circulation.

The basic features of the zonal winds are very familiar. At the surface, we see easterlies in low latitudes, westerlies in midlatitudes, and weak easterlies again near the poles. The surface pressure distribution is in approximate geostrophic balance with these zonal winds, with subtropical highs and subpolar lows. The zonal force balance will be of greater interest to us than the meridional force balance, however. The zonal component of the pressure gradient averages to zero when integrated around a latitude circle. Therefore, the dominant terms in the zonally averaged zonal component of the force balance near the surface are the Coriolis force resulting from north-south motion and the frictional torques that retard the zonal winds. Therefore, the existence of an eastward frictional torque in the tropical region of surface easterlies requires, in a steady state, a zonally averaged equatorward flow, in order for the Coriolis force on this flow to balance the frictional torques. The return flow occurs near the tropopause. This tropical circulation is referred to as the Hadley cell. Over the midlatitude westerlies, the same argument requires the opposite sense of meridional overturning, referred to as the Ferrel cell. A weak polar cell over the polar surface easterlies completes the three-cell picture (see figure 3). The mass transports in the surface branches of these atmospheric meridional circulations are simply the boundary layer *Ekman mass transports* that balance the surface stresses.

In addition to the surface winds, we see in Figure 2 the familiar increase in the westerlies with height, consistent with the thermal wind equation and the north-south temperature gradient. If we have an understanding of the surface wind field and of the north-south temperature gradients, we have an understanding of the upper tropospheric flow as well.

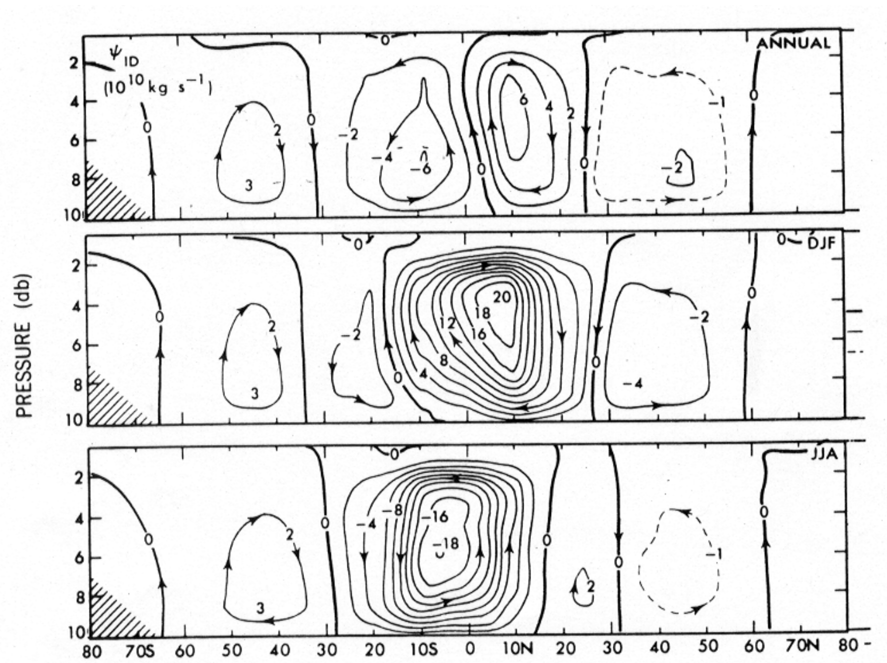

Figure 3: The traditional overturning streamfunction, shown for the annual mean and the summer and winter seasons. The contour interval is 10 Sverdrups; from [14]

A key feature in the mean circulation is the upper tropospheric subtropical jet, which can usefully be thought of as marking the boundary between the tropics and extratropics. Its latitude, to a good approximation, is also the latitude of the boundary between surface easterlies and westerlies.

The angular momentum per unit mass, M, is defined as

$$M = (u + \Omega a \cos\theta) a \cos\theta \tag{1}$$

where Ω is the angular velocity of the Earth, u the zonal flow of the atmosphere with respect to this solid body rotation, a the radius of the Earth, and θ is latitude. The first term is the contribution from the solid body rotation and the second term the contribution due to departures from solid body rotation. (We have made a thin shell approximation by assuming that the distance to the axis of rotation is simply $a \cos\theta$, independent of the height of the parcel above the surface.) The total angular momentum integrated over the atmosphere does not vary in time when averaged over a long enough time period

$$\frac{\partial}{\partial t} \int \rho M dV \approx 0$$

Due to surface frictional stresses, and *form drag* due to correlations between the surface pressure and the east-west slope of the surface, both of which tend to oppose the near-surface winds, angular momentum is transferred from the atmosphere to the Earth in midlatitudes and from the Earth to the atmosphere in the tropics and the polar regions. To obtain a steady state, angular momentum must be transported from the tropics and the polar regions to midlatitudes. The dominant path by far is that from the tropics to midlatitudes.

The poleward flux of angular momentum can be broken into contributions from the zonal mean flow (denoted by an overbar) and deviations from this mean (denoted by a prime)

$$\overline{vM} = \overline{v}\overline{M} + \overline{v'M'}$$

As first hypothesized by Jefferies in the 1920's and confirmed by observations when upper air data became available after WWII, this angular momentum transfer is mainly accomplished by large-scale eddies. Most of this eddy momentum flux occurs in the upper troposphere. The mean meridional circulation is too weak to make a significant contribution, except in the deep tropics. Understanding the distribution of surface winds and surface stresses is equivalent (outside of the deep tropics) to understanding the convergence of the eddy fluxes of angular momentum.

As a simple starting point, assume the upper troposphere can be modeled as a homogeneous $\rho = const$, nearly inviscid, nondivergent flow $\nabla \cdot \mathbf{u} = 0$ in a infinitesimally thin spherical shell. We first need to review two basic facts about vorticity: Stokes' Theorem (a kinematic result from vector calculus) and Kelvin's Circulation Theorem (a dynamic result for this homogeneous model).

By Stokes' Theorem, the circulation, the line integral of the velocity, around any closed loop is equal to the normal component of the vorticity, the curl of the velocity, integrated over any surface bounded by the loop:

$$\oint \mathbf{u} \cdot \mathbf{dl} = \int \int \omega \cdot \hat{\mathbf{n}} dA$$

where ω is the vorticity. If the loop is a latitude circle, we see that the circulation is just the zonally-averaged flow $\bar{u}$ times the length of a latitude circle. By choosing the bounding surface to be the surface of the sphere itself, we also see that the zonally averaged flow at some latitude is also the radial component of vorticity integrated over the polar cap bounded by this latitude circle, divided by the length of the latitude circle.

Kelvin's Circulation Theorem states, for our homogeneous, incompressible and inviscid fluid, that the circulation around a material loop (one moving with the flow), does not change in time. By Stokes' Theorem, it follows that the time derivative of the surface integral of the normal component of the vorticity, over any surface bounded by this material loop, is also equal to zero. If we specialize to the case of an infinitesimal loop, we have

$$\frac{D}{Dt}(\omega \cdot \mathbf{n} dA) = 0.$$

where $\mathbf{n}$ is the normal to the loop and dA is its area. For the nondivergent flow on a sphere considered here, the area of a material loop cannot change. Thus, the radial component of vorticity is conserved following parcels on a spherical surface.

In solid body rotation with angular velocity $\mathbf{\Omega}$, the radial component of the vorticity is $2\Omega sin(\theta)$. It is fundamental to large-scale atmospheric and oceanic flows that this is a monotonically increasing function of latitude, from the south pole to the north pole.

Suppose we stir the system with a 'spoon' at midlatitudes such that no net zonal momentum is added to the upper troposphere. (The latter restriction is meant to focus attention on horizontal redistribution of angular momentum; as we will see, the appropriate model of this stirring will, in fact, include a net tranfer of angular momentum from the upper to the lower troposphere.) Consider a latitude circle outside of the stirring region. If the disturbance reaches this location, the line of particles initially on the latitude circle will distort, conserving the total area encompassed by this line. When the particles move off the latitude circle, they conserve the radial component of their vorticity. Assume that the initial condition is close enough to solid body rotation that the vorticity distribution is monotonic. Then when the ring of particles deforms, high vorticity particles move south across the latitude circle and low vorticity particles move north. The net effect is an equatorward flux of vorticity at this latitude.

Thus, integrated over the polar cap bounded by this latitude circle, the radial component of vorticity will decrease. Therefore, by Stokes' Theorem, we know that $\bar{u}$ will also decrease (i.e. accelerate westward). Since we assumed no net angular momentum is imparted by the spoon to the layer as a whole, and since there is westward acceleration everywhere except in the region directly stirred by the spoon, the region stirred by the spoon must be accelerating eastward. *Angular momentum converges into the 'stirred' region.*

How can this convergence be maintained? Suppose that the stirring is episodic. After a burst of eddy activity, the flow might reversibly return to its initial condition, in which case the vorticity transport and associated convergence of angular momentum would be reversed during the decay of the disturbance. However, if the stirring generates irreversible mixing (i.e. pieces of vorticity break off and mix into the surrounding air), we cannot return to the initial state. Thus, we need to include some irreversibility in our model, we need to mix vorticity, to create sustained convergence of angular momentum.

Additionally, as this irreversible mixing continues, the gradients of vorticity will decrease and eventually disappear where this mixing is strongest, after which there will be no more vorticity transport and momentum flux convergence in these regions. If we are to take this barotropic model seriously as a model of a statistically steady state, we would need some sort of mechanism to restore the vorticity gradient. At this point, this model becomes a bit artificial – in the more realistic baroclinic case, as we will see, it is the fluxes of potential vorticity that play the role that vorticity fluxes play here, and (unlike vorticity) potential vorticity gradients can be restored by radiative heating.

This barotropic model is meant as a model of the upper troposphere, where the eddy momentum fluxes are concentrated. In midlatitudes there are no other mechanisms besides these upper tropospheric eddy fluxes that transfer significant amounts of angular momentum meridionally. The converging angular momentum in the midlatitude (stirred) region must be removed somehow. The only mechanism available is removal at the surface. How the momentum gets to the surface from the upper troposphere will be a dominant theme later in these lectures, For the time being we need only realize that it must get there somehow so as to be removed by surface torques.

Therefore, we should think of the midlatitude westerlies as existing so as to remove the angular momentum being sucked into these latitudes in the upper troposphere. This sucking is created by the meridional spreading of eddies away from their region of excitation in midlatitudes. *Westerlies appear under the stirred region.*

These upper tropospheric eddy momentum fluxes force the surface stress distribution and the associated low level wind field. They do not explain the upper tropospheric flow, which is more strongly controlled by the vertical shears. It is a common mistake to talk of how horizontal eddy momentum fluxes "maintain" the upper level winds. One should think of the upper level flow as the sum of the surface winds and the vertical wind shear. Only the former is closely tied to the eddy momentum fluxes.

We gain some intuition about momentum fluxes, not by thinking about the momentum equation directly, but by thinking about vorticity fluxes and changes in circulation. It is because vorticity is conserved following fluid parcels, in this barotropic model, that we can gain some intuitive feel for the structure of the eddy vorticity fluxes.

This model does not explain why the stirring occurs preferentially in midlatitudes on our Earth. In fact, as the rotation rate is increased in Earth-like climate models, one generally finds that the three-cell circulation is replaced by a five-cell pattern, with two regions of surface westerlies per hemisphere [22]. In idealized quasi-geostrophic models, it is easy to generate a dozen or more bands of surface westerlies [12]. The "stirring" evidently organizes itself in these models into multiple bands when the eddy scale is much smaller than the size of the domain. On Earth, the eddy scale is large enough that we have only one such band.

2 A linear perspective on eddy momentum fluxes

The discussion of meridional angular momentum redistribution in Lecture 1 did not make any reference to Rossby waves, or to any linearized dynamics for that matter. But by considering linear dynamics on a stable zonal flow we can gain additional perspective on the central result that a rapidly rotating flow, when stirred in a localized region, will converge angular momentum into this region.

The disturbances that carry angular momentum meridionally in the upper troposphere are, in fact, rather wavelike. We can decompose the covariance between u and v into contributions from different space and time scales by Fourier decomposing both fields in the zonal direction and in time. We can write the resulting co-spectrum as a function of zonal wavenumber k and frequency ω, or, equivalently, as a function of wavenumber and phase speed, $c \equiv \omega/k$:

$$\overline{u'v'} = \sum_k \int dc M(k,c)$$

Shown in Figure 4 is $M(k,c)$ integrated over all wavenumbers on the 200mb pressure surface at 38°N for the months of December thru March, plotted as a function of latitude. (Not evident because of the integration over wavenumber is the fact that zonal wavenumbers 4-7 dominate the flux.) The time-averaged zonal wind at this level and season is also shown as a heavy solid line. The zonal propagation of the eddies dominating the momentum flux is always westward with respect to the flow at this level (although typically eastward with respect to the ground), as we would expect from our simplest theory for Rossby waves.

Consider a system linearized about some zonal mean flow that is stirred with a particular space-time spectrum, localized in midlatitudes. Start with the barotropic vorticity equation for 2D flow on a β-plane, (ignoring spherical geometry for the moment)

$$\frac{\partial \zeta}{\partial t} = -u\frac{\partial \zeta}{\partial x} - v\frac{\partial \zeta}{\partial y} - \beta v + S - D \tag{2}$$

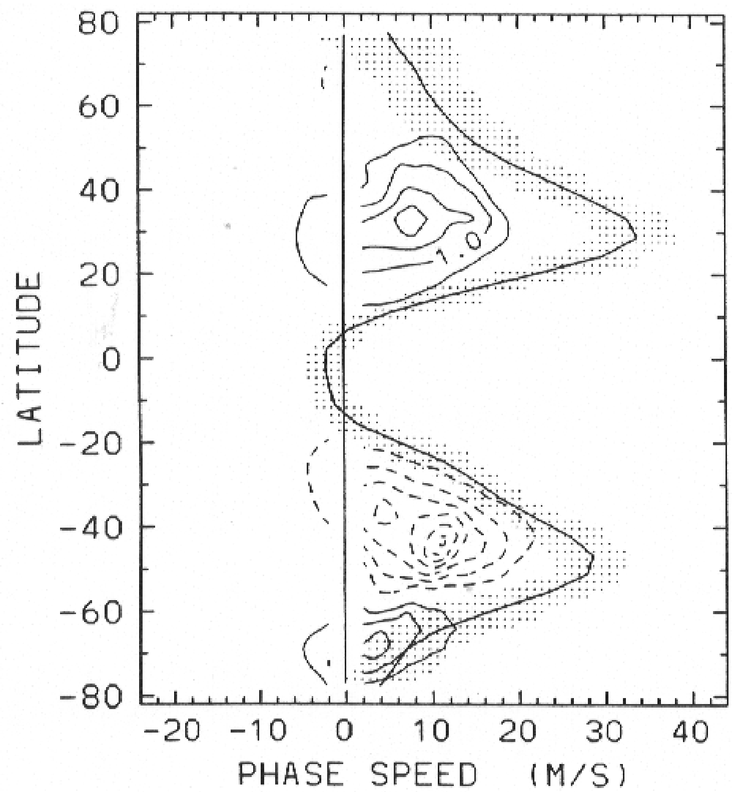

Figure 4: Phase speed spectrum of eddy momentum flux, $\overline{u'v'}$ at 200mb, 38°N, averaged over Dec.-Mar., from [15]. Solid contours → northward flux; Dashed contours → southward flux. In the Northern hemisphere, the convergence into midlatitudes is almost entirely one-sided (from the south); in the Southern hemisphere, there is some flux from the poleward side of the source region as well, implying significant poleward propagation

Here S is the stirring and D is some unspecified damping, and

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi \tag{3}$$

is the relative vorticity of the flow, the zonal and meridional components of the horizontal velocity being

$$(u, v) = \left(-\frac{\partial \psi}{\partial y}, \frac{\partial \psi}{\partial x} \right) \tag{4}$$

Linearizing about the zonal flow $\overline{u}$, we have

$$\frac{\partial \zeta'}{\partial t} = -\overline{u}\frac{\partial \zeta'}{\partial x} - \gamma v' + S' - D' \tag{5}$$

where

$$\gamma = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} \tag{6}$$

This stirring will generate waves that propagate away from the source, conserving their phase speed (on a sphere, their angular phase speed) and their zonal wavenumber. We can solve for each wavenumber and phase speed independently.

As the simplest case, suppose the background flow is uniform and neglect damping for the moment. Then, away from the source, for each space-time component the solution will simply be a plane Rossby wave,

$$\psi = A\cos(kx + ly - \omega t)$$

where A is a constant, and l must be determined, for given k and c, so as to satisfy the familiar Rossby wave dispersion relation:

$$\omega = ck = \overline{u}k - \frac{\beta k}{k^2 + l^2} \equiv \Omega(k, l)$$

There are two solutions for l:

$$l = \pm(\frac{\beta}{\overline{u} - c} - k^2)^{1/2}$$

both of which are real, assuming that propagation is allowed (i.e., as long as the quantity within the square root is positive).

By substituting the plane-wave form for ψ into the expressions for u' and v', one finds that the eddy momentum flux in this wave component is

$$\overline{u'v'} = -A^2\frac{kl}{2} \tag{7}$$

The product kl determines the direction of the eddy momentum transport: if $kl < 0$ then the eddy momentum flux is northward (positive $\overline{u'v'}$) and if $kl > 0$ then the eddy momentum flux is southward (negative $\overline{u'v'}$).

The appropriate choice of signs to the north and south of the source can be determined from consideration of the meridional group velocity of a Rossby wave.

$$G_y = \frac{\partial \Omega}{\partial l} = \frac{2\beta kl}{(k^2 + l^2)^2} \tag{8}$$

The meridional group velocity has the sign of kl. *The meridional flux of zonal momentum in Rossby waves is in the opposite direction to the group velocity.* North of the source, we satisfy the radiation condition by choosing $kl > 0$ so that G_y is positive. South of the source we choose $kl < 0$. The resulting Rossby waves converge positive zonal momentum into the source region from both directions, as we already know they must from the more general considerations in the first Lecture.

For those of unfamiliar with this use of a *radiation condition* in wave propagation problems, let's redo this computation in the presence of a small amount of damping of the form $D' = \kappa\zeta'$. Adding this damping is equivalent to adding a imaginary part to the frequency; outside of the source region, the plane wave solution must now satisfy

$$\omega + i\kappa = \Omega(k, l)$$

We can compute l by assuming that it is close to one of the solutions to the inviscid dispersion relation $l = l_0 + l_1$, with $\omega = \Omega(k, l_0)$. Using a Taylor expansion around l_0,

$$\omega + i\kappa = \Omega(k, l_0) + \frac{\partial\Omega(k, l_0)}{\partial l} l_1$$

From the definition of group velocity we have

$$l_1 = \frac{i\kappa}{G_y(k, l_0)}$$

Substituted back into the original wave solution, the result is

$$\tilde{\psi} = Re\left[exp\left(il_0 y - \frac{\kappa y}{G_y}\right)\right]$$

In the presence of damping the solution must decay away from the source. Therefore, to the north (south) of the source we must choose the solution that results in positive (negative) G_y.

With $kl > 0$ north of the source, lines of constant phase ($kx + ly =$ const) are tilted NW/SE. South of the source, lines of constant phase are tilted NE/SW. The equatorward propagation and NE/SW tilt are predominant in our atmosphere, with most of the convergence of flux into midlatitudes coming from the equatorward side. If we add on the mean zonal flow, the result is the "tilted trough" structure familiar to meteorologists (see Figure 1).

If $\overline{u}$ is not uniform, the picture does not change if we can make a WKB approximation. Eventually, as a wave propagates equatorward, for example, it may reach a latitude at which its (angular) phase speed matches that of the zonal flow – beyond which point no Rossby wave propagation is possible. If WKB remains valid, l increases rapidly as this point is approached, implying that the group velocity decreases, so that any damping that is present will have a long time to act and the wave will be dissipated. This argument is flawed, as one can show that WKB always breaks down before this "critical latitude" is reached. However, the conclusion is still correct – if linear theory is valid, and in the limit of infinitesimally small dissipation, the effects of dissipation are localized at each wave's critical latitude, where the waves are absorbed. We therefore expect waves with smaller phase speeds to propagate further into the tropics, as is clearly seen in the observations shown in Figure 4.

In this Rossby wave picture, the focus is on the momentum flux and the sign of the group velocity, whereas in the first lecture the focus was on the vorticity flux and Kelvin's circulation theorem, and no reference at all was made to a dispersion relation. We need to try to understand the relationship between these two arguments. For starters, we need to understand that the fluxes of (angular) momentum and vorticity are simply related. For our non-divergent flow (still ignoring spherical geometry) we have

$$\zeta' v' = (\frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}) v' = \frac{1}{2}\frac{\partial v'^2}{\partial x} - \frac{\partial u'v'}{\partial y} + u'\frac{\partial v'}{\partial y} = \frac{\partial}{\partial x}(\frac{1}{2}(v'^2 - u'^2)) - \frac{\partial u'v'}{\partial y}$$

Averaging over x leaves

$$\overline{\zeta' v'} = -\frac{\partial}{\partial y}(\overline{u'v'}) \tag{9}$$

so that the eddy vorticity flux is the divergence of the eddy momentum flux.

The equation of motion in the x-direction is

$$\frac{\partial u}{\partial t} = fv - \frac{\partial(u^2)}{\partial x} - \frac{\partial(uv)}{\partial y} - \frac{1}{\rho}\frac{\partial p}{\partial x}$$

Zonally averaged we get

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial}{\partial y}(\overline{u'v'}) \tag{10}$$

or, from (9),

$$\frac{\partial \overline{u}}{\partial t} = \overline{v'\zeta'} \tag{11}$$

where we have used the fact that $\overline{v} \equiv 0$

All of these results carry over to the sphere. For example. the spherical coordinate version of this relationship between vorticity and angular momentum fluxes is

$$\overline{v'\zeta'}\cos(\theta) = -\frac{1}{a\cos(\theta)}\frac{\partial}{\partial\theta}(cos^2(\theta)\overline{u'v'}) \tag{12}$$

The northward vorticity flux must sum to zero when integrated over latitude; in particular, if not identically zero, it must take on both positive and negative values.

Return to the linearized barotropic vorticity equation (5), multiply it by ζ'/γ and average over x. The result is:

$$\frac{\partial P}{\partial t} = -\overline{v'\zeta'} + \frac{\overline{S'\zeta'}}{\gamma} - \frac{\overline{D'\zeta'}}{\gamma} \tag{13}$$

where

$$P \equiv \frac{1}{2\gamma}\overline{\zeta'^2} = \frac{\gamma}{2}\overline{\eta^2} \tag{14}$$

The second form depends on the definition $\eta \equiv -\zeta'/\gamma$. One can think of η as the meridional parcel displacement that would create the perturbation vorticity, given the environmental vorticity gradient γ. P is referred to as the *pseudomomentum* density of the waves. (Pseudomementum is, strictly speaking, $-P$, but we will ignore this!) Since the vorticity flux integrates to zero, the pseudomomentum integrated over the domain is conserved, in the absence of sources and sinks, and we can write the mean flow modification as

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial P}{\partial t}$$

Consider a latitude circle at $y = y_0$ on which there are no eddies at $t = 0$. At time t, when there are eddies present, we simply have $\overline{u(y,t)} - \overline{u(y,0)} = -P(y,t)$.

If $\gamma > 0$, P is positive definite. When a wave enters a previously quiescent region, or, more precisely, if P increases, the mean flow will decrease. This is a restatement of our previous results in this linear context, in which we see that P is *the* measure of wave amplitude that is directly related to the mean flow modification.

If γ is positive everywhere, simultaneous growth of the eddy throughout the domain (as in a normal mode instability) is precluded, for this would require P to increase and the vorticity flux to be negative (downgradient) everywhere, which is impossible. This is why we specify external stirring to generate our eddies in this model; barotropic instability, resulting from a changes in sign of γ, is not the dominant process generating eddies in the upper troposphere.

To understand the expression for P, consider the fluid particles on a latitude circle, $y = y_0$, at time t. Look backwards in time, to $t = 0$, and locate these same particles, which trace out a curve $y - y_0 = \eta(x)$. For simplicity assume that this curve is gentle enough, as it would be for linear waves, that η is a single valued function of x. The vorticity flux through this latitude, integrated from 0 to t, can simply be computed from the original vorticity distribution, by integrating the vorticity over the regions that end up moving through the latitude in question. If η is small enough we get

$$\overline{\int_{y_0}^{y_0+\eta} (\zeta(0) + \gamma y) dy} = \frac{\gamma}{2}\overline{\eta^2} \tag{15}$$

(Question: is this η the same as that in the definition of P above?) For further reading on pseudomomentum, see [1], [2]. The construction leading to (15) suggests one path towards generalizing the definition of P to nonlinear disturbances.

The vorticity perturbation for a plane wave is $(-k^2 + l^2)\psi'$, so the pseudomomentum density is

$$P = \frac{1}{4\gamma}(k^2 + l^2)^2 A^2 \tag{16}$$

Using the expressions for the meridional group velocity of the Rossby wave (8) and the eddy momentum flux (7), we have the simple result

$$-\overline{u'v'} = G_y P. \tag{17}$$

This relationship continues to be valid when the flow is slowly varying rather than constant.

In general, in the presence of forcing and dissipation in the wave equation,

$$\frac{\partial P}{\partial t} = -\frac{\partial(-\overline{u'v'})}{\partial y} + sources - sinks \tag{18}$$

In a statistically steady state, the eddy momentum flux convergence is determined by the sources and sinks of pseudomomentum.

Eddy energy, unlike pseudomomentum, is not conserved by this linear model, but rather the wave can exchange energy with the mean flow. Start with (10), multiply both sides by $\overline{u}$ and integrate over y,

$$\frac{\partial}{\partial t}\int \frac{1}{2}\overline{u}^2 dy = \int \overline{u'v'}\frac{\partial \overline{u}}{\partial y}dy$$

Since the kinetic energy of the flow must be conserved, changes in zonal kinetic energy must be compensated by changes in eddy energy

$$\frac{\partial}{\partial t}\int \frac{1}{2}\overline{u'^2 + v'^2}dy = -\int \overline{u'v'}\frac{\partial \overline{u}}{\partial y}dy$$

The globally integrated eddy kinetic energy decays as eddies propagate from regions of large zonal flow to regions of weaker flow, producing a countergradient flux of momentum. On the sphere, the meridional gradient of angular velocity plays the role of $\partial_y \overline{u}$:

$$\frac{\partial}{\partial t}\int \frac{1}{2}\overline{u'^2 + v'^2}dy = -\int \overline{u'v'}\cos(\theta)\frac{\partial}{\partial \theta}\left(\frac{\overline{u}}{\cos(\theta)}\right)d\theta$$

The term "negative viscosity" is sometimes encountered (less often in recent years, fortunately) when describing the poleward flux of momentum in the subtropics. It should be avoided. It is sometimes used simply as another term for the inverse energy cascade in 2D flows, and as such it is harmless but unnecessary. (The inverse energy cascade will be discussed later on in these lectures.) Transfer of energy to the zonal flow cannot, in general, be equated with transfer to larger scales. (A barotropic instability of a zonal flow transfers energy from the zonal mean to eddies while, like any 2D flow, moving energy to smaller total horizontal wavenumbers.) We have seen that the pattern of momentum fluxes is most easily understood as a consequence of downgradient vorticity fluxes (or, more properly, as we will see, potential vorticity fluxes). Whether the momentum fluxes are up or down the angular velocity gradient is somewhat incidental.

Note also that the dominant poleward eddy angular momentum fluxes in the subtropics are up the angular *velocity* gradient but down the angular *momentum* gradient; $\overline{u}/\cos(\theta)$ increases with increasing latitude but the angular momentum decreases. The latter becomes important when we discuss the possibility of equatorial superrotation later in these lectures.

We need to know the sinks as well as the sources of pseudomomentum to understand the distribution of surface stress in midlatitudes. In the upper tropospheric layer under consideration, we are generally far removed from any direct effects of boundary layer turbulence, so "dissipation" of waves in this context should be thought of as a consequence of "wave breaking". If Rossby waves are large enough they cause the vorticity contours to overturn in the horizontal plane, causing a cascade of vorticity variance to small scales where it is eventually dissipated. Consider a wave that produces eddy zonal winds of magnitude u' at a particular latitude. Suppose that the wave amplitude is steady. Move to a reference frame translating with the wave so that the total flow is independent of time and so that streamlines, vorticity contours and particle trajectories all coincide. Being a Rossby wave, the mean flow in this reference frame, $\overline{u} - c$ is positive. But if

$$u' > \overline{u} - c \tag{19}$$

in part of the wave, streamlines and vorticity contours will overturn. Since the wave is not steady in reality, these overturning contours will result in wave breaking, i.e, mixing. If the wave is infinitesimal, it will break only when it reaches its critical latitude, where $\overline{u} = c$. More realistically, it will break before it reaches this latitude, by an amount dependent on the wave amplitude. One can sense this kind of behavior in Figure 5 above.

3 A shallow water model

So far, we have used a non-divergent two-dimensional model in our discussions. We now describe a model for fluid of finite depth flowing over topography. This system provides a simple framework for introducing some important concepts.

Consider a layer of homogeneous, inviscid and incompressible fluid with a rigid lid at $z = H_0$ and a rigid undulating lower boundary defined by $z = h(x, y)$. The thickness of the layer is $H(x, y) = H_0 - h(x, y)$. The thickness perturbation is $H'(x, y) = -h(x, y)$. The layer of fluid is assumed to have a small aspect ratio, i.e., the average depth H_0 of the layer is small compared with the horizontal length scale of the fluid motion. In this case, the vertical acceleration of the fluid can be neglected and we can make the hydrostatic approximation

$$\frac{\partial p}{\partial z} = -\rho_0 g$$

where ρ_0 is the constant density of the fluid. It follows by integrating from the top down that the horizontal pressure gradient is independent of depth, and equal to the pressure gradient on the rigid lid, so we can look for solutions for which the horizontal velocity is also independent of depth. The vertical velocity varies linearly with depth from its value at the lower boundary $w = \mathbf{v} \cdot \nabla h$ to zero at the top.

The equations that result from making the above assumptions are the shallow water equations, which we are specializing here to the case of a rigid lid. Kelvin's circulation theorem remains valid (although consistent with the hydrostatic approximation we ignore the contribution of vertical motion to the circulation.) As before, using Stokes theorem for an infinitesimal material loop with area δA,

$$\frac{D}{Dt}(\omega \cdot \hat{\mathbf{n}} \delta A) = \frac{D}{Dt}\left(\frac{\omega \cdot \hat{\mathbf{n}}}{H}(H\delta A)\right) = 0$$

Note that $H\delta A$ is the mass of the vertical column of fluid whose cross section is defined by the surface element δA; this mass must be conserved, so

$$\frac{D}{Dt}\left(\frac{\omega \cdot \hat{\mathbf{n}}}{H}\right) = \frac{D}{Dt}\left(\frac{f + \zeta}{H}\right) = 0$$

The potential vorticity

$$Q \equiv \frac{f + \zeta}{H} \tag{20}$$

is conserved following the flow.

We now make the quasi-geostrophic approximation, based on the assumptions that the flow is on a β-plane with $f = f_0 + \beta y$ and $\beta y << f_0$, that the vorticity of the motion relative to solid body rotation is small compared to the vorticity of the solid body itself, $\zeta << f_0$ and that the amplitude of the perturbation height is small compared with the thickness of the layer $h << H$. We can then expand Q in these three small parameters and obtain:

$$Q \approx \frac{f_0}{H_0} + \frac{1}{H_0}\left(\beta y + \zeta + \frac{f_0 h}{H_0}\right)$$

We define

$$q = \beta y + \zeta + \frac{f_0 h}{H_0}. \tag{21}$$

The effect of topography is now embedded in the conservation law for q:

$$\frac{\partial q}{\partial t} = -u\frac{\partial q}{\partial x} - v\frac{\partial q}{\partial y} \tag{22}$$

The fact that the flow is non-divergent to first approximation, due to assumptions of small Rossby number and $\beta y << f_0$, allows us to define a streamfunction ψ, as in the non-divergent case:

$$\left(-\frac{\partial \psi}{\partial y}, \frac{\partial \psi}{\partial x}\right) = (u, v)$$

The equation for the evolution of the zonally averaged flow in a shallow water model can be obtained by starting with the horizontal momentum equation:

$$\frac{\partial u}{\partial t} = fv - u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} - \frac{1}{\rho}\frac{\partial p}{\partial x}$$

or

$$\frac{\partial u}{\partial t} = (f + \zeta)v - \frac{\partial}{\partial x}\left(\frac{p}{\rho_0} + \frac{1}{2}(u^2 + v^2)\right)$$

Averaging zonally, we have

$$\frac{\partial \overline{u}}{\partial t} = \overline{(f + \zeta)v}$$

With the small Rossby number approximation of QG theory this reduces to

$$\frac{\partial \overline{u}}{\partial t} = f_0\overline{v} + \overline{v'\zeta'} = f_0\overline{v} - \frac{\partial \overline{u'v'}}{\partial y} \tag{23}$$

The replacement of the vorticity flux by the momentum flux convergence is exact if the flow is non-divergent. We can still make this substitution within the quasi-geostrophic approximation because the flow is non-divergent to lowest order.

Since f_0 is large compared to the relative vorticity, the meridional advection of the vorticity of solid body rotation by the (ageostrophic) mean meridional flow must be retained. It is important that $\overline{v}$ *need not be zero, even though the fluid is confined between rigid top and bottom boundaries*. Mass conservation does tell us that $\partial_y(\overline{vH})$ must be zero, since we cannot converge mass into any region with rigid lids. Assuming that the mass flux vanishes at northern or southern boundaries, $\overline{vH}$ must vanish everywhere, so that

$$\overline{vH} = \overline{v}\overline{H} + \overline{v'H'} = 0$$

So $\overline{v}$ can be non-zero if the eddy north-south flow is correlated with the thickness of the layer.

Linearizing the potential vorticity equation about a zonally symmetric mean state as before, we have

$$\frac{\partial q'}{\partial t} = -\overline{u}\frac{\partial q'}{\partial x} - v'\frac{\partial \overline{q}}{\partial y} - D$$

where the term D has been added to represent dissipation. We do not force the eddies with external stirring anymore – the zonally asymmetric topography is the source of the eddies. On multiplying by q' and taking the zonal average, we get

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial t}\left(\frac{\overline{q'^2}}{2\gamma}\right) = -\overline{v'q'} - \overline{Dq'}/\gamma$$

where $\gamma \equiv \partial_y \overline{q}$. This is the pseudomomentum conservation law for this system. In a statistically steady state, the value of the potential vorticity flux is determined by the dissipation; in the regions where $D = 0$, $\overline{v'q'}$ must be zero as well.

The eddy flux of potential vorticity is

$$\overline{v'q'} = \overline{v'\zeta'} - \frac{f_0\overline{v'H'}}{H_0} \tag{24}$$

This flux is comprised of two parts: the vorticity flux $\overline{v'\zeta'}$; and the contribution from the mass flux $\overline{v'H'}$. In a statistically steady state, and in the absence of dissipation, the sum of the two terms on the RHS must vanish. The individual terms in this expression need not vanish.

Now consider the situation in which the orography is confined to a particular zone of latitudes. Rossby waves will be generated by the flow over this orography and propagate meridionally away from this zone, into the region where the lower boundary is flat. Assume that the situation is statistically steady. Assume also that the wave amplitudes and horizontal shears are such that there is no wave breaking in the region of the orography; rather the waves break outside of this region. Therefore, the potential vorticity flux vanishes in the source region, but is negative in the region of wave dissipation far from the source. As we have seen, Rossby waves radiating away from a source region transport zonal momentum into this region. Therefore, within the source region, the momentum flux convergence must be balanced by an eddy mass flux:

$$\frac{f_0\overline{v'H'}}{H_0} = -\frac{\partial\overline{u'v'}}{\partial y}$$

The eddy mass flux must be poleward above the topography. Since there can be no total mass flux, the mass transport by the "mean meridional circulation", $\overline{v}$, must be equatorward *just as in the upper tropospheric branch of the Ferrell cell.*

This is not a coincidence – this is the essence of the Ferrell cell! Once again we need to consider this flow as a model of an upper tropospheric layer of fluid, forced by undulations in its lower boundary. Note first that nothing in the argument depends on the undulations being fixed in time. More precisely, as long as the zonal mean of the orography does not change, everything in this argument carries through unmodified. If the zonal means do change in time, then we need only add on the meridional mass transport required by mass conservation.

We can write the mean flow acceleration in terms of the potential vorticity flux by adding and subtracting the mass flux:

$$\frac{\partial \overline{u}}{\partial t} = f_0 \overline{v} + \frac{f_0}{H_0} \overline{v'H'} + \overline{v'q'} = f_0 \overline{v}^* + \overline{v'q'} \tag{25}$$

The residual circulation $\overline{v}^*$ is defined so as to be proportional to the total mass flux:

$$\overline{v}^* = \frac{\overline{vH}}{H_0} = \overline{v} + \frac{\overline{v'H'}}{H_0} \tag{26}$$

As we have already seen, from mass conservation in this simple rigid lid system, $\overline{v}^* = 0$.

It is interesting, and fundamental, that in this simple system it is the potential vorticity flux, rather than the vorticity flux, that determines the zonal acceleration of the mean flow. One way of thinking about this is to start with the zonal mean potential vorticity equation:

$$\frac{\partial \overline{q}}{\partial t} = -\frac{\partial \overline{v'q'}}{\partial y}$$

Since the rigid lids require that the mean thickness of each layer is unchanged,

$$\frac{\partial \overline{q}}{\partial t} = -\frac{\partial^2 \overline{u}}{\partial t \partial y}$$

One integration in latitude then yields, as before,

$$\frac{\partial \overline{u}}{\partial t} = \overline{v'q'} = \overline{v'\zeta'} - \frac{f_0 \overline{v'H'}}{H_0}$$

The final term of the RHS of this expression has an important physical interpretation: it is (the quasi-geostrophic approximation to) the force that the topography exerts on the fluid. The pressure force exerted by the atmosphere on the surface is normal to the boundary, so it has a zonal component when the surface has an east-west slope. If the slope is $\partial_x h \equiv \tan(\phi)$, then the eastward component of the pressure force is $p \sin(\phi)$. Making the small angle approximation (inherent in the shallow water model), we can ignore the difference between sin and tan so that the zonal component, averaged around a latitude circle, is

$$\overline{p_s \frac{\partial h}{\partial x}} = -\overline{h \frac{\partial p_s}{\partial x}}$$

where p_s is the surface pressure, related to the interior pressure by the hydrostatic relation:

$$p_s = p + \rho_0 g (z - h)$$

Since

$$\overline{h \frac{\partial h}{\partial x}} \equiv 0,$$

we can replace p_s by p, the pressure at a fixed height, in the expression for the zonal force on the surface. We can then use geostrophy,

$$\frac{\partial p'}{\partial x} = f_0 \rho_0 v'$$

to rewrite the force as

$$-f_0 \rho_0 \overline{v'h} = f_0 \rho_0 \overline{v'H'}.$$

The force exerted by the surface on the atmosphere is equal in magnitude and opposite in sign. Dividing this force per unit horizontal area by the mass per unit area, $\rho_0 H_0$ to get an acceleration we find the desired expression.

$$\frac{\partial \overline{u}}{\partial t} = ... - \frac{f_0}{H_0}\overline{v'H'}$$

When localized topography forces waves that propagate away from the region of excitation without being dissipated in this region, the resulting eddy momentum flux convergence is exactly balanced by the deceleration of the flow by the "form drag", or "mountain torque" exerted by the topography on the atmosphere! It is vitally important here that the "stirring" creating the eddies is not, as in Lectures 1 and 2, an externally specified vorticity source; the vorticity source here, the stretching associated with flow up and down the surface topography, is flow dependent. It is striking that this flow dependence produces this non-acceleration result for steady, inviscid waves. Potential vorticity dynamics is clearly the simplest way of seeing where this result comes from.

4 A Two Layer Model

Two layer models are a classical and natural starting point for thinking about the baroclinic dynamics of the troposphere. We choose to consider two isentropic layers of a compressible ideal gas, to make our model a bit more meteorological, rather than the more traditional two incompressible layers with fixed densities.

Evolution in the troposphere is sufficiently slow as compared to the time scales at which local thermodynamic equilibrium is maintained that we can assume that the entropy of a fluid parcel is conserved except for explicit diabatic effects, the most important of which is radiative heating. (We continue to consider a dry model.) The entropy per unit mass, s, of an ideal gas, with the equation of state $p = \rho R T$, can be written as

$$s = c_p \ln(\Theta); \;\; \Theta \equiv T/\Pi; \;\; \Pi \equiv \left(\frac{p}{p_*}\right)^{\kappa}; \;\; \kappa \equiv \frac{R}{c_p}.$$

Here T is the temperature, Θ the potential temperature, p_* is a reference pressure, and c_p the heat capacity at constant pressure.

Figure 5 shows the observed zonal mean isentropes in the troposphere. The potential temperature increases with height, at about 3K/km, indicating gravitational stability to dry convection. Given this vertical stability and the decrease in temperature with increasing latitude, isentropic surfaces tilt upwards with latitude in the troposphere. It is interesting that, roughly speaking, this slope is such that it takes one from the surface in the tropics to the tropopause in high latitudes. That is, the potential temperature difference between equator and pole is roughly the same as that between the surface and the tropopause. Is this a coincidence?

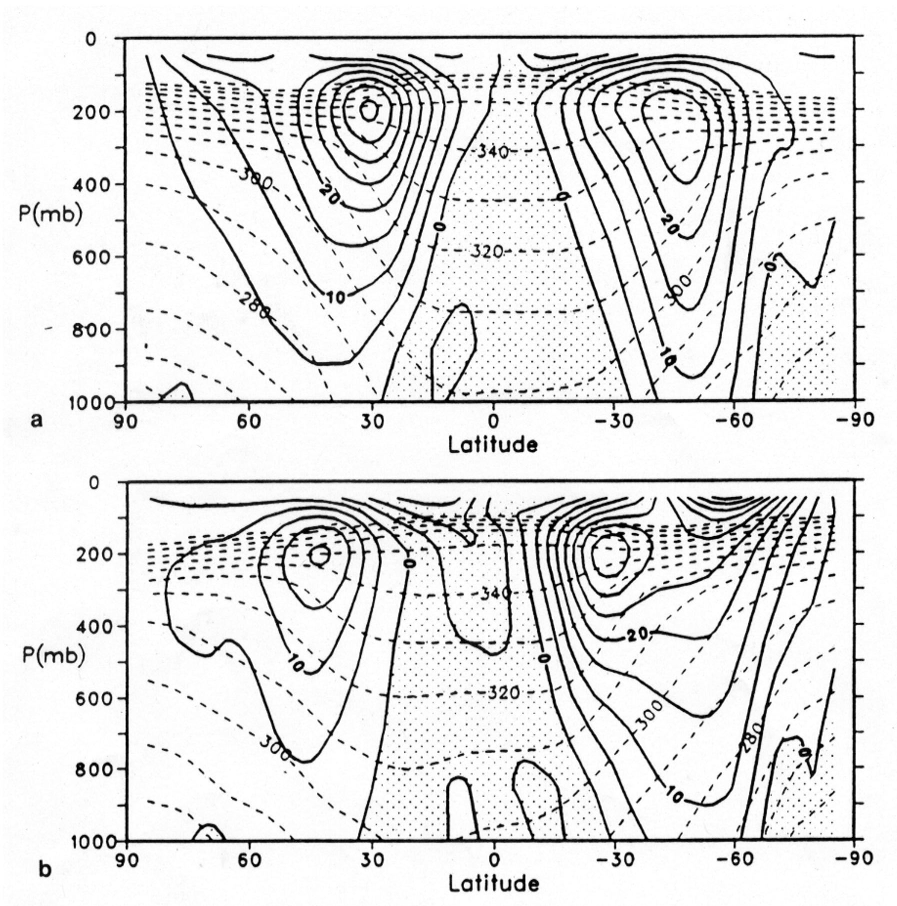

Figure 5: Observed zonal mean isentropes (dashed) and zonal winds (solid). The upper panel is DJF; the lower panel JJA. from [8]

We continue to focus on midlatitudes, and idealize the troposphere as consisting of two isentropic layers separated by one interface. We work in pressure coordinates, and speak of the pressure, rather than the height, of this interface. We allow the pressure at the surface (assumed to be at constant height) to vary, but we fix the pressure at the top of the upper layer. The subscripts 1 and 2 refer to the upper and lower layers respectively. The two potential temperatures are Θ_1 and Θ_2, with $\Theta_1 > \Theta_2$. p_B, p_M, and p_T are, respectively, the pressures at the bottom of the lower layer, the interface between the layers, and the top of the upper layer.

The hydrostatic approximation for an ideal gas is

$$\frac{\partial \Phi}{\partial \Pi} = -c_p \Theta$$

where Φ is the geopotential. If we integrate from the bottom upwards we get in lower and upper layers, respectively,

$$\begin{aligned}\Phi_2 &= c_p\Theta_2(\Pi_B - \Pi)\\ \Phi_1 &= c_p\Theta_2(\Pi_B - \Pi_M) + c_p\Theta_1(\Pi_M - \Pi)\end{aligned}$$

so that, computing the gradient at fixed pressure

$$\begin{aligned}\nabla\Phi_2 &= c_p\Theta_2\nabla\Pi_B\\ \nabla\Phi_1 &= c_p\Theta_2\nabla\Pi_B + c_p(\Theta_1 - \Theta_2)\nabla\Pi_M\end{aligned}$$

These gradients are independent of pressure within each layer, so we can look for solutions in which the horizontal flow is independent of pressure, analogous to shallow water theory.

Geostrophic motion in pressure coordinates takes the form

$$f(u,v) = (-\frac{\partial\Phi}{\partial y}, \frac{\partial\Phi}{\partial x})$$

So the thermal wind equation for the zonal flow, for example, is given by

$$f(u_1 - u_2) = -c_p(\Theta_1 - \Theta_2)\frac{\partial\Pi_M}{\partial y}$$

In the context of this two layer model, radiative heating can be thought of as converting some of the mass of the lower layer into mass of the upper layer, that is, as lowering the interface. Radiative cooling raises the interface. Keeping in mind that fixing the static stability is somewhat artificial in this context, radiative fluxes can be modeled by relaxing the interface to some steep "radiative equilibrium" value. It is the competition between the steepening due to the radiation drive and the shallowing of the interface due to the circulation that we are interested in. It should be clear from the start that the circulation must transport mass poleward in the upper layer, and equatorward in the lower layer, if it is to maintain a slope that is less steep than "radiative equilibrium".

Kelvin's circulation theorem remains valid for an inviscid ideal gas as long as the material circuit is confined to an isentropic surface. This centrally important result is responsible, to a great extent, for the utility of isentropic analyses in the atmosphere. In our case, we need only restrict ourselves to circuits that reside entirely in one or the other layer. Just as for the shallow water model of Lecture 3, we have

$$\frac{D}{Dt}(\omega\cdot\hat{\mathbf{n}}dA) = 0.$$

We use the notation

$$H \equiv \frac{\Delta p}{g}$$

for the mass per unit area of a layer, where Δp is the pressure thickness of the layer ($p_T - p_M$ in the upper layer and $p_M - p_B$ in the lower layer). Conservation of mass

$$\frac{D}{Dt}(HdA) = 0$$

implies

$$\frac{D}{Dt}\left(\frac{f+\zeta}{H}\right)=0$$

just as before. Similarly, under the quasi-geostrophic approximation, in each layer,

$$q=\beta y+\zeta-\frac{f_o H}{H_o}.$$

Also the mean zonal acceleration in each layer is again given by $\overline{v(f+\zeta)}$, and the same manipulations, using quasi-geostrophic scaling, yields the mean momentum equation in each layer in the "Transformed Eulerian Mean" form ([1], [2]).

$$\begin{aligned}\frac{\partial\overline{u_1}}{\partial t} &= fv_1^*+\overline{v_1'q_1'} &(27)\\ \frac{\partial\overline{u_2}}{\partial t} &= fv_2^*+\overline{v_2'q_2'}-\kappa\overline{u_2}. &(28)\end{aligned}$$

We have now included frictional drag in the lower layer, with κ being a linear drag coefficient. The northward mass flux in each layer is $\overline{v_i'H_i'}\equiv H_{0i}v_i^*$.

The QG potential vorticity flux can again be written as a sum of a vorticity flux, or momentum flux convergence, and a mass flux. In the lower layer, for example,

$$\overline{v_2'q_2'}=\overline{v_2'\zeta_2'}-\frac{f_o}{H_{o,2}}\overline{v_2'H_2'}$$

Focusing on the thickness flux,

$$\overline{v_2'H_2'}=\frac{\overline{v_2'(p_B'-p_M')}}{g}$$

By geostrophy

$$\overline{v_2'p_B'}=\frac{c_p\Theta_2}{f}\overline{\frac{\partial\Pi_B'}{\partial x}p_B'}.$$

However, Π_B is simply a function of p_B, so $\overline{v_2'p_B'}=0$ and

$$\overline{v_2'H_2'}=-\frac{\overline{v_2'p_M'}}{g}.$$

Meanwhile, in the upper layer,

$$\overline{v_1'H_1'}=\frac{\overline{v_1'(p_M'-p_T')}}{g}=\frac{\overline{v_1'p_M'}}{g},$$

since p_T is assumed to be constant. Finally, by the thermal wind relation,

$$\overline{(v_1'-v_2')p_M'}=\overline{\frac{\partial F(p_M')}{\partial x}p_M'}=0$$

where F is a function of p_M so that

$$\overline{v_1'p_M'}=\overline{v_2'p_M'}$$

and

$$\overline{v_1' H_1'} = -\overline{v_2' H_2'}.$$

The eddy mass fluxes in the two layers cancel. It might appear that we could have obtained this directly from mass conservation, but mass conservation only tells us that, in a steady state, the total mass fluxes in the two layers must cancel. Given this relation between the eddy fluxes, we know that the mean mass fluxes must also be equal and opposite in a steady state.

This result can once again be related to the form drag exerted by one layer on the other layer. Given an interface with height z_M, or slope $\partial_x z_M$, the drag per unit mass exerted on the lower layer by the upper layer is

$$F = \frac{1}{H_{o,2}} \overline{p_m' \frac{\partial z_M}{\partial x}} = \frac{1}{gH_{o,2}} \overline{p_M' \frac{\partial \phi_M}{\partial x}}$$

Using geostrophy

$$f v_2' = c_p \Theta_2 \frac{\partial \Pi_B'}{\partial x}$$

and the expression for Φ_M

$$\Phi_M = c_p \Theta_2 (\Pi_B - \Pi_M)$$

this yields

$$F = \frac{f_o}{gH_{o,2}} \overline{v_2' H_2'}.$$

Thus, the statement that $\overline{v_1' H_1'} = -\overline{v_2' H_2'}$ is, within the quasi-geostrophic approximation, a statement of Newton's Third Law.

Recall that in a non-divergent barotropic model the northward eddy vorticity flux must sum to zero when integrated over latitude. Using our result that the eddy fluxes of mass cancel in the two layers, and that the vorticity flux is the convergence of the eddy momentum flux in each layer, we have, as a generalization, that

$$\sum_{i=1}^{2} \int H_{o,i} \overline{v_i' q_i'} dy = 0, \tag{29}$$

where now we must sum over the two layers as well as integrate over latitude.

Because potential vorticity is conserved following the flow, linear conservative eddies that are growing must transport potential vorticity down the mean gradient. That is, multiplying

$$\frac{\partial q'}{\partial t} = -\overline{u} \frac{\partial q'}{\partial x} - v' \frac{\partial \overline{q}}{\partial y}$$

by q' and averaging zonally,

$$\frac{1}{2} \frac{\partial \overline{q'^2}}{\partial t} = -\overline{v' q'} \frac{\partial \overline{q}}{\partial y}$$

Coupled with the fact that the potential vorticity flux integrates to zero over the domain, this results in the Charney-Stern-Pedlosky criterion that the mean potential vorticity gradient must take on both signs if there is to be simultaneous growth everywhere, as in a normal mode instability. For nonlinear eddies, the variance of potential vorticity can be advected from one latitude to another, so the local relationship between eddy growth and downgradient flux can be broken:

$$\frac{1}{2}\frac{\partial \overline{q'^2}}{\partial t} = -\overline{v'q'}\frac{\partial \overline{q}}{\partial y} - \frac{1}{2}\frac{\partial \overline{v'q'^2}}{\partial y}$$

But integrating in y this nonlinear term does not contribute

$$\frac{1}{2}\frac{\partial}{\partial t}\int \overline{q'^2}dy = -\int \overline{v'q'}\frac{\partial \overline{q}}{\partial y}dy.$$

In a statistically steady state, the production of variance due to downgradient fluxes is balanced by the loss of variance resulting from a cascade to small scales where dissipative processes act. If we assume that a) all (or most, at least) nonconservative effects in the large-scale potential vorticity equation are dissipative, then the eddy potential vorticity flux must be downgradient on average. If, in addition, we assume that b) the advection of variance is small, then the flux must be downgradient at all latitudes. While there are a couple of assumptions required, the local downgradient restriction is a very useful one when developing models of the zonally averaged eddy potential vorticity flux. (Without zonal averaging, in a zonally asymmetric climate, the advection of eddy variance by the mean flow becomes much more important, and the problem is more complicated.)

Since the eddy fluxes cannot be everywhere of the same sign, given locally downgradient fluxes the zonal mean potential vorticity gradient cannot be everywhere positive. This is a (crude) generalization of the Charney-Stern-Pedlosky criterion to nonlinear statistically steady states, requiring the two assumptions a) and b) above.

The mean potential vorticity gradients are given by:

$$\frac{\partial \overline{q_i}}{\partial y} = \beta - \frac{\partial^2 \overline{u_i}}{\partial y^2} - \frac{f_o}{H_{o,i}}\frac{\partial H_i}{\partial y}.$$

Either the horizontal curvature or the thickness gradient must overcome β. If horizontal curvature is the primary reason for the sign change, we speak of barotropic eddy production; if it is the thickness gradient we speak of baroclinic production. In the troposphere, baroclinic eddy production is the dominant process. If the slope of the interface is sufficiently steep, the thickness gradient in the lower layer will overcome β, and the lower layer potential vorticity gradient will be negative. In a statistically steady state, we then expect poleward potential vorticity fluxes in the lower layer and equatorward fluxes in the upper layer. These potential vorticity fluxes are dominated by the eddy thickness fluxes, as the form drag at the interface transfers zonal momentum from the upper to the lower layer.

The acceleration due to the upper layer flux requires, in a steady state, a mass flux, i.e., a residual circulation v^*, that is poleward, so that the Coriolis force on this circulation balances the torque due to the potential vorticity flux (primarily the form drag). The residual circulation in the lower layer must carry the same mass flux equatorward. This circulation is consistent with the idea that the mass fluxes associated with the atmospheric circulation will be such as to fight the steepening due to the radiative driving.

In the lower layer, the mass transport is in the opposite direction to the Ekman drift associated with the surface westerlies, ie., it is opposite to the direction of the Ferrel cell.

As eddy momentum fluxes are small in the lower troposphere, this is equivalent to saying that the form drag on the upper surface of the lower layer, which is accelerating the zonal winds in that layer, must be larger than the surface drag, which is decelerating these winds. For the mass flux to be opposite in sign to the mean meridional circulation, it must be the case that the eddy mass flux (proportional to the form drag) is not only opposite in sign to the mass flux due to the mean circulation, but also larger in amplitude.

The Ferrel cell disappears when one considers mass fluxes rather than averages at constant height, pressure, or entropy. (The Ferrel cell is still present when averaging at constant entropy, despite the statements one sometimes hears to the contrary; it is the mass weighting that is crucial.) But this does not mean that the midlatitude overturning merges smoothly into the tropical overturning in the Hadley cell. Figure 6 is a plot of the mass flux, in the (entropy, latitude) plane computed from a GCM climate simulation for January. The atmosphere is divided into a large number of isentropic layers and the poleward mass flux is computed in each one; the cross-isentropic mass flux, due to radiation and latent heat release, is then computed from mass conservation. Also shown is the mean surface potential temperature. The key point is that the tropical poleward mass flux, ending at the latitude of the subtropical jet, occurs within very different isentropic layers than the shallower midlatitude overturning. The Hadley cell still has a well-defined meridional extent despite the fact that the flow in the upper troposphere is everywhere polewards.

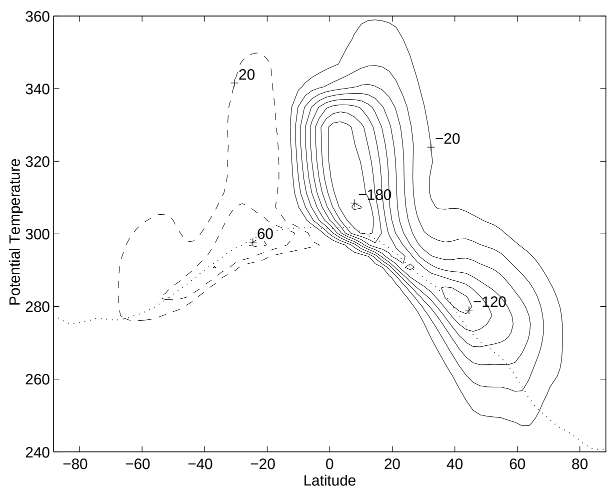

Figure 6: The mass transport streamfunction as a a function of potential temperature and latitude, in January, as computed from a GCM, from [6]. The dotted line is the median surface temperature.

If we sum the zonal mean equations of motion in the two layers, in their Transformed Eulerian form, assume a steady state, and use the fact that the vertically integrated poleward mass flux must be zero,

$$v_1^\star H_{o,1} + v_2^\star H_{o,2} = 0.$$

we have

$$H_{o,1}\overline{v_1' q_1'} + H_{o,2}\overline{v_2' q_2'} = \kappa \overline{u_2} H_{o,2} \tag{30}$$

The lower layer wind is determined by the vertical integral of the eddy potential vorticity fluxes.

Suppose, as is reasonable, that the potential vorticity fluxes are everywhere equatorward in the upper layer and poleward in the lower layer. The magnitude of the fluxes must be equal when integrated over latitude, but in order to produce surface westerlies in midlatitudes, we must now require the lower layer flux to be more sharply peaked in midlatitudes, and the upper layer flux to be more broadly distributed. Figure 7 shows this situation schematically, along with the opposite case, with a more broadly distributed lower layer flux that would produce midlatitude surface easterlies!

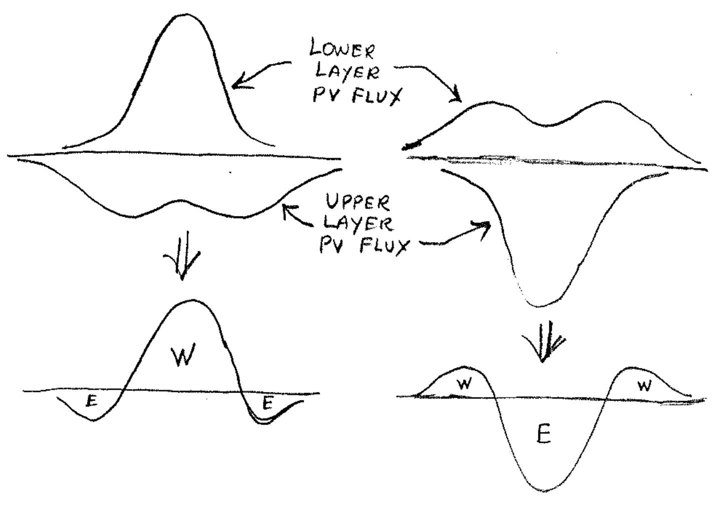

Figure 7: Schematic of two-layer potential vorticity flux distributions that would result in surface westerlies and surface easterlies underneath the region of baroclinic eddy production.

We now need a physical explanation for why we expect to see a broader flux in the upper layer. We expect that the eddies will form and then emanate from the source region until they become nonlinear, break, and mix potential vorticity. Heuristically, we would like to argue that eddies will be more linear in the upper layer, and thereby propagate more readily away from the source region at upper levels. Surface friction is part of the story, but I do not think that it is the dominant part. As a measure of linearity, let us again use the quantity $u'/(\overline{u} - c)$. When this quantity becomes large we expect wave breaking and mixing. Baroclinic instability theory (not discussed here) suggests that when β is small and the vertical shear and stratification are uniform, the steering level of the unstable waves (the height at which $\overline{u} = c$), is in the midtroposphere. However, the β-effect naturally encourages westward propagation, moving the steering level to lower levels. This implies that $\overline{u} - c$ is larger in the upper troposphere than in the lower troposphere, on average, so that the upper layer supports linear waves better than does the lower layer. Thus, the linear waves are able to propagate further away from the source region and the potential vorticity fluxes occur over a broader latitude band.

A positive feedback results from the generation of surface westerlies, since the presence of these westerlies itself causes $\overline{u} - c$ to decrease at the center of the storm track in the lower layer and to increase in the upper layer, accentuating this asymmetry in the nonlinearity. The low level component of a developing baroclinic wave breaks, or "occludes", in the source region, without propagating meridionally appreciably; upper level disturbances do disperse to some extent. Surface friction contributes to this asymmetry as well. We see momentum fluxes in the upper troposphere but relatively little in the lower troposphere. It is this asymmetry that motivates the stirred upper tropospheric model of the first few lectures.

Figure 8 shows the upper level potential vorticity in an idealized two-layer model of an baroclinically unstable jet on a β-plane. There is relatively little breaking at the jet core, marked by the concentrated red potential vorticity contours near the center of the channel, as compared to the breaking, or rolling up of potential vorticity, seen on both flanks. Indeed, it is this breaking and homogenization on each side of the jet that creates the sharp potential vorticity boundary between them, sharpening the jet. A similar figure at low levels shows maximum mixing at the center of the channel.

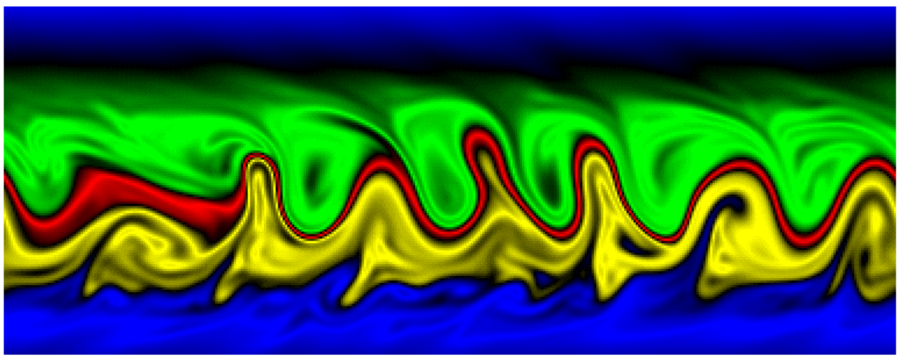

Figure 8: A baroclinically unstable jet

(Without spherical geometry it is difficult for baroclinic waves to propagate significantly away from their source region. On the sphere, as one moves equatorward, the Coriolis parameter in the thermal wind equation gets smaller and one enters a region of very weak temperature gradients and weak baroclinic eddy production even though the upper level winds are still strong enough to support wave propagation. In a QG model on a β-plane, if one concentrates the vertical shear in a jet, the upper level flow is not strong enough away from the jet to support significant propagation. So Figure 8 would be modified signficantly on the sphere, with the breaking on the equatorward side extended more broadly.)

We can think of eddy production and decay in midlatitudes as a two-step process. In the first step eddies grow baroclinically, reducing the vertical shear, decelerating the upper layer and accelerating the surface westerlies through form drag. In the second step, the upper level disturbance disperses meridionally, eventually breaking and mixing potential vorticity in a broad region, while the lower level eddy field occludes and dissipates. The upper layer deceleration is thereby compensated somewhat, as the dispersion of eddies, primarily into the tropics, redistributing the upper level deceleration towards the tropics.

The critical thickness gradient needed to allow baroclinic instability in this two-layer model is determined by setting the lower layer potential vorticity gradient to zero.

$$0 \sim \frac{1}{q}\frac{\partial q}{\partial y} \sim \frac{1}{f}\frac{\partial f}{\partial y} - \frac{1}{H}\frac{\partial H}{\partial y}$$

Since

$$\frac{1}{f}\frac{\partial f}{\partial y} \sim \frac{\overline{\beta}}{f} \sim \frac{\cot\theta}{a}$$

in midlatitudes the critical gradient is

$$\left.\frac{\partial H}{\partial y}\right|_c \sim \frac{H}{a}$$

The critical interface slope in this two-layer model is roughly that needed to carry one from the bottom of the model in the tropics to the top of the model near the poles. This is why we never use two-layer models when studying the entire sphere simultaneously. If the flow is baroclinically unstable in midlatitudes, the lower layer will awkwardly disappear in low latitudes, and the upper layer in high latitudes, leaving no vertical resolution over much of the globe. The two-layer model is only useful when isolating mid-latitude, or tropical, dynamics. (Two-*level* models are very common, but these are not models of a physically realizable layered system; they are just a crude finite differencing of the continuous equations.)

Suppose that the radiative drive is weak. Then we might expect to see an isentropic slope in the final statistically steady state that is close to this critical slope. This picture is often referred to as *baroclinic adjustment*. This seems like an attractive simple model for the observed isentropic slope in the troposphere. Among the difficulties with this picture is the simple fact that the atmosphere is not a two-layer system but has, instead, continuous stratification. Baroclinic instability theory tells us that the critical shear for instability, in inviscid flow, disappears in the simplest continuously stratified models. (One can think of the number of layers increasing and the thickness of the lowest layer decreasing, so that the isentropic slope required to produce the same *fractional* thickness gradient decreases.) Two-layer baroclinic adjustment is enticing, but unsatisfying.

5 Continuous stratification

We still have a problem explaining the mean overturning circulation in midlatitudes in the continuous limit.

Potential vorticity gradients are generally positive everywhere in the troposphere, because the β-effect dominates. Therefore, baroclinic instability is not generated by a reversal of the potential vorticity gradient in the most obvious way. As Charney showed, in a continuously stratified atmosphere the baroclinic instability arises because of the existence of a temperature gradient at the surface, with temperatures decreasing polewards. This surface temperature gradient plays the role of the lower layer in our previous model while the entire atmosphere acts as the upper layer! While the mathematics of quasi-geostrophic theory points to this interpretation, its physical implications are not that easily appreciated. Most important is the need to understand the implications of this picture for the overturning circulation in midlatitudes.

In the upper layer of the two-layer model, mixing results in an equatorward potential vorticity flux and a poleward mass flux. Why can't we apply this same argument to any infinitesimal isentropic layer within the troposphere, which would require a poleward flux in every such layer? Where is the return flow? Quasigeostrophic theory indicates that the return flow is confined to a δ-function at the surface(!). We need to develop a physical understanding of this theoretical construct.

When isentropic layers intersect the surface, the thickness of the layer goes to zero and we have some difficulty in defining potential vorticity. But as long as we are examining a layer that is safely removed from the surface despite being buffeted by baroclinic eddies, we have no such difficulty. In uninterrupted layers, potential vorticity is well defined and, we assume, increases poleward within the layer in the mean. Mixing then produces an equatorward potential vorticity flux and a poleward mass flux. It seems that we have to look to the interrupted layers for the equatorward mass flux. This idea is supported by Figure 7, where one sees that much of the return flow occurs within isentropic layers that are close to, and to a large extent, colder than, the mean surface temperature. Nearly all of this flux is confined to interrupted layers, as discussed in [6]. *The return flow can be thought of as occurring in cold air outbreaks near the surface.*

Consider the atmosphere at a particular latitude. Divide the atmosphere into two parts: isentropic layers that sometimes intersect the surface at that latitude, and isentropic layers that never (or hardly ever) intersect the surface. We can estimate the total depth of the interrupted layers, h_I as roughly

$$h_I \approx \frac{\Theta'}{\partial\overline{\Theta}/\partial p}.$$

An estimate of the RMS of Θ' in midlatitudes is $10°K$, and $\partial\overline{\Theta}/\partial p$ is about $3°K/km$. Thus, the interrupted layer extends through about one-third of the troposphere. In quasi-geostrophic theory this thickness is assumed to be infinitesimal! (It is assumed that static stability variations are small compared to the mean static stability, but the former are assumed to scale as the temperature perturbation divided by the total depth of the troposphere, H, which implies that $h_I << H$).

Assume that surface temperature is distributed symmetrically about the mean, as seems in fact to be a fairly good approximation in the atmosphere. The probability distribution $P(\Theta_s)$ is normalized so that

$$\int_0^\infty P(\Theta_s)d\Theta_s = 1.$$

Assume, in addition, so as to provide a very simple way of thinking about mass fluxes in the interrupted layers, that the surface potential temperature is perfectly correlated with the meridional velocity throughout these layers:

$$v' = \Theta_s'\alpha.$$

This actually holds fairly well in our atmosphere – when the surface wind is from the south the air is warmer than average, a consequence of the fact that surface temperatures are very strongly forced and the mixing induced by baroclinic waves is not able to deform this distribution profoundly.

Examine the mass flux in an infinitesimally thin layer of potential temperature Θ. Assume that the static stability is uniform, so that whenever the layer exists (i.e. $\Theta > \Theta_s$), it has a constant thickness H. Thus, the instantaneous mass flux through the layer is Hv'. However, we also need to take into account how often the layer is present, so we weight the instantaneous mass flux by the probability density. The average mass flux M through the layer is

$$M(\Theta) = H \int_0^\Theta P(\Theta_s) v' d\Theta_s. = H\alpha \int_0^\Theta (\Theta_s - \bar{\Theta}_s) P(\Theta_s) d\Theta_s. \tag{31}$$

From this expression, we see that this mass flux decays to zero when either the surface temperature is much colder or much warmer than the mean. When the temperature is too warm, the layer is almost always present, but both poleward and equatorward flows occur and cancel, resulting in very little net mass flux. When the temperature is too cold, the layer is never present so, once again, there is no flux. The maximum mass flux is expected to occur in layers with temperatures that are close to the mean surface temperature. In fact, if the probability distribution is a Gaussian, and accepting the perfect correlation between perturbation temperature and meridional wind, the distribution of the mass flux will also be a Gaussian, centered at the mean surface temperature. Half of the mass transport will occur in layers that are colder than the mean.

We have neglected the planetary boundary layer (PBL) in determining the surface layer mass transport. This is a problem because the PBL is well-mixed, so there is no vertical entropy gradient. To account for the PBL, we need to recognize first that the argument above applies to the interrupted layers above the PBL, and that the static stability referred to above is the static stability in the free troposphere. If we assume that the PBL is always present, and if its thickness is not correlated with surface temperature, then there will be no net mass flux in the layer, although it will contribute poleward transport to those isentropic layers that are above the mean and equatorward transport to those below the mean. This will simply have the effect of shifting the distribution of the return flow to colder temperatures. (See [6] to read more about this confusing point.) Additional effects could arise due to correlations between the thickness of the PBL and the surface temperature, however.

Potential temperature is conserved along the surface in the absence of diabatic forcing, and midlatitude eddies try to mix the potential temperature (although they are far from successful in homogenizing it) causing a downgradient flux in the usual way. If we assume that the meridional flow has little vertical structure throughout this layer and that the mean static stability is uniform (both of these assumptions are motivated by the quasi-geostrophic scaling limit) then the mass flux is simply

$$\frac{1}{g}\frac{\overline{v'(\Theta_I - \Theta_s')}}{\partial\overline{\Theta}/\partial p} = -\frac{1}{g}\frac{\overline{v'\Theta_s'}}{\partial\overline{\Theta}/\partial p}$$

Here Θ_I is a constant that marks the top of the interrupted layer at the latitude of interest. The mass flux in this layer is equal, but opposite in sign, to the surface heat flux divided by the static stability above the planetary boundary near the surface. *The interrupted layers play the role of the lower layer in the two-layer model*, with the irreversible mixing of potential temperature along the surface generating an equatorward flow balancing the poleward flow forced by potential vorticity mixing in the uninterrupted interior layers.

The Coriolis force on this mass transport balances the form drag at the top of this layer. There is, in addition, mass transport in the surface layers needed to balance the surface stresses. The former is due to geostrophic eddies, the latter to an ageostrophic Ekman drift.

6 The Hadley cell

Return now to the budget of angular momentum $M = (\Omega a \cos\theta + u) a \cos\theta$ in the latitude-pressure plane. In pressure coordinates ($\omega \equiv Dp/Dt$):

$$\frac{\partial \overline{M}}{\partial t} = -\frac{1}{a\cos\theta}\frac{\partial\left(\cos\theta\overline{vM}\right)}{\partial\theta} - \frac{\partial\left(\overline{\omega M}\right)}{\partial p} \tag{32}$$

or

$$a\cos\theta\frac{\partial\overline{u}}{\partial t} = -\frac{1}{a\cos\theta}\frac{\partial\left(\cos\theta\overline{v}\,\overline{M}\right)}{\partial\theta} - \frac{\partial\left(\overline{\omega}\,\overline{M}\right)}{\partial p} \tag{33}$$

$$-\frac{1}{a\cos\theta}\frac{\partial\left(\cos\theta\overline{v'M'}\right)}{\partial\theta} - \frac{\partial\left(\overline{\omega' M'}\right)}{\partial p} \tag{34}$$

A bit of algebra, utilizing the relation between the mean absolute vorticity and meridional gradient of the mean angular momentum,

$$f + \overline{\zeta} = -\frac{1}{a^2 cos(\theta)}\frac{\partial\overline{M}}{\partial\theta} \tag{35}$$

yields

$$\frac{\partial\overline{u}}{\partial t} = \left(f + \overline{\zeta}\right)\overline{v} - \overline{\omega}\frac{\partial\overline{u}}{\partial p} - \frac{1}{a\cos^2\theta}\frac{\partial\left(\cos^2\theta\overline{u'v'}\right)}{\partial\theta} - \frac{\partial\overline{\omega' u'}}{\partial p} \tag{36}$$

For the purpose of developing a qualitative picture of the circulation, we can think of the final term on the RHS, the vertical eddy flux of zonal momentum, as only important in the planetary boundary layer.

One could, at this point, perform analogous manipulations to those used in our discussion of the two-layer quasi-geostrophic model, to express the mean flow modification in terms of the residual circulation. But we now want to discuss the low latitude circulation, and eddy thickness fluxes and the associated form drag drop to very small values rather quickly as one moves equatorward of the subtropical jet. The potential vorticity flux is dominated by the momentum flux convergence, and the residual circulation and the Eulerian mean circulation are nearly the same. So we will not bother with this reorganization here.

If we also ignore the transport of relative angular momentum by the mean meridional circulation, we have, outside of the boundary layer,

$$\frac{\partial\overline{u}}{\partial t} = 0 = f\overline{v} - \frac{1}{a\cos^2\theta}\frac{\partial\left(\cos^2\theta\overline{u'v'}\right)}{\partial\theta}$$

which is the expression that formed the basis of our discussion of the momentum balance in midlatitudes.

The low Rossby number assumption ($\zeta << f$) breaks down as one moves to lower latitudes. This results in a fundamental change in the way in which the overturning circulation in the troposphere is controlled. In the low Rossby number limit the overturning circulation is a slave to the eddy stresses in a steady state, as the Coriolis force on the meridional circulation must balance the stress. In low latitudes, one intuitively expects that the thermal forcing must exert a direct effect on the overturning. Since the eddy stresses are due to potential vorticity mixing associated with disturbances generated in midlatitudes and spreading into the tropics, the connection between these stresses and tropical forcing is subtle at best, so it is difficult to see how this effect can be mediated through changes in these eddy stresses. To clarify this point, it is useful to simplify the problem by dropping the vertical advection of momentum by the mean circulation and retaining only the horizontal advection:

$$0 = \frac{\partial \overline{u}}{\partial t} = (f + \overline{\zeta})\overline{v} - S \tag{37}$$

where S is the eddy stress. While not obviously justifiable by a scale analysis, it turns out that the vertical advection term would modify our arguments in only a modest way. (It just so happens that the strongest vertical mean flows in the tropics occur where the vertical shear is weak. If we consider "superrotating" atmospheres, in which there is strong vertical shear in the deep tropics, then this term does become important – see [19].)

The overturning is no longer tied to the eddy stresses. If the thermal forcing in the tropics is increased, for example, one can imagine that $\overline{v}$ will increase, even if the eddy stresses in the upper troposphere are not changed. One can still satisfy the momentum balance if the absolute vorticity is reduced in the upper troposphere. Since the absolute vorticity is just the gradient of angular momentum, one is simply saying that, in the presence of a stronger circulation, the same stresses do not have as much time to extract zonal momentum from the poleward moving air.

In the extreme case in which the eddy stresses are reduced to zero, the mean meridional circulation has two options. Either $\overline{v}$ vanishes (the only possibility in the low Rossby number limit), or $f+\overline{\zeta}$ vanishes. In the latter case, a mean flow still exists, that conserves its angular momentum as it moves polewards. The circulation must then clearly be determined by the thermal forcing.

This is, in a nutshell, the key dynamical distinction between the tropical and midlatitude general circulations. In the tropics, the overturning circulation can and does respond to changes in thermal forcing, independently of changes in eddy stress. In midlatitudes, the overturning cannot be altered significantly without altering the eddy stresses. What determines the size of the "tropics" from this perspective?

Consider the upper layer of a two layer model, assuming homogeneous incompressible layers for simplicity. The zonal flow in the lower layer is assumed to be negligible due to surface friction. The flow is driven by "radiative forcing" that produces mass fluxes between the two layers which relax the interface towards some radiative equilibrium shape. The pressure at the upper boundary is assumed to be constant. The steady-state equations for axisymmetric flow in the resulting reduced-gravity shallow water model are (dropping the overbars)

$$
\begin{aligned}
0 &= \frac{\partial u}{\partial t} = (f+\zeta)v - S & (38)\\
0 &= \frac{\partial v}{\partial t} = -fu - \frac{g^\star}{a}\frac{\partial H}{\partial \theta} - \frac{v}{a}\frac{\partial v}{\partial \theta} - u^2\frac{\tan(\theta)}{a} & (39)\\
0 &= \frac{\partial H}{\partial t} = -\frac{1}{a\cos(\theta)}\frac{\partial}{\partial \theta}(\cos(\theta)vH) - \frac{H - H_{eq}(\theta)}{\tau} & (40)
\end{aligned}
$$

where H is the thickness of the upper layer, $H_{eq}(\theta)$ the radiative equilibrium thickness, τ the radiative relaxation time, and $g^\star$ the reduced gravity. We have ignored the vertical momentum transfer associated with the mass transfer between layers, equivalent to neglecting the vertical advection of momentum in the discussion above. (Again, this is adequate for circulations similar to that existing on Earth but not for flows with strong vertical shear at the equator.) For the parameter range we are interested in the meridional flow is much weaker than the zonal flow, and we can ignore the meridional advection of the meridional flow itself in the v-equation. (One can check after the fact that our solutions would be modified by this term only in a very small region around the equator extending to only $1 - 2^\circ$ latitude.) This would leave us with gradient wind balance, but neglecting the term proportional to u^2 is also justified as long as the zonal flows are weak compared to Ωa. The meridional equation of motion is then simply geostrophic balance.

$$
0 = \frac{\partial v}{\partial t} = -fu - \frac{g^\star}{a}\frac{\partial H}{\partial \theta}
$$

We still need to specify S. Let's assume that $S = \kappa s(\theta)$, with $s > 0$ everywhere except at the equator where $s = 0$. κ is a parameter that allows us to pass to the inviscid limit. We will be considering solutions that are forced symmetrically about the equator, so that $v = 0$ at the equator. If $u = 0$ at the equator initially, this will be true for all times, and we will only look for solutions of this type. If S does not vanish at the equator, this model could not reach a steady state, since we have neglected the vertical momentum exchange that would be required in this model to balance any imposed stresses at the equator! As an additional simplification, we can linearize the thickness equation about the mean thickness, H_0. The only important nonlinearity in this model is in the zonal momentum equation.

The maximum value of angular momentum must be attained at the equator in this model. Suppose this were not the case, and there were a point off the equator at which M were a local maximum. The absolute vorticity would vanish at this point, but there would be nothing balancing non-zero eddy stresses. Therefore, we must have $u \leq u_M(\theta)$, where u_M is the wind field obtained by conserving angular momentum, stating with $u = 0$ at the equator.

When $\kappa = 0$, one solution of these equation is simply radiative equilibrium: $H = H_{eq}; u = u_{eq}, v = 0$. But this cannot be the limit of the solutions for non-zero κ as $\kappa \to 0$, since u_{eq} does not vanish at the equator. This is a special case of a centrally important fact: radiative equilibrium temperatures cannot, typically, be realized close to the equator, for the temperature gradients would force unphysical upper tropospheric winds. This is even more self-evident if we allow radiative equilibrium to have a meridional gradient at the equator, for then the radiative equilibrium winds are infinite!

Let's assume the equilibrium height field is parabolic close to the equator:

$$H_{eq} \approx H_o(1 - \alpha\theta^2)$$

Then,

$$fu_{eq} = -\frac{g^*}{a}\frac{\partial H_{eq}}{\partial \theta} \approx \frac{2\alpha g^* H_o}{a}\theta$$

Since $f = 2\Omega\theta$ near the equator, we have

$$\frac{u_{eq}}{\Omega a} \approx \alpha R \Delta_V$$

where

$$R = \frac{gH_o}{\Omega^2 a^2} \tag{41}$$

We have set $g^* = g\Delta_V$, where Δ_V is the fractional change in potential temperature in the vertical.

This expression gives the misleading impression that these radiative equilibrium winds depend on the static stability of the atmosphere. To see why this is not the case, assume, for example, that the radiative-equilibrium potential temperature field is of the form

$$\frac{\Theta_{eq}}{\Theta_0} = 1 - \Delta_H \sin^2(\theta) + \Delta_V \frac{z}{H} \approx 1 - \Delta_H \theta^2 + \Delta_V \frac{z}{H}$$

Think of the bottom of the upper layer as a surface of constant Θ. Then the thickness of the upper layer of our two-layer model, in radiative equilibrium, would be

$$\frac{H_{eq} - H_0}{H_0} \approx 1 - \frac{\Delta_H}{\Delta_V}\theta^2$$

Therefore, $\alpha \approx \Delta_H/\Delta_V$, and we find

$$\frac{u_{eq}}{\Omega a} \approx R\Delta_H$$

It is the horizontal temperature gradients in radiative equilibrium, not the static stability, that enters this expression. For the Earth, a relevant value might be $\Delta_H \approx 0.3$ (implying a pole-to-equator temperature difference in radiative equilibrium of $100K$) and $R \approx 0.7$ so that $R\Delta_H \approx 0.2$

Whatever the value of κ, a circulation must exist to cause the flow to depart from this radiative equilibrium throughout the region in which $u_{eq} > u_M$. Continuing to make the small angle approximation,

$$u_M \approx \Omega a \theta^2.$$

a circulation must exist at least from the equator up to the latitude

$$\theta_H \approx \sqrt{R\Delta_H}$$

In the limit $\kappa \to 0$, since $v \neq 0$ we must have zero absolute vorticity, or $u = u_M$ in this region.

To be more precise, assume that as $\kappa \to 0$ there is a latitude θ_H equatorward of which $v \neq 0$ and $u = u_M$, and poleward of which $v = 0$ and $u = u_{eq}$. We refer to the region $\theta < \theta_H$ as the "tropics". Within the tropics, we can compute the thickness gradient consistent with angular momentum conservation:

$$\frac{g^\star}{a}\frac{\partial H}{\partial \theta} \approx -2\Omega^2 a \theta^3$$

Integrating from the equator, we obtain the thickness distribution

$$\frac{H(\theta) - H(0)}{H_0} = -\frac{1}{2R\Delta_H}\theta^4 \tag{42}$$

This height field is extremely flat in the deep tropics, so as to keep the upper tropospheric winds from increasing faster than allowed by angular momentum conservation, or, equivalently, by the criterion of inertial stability. (A balanced flow with angular momentum increasing as one moves towards the axis of rotation – i.e., polewards on the sphere – is inertially unstable.) Keep in mind that this is an *upper* limit on the height gradients in the tropics; significant eddy stresses would reduce this gradient even further.

We need two constraints to allow us to solve for θ_H and $H(0)$ simultaneously. The constraints are continuity of H at $\theta = \theta_H$ and mass conservation, which requires that

$$\int_0^{\theta_H} (H(\theta) - H_{eq})d\theta = 0.$$

The "equal-area" construction in Figure 9 allows one to picture how these two constraints are satisfied simultaneously. One sees from the figure that the height gradient at the boundary of the tropics will be discontinuous, and that the zonal wind overshoots the radiative equilibrium value. In the small angle approximation, the result of this construction is $\theta_H = \sqrt{\frac{5}{3}R\Delta_H}$. For $\Delta_H R \approx 0.2$ one gets $\theta_H \approx 0.6 \approx 30°$. The small angle approximation is evidently only marginally adequate for these parameters, but calculations on the sphere do not change things much. See [4].

Figure 10 shows numerical solutions to this set of equations on the sphere, in which $H_{eq} = H_{eq}(0) - \alpha \sin^2(\theta)$ and with $S = -\kappa u$. This form for S is not meant to capture any of the physics of the actual drag in the upper troposphere; it is just a simple way of illustrating the essential distinction between the tropical and extratropical balances in the model.

This simple model of the Hadley cell is easy to criticize as unrealistic. For one thing, if an angular momentum conserving flow extended to the subtropics, winds in excess of $130m/s$ would be generated. (Most of these strong winds are generated after one moves past $\approx 20°$) The observed zonally averaged subtropical jet is less than $40m/s$. Obviously, eddy stresses are substantial once the flow approaches the subtropics. There are a number of open questions related to modeling the eddy stresses in a more physically relevant way and then studying how these stresses interact with the Hadley cell. One starting point might be linear "Rossby wave chromatography", in which a a phase speed spectrum of the input of pseudomentum in midlatitudes is prescribed, and then each wave deposits its drag at its critical latitude. However, this cannot be appropriate when the model is close to its

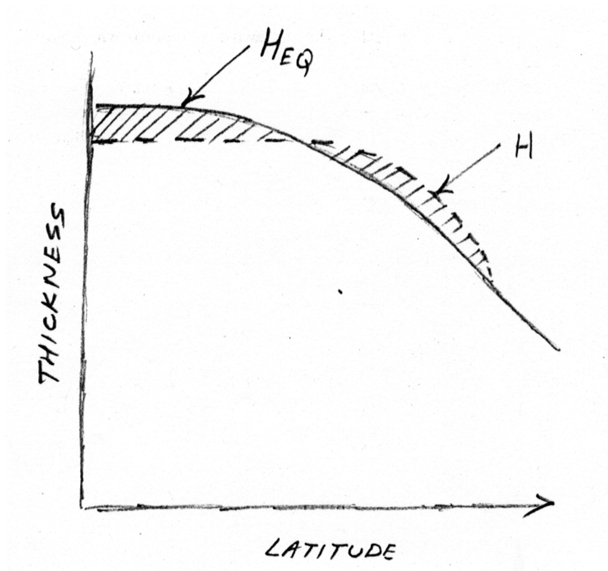

Figure 9: Simultaneously satisfying continuity of temperature at the subtropical boundary of the Hadley cell and mass conservation

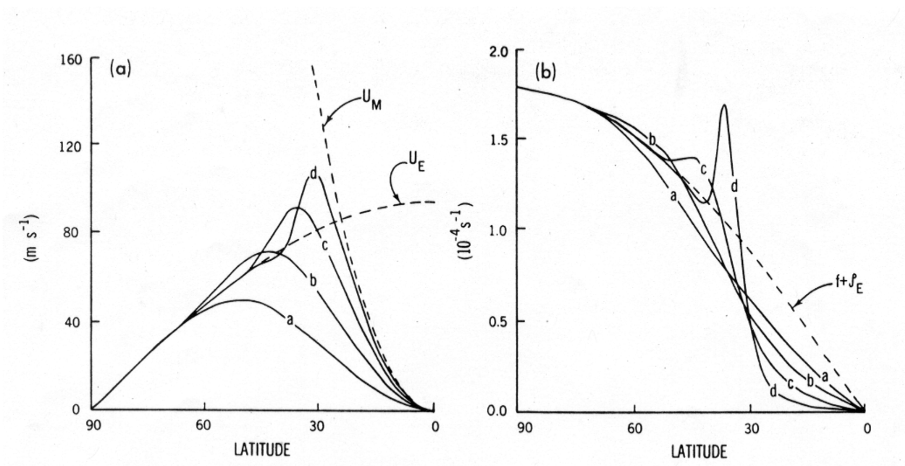

Figure 10: Solutions to the idealized Hadley cell model as the eddy stresses in the zonal momentum equation are reduced in strength. The zonal winds and the absolute vorticity are shown.

nearly inviscid limit, in which there is little vorticity gradient in the tropics, so that Rossby waves are not able to propagate past the subtropical jet to get to their critical latitude. This suggests an interesting "bootstrap" feedback in which the eddy stresses themselves create the vorticity gradient that supports their propagation into the tropics.

An alternative, and more traditional, view of how the width of the Hadley cell is determined is that angular-momentum conservation continues polewards until the resulting vertical shears become baroclinically unstable. Suppose we use the two-layer model's criterion for instability for this purpose:

$$\beta - \frac{f}{H_0}\frac{\partial H}{\partial y} = 0 \Rightarrow \frac{\partial H}{\partial y} = \frac{\beta H_0}{f}.$$

From thermal wind balance,

$$g^* \frac{\partial H}{\partial y} = f(u_1 - u_2)$$

With $u_2 \approx 0$, then, in the upper troposphere

$$u_1 \approx \beta \lambda^2 = \beta \frac{g^* H_0}{f^2}$$

or

$$\frac{u}{\Omega a} \approx R\Delta_V \frac{\cos(\theta)}{\sin^2(\theta)}.$$

On the other hand, for the angular momentum conserving wind we have:

$$\frac{u_M}{\Omega a} \approx \frac{\sin^2(\theta)}{\cos(\theta)}.$$

Invoking the small angle approximation for algebraic simplicity once again. we solve the equation $u_1 = u_M$ to obtain the latitude , θ_{BC}, at which the two wind profiles coincide, in the small angle approximation,

$$\theta_{BC} \approx (R\Delta_V)^{\frac{1}{4}}.$$

The relationship between the axisymmetric theory and this instability argument is

$$\theta_{BC}^4 = \frac{\Delta_V}{\Delta_H}\theta_H^2$$

One could combine the two theories by saying that the Hadley cell stops at the smaller of θ_{BC} and θ_H. If $\theta_{BC} < \theta_H$, the flow would become unstable before reaching the axisymmetric limit. If $\theta_H > \theta_{BC}$, on the other hand, the Hadley cell would terminate before becoming unstable. Is the latter situation physically realizable? In an atmosphere heated from below, what would maintain the static stability in the region between the edge of the Hadley cell and the start of the baroclinically unstable zone? Would the isentropic slope, Δ_V/Δ_H, simply adjust so as to remove this gap?

One could use some other instability criterion in this argument, rather than rely on the existence of a critical shear in the two layer model, the physical relevance of which is obscure at best. For example, an estimate of growth rate ω_I from the Eady or Charney models is $fRi^{-1/2} \approx fu/NH$, or

$$\omega_I \approx \frac{fu}{\sqrt{gH\Delta_V}} \approx 2\frac{u}{a\sqrt{R\Delta_V}}\theta$$

If we suppose that this growth rate must be larger than some frictional damping κ, due to boundary layer drag, for example, then the angular momentum conserving wind would reach this value at

$$\theta_\kappa^3 \approx \frac{\kappa}{2\Omega}\sqrt{R\Delta_V}.$$

Even better would be to use the kind of diffusive closure theory described below, and ask where the flux divergence due to the resulting heat flux becomes comparable to the radiative forcing. The details are different, but the qualitative behavior is the same.

7 A Moist Hadley Cell

People are often amazed that one would try to talk about the Hadley cell without talking about moist convection, as in the previous section. How does moist convection change this picture? Although less often remarked upon, it should also seem strange that one would talk about the Hadley cell without taking into account the oceanic circulation in the tropics, which transports more energy polewards than does the atmospheric circulation. These two issues are related.

To approach this subject we look first at the energy balance at the top of the atmosphere. For our purposes, we can think of the stratosphere as in radiative equilibrium, so we are really thinking about the energy fluxes at the tropopause. Let's also start by ignoring any energy transport by the ocean so that, in a steady state, there are no fluxes through the surface, and the energy fluxes at the tropopause are the only sources of energy for the troposphere.

The incoming solar radiation $S(\theta)$ has to balance the outgoing infrared radiation $I(T)$. We assume that S is only a function of latitude and I only a function of T. These assumption obviously ignore all sorts of complexity associated with the effects of clouds and water vapor on these radiative fluxes. Integrated over the Hadley cell, and ignoring any energy fluxes out of the Hadley cell into midlatitudes, we require

$$\int_0^{\theta_H} (S(\theta) - I(T))\cos(\theta)d\theta = 0 \tag{43}$$

Assuming once again that the zonal mean flow is in thermal wind balance,

$$f\frac{\partial u}{\partial z} = -\frac{g}{T}\frac{\partial T}{\partial y}. \tag{44}$$

and integrating this equation from the surface up to the tropopause, assuming that $u(0) \approx 0$ we get

$$u(H) = -\frac{gH}{fT_0}\frac{\partial [T]}{\partial y}$$

where $[T]$ is the vertically averaged temperature. We are also assuming that the height of the tropopause is independent of latitude. We assume that we can think of I as a function of this vertically averaged temperature, $[T]$, i.e., that variations of the temperature field are always sufficiently coherent in the vertical within the tropics. We drop the brackets in the notation for this vertically averaged temperature from here on.

The Hadley cell theory can now proceed exactly as before. Angular momentum conservation implies a temperature profile $T(\theta) - T(0)$ and a profile of outgoing radiation. Continuity of T at the boundary of the cell and conservation of energy result in the same equal area construction for $T(0)$ and the width of the cell. The solution includes the energy transport by the circulation. It does not tell us its strength, that is, the mass transport. To obtain the mass transport, we need additional information. It is the ratio of energy to mass transport that is affected by the presence of water vapor.

We need to digress a bit to talk about energy conservation in a compressible fluid. Start with the conservation of kinetic and potential energy,

$$\rho \frac{D}{Dt} \frac{1}{2} |\mathbf{v}|^2 = -\rho \mathbf{v} \cdot \mathbf{v} \Phi - \mathbf{v} \cdot \nabla p$$

or

$$\rho \frac{D}{Dt} \left(\frac{1}{2} |\mathbf{v}|^2 + \Phi \right) = -\nabla \cdot (p \mathbf{v}) + p \nabla \cdot \mathbf{v} \tag{45}$$

Similarity for internal energy,

$$\rho \frac{De}{Dt} = Q - p \nabla \cdot \mathbf{v} \tag{46}$$

Here $\Phi = gz$ is the geopotential and Q is the heating rate. Adding the kinetic, potential, and internal energies, we obtain the expression for energy conservation:

$$\rho \frac{D}{Dt} \left(\frac{1}{2} |\mathbf{v}|^2 + \Phi + e \right) = -\nabla \cdot (p \mathbf{v}) + Q. \tag{47}$$

Using the continuity equation, we may write this in flux form:

$$\frac{\partial}{\partial t} \rho \left(\frac{1}{2} |\mathbf{v}|^2 + \phi + e \right) = -\nabla \cdot \left(\rho \mathbf{v} \left(\frac{1}{2} |\mathbf{v}|^2 + \phi + h \right) \right) + Q, \tag{48}$$

where $h = e + p/\rho$ is the enthalpy. For our ideal gas, $e = c_v T$ and $h = c_p T$. Note that e appears on the LHS and h on the RHS. We also define the dry static energy s

$$s = \phi + h = c_p T + gz.$$

The kinetic energy is a very small part of the total energy in the atmosphere (recall that $c_p \approx 10^3 m^2/(s^2 K)$). so we can compute the energy flux from the flux of dry static energy. In a steady state

$$\nabla \cdot (\rho \mathbf{v} s) = Q \tag{49}$$

If the dry stability of the atmosphere is positive, then the dry static energy increases with height:

$$\frac{\partial s}{\partial z} = c_p \frac{\partial T}{\partial z} + g = c_p \frac{T}{\Theta} \frac{\partial \Theta}{\partial z}.$$

The north south flow in the Hadley cell is concentrated near the surface and near the tropopause. We refer to the dry static energy characterizing the surface air as $s(0)$ and that of the air near the troposphere as $s(H)$. Vertically integrating, we have

$$\nabla \cdot (V\Delta_D) = \int Q \tag{50}$$

where V is the mass flux in either the poleward or equatorward branch of the cell and $\Delta_D \equiv s(H) - S(0)$ is referred to as the *gross dry stability* of the overturning circulation. It is simply the ratio of the dry static energy transport to the mass transport. In the tropics, $\Delta_D \approx (45^\circ K)c_p$.

The heating rate Q is comprised of the vertical divergence of radiative and sensible heat fluxes, and latent heat release. Therefore, the vertical integral of Q can be written as the sum of three terms: 1) the difference between the radiative fluxes at the top and bottom of the atmosphere; 2) the sensible heat flux at the surface; and 3) the total latent heating.

$$\nabla \cdot (V\Delta_D) = R_T + R_B + S + LP$$

where L the latent heat of condensation, P the precipitation, and fluxes are positive when directed into the atmosphere. However, the total latent heating is also related to the convergence of vapor in the atmosphere and the evaporation E. In a steady state, conservation of water vapor implies that

$$\nabla \cdot (\rho \mathbf{v} q) = E - P$$

where q is the mixing ratio for water vapor. We can define a *moist static energy*, $m \equiv s + Lq$, and a gross moist stability, $\Delta_m = m(H) - m(0)$, so that

$$\nabla \cdot (V\Delta_M) = R_T + R_B + S + LE$$

But $R_B + S + LE = 0$ if there is no flux through the surface, so

$$\nabla \cdot (V\Delta_M) = R_T = S(\theta) - I(T) \tag{51}$$

The energy transport by the Hadley cell is determined by the requirement that it create a temperature distribution that is flat enough to be consistent with upper level winds that are, at most, angular momentum conserving. The mass transport in the cell can then be determined if one has a theory for the gross moist stability (or, conversely, the gross moist stability can be determined if one has a theory for the mass transport!)

The gross moist stability in the tropics is much smaller than the dry stability because nearly all of the water vapor resides at low levels. Note that the energy transport must be smoothly varying in latitude in this theory, since T varies smoothly. The only way to create sharp structure in the mass flux, as at an ITCZ (a region of very concentrated upward motion) is through structure in Δ_M, and, in particular, by having $\Delta_M \approx 0$ in some region.

In reality, there is a lot of energy flowing from atmosphere to ocean in the tropics, and this energy is removed by transport polewards in the oceans. Indeed there is something we can call the *oceanic Hadley cell* in the tropics. Much of the circulation in the tropical oceans is wind-driven. The easterlies in low latitudes cause a poleward Ekman drift in the oceanic mixed layer, which results in equatorial upwelling. This circulation is closed by subduction of water in the subtropics which returns by complicated routes to supply the equatorial upwelling. Ekman transport is defined as the mass transport required in the boundary layer so that the resulting Coriolis force balances the surface stress. Since the surface stress on the two media is by definition equal and opposite, *Ekman mass transports are also equal and opposite in atmosphere and ocean.* If we can ignore stresses on land, and if the Ekman transport is the main thing going on in the surface layers of both atmosphere and ocean, then we can expect the mass transport in the oceanic Hadley cell to be comparable to that in the atmospheric Hadley cell (the latter is about 60 Sverdrups in the annual mean). This seems to be the case in the tropics.

We can define an oceanic gross stability, $\Delta_O = c_O \delta T$, where c_O is the heat capacity of water and δT the characteristic temperature difference between the poleward and equatorward moving water. Since the oceanic interior is adiabatic, we can assume that δT is also the characteristic *horizontal* temperature difference across the Hadley cell. The atmospheric gross stability is also closely related to this surface temperature gradient; since the atmosphere is nearly saturated at low levels, one can obtain the moisture gradient, which is the key term in the gradient of moist stability, from the temperature gradient.

The ratio of the total energy transport in the system to the mass transport in either ocean or atmosphere is determined by the sum of the gross stabilities of the two media, $\Delta_{tot} \equiv \Delta_O + \Delta_M$. In the deep tropics one finds that most of the stability of the system resides in the ocean, so the ocean provides most of the energy transport. It is the energy transport in the atmosphere-ocean system as a whole that is determined by our equal area argument. The contributions of atmosphere and ocean to Δ_{tot} determine how this energy transport is partitioned between the two fluids.

8 Superrotation

Suppose that there is a local maximum of angular momentum M somewhere in the atmosphere away from the surface. Surround this point by a surface on which the mean M is a constant. By conservation of mass, the mean transport of M into or out of this volume is zero. If there is any downgradient molecular (or small scale turbulent) vertical transport at all, this will remove M from this region; to maintain a steady state we need countergradient eddy fluxes. This does not typically occur in the troposphere, although it does in the stratosphere in the westerly phase of the QBO. where eastward propagating Kelvin and/or gravity waves provide a countergradient vertical flux. In the troposphere, as we have seen, the dominant large scale flux is horizontal and down the angular momentum gradient, from the tropics towards midlatitudes. It is an interesting fact that on Jupiter, Saturn, and the Sun the atmosphere near the equator rotates faster than the planetary interior, as measured by the rotation of the magnetic fields or by helioseismology, so maybe countergradient fluxes are the rule, rather than the exception. Could the upper troposphere superrotate?

Remarkably, some idealized models of the atmosphere do produce an abrupt transition to a superrotating state as one varies certain parameters. Figure 11 shows perhaps the best example [21].

This result is from a dry two-*level* model on the sphere, forced by linear relaxation to a specified radiative equilibrium temperature distribution. One starts with zonally symmetric forcing and then perturbs the system by modifying radiative equilibrium in the tropics so as

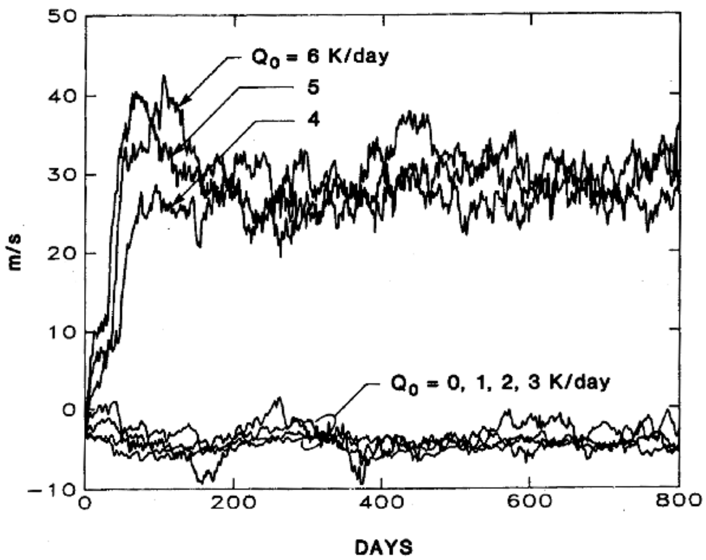

Figure 11: Zonal mean wind at the equator in the upper layer of a two layer model, as a function of time, for different values of the strength of the asymmetric component of tropical heating, from [21]

to be more and more zonally asymmetric. At a certain strength of this asymmetric forcing, the model atmosphere makes a transition from a perfectly normal circulation, with near zero equatorial flow and strong subtropical jet streams, to one with very strong equatorial superrotation, with an upper tropospheric wind field that is closer to solid body rotation.

The starting point for understanding this behavior is that the asymmetric tropical forcing excites a stationary disturbance that will propagate out of the tropics in the upper troposphere, to the extent that conditions are favorable for the existence of Rossby waves. Waves emanating from the tropics are presumably dissipated in various complex ways once they enter midlatitudes. So there is a source of Rossby wave pseudomomentum in the tropics and a sink in midlatitudes. The result, as we have seen, will be a convergence of eddy angular momentum flux and acceleration of the tropical flow. But why should there is such an abrupt transition from the normal to the superrotating state?

There is a fascinating feedback between the strength of the westerlies in the tropical upper troposphere and the propagation and dissipation of the waves excited in *midlatitudes* by baroclinic instability. We have seen that disturbances excited in midlatitudes will propagate into the tropics, the extent of this propagation depending on how much wave breaking occurs along the way. Wave breaking, in turn, depend in part on the ratio of the wave's zonal wind perturbations as compared to the mean wind in the reference frame moving with the wave's phase speed. The dominant phase speed c of these waves is of the order of 5-12 m/s. In the normal state, $\overline{u}$ decreases to 0 as one approaches the equator, so breaking is inevitable before one reaches into the deep tropics. But if one can accelerate the tropical upper level winds, breaking will be less likely to occur. This is a positive feedback on the tropical winds because one is thereby losing the deceleration associated with the breaking. Eventually, once the equatorial winds are strong enough, the tropics evidently becomes rather transparent to this Rossby wave activity, with little breaking occurring there. This loss of wave breaking and deceleration is strongest as the mean flow rises above the dominant phase speeds in the wave field, and this seems to be the primary cause of the bifurcation in this model.

The problem is actually quite a bit more complicated than this, however. There are at least two other potential feedback mechanisms. First, the tropically forced wave itself is sensitive to the existence of the upper level westerlies. For this forced stationary wave to propagate out of the tropics, westerlies are required, at least in a linear picture. If these westerlies are too weak, Rossby waves will have difficulty emerging from the tropics. As the westerlies increase, the tropically forced wave will emerge more easily, providing greater acceleration near the equator, which is required to get the ball rolling. The quantitative importance of this feedback is unclear. It does not appear to be required, for one can simply omit the zonally asymmetric heating and accelerate the upper troposphere with a prescribed zonally symmetric force. Varying the strength of this force, one still finds a similar bifurcation [17]. The feedbacks related to eddies generated in midlatitudes and in the tropics are illustrated schematically in Figure 12.

Secondly, the Hadley cell itself can feed back positively on the upper level equatorial westerlies. Given a torque at the equator, what provides the compensating deceleration in a steady state? Presumably the transport of momentum from the lower to the upper troposphere by the Hadley cell is the dominant factor:

$$w\frac{\partial \overline{u}}{\partial z}(\theta = 0) \approx torque$$

However, as the westerlies increase, one expects the Hadley cell to decrease in strength. Upper level equatorial westerlies imply a warmer troposphere by thermal wind balance (taking into account the change in sign of f at the equator). The adiabatic cooling at the equator must be reduced to be consistent with these warmer temperatures. (This is not the whole story, as one also expects the Hadley cell to shrink, since angular momentum conservation, starting with non-zero wind at the equator, will generate stronger temperature gradients). Again, it is not clear if this mechanism plays an important role in the kind of calculation pictured above. But in isolation this mechanism is capable of creating a bifurcation in axisymmetric models in which there are no large scale eddy stresses, as one varies the strength of a zonally symmetric equatorial torque [19].

In a moist atmosphere, one might expect superrotation to be harder to achieve if only because the Hadley cell is stronger, so one might imagine that the vertical advection opposing the westerly torques would be stronger. But this is not self-evident. The problem is that there really is no "large scale" upward motion in a moist Hadley cell. All of the upward motion is presumed to occur in small-scale convective towers. Between these towers, disregarding vertical motion associated with "synoptic" eddies of various kinds, we expect subsidence driven by radiative cooling. The implication is that there is no strong reason to believe that $w\partial\overline{u}/\partial z$, where w is an average over some scale (say a few hundred kilometers), is a good model of momentum transport. The net effect of the convective motions and the environmental subsidence could be a flux that is smaller than the "large-scale" advective value; note that the environment alone transports momentum in the opposite sense. It is not clear that moist models should be less likely to superrotate than dry models.

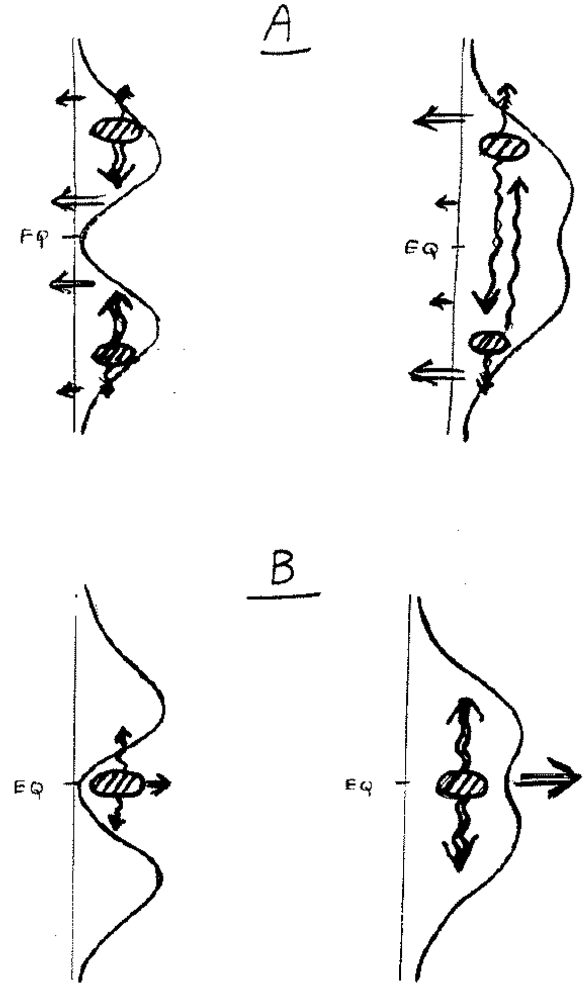

Figure 12: Schematic of the two eddy feedbacks

In any case, there is as yet no evidence that comprehensive climate models undergo an abrupt bifurcation of this sort. Why this is so is unclear. It does seem that two-layer models are more susceptible to this bifurcation, so it may simply be a matter of vertical resolution, but there are no compelling arguments for why vertical resolution should matter that much. We often think in terms of two-level or two-layer models to help us get our minds around the complexity of the general circulation, hoping that we have not thrown out the baby with the bath water. The critical shear for baroclinic instability, which typically decreases to zero as the vertical resolution increases, is but one example of how one can be misled by thinking in terms of too idealized a model; the superrotation bifurcation may provide another example.

9 Turbulent Diffusion in Midlatitudes

Turning away from the tropics, we can no longer avoid the basic question of how the lower tropospheric eddy heat transport through midlatitudes is controlled. It is this heat transport that in turn controls the equator-to-pole temperature gradient on the Earth. As we have seen, given this heat transport we can also compute the equatorward mass flux in the "interrupted" surface layers, or the returning mass flux aloft (to within corrections due to the Ekman drift associated with the surface drag generated by eddy momentum fluxes, and assuming that we are given the tropospheric static stability). Equivalently, the heat flux is also the form drag that transfers angular momentum from the upper to the lower troposphere, angular momentum that is constantly replenished by transport from lower latitudes.

One can make some progress by thinking of these fluxes as diffusive

$$\overline{v'T'} \sim -\mathcal{D}\frac{\partial T}{\partial y}.$$

At the simplest level, this is just dividing the flux by the gradient, which leaves the convenient units of $(length)^2(time) = (velocity)(length)$. The value of diffusivity one finds in midlatitudes is $1-2\times 10^6 m^2/s$. The temperature field is not strongly distorted by the eddies in the lower troposphere, being rather strongly forced, so it seems plausible to think of the eddies as riding on a well-defined environmental gradient (as contrasted with the situation in the upper troposphere, where the eddies are more efficient at distorting and mixing the potential vorticity and where, therefore, one is not inclined to divide the flux by the gradient.) Perhaps more importantly, one can try to argue that finite amplitude baroclinic eddy production is *fundamentally local* and therefore effectively diffusive.

Think of a standard Benard convection problem: fluid confined between horizontal plates with fixed temperatures, and of horizontal extent much larger than the vertical extent. Diffusive theories cannot have any great value in this context because the scale of eddies transporting heat is determined by the vertical size of the domain, so there is no scale separation between eddies and mean flow. A theory for heat transport in such a geometry must be fundamentally global.

But now consider an alternative geometry, pictured in Figure 13, in which the aspect ratio of the the apparatus has been altered so that the horizontal scale is much smaller than the vertical scale. One might now expect that, in a turbulent flow, in which eddy statistics in different directions tend to be isotropized by nonlinear advection, that the small horizontal scale would set the dominant eddy scale. The result would be a separation of scales between this mixing length and the larger vertical scale imposed by the geometry. The heat flux through this kind of system would be diffusive in character. In particular, the resulting temperature profile, rather than being more or less homogenized in the interior might be more or less linear, as in the simplest solutions of the diffusion equation. The essential distinction is that, in this second case, *the scale of the eddies is not determined by the scale of the mean inhomogeneity in the direction of the transport, as in most familiar turbulent flows, but rather by mean flow inhomogeneity in a direction perpendicular to the transport.*

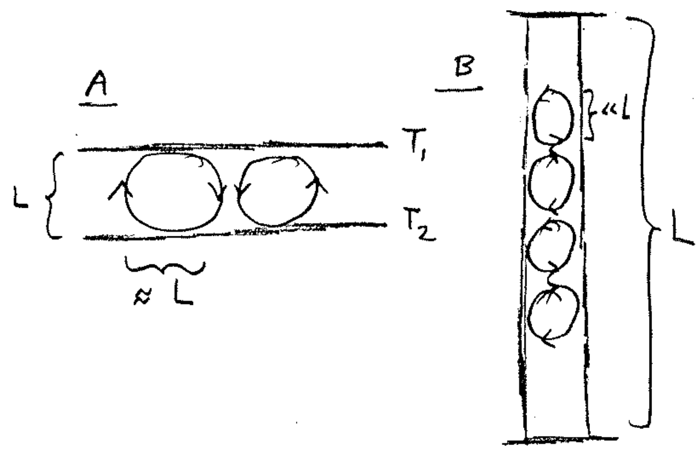

Figure 13: How turbulent convection in a domain with height/width $\ll 1$ is different from convection in a domain with height/width $\gg 1$; the latter is more "diffusive".

Why is baroclinic eddy production "diffusive" in this sense? In linear barotropic shear flows, the zonal scale of the instability is simply determined by the meridional scale of the mean shear variations. At finite amplitude, if sufficiently nonlinear, we expect the mixing length across the shear will also be of this scale. So no scale separation is possible here. In linear baroclinic problems, however, it is the *vertical* scale of the mean shears, multiplied by the Prandtl ratio N/f, that sets the zonal scale of the eddies. We speak of the zonal scale being the "radius of deformation" NH/f where we can typically think of H as proportional to the depth of the troposphere. Once again, we can expect nonlinear isotropization in the horizontal to turn this scale into a meridional mixing length. If NH/f is small compared to the scale of horizontal inhomogeneity, then we do have the potential for scale separation. Even if the separation is not pronounced, we can hope, as in WKB theory, that the local theory is qualitatively useful even when the ratio of intrinsic to environmental scales is order unity.

To make this claim more precise, one can define a (numerical) apparatus to measure the diffusivity of turbulent baroclinic eddies as a function of the environmental gradients, just as one sets up a laboratory apparatus to measure the molecular diffusivity of a gas or the resistivity of a metal. Working in QG theory, the procedure is to specify the environmental potential vorticity gradients at each height, and then assume that the departures from these environmental values are doubly periodic in the horizontal:

$$\frac{\partial q}{\partial t} = -J(\psi, q) - U\frac{\partial q}{\partial x} - \frac{\partial \psi}{\partial x}\frac{\partial Q}{\partial y} - D(q)$$

Here q and ψ are doubly periodic, U and $\partial Q/\partial y$ are the horizontally uniform environment, and D is the dissipation. This equation for the eddies is horizontally homogeneous (but anisotropic). If the environment is unstable, eddies will grow and generate homogeneous geostrophic turbulence that will produce a downgradient flux of potential vorticity in the interior and a downgradient flux of heat along the lower surface. If one tried to set up such a problem for standard Benard convection, the model would "short circuit"; by maintaining the environment (analogous to maintaining a potential drop across a piece of metal) one is providing an infinite supply of energy. Only if a mixing length does develop will a meaningful answer emerge from such a geometry (just as the mean free path of electrons must be smaller than the apparatus if we are to measure the resistance of the metal rather than create an accelerator!) When we perform such a calculation, we find that baroclinic eddies do generate a mixing length that is independent of horizontal mean flow inhomogeneities.

Given the diffusivity determined by such a "theory", which will be dependent on the environmental shears and the form of the dissipation, we can then use this diffusivity in theories for the large-scale flow. Figure 14 provides some evidence that this kind of local diffusive theory actually works in some idealized cases (see [13]). A two-layer model, in which the interface slope is relaxed to some zonally symmetric radiative equilibrium, has been integrated to statistically steady state, for different values of the width of the unstable region. The upper level potential vorticity fluxes are shown as a function of latitude. Also shown are the fluxes obtained by using a diffusive flux closure obtained from a horizontally homogeneous diffusivity-measuring apparatus. (In this model, jet formation is not very strong, due in part to strong surface friction, so momentum fluxes are weak and potential vorticity fluxes in the upper layer are dominated by thickness fluxes.) Even when the baroclinic zone is only a few radii of deformation wide, the "theory" looks useful. But even if only qualitative, it provides a heuristic way of thinking of the problem as divided into two parts: a local theory for diffusivities; and the global response to the resulting eddy fluxes.

Let's focus on possible theories for the diffusivity. We need a velocity scale V and a length scale L. A good starting point [20] is that $L \sim NH/f$ and $V \sim U \sim H\partial u/\partial z$, so that

$$\mathcal{D} \sim \frac{NH^2}{f}\frac{\partial u}{\partial z}$$

How can one justify the assumption that $V \sim U$, that the eddy kinetic energy is comparable to the zonal mean kinetic energy?

Baroclinic instability extracts available potential energy (APE) from the mean state. It is plausible that an eddy of scale L can only extract the APE contained with a domain of scale L. In QG theory, defining the buoyancy $b = g\Theta/\Theta_0$ we have for this energy

$$V^2 \sim APE \approx \frac{(\Delta b)^2}{N^2} \approx \left(\frac{L}{N}\frac{\partial b}{\partial y}\right)^2 \approx \left(\frac{fL}{N}\frac{\partial u}{\partial z}\right)^2.$$

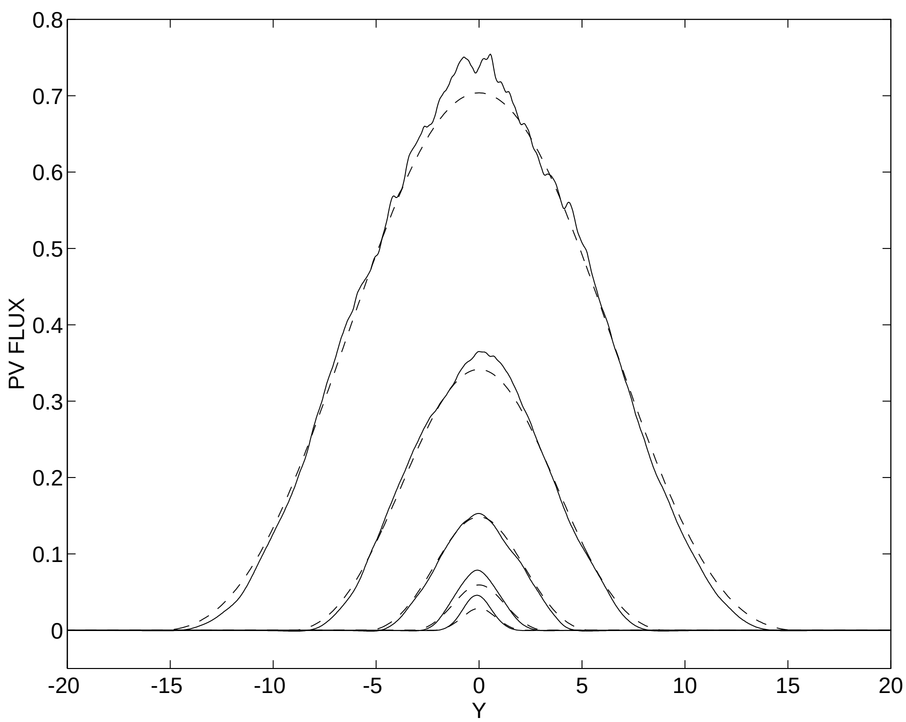

Figure 14: Solid: potential vorticity fluxes in the upper layer of a two-layer model. Dashed: diffusive closure theory with diffusivity determined numerically by integrating homogeneous model, from [13]

If $L = \lambda \equiv NH/f$, this yields $V \sim U$ as desired. More generally,

$$V \sim \frac{L}{\lambda} U.$$

(For an alternative derivation, see [5].) Whatever L, this expression for V leads to the time scale

$$T \sim \frac{L}{V} \sim \frac{\lambda}{U} \sim \frac{N}{f \partial U / \partial z} \sim \frac{Ri^{1/2}}{f}, \tag{52}$$

which happens also to be the e-folding time for baroclinic instability in the Eady and Charney models. Note that T is independent of L. By these arguments, the diffusivity is

$$\mathcal{D} \sim \frac{L^2}{\lambda} U \tag{53}$$

If $L \sim \lambda$, then the diffusivity is proportional to the horizontal temperature gradient, so the heat flux is proportional to the square of the temperature gradient. In this case the heat flux is also directly proportional to N, which is counterintuitive.

A familiar and fundamental property of two-dimensional flows is that energy cascades to larger horizontal scales, rather than smaller scales, as in 3D turbulence. In QG flows, the generic behavior is that the flow evolves to both larger vertical and horizontal scales, but that once the vertical scale has been maximized by the flow becoming more and more barotropic, the cascade to larger horizontal scales continues. If there is room and if nothing stops the cascade, then we expect $L > \lambda$.

There are a number of mechanisms that can stop the inverse energy cascade. One is surface friction. Suppose that surface friction results in a spin-down time of κ^{-1}. One might suspect that the time scale of the eddies would increase as L increased, until this time scale reached κ^{-1}, at which point the cascade would stop. But this does not occur; the time scale is independent of L! This is a consequence of an important feedback: as the eddy length scale increases, the mixing length, the APE extracted by a typical eddy, and the eddy velocity scale all increase. So it does not appear that friction can stop this cascade. However, a better model of frictional damping in the atmosphere follows from assuming that the surface stress is quadratic in the geostrophic velocity, $\propto C_D V^2$, where C_D is a non-dimensional drag coefficient. The value of C_D depends on the roughness of the surface; typical values range from 10^{-3} over the ocean to 10^{-2} over very rough land surfaces. Since this stress must be distributed over the vertical depth H to obtain an acceleration, the effective frictional time scale is

$$\kappa^{-1} \sim \frac{H}{C_D V}$$

Setting $\kappa T \sim 1$ yields the very simple result $L \sim H/C_D$! This scale is larger than the radius of deformation as long as C_D is smaller than f/N. The resulting diffusivity would still be proportional to the temperature gradient, but independent of the static stability and proportional to C_D^{-2}

An alternative to stopping the inverse cascade with friction is to stop it, or slow it down, with the β-effect [16]. Longer Rossby waves have higher frequencies, so as the inverse cascade proceeds one has the potential to excite faster and faster waves. Once these frequencies become comparable to the inverse of the eddy time scale, one expects the nonlinear cascade to slow down and be replaced by more selective anisotropic interactions. Equating the Rossby wave frequency βL with T^{-1} results in $L \sim 1/(\beta T)$ and a diffusivity

$$\mathcal{D} \sim \beta^{-2} T^{-3}.$$

This diffusivity has a very strong cubic dependence on horizontal temperature gradient, resulting in a heat flux proportional to the fourth power (!) of the gradient. Experiments with horizontally homogeneous two-layer models on a β-plane, with linear friction, show a dependence of the flux on the temperature gradient which, if anything, is even stronger than this [5]. The expression for L can be rewritten as

$$\frac{L}{\lambda} L \sim \frac{U}{\beta \lambda^2} \sim \frac{a}{H} I_\Theta$$

where I_Θ is the isentropic slope and we have evaluated f and β in midlatitudes. If the isentropic slope is such that it takes one from the top to the bottom of the troposphere as one moves from the pole to the equator (as observed), then we evidently should not expect to cascade to larger scales than the radius of deformation before we are stopped by β [18].

These arguments are too crude to determine whether quadratic friction or the β-effect are dominant in stopping the cascade in the atmosphere (both are probably important). If nothing stops the cascade before its generates mixing lengths comparable to the radius of the Earth, we can set $L \sim a$, which yields the diffusivity proposed in [3].

10 The Extratropical Static Stability

In addition to the amplitude of the eddy fluxes, we also need to understand the vertical structure of the potential vorticity fluxes in order to discuss the maintenance of the extratropical static stability. How high does the eddy potential vorticity mixing extend? Much of this mixing admittedly occurs near the tropopause, partly because the eddies have been relatively successful at homogenizing the mid-troposphere. (These waves have steering levels in the mid-troposphere, so this is where the largest meridional particle displacements are to be expected). But what determines the height of the tropopause?

To understand what some of the issues are, consider a Boussinesq atmosphere with no upper boundary and with uniform $N(z)$ and uniform vertical shear $\overline{u} = \Lambda z$ on a β-plane. The potential vorticity gradient in the interior is simply β, with the surface temperature gradient creating baroclinic instability. What is the horizontal scale of the most unstable wave in this system? One is tempting to say Nh/f, but what is the appropriate vertical scale h? The answer is

$$h \approx \frac{f^2 \Lambda}{\beta N^2}; \;\; L \approx \frac{f \Lambda}{\beta N} \tag{54}$$

One can obtain this answer in several ways: by examining the linear system and nondimensionalizing (there is only one way of doing this); or by using the counter-propagating Rossby wave picture of baroclinic instability (which we haven't discussed here, see [9]) and finding that height h at which a Rossby wave with scale Nh/f would propagate westward with the right phase speed so as to match the eastward propagation of the "Eady edge wave" with respect to the flow at the ground.

Notice that the vertical penetration of the wave is proportional to the vertical shear, or, more precisely, to the isentropic slope: $h \approx fI_\Theta/\beta \approx I_\Theta a$. *The eddies adjust themselves so that the fractional variations in buoyancy, or potential temperature, in the vertical over the scale of the wave itself are comparable to the variations in the horizontal over the scale of the planet!*

There is an analogous barotropic problem, the eastward point jet on a β-plane. Here the flow is simply $\overline{u} = \Lambda|y|$. The most unstable wave in this system has the scale Λ/β in x and extends roughly this same distance in y. The baroclinic result is identical except that it is scaled by the Prandtl ratio f/N as appropriate.

What is the relevance of this argument for statistically steady states, as opposed to linear theory? Consider the barotropic jet problem first. The eddies will try to stabilize the flow, but if we use the Rayleigh-Kuo necessary condition for instability as a guide, to do this they need to remove the destabilizing curvature at $y = 0$. What is the minimum distance out to which vorticity must be homogenized so at to insure that the vorticity gradient is no longer negative? If we homogenize the vorticity up to distance Λ/β and assume continuity of $\overline{u}$ we can remove the negative vorticity gradient, but this requires the mean flow to be accelerated everywhere, which does not conserve momentum. To conserve momentum, we have to mix somewhat farther, but, in any case, this distance scales like Λ/β. We can refer to this as "barotropic adjustment", and can picture the eddies trying to achieve this state and coming to some accommodation with the forcing of the mean flow in which these same scales remain relevant. The same picture can be drawn for the baroclinic problem by simply changing y to z, even though the nonlinear dynamics is profoundly different, (and we need only that part of the domain with $z > 0$) resulting in a construction that can be thought of as the continuous version of "baroclinic adjustment" ([10]), in which the eddies try to destroy the surface temperature gradient and the potential vorticity gradients up to the height h. (They do not succeed entirely, especially near the surface where the forcing is too strong, but also at upper levels if a jet is present which will encourage mixing on its flanks and the formation of a mixing barrier at the jet core, as in Figure 8.)

The scaling in (54) assumes a Boussinesq atmosphere. Once this scale becomes comparable to the scale height H, dimensional arguments alone are insufficient to determine how these scales depend on h/H. The baroclinic adjustment construction can be generalized to the non-Boussinesq case; the depth of the eddy fluxes increases logarithmically with increasing horizontal temperature gradient, rather than linearly as above. Whether or not this prediction is reliable is unclear.

Given a theory for the vertical extent of the baroclinic eddy fluxes, can we complete our picture by providing a theory for the tropospheric static stability?

We have to return at this point to thinking about radiative forcing, which after all is ultimately driving everything. Think of the atmospheric absorber distribution as fixed for simplicity. Pure radiative equilibrium produces a very unstable vertical temperature profile. One simple way of fixing this is with a convective adjustment. Whenever the tropospheric lapse rate exceeds the dry adiabatic value, adjust it back to this value while conserving energy. This requires a vertical energy flux. The result of this construction is a "troposphere" extending from the ground up to some height H_T, in which the lapse rate is the dry adiabatic value, bounded above by a "stratosphere" in radiative equilibrium. Now suppose one has a theory for how the circulation maintains a lapse rate that is more stable than this dry adiabatic value. One can use this value in one's adjustment algorithm and generate a model in which the troposphere is forced to possess this prescribed static stability. As shown in Figure 15, this will result in a deeper troposphere. In general, one can think of H_T as a function of N. We can refer to this as the "radiative constraint" between H_T and N.

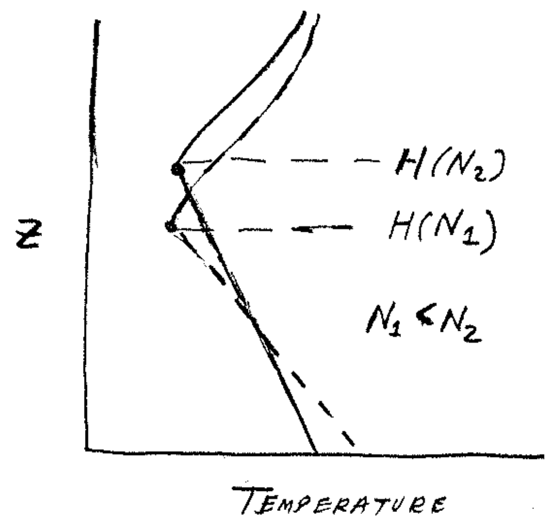

Figure 15: Schematic of temperature profiles obtained from a convective-adjustment algorithm, for different values of the prescribed tropospheric static stability

We need another relation, a dynamical constraint, between these two parameters in order to determine their values. Usually this kind of convective adjustment model is applied to the global mean temperature profile, but we can usefully apply it to the tropics and extratropics separately; indeed we must, since the kinds of eddies transporting heat vertically are so different in these two regions. (Technically, we now need to include the transport of heat from the tropics to midlatitudes in our computations, in both atmosphere and ocean; formally, we can just bundle these fluxes together with the radiative forcing, but this does not change the underlying picture.)

In the tropics, moist convection is dominant in the vertical energy fluxes. We can assume that moist convection more or less forces the moist static energy near the tropopause to be close to that near the surface in the most strongly convecting regions, or, averaged over the troposphere,

$$N^2 \approx \frac{Lq(0)}{c_p T(0)} \frac{g}{H_T}$$

which serves as our dynamical constraint. We can think of the mixing ratio for water vapor near the surface, $q(0)$, as given, or, better, we can imagine computing it from the surface temperature predicted by our radiative-convective model, assuming saturation close to the ground. Figure 16 illustrates how the radiative and dynamical constraints would then combine to determine N and H_T.

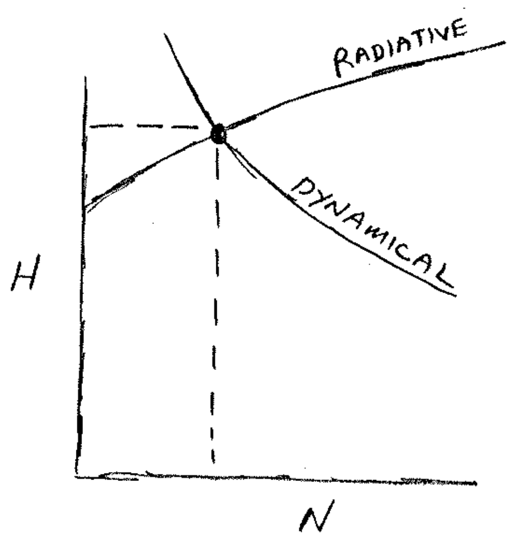

Figure 16: The radiative and dynamical constraints combine to determine the tropospheric static stability and the height of the tropopause.

In the extratropics, we need an analogous theory for the depth of penetration of the dynamical heating as a function of N. A theory of the type outlined above, leading to the non-Boussinesq generalization to (54) is a candidate for the appropriate constraint. It once again predicts that the tropopause height increases with decreasing static stability, so the resulting picture is similar to that above.

To summarize, in midlatitudes the low level eddy heat flux is more or less diffusive, with a diffusivity that depends on horizontal and vertical potential temperature gradients and on the strength of surface drag. Horizontal temperature gradients are determined by the balance between these diffusive fluxes and the radiative driving, as in the simplest diffusive energy balance models. Given this horizontal gradient, the tropopause height and the tropospheric static stability adjust to simultaneously satisfy the radiative and dynamical constraints, with both feeding back on the diffusivity to obtain a self-consistent picture. The missing ingredient that we have not discussed is the effect of moist convection in the warm sectors of cyclones on both the static stability of the troposphere and on eddy diffusivities.

In the tropics, vertical structure can be determined by a radiative-convective model in which the adjustment is made to the moist adiabat, using moisture levels from the region with the largest moist static energy, or moist entropy, at low levels. A circulation is generated so as to maintain small temperature gradients in the free troposphere, the width of this tropical circulation being determined by the region within which, in its absence, the upper levels winds would be too large to be physical realizable. This circulation forces the entire tropics to lie close to this moist adiabat and to possess the same tropopause height, resulting in a discontinuity in the tropopause at the subtropical jet. The inconsistency between this moist adiabat and surface temperatures is taken up at the trade wind inversion.

References

[1] Andrews, D. G., and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear. The Generalized Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031-2048.

[2] Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600-2616. Erratum: **38**, 1115.

[3] Green, J. S. A., 1970: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **96**, 157-185.

[4] Held, I. M. and A. Y. Hou, 1980: Nonlinear axisymmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.*, **37**, 515-533.

[5] Held, I. M., and V. D. Larichev, 1996: A scaling theory for horizontally homogeneous, baroclinically unstable flow on a β-plane. *J. Atmos. Sci.*, **53**, 946-952.

[6] Held, I. M., and T. Schneider, 1999: The surface branch of the zonally averaged mass transport circulation in the troposphere. *J. Atmos. Sci.*, **56**, 1688-1697.

[7] Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Am. Meteor. Soc.*, **75**, 1825-1830.

[8] Hoskins, B. J., H. H. Hsu, I. N. James, M. Masutani, P. D. Sardeshmukh, and G. H. White, 1989: *Diagnostics of the Global Atmospheric Circulation based on ECMWF Analyses 1970-1980.* WCRP-27, WMO/TD No. 326, 217 pp.

[9] Hoskins, B. J., M. E. McIntyre, A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111**, 877-946.

[10] Lindzen, R. S., and B. Farrell, 1980: The role of polar regions in global climate; and a new parameterization of global heat transport. *Mon. Weather. Rev.*, **108**, 2064-2079

[11] Palmen, E., and C. W. Newton, 1969: *Atmospheric Circulation Systems*, Academic Press, New York, 603pp.

[12] Panetta, R. L., 1983: Zonal jets in wide baroclinically unstable regions: persistence and scale selection. *J. Atmos. Sci.* **40**, 2073-2106.

[13] Pavan, V., and I. M. Held, 1996: The diffusive approximation for eddy fluxes in baroclinically unstable jets. *J. Atmos. Sci.*, **53**, 1262-1272.

[14] Peixoto, J. P. and A. H. Oort, 1992: *Physics of Climate.* American Institute of Physics, New York, pp. 520.

[15] Randel, W. J., and I. M. Held, 1991: Phase speed spectra of transient eddy fluxes and critical layer absorption. *J. Atmos. Sci.*, **48**, 688-697.

[16] Rhines, P. B., 1975: Waves and turbulence on a β-plane. *J. Fluid Mech.*, **69**, 417-443.

[17] Saravanan, R., 1993: Equatorial superrotation and maintenance of the general circulation in two-level models, *J. Atmos. Sci.* **50**, 1211-1227.

[18] Schneider, T. , personal communication, 1998.

[19] Shell, K., this volume

[20] Stone, P. H., 1972: A simplified radiative-dynamical model of the static stability of rotating atmospheres. *J. Atmos. Sci.*, **29**, 405-418.

[21] Suarez, M. J. , and D. G. Duffy, 1992: Terrestrial superrotation: a bifurcation of the general circulation. *J. Atmos. Sci.* **49**, 1541-1554.

[22] Williams, G. P., and J. L. Holloway, Jr., 1982: The range and unity of planetary circulations. *Nature*, **297**, 295-299.