Generalized dispersion relation and critical lines for a conservative, finite-amplitude Rossby wave in slowly varying barotropic shear flow

Noboru Nakamura and Clare S-Y Huang

Department of the Geophysical Sciences, University of Chicago

Corresponding author address: Noboru Nakamura, Department of the Geophysical Sciences, The University of Chicago, 5734 S. Ellis Ave, Chicago, IL 60637. Email: nnn@uchicago.edu

Abstract

A form of dispersion relation is proposed for a conservative, finite-amplitude, barotropic Rossby wave in slowly varying parallel shear flow. The relation is expressed in terms of pseudomomentum and pseudoenergy densities of the wave whose exact conservation laws are known. The zonal phase speed is given by the functional derivative of pseudoenergy density with respect to pseudomomentum density, wherein the effects of wave-mean flow interaction and the amplitude dependence of the phase speed are implicit. This theoretical prediction is compared with the observed phase speed in a numerical simulation of nonlinear barotropic decay on a sphere. The theory agrees well with the observed phase speed of the wave except in regions where the meridional eddy momentum flux changes sign and/or where the phase speed and phase tilts change abruptly. In a significant departure from standard theory, multiple critical lines are identified on each flank of the jet in the simulation, with significant wave amplitude through them. It is shown that wave-mean flow interaction in the generalized dispersion relation renders critical lines nonsingular and permits the Rossby wave to be transmitted through them, even into the region where the phase speed of the wave exceeds the zonal-mean zonal wind.

1. Introduction

The Rossby wave plays a central role in the large-scale circulation of the Earth's atmosphere. While early linear theory illuminated the fundamental properties of the atmospheric Rossby wave (Rossby[1] 1939; Haurwitz 1940; Charney and Drazin 1961; Longuet-Higgins 1964, 1965; Platzman 1968; Hoskins and Karoly 1981; Held 1983), extensive inquiries were also directed toward the wave's interaction with the mean flow (Dickinson 1969; Matsuno 1971; Andrews and McIntyre 1976; Boyd 1976; McIntyre and Palmer 1983; Killworth and McIntyre 1985; Haynes and McIntyre 1987). It is now widely recognized that the Rossby wave can drive the mean state of the atmosphere through its radiation stress, represented by the generalized Eliassen-Palm flux in the quasigeostrophic limit of the Transformed Eulerian Mean (TEM) formalism (Andrews and McIntyre 1976, Edmon et al. 1980, Andrews et al. 1987).

Although the response of the mean field to the wave forcing is well described by the TEM set, the corresponding theory for the finite-amplitude Rossby wave in shear flow is less well developed. General description of finite-amplitude wave-mean flow interaction was considered mainly in the field theory and fluid mechanics literature (Sturrock 1961; Whitham 1965; Bretherton and Garrett 1968; Hayes 1970) and culminated in the Generalized Lagrangian Mean (GLM) formalism of Andrews and McIntyre (1978ab, see also Bühler 2009). Unfortunately these theories are not readily verified by (or applied to) atmospheric data because key quantities

[1] The original Rossby paper, while almost exclusively dealing with the atmospheric Rossby wave, predated the launch of the *Journal of Meteorology* (the predecessor of the *Journal of the Atmospheric Sciences*) and was published in the *Journal of Marine Research*.

such as pseudomomentum and pseudoenergy densities are difficult to diagnose at finite amplitude. (A notable exception is the formalism developed by Killworth and McIntyre 1985, McIntyre and Shepherd 1987, and Haynes 1988.)

Recently Nakamura and Zhu (2010) derived an exact conservation law for finite-amplitude pseudomomentum (wave activity) density applicable to Rossby waves and balanced eddies, extending the generalized Eliassen-Palm theorem and hence the TEM set to arbitrary eddy amplitude. For conservative barotropic flow on a sphere, this can be written as (Nakamura and Solomon 2010, Solomon and Nakamura 2012)

$$\frac{\partial}{\partial t}\left(Aa\cos\phi\right)=\frac{\partial}{\partial \mu}\left(\overline{u'v'}\left(1-\mu^2\right)\right), \tag{1}$$

where t is time, a is the radius of the sphere, $(u,v)=(\dot{\lambda}a\cos\phi, a\dot{\phi})$ is the wind vector in longitude λ and latitude ϕ, $\mu\equiv\sin\phi$, and the overbar and prime denote longitudinal average and departure from it, respectively.[2] The angular (zonal) pseudomomentum density $Aa\cos\phi$ is defined by

$$\begin{aligned} A(\mu,t)a\cos\phi &= \frac{1}{2\pi}\left(C\left(Q(\mu,t)\right)-\bar{C}(\mu,t)\right) \\ &= \frac{1}{2\pi}\left(\iint_{D(Q)} qa^2\, d\lambda\, d\mu - \iint_{D(\mu)} qa^2\, d\lambda\, d\mu\right), \end{aligned} \tag{2}$$

where $C(Q(\mu,t))$ is Kelvin's circulation around the wavy contour of absolute vorticity $q(\lambda,\mu,t)=Q$ that encloses the same area as the polar cap north of μ, whereas $\bar{C}(\mu,t)=2\pi a\cos\phi\left(\bar{u}(\mu,t)+\Omega a\cos\phi\right)$ is Kelvin's circulation around the

[2] Throughout the text $\cos\phi$ is used in place of $(1-\mu^2)^{1/2}$ where it simplifies the notation.

zonal circle at latitude μ, and Ω is the sphere's rotation rate. The last line in (2) uses Stoke's theorem and $D(Q)$ and $D(\mu)$ denote the regions delimited by the Q-contour and by the latitude circle, respectively. It is readily shown (Nakamura and Zhu 2010) that $A \geq 0$ and that A converges to the familiar linear pseudomomentum density [Held 1985, see also (29a) below] in the small-amplitude limit. To evaluate (2) one only needs instantaneous distribution of absolute vorticity q.

Under conservative dynamics the first term on the right hand side of (2) is a constant of motion due to Kelvin's circulation theorem, so the time derivative of (2) leads to

$$\frac{\partial}{\partial t} U_{REF}(\mu) = 0, \qquad U_{REF} = \overline{u}(\mu,t) + A(\mu,t). \tag{3ab}$$

The reference state velocity U_{REF} is the zonal velocity of the flow that would emerge if the wavy q contour were 'zonalized' without changing the enclosed area or Kelvin's circulation (Nakamura and Zhu 2010, Solomon and Nakamura 2012). Note that U_{REF} is steady even when $\overline{u}$ and A are not. This fact can be exploited to derive conservation of pseudoenergy density $E(\mu,t)$, the difference between the zonally averaged energy density and the energy density of the reference state, both measured relative to the rotating sphere:[3]

$$\frac{\partial}{\partial t} E = -\frac{\partial}{\partial \mu} \overline{\left(\overline{u} u' + \frac{p'}{\rho_0} + e \right) \left(\frac{v' \cos\phi}{a} \right)}\ , \tag{4}$$

[3] Unlike angular pseudomomentum density $A\cos\phi$, pseudoenergy density E does not obey rotational invariance (analogous to Galilean invariance on the beta plane), so it is important to define the reference.

$$E(\mu,t)=\frac{1}{2}\overline{\left((\overline{u}+u')^2+v'^2\right)}-\frac{1}{2}U_{REF}^2=\overline{e}-U_{REF}A+\frac{A^2}{2}, \qquad (5)$$

where $e=(u'^2+v'^2)/2$ is eddy kinetic energy density, p' is pressure perturbation and ρ_0 is constant density. Note we used (3b) to derive the last expression in (5).

Both A and E vanish if the flow is zonally symmetric and the gradient of absolute vorticity is single-signed, so they are properties of eddy (Nakamura and Zhu 2010). These quantities are easy to evaluate from instantaneous data and all the foregoing results are valid for arbitrary amplitude and shape of eddies, whether wavelike or turbulent. Furthermore, it is clear from (1) and (4) that the domain integrals of angular pseudomomentum and pseudoenergy are invariant in time, since the eddy fluxes on the right-hand side vanish at the poles.

In this article we will utilize the above conservation laws to derive generalized dispersion relation for finite-amplitude Rossby waves in barotropic shear flows. By 'generalized' we mean that the phase speed is expressed as a functional of A and E, fundamental conserved properties of the wave at arbitrary amplitude, instead of prescribed wavenumbers. This allows one to incorporate implicitly the dependence of the phase speed on wave amplitude and the effects of wave-mean flow interaction. Since both A and E are evaluable from data, the theory can be tested by numerical simulations of finite-amplitude Rossby waves.

The next section outlines the derivation of the generalized dispersion relation, emphasizing the assumptions made. In section 3 the theoretical prediction of phase speed is tested with a numerical simulation of a freely decaying Rossby wave in a shear flow on a rotating sphere (Held and Phillips 1987). The accuracy of the theory will be demonstrated along with some limitations. Section 4 closely examines the phase speed structure and critical lines in the simulation. It will be shown that the geometry of critical lines differs significantly from classical theory. We will see that nonlinearity in the generalized dispersion relation removes singularity from critical lines and enables finite-amplitude Rossby wave to propagate through them, even into a region where the phase speed is eastward with respect to the zonal mean flow. The final section summarizes the results.

2. Generalized dispersion relation

A well-established strategy for finding the dispersion relation and conservation laws for a slowly modulated, almost plane wave is to apply variational principle to the *phase-averaged Lagrangian density* (Sturrock 1961, Whitham 1965, 1970, Bretherton and Garrett 1968, Hayes 1970, Grimshaw 1984). Normally this procedure entails finding an appropriate Lagrangian, taking the average over phase, and then requiring stationarity of the average action with respect to amplitude and phase to obtain the dispersion relation and conservation laws, respectively. However it is not always easy to find a suitable Lagrangian for a finite-amplitude wave and its physical interpretation is confounded by the fact that there may be more than one Lagrangian that generate the same governing equation. To date, average Lagrangian densities for the Rossby wave have been found only in the small-amplitude limit (Seliger and Whitham 1968, Buchwald 1972).

Instead of attempting to identify the exact form of Lagrangian density, in what follows we *assume* that an average Lagrangian density for the Rossby wave exists as a function of local wavenumbers and frequency. Then by requiring stationarity of the averaged action with respect to phase, we obtain the conservation laws in terms of Lagrangian density (Whitham 1965, Bretherton and Garrett 1968). Since the derived conservation laws should correspond to (1) and (4), by matching the terms we can determine the relationships among the (unknown) Lagrangian density, (known) pseudomomentum and pseudoenergy densities A, E [(2)(5)], and the (unknown) phase speed of the Rossby wave. Finally by eliminating the Lagrangian density from these relationships we will obtain the expression for the phase speed in terms of A and E, which we call the generalized dispersion relation.

Barotropic flow on a sphere is described by streamfunction $\psi(\lambda,\mu,t)$ such that $\partial\psi/\partial\lambda = va\cos\phi$, $\partial\psi/\partial\mu = -ua/\cos\phi$. Let

$$\psi(\lambda,\mu,t) = \psi' + \Delta\bar{\psi} + \psi_{REF}, \tag{6}$$

where

$$\psi_{REF}(\mu) = -a\int \frac{U_{REF}}{\cos\phi} d\mu \tag{7}$$

is streamfunction of the reference state defined by (3) and

$$\Delta\bar{\psi}(\mu,t) = a\int \frac{A}{\cos\phi} d\mu \tag{8}$$

is the change in the zonal-mean streamfunction due to eddy, where A is defined in (2). The eddy streamfunction ψ' is assumed to be nearly plane, that is, a periodic function of phase function $\theta(\lambda,\mu,t)$ such that

$$\psi'(\theta) = \psi'(\theta + 2\pi). \tag{9}$$

It is understood that the average of ψ' over 2π vanishes. The local wavenumbers and frequency are defined as the gradients of phase function

$$k \equiv \theta_\lambda, \quad l \equiv \theta_\mu, \quad \sigma \equiv -\theta_t, \tag{10}$$

where the subscripts denote partial derivative. For simplicity, we only consider waves for which (k,l,σ) is independent of longitude λ (compatible with the fact that the background flow $\Delta\bar{\psi} + \psi_{REF}$ is independent of λ). Then

$$\begin{gathered} k_\lambda = 0,\ l_\lambda = \theta_{\mu\lambda} = k_\mu = 0,\ \sigma_\lambda = \theta_{t\lambda} = -k_t = 0 \\ \Rightarrow \quad k:\ \text{constant}, \quad l = l(\mu,t), \quad \sigma = \sigma(\mu,t). \end{gathered} \tag{11}$$

To ensure periodicity (9), neither the background flow $\Delta\bar{\psi} + \psi_{REF}$ nor the wave properties can change significantly over one cycle (2π) of phase. Thus, we require

$$l \gg \mu_b^{-1}, \qquad \sigma \gg t_b^{-1} \text{ (except for steady state),} \tag{12}$$

where μ_b and t_b are the meridional and temporal scales of the background state, respectively. [A more formal two-scale analysis is spared for the sake of space—see for example Whitham (1970, 1974 ch.14).]

Now we assume that the barotropic equations of motion arise from stationarity of action

$$\delta \iiint L \, d\lambda d\mu dt = 0, \tag{13}$$

where L is the Lagrangian density of the flow. Finding the exact form of L is beyond our scope, but it is safe to assume that L depends on the first derivatives of ψ. With the waveform given by (9) and (10), the zonal average of L, denoted by $\bar{L}$, should depend on $(\theta_t, \theta_\lambda, \theta_\mu) = (-\sigma, k, l)$ but not on θ or λ because any derivative of ψ' is written as ψ'_θ times the gradient of θ [(9), (10)], whereas the zonal averaging eliminates the dependence on θ or λ (due to the zonal periodicity of the flow and constant k, zonal averaging has the same effect as the phase averaging). Likewise

$\bar{L}$ should depend on the amplitude of ψ'_θ, $\alpha^2(\mu,t) \equiv \overline{\psi'^2_\theta}$, as well as $A/\cos\phi$ through $\Delta\bar{\psi}$ [(8)] and $U_{REF}(\mu)/\cos\phi$ through ψ_{REF} [(7)], but not explicitly on time (the time dependence enters only through modulation of $\nabla\theta$, α^2 and A). Let us then define *perturbation* Lagrangian density $\mathcal{L}$ as

$$\mathcal{L}(\theta_t, \theta_\lambda, \theta_\mu, \alpha^2, A, \mu) \equiv \bar{L} - L_{REF}, \tag{14}$$

where $L_{REF}(\mu)$ is the Lagrangian density of the reference state. Again $\mathcal{L}$ does not depend explicitly on λ or t because of the zonal and temporal symmetries of the reference state.

Conservation laws and dispersion relation for the wave are derived by requiring that the average action be stationary (Whitham 1965, Bretherton and Garrett 1968)

$$\delta \iint \mathcal{L}\, d\mu dt = 0 \tag{15}$$

with respect to independent variables of the wave, normally phase θ and amplitude. Of the list of variables in (14), both α^2 and A concern wave amplitude, one Eulerian and the other Lagrangian, and therefore they are interdependent. Furthermore, even though neither α^2 nor A depends on θ, at finite amplitude they may be related to frequency $-\theta_t$ (Sturrock 1961; Whitham 1974 ch.14). Therefore separating phase and amplitude is not trivial at finite amplitude; for this reason we only consider variation with respect to θ as an independent variable. (In linear theory, α^2 and A are proportional to each other and they are independent of frequency. Thus variation with respect to α directly yields a dispersion relation: Whitham 1965; Bretherton and Garrett 1968; Seliger and Whitham 1968.)

Variation with respect to θ at fixed μ leads to the following conservation laws for pseudomomentum and pseudoenergy in the form of Euler-Lagrange equations (e.g. Salmon 2013):

$$\frac{\partial}{\partial t}\left(\theta_\lambda \frac{\partial \mathcal{L}}{\partial \theta_t}\right)+\frac{\partial}{\partial \mu}\left(\theta_\lambda \frac{\partial \mathcal{L}}{\partial \theta_\mu}\right)=0\,, \tag{16}$$

$$\frac{\partial}{\partial t}\left(\theta_t \frac{\partial \mathcal{L}}{\partial \theta_t}-\mathcal{L}\right)+\frac{\partial}{\partial \mu}\left(\theta_t \frac{\partial \mathcal{L}}{\partial \theta_\mu}\right)=0\,. \tag{17}$$

These conservation equations arise from the fact that $\mathcal{L}$ does not depend explicitly on λ or t (i.e., the zonal and temporal symmetries of the reference state). Equations (16) and (17) correspond to (1) and (4), respectively, for a slowly modulated, near plane Rossby wave. Matching of terms between (1) and (16) and between (4) and (17) suggests

$$Aa\cos\phi=\theta_\lambda \frac{\partial \mathcal{L}}{\partial \theta_t}=-k\frac{\partial \mathcal{L}}{\partial \sigma}\equiv-\frac{\partial \mathcal{L}}{\partial \omega}, \qquad \omega=\frac{\sigma}{k} \tag{18}$$

$$E=\theta_t \frac{\partial \mathcal{L}}{\partial \theta_t}-\mathcal{L}=\sigma\frac{\partial \mathcal{L}}{\partial \sigma}-\mathcal{L}=\omega\frac{\partial \mathcal{L}}{\partial \omega}-\mathcal{L}=-\omega Aa\cos\phi-\mathcal{L}\,, \tag{19}$$

$$-\overline{u'v'}\left(1-\mu^2\right)=\frac{kl(1-\mu^2)\alpha^2}{a^2}=\theta_\lambda \frac{\partial \mathcal{L}}{\partial \theta_\mu}=k\frac{\partial \mathcal{L}}{\partial l}\ , \tag{20}$$

$$\overline{\left(\bar{u}u'+\frac{p'}{\rho_0}+e\right)\left(\frac{v'\cos\phi}{a}\right)}=\theta_t \frac{\partial \mathcal{L}}{\partial \theta_\mu}=-\sigma\frac{\partial \mathcal{L}}{\partial l}=-\omega k\frac{\partial \mathcal{L}}{\partial l}\,, \tag{21}$$

where ω is angular (zonal) phase speed of the wave. (In what follows we will prefer to work with $c\equiv\omega a$, equivalent phase speed at the equator.) From (18)

$$\mathcal{L} = -\int (Aa\cos\phi)d\omega = -\int (A\cos\phi)dc,^4 \tag{22}$$

and from (19)

$$\begin{aligned} E &= -\omega Aa\cos\phi - \mathcal{L} = -cA\cos\phi + \int (A\cos\phi)dc \\ &= -\int c\, d(A\cos\phi). \end{aligned} \tag{23}$$

The derivative form of (23) is

$$\begin{aligned} c(\mu,t) &= -\left.\frac{\partial E}{\partial(A\cos\phi)}\right|_{k,l,\mu \text{ fixed}} \\ &= \frac{U_{REF} - A}{\cos\phi} - \frac{\partial \overline{e}}{\partial(A\cos\phi)} = \frac{\overline{u}}{\cos\phi} - \frac{\partial \overline{e}}{\partial(A\cos\phi)}, \end{aligned} \tag{24}$$

where we used (5) to derive the second expression. In (24) the phase speed of the Rossby wave c is given by the functional derivative of pseudoenergy density E with respect to angular pseudomomentum density $A\cos\phi$. We call (23) and (24) the generalized dispersion relation, since the definitions and conservation of $A\cos\phi$ and E [(1)-(5)] do not depend on specific shape of the wave. Admittedly we assumed a rather restrictive waveform (9)-(11) to derive the dispersion relation, but the derived result (24) may be used to find the phase speed even when the exact waveform is not known (details to be outlined below). All information about the properties of the wave and the mean flow that affect the phase speed is included in E and $A\cos\phi$, quantities easily evaluable from data.

The generalized dispersion relation is nonlinear in two important ways. First,

---

[4] An arbitrary function of k and l may be added to the right-hand side as a constant of integration, but this function is actually zero since (19) requires $\mathcal{L} \to 0$ as $E, A \to 0$.

the phase speed c depends on angular pseudomomentum density $A\cos\phi$ and therefore it is amplitude-dependent. This is why the right-hand side of (23) takes an integral form, and (24) a derivative form: in other words, the $A\cos\phi$ - E relation is nonlinear. Second, the dispersion relation includes the effect of wave-mean flow interaction [(3)] through the quadratic term in A [(5)]. This term is negligible at small amplitude because in that limit both $\bar{e}$ and $A\cos\phi$ are quadratic in (the small) wave amplitude (see section 2b below). Before evaluating (24) from data, we shall discuss meridional propagation of wave amplitude and a few useful limiting cases.

(2a) *meridional propagation of amplitude*

If rewrite (1) and (4) as

$$\frac{\partial}{\partial t}(Aa\cos\phi)+\frac{1}{a}\frac{\partial}{\partial\mu}(c_A Aa\cos\phi)=0\,, \tag{25}$$

$$\frac{\partial}{\partial t}E+\frac{1}{a}\frac{\partial}{\partial\mu}(c_E E)=0\,, \tag{26}$$

where c_A and c_E are the effective meridional transport velocities of angular pseudomomentum and pseudoenergy densities, respectively, then from (18)-(21)

$$c_A=\frac{-\overline{u'v'}(1-\mu^2)}{A\cos\phi}=\frac{kl\,(1-\mu^2)\alpha^2}{a^2A\cos\phi}=-a\frac{\partial\mathcal{L}/\partial l}{\partial\mathcal{L}/\partial\sigma}, \tag{27}$$

$$c_E=E^{-1}\overline{\left(\bar{u}u'+\frac{p'}{\rho_0}+e\right)v'\cos\phi}=-a\frac{\sigma\partial\mathcal{L}/\partial l}{\sigma(\partial\mathcal{L}/\partial\sigma)-\mathcal{L}}\,. \tag{28}$$

As we will see below, in the small-amplitude limit $c_A\approx c_E$ and they are reduced to the meridional group velocity of the Rossby wave.

(2b) *small-amplitude limit*

In the small-amplitude limit, A and E are approximated as (e.g., Held 1985; Nakamura and Solomon 2010, appendix)

$$A\cos\phi \approx \frac{a}{2}\frac{\overline{q'^2}}{\partial\overline{q}/\partial\mu}; \qquad E \approx \overline{e} - \overline{u}A \tag{29ab}$$

(the A^2 term in E is negligible in this limit). Since both E and $A\cos\phi$ are quadratic in small wave amplitude, $\partial E/\partial(A\cos\phi)$ in (24) may be approximated as $E/A\cos\phi$:

$$c(\mu,t) = -\frac{E}{A\cos\phi} \approx \frac{\overline{u}}{\cos\phi} - \frac{\overline{e}}{A\cos\phi}. \tag{30}$$

Therefore (23) becomes

$$E \approx -c\,A\cos\phi = -\omega A a\cos\phi; \qquad \mathcal{L} \approx 0, \tag{31}$$

namely, the amplitude dependence of phase speed in (23) is negligible and the Lagrangian density vanishes. The vanishing $\mathcal{L}$ renders $c_E = c_A$ [(28), (27)] and also

$$0 \approx \delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\sigma}\delta\sigma + \frac{\partial\mathcal{L}}{\partial l}\delta l. \tag{32}$$

Hence with (27) and (32)

$$c_E = c_A = -a\frac{\partial\mathcal{L}/\partial l}{\partial\mathcal{L}/\partial\sigma} \approx a\frac{\delta\sigma}{\delta l} = c_g, \tag{33}$$

namely c_E and c_A both approach the meridional group velocity of the Rossby wave.

(2c) *Steady amplitude limit*

When the wave amplitude is steady, the tendency terms in (16) and (17) vanish and thus the fluxes are invariant with μ. From (16) and (20)

$$k\frac{\partial \mathcal{L}}{\partial l} = -\overline{u'v'}(1-\mu^2) = \frac{kl\,(1-\mu^2)\alpha^2(\mu)}{a^2} = \text{indep. of } \mu$$
$$\Rightarrow \quad \left[l(1-\mu^2)\right]^{-1} \propto \alpha^2(\mu), \qquad (34)^5$$

and since l is invariant in time, from (10) $l_t = \theta_{\mu t} = -\sigma_\mu = -k\omega_\mu = 0$. Thus c is independent of latitude. The dispersion relation (24) becomes

$$c = \frac{\overline{u}}{\cos\phi} - \frac{\partial \overline{e}}{\partial (A\cos\phi)}, \qquad (35)$$

where c is constant but the terms on the right-hand side are slowly varying functions of μ. One may consider (34)-(35) as the finite-amplitude extension to the Wentzel-Kramers-Brillouin (WKB) dispersion relation for the steady Rossby wave.

(2d) *Beta-plane approximation*

Translating the foregoing results for the beta plane is straightforward: one simply needs to replace $(a\lambda,\ a\mu,\ \cos\phi,\ 1-\mu^2)$ with $(x, y, 1,\ 1)$. For example, (24) becomes

$$c(y,t) = -\left.\frac{\partial E}{\partial A}\right|_{k,\,l,\,y\ \text{fixed}} = \overline{u}(y,t) - \frac{\partial \overline{e}}{\partial A}. \qquad (36)$$

It is well known that plane wave

$$\psi'(x,y,t) = B\cos\big(k(x-ct)+ly\big) \qquad (37)$$

is an exact solution of the nonlinear barotropic vorticity equation on the infinite beta plane if the background flow is uniform ($\overline{u} = u_0$) and if the phase speed c is given by

[5] If the wave amplitude is steady globally, $\overline{u'v'}(1-\mu^2) = 0$ everywhere since the flux vanishes at the poles. Thus (34) should be applied only locally. (This is not the case for an infinite beta plane.)

$$c = u_0 - \frac{\beta}{k^2 + l^2}, \tag{38}$$

where β is the constant gradient in the Coriolis parameter. Because of the symmetry along the phase lines, it is straightforward to show that (the Cartesian version of) (2) leads to (appendix A)

$$A = \frac{\overline{(\nabla^2 \psi')^2}}{2\beta} = \frac{(k^2 + l^2)^2 B^2}{4\beta} , \tag{39}$$

i.e., exactly the same form as the linear pseudomomentum density [(29a)]. Since

$$\overline{e} = \frac{\overline{(\nabla \psi')^2}}{2} = \frac{(k^2 + l^2) B^2}{4} = \frac{\beta}{k^2 + l^2} A, \tag{40}$$

substitution in (36) recovers (38). Therefore, the well-known dispersion relation for the Rossby wave is derived from the generalized dispersion relation based on pseudomomentum and pseudoenergy densities. In particular, the 'beta-effect' is introduced through A [(39)].

(2e) *Diagnosing the phase speed of nonsteady wave from A and E*

The foregoing Lagrangian-based formalism is in fact applicable to a wide class of slowly modulated near-plane waves, not just Rossby waves. Results analogous to (18)-(33) have been reported for example by Sturrock (1961), Bretherton and Garrett (1968), and Grimshaw (1984). What is new here is that, since A and E for the Rossby wave (and balanced eddies) may be evaluated with data according to (2) and (5), we can test the validity of the finite-amplitude dispersion relation. A final obstacle in doing so is that (24) is written in terms of functional derivatives that cannot be evaluated explicitly because the precise functional dependence of $\overline{e}$ on A

is unknown except for special cases like (40) in which the two quantities are proportional to each other.

To circumvent this difficulty, we rewrite (16) and (17) using (18)-(21) as

$$\frac{\partial}{\partial t}\left(Aa\left(1-\mu^2\right)^{1/2}\right)=-\frac{\partial}{\partial \mu}\left(k\frac{\partial \mathcal{L}}{\partial l}\right), \tag{41}$$

$$\frac{\partial E}{\partial t}=\frac{\partial}{\partial \mu}\left(\omega k\frac{\partial \mathcal{L}}{\partial l}\right). \tag{42}$$

Since the fluxes on the right-hand side vanish at the poles [(20), (21)], c may be evaluated as

$$c(\mu,t)=\omega a=-\frac{\partial E^\dagger/\partial t}{\partial A^\dagger/\partial t},$$

$$E^\dagger(\mu,t)\equiv\int_{-1}^{\mu}E(\mu',t)d\mu', \quad A^\dagger(\mu,t)\equiv\int_{-1}^{\mu}A(\mu',t)\left(1-\mu'^2\right)^{1/2}d\mu'. \tag{43}$$

Unlike the linear limit, the phase speed is not determined locally but involves integrals of $A\cos\phi$ and E over latitude. [In (43) the integrals are defined over $[-1,\mu]$ but they can be also defined over $[\mu,1]$ without affecting the result.] However where $\partial c/\partial\mu$ or $\overline{u'v'}$ vanishes it may be evaluated locally:

$$c^*\equiv-\frac{\partial E(\mu,t)/\partial t}{(1-\mu^2)^{1/2}\,\partial A(\mu,t)/\partial t}=c+\frac{(\partial c/\partial\mu)\left(\overline{u'v'}(1-\mu^2)\right)}{a(1-\mu^2)^{1/2}\,\partial A(\mu,t)/\partial t},$$

$$c^*=c \quad if \quad \frac{\partial c}{\partial \mu}=0 \quad \text{or} \quad \overline{u'v'}=0. \tag{44}$$

In the next section we will use (43) to evaluate the phase speed of a numerically simulated finite-amplitude Rossby wave and compare the result with the directly measured phase speed to verify the accuracy of the theory.

3. Rossby wave phase speed during nonlinear barotropic decay on a sphere

The generalized dispersion relation derived above is now tested in a numerical simulation of nonlinear barotropic decay on a sphere. The experimental setup is identical with that of Held and Phillips (1987), HP87 hereafter, except that we use a T170 spectral transform model with a 12$^{\text{th}}$-order hyperdiffusion that damps the shortest wave with the e-fold time of 0.3 day. A wave with zonal wavenumber $k = 6$ whose amplitude is initially centered at 45° N, slightly north to a midlatitude jet, is allowed to evolve freely thereafter. This may be thought of as a crude model of decaying synoptic eddies in the upper troposphere. The model solves barotropic vorticity equation and angular pseudomomentum density $A\cos\phi$ is computed at a regular time interval (6000s) using the method described in appendix B. Together with the velocity output, we compute pseudoenergy density E according to (5).

Figure 1 samples snapshots of absolute vorticity in the Northern Hemisphere at six stages of a simulation with initial wave amplitude (ζ_0 in HP87) of $4\times10^{-5}s^{-1}$. This amplitude is chosen because the result exhibits a wealth of wave behaviors before dissipation renders the flow nonconservative. The wave initially centered at 45° N creates large meandering of vorticity contours (Fig.1a), but subsequently the meanders migrate southward with a characteristic eastward tilt with increasing latitude (Figs.1bc). Eventually the wave breaks around 25° N and forms a critical layer in which vorticity is quickly stirred (Figs.1d-f, see section 4 for more on critical lines). By day 15 zonal wavenumber 12 emerges as a prominent structure in 20° - 30° N (Fig.1f). There is also poleward intrusion of filaments of low absolute vorticity, although it is less wavelike because the filaments are strained by the cutoff vortices that already exist in the initial condition.

Figures 2a and 2b show $A\cos\phi$ and E as functions of time and latitude. The former is everywhere positive by definition, whereas the latter proves predominantly negative because of the second term in the last expression of (5). The two quantities are highly correlated and they demonstrate clearly that the initially centralized wave activity separates into equatorward and poleward packets, until they congregate around 24° and 55° N. At these latitudes $A\cos\phi$ and E attain maximum amplitudes around day 9-10. Up to ~day 6 the transport velocity for pseudomomentum c_A is characterized by large negative (equatorward) values to the south of 45° N and small positive (poleward) values to the north (Fig.2c). Then c_A changes sign to positive in the midlatitude around day 8.5. This coincides with a reversal of phase tilt due to rapid reconfiguration of phase lines (see Fig.11 below). Such transition hints at reflection of the wave from the critical lines and/or interference of multiple waves (though this is hardly visible in Figs.2a and 2b because of very small wave activities in the midlatitude), at which point the generalized dispersion relation becomes invalid since the theory is built on the assumption of single wave. Figures 3a-c show $\overline{u}/\cos\phi$ and its departure from the initial condition, as well as $U_{REF}/\cos\phi$ in the same coordinate as Fig.2. The zonal-mean angular velocity $\overline{u}/\cos\phi$ is accelerated around 45° N as the wave exits the source region, whereas it is decelerated where the wave packets arrive (compare Fig.3b with Fig.2a). Since the region of acceleration is north to the axis of the initial jet (35° N), the jet is displaced northward: by day 6, the axis moves to 42° N. This shift does not occur in $U_{REF} / \cos\phi$, consistent with (3a), demonstrating that the displacement of the jet is due to repartitioning of $\overline{u}$ and A.

Global conservation of $A\cos\phi$ and E are examined in Fig.4. The domain averages of these quantities are invariant under conservative dynamics and indeed they remain constant up to ~day 6, after which the average $A\cos\phi$ (E) decreases (increases) significantly. This is largely due to dissipation of pseudomomentum density by enhanced diffusive flux of vorticity (Nakamura and Zhu 2010) in the critical layer. We therefore expect (43) to hold for at least first six days of the simulation.

The phase speed in (43) is evaluated as the slope of the $A^{\dagger}$-$E^{\dagger}$ curve at each latitude. Figure 5 shows scatter plots of $E^{\dagger}$ (ordinate) versus $A^{\dagger}$ (abscissa) at ten different latitudes through the first seven days of simulation. These quantities are computed by integrating $A\cos\phi$ and E meridionally [(43)], from the South Pole for $15^{\circ} - 45^{\circ}$ N and from the North Pole for $50^{\circ} - 60^{\circ}$ N. As one might expect from Figs.2a and 2b, the two quantities are tightly correlated and form a compact curve at all latitudes. Each curve consists of 101 data points and they proceed in time from the upper left to the lower right. The length of the curve is short at 15° and 60° N because the wave amplitude is consistently small in the domain of integration throughout the seven-day period (Figs. 2a and 2b). The curve length increases dramatically from 15° to 35° N where the wave amplitude undergoes significant growth over time. The curve shortens again from 35° to 45° N, as the wave amplitude here decreases in time and this partially cancels the increasing trend in the lower latitudes upon meridional integration. The slope of the curves in Fig.5 [(43)] equals the phase speed $c = \omega a$. To the lowest order the curves appear close to being linear with similar slopes, but weak curvatures are also recognizable particularly during the early stage of the simulation, suggesting that the phase speed does depend on the wave amplitude.

The phase speed estimated from Fig.5 will be shown shortly, but for a validation purpose we need direct 'observation' of the phase speed, which we compute with the meridional velocity (v) based on the formula

$$c_{obs}(\mu,t) \approx -a\frac{\overline{(\partial v/\partial t)(\partial v/\partial \lambda)}}{\overline{(\partial v/\partial \lambda)^2}}, \tag{45}$$

where overbar denotes zonal average. Unlike spectral analysis, (45) may be evaluated instantaneously (the tendency term is evaluated from the difference between two consecutive outputs), and if v is simply translating in longitude at a uniform phase speed, i.e., $v = v(\lambda - \varphi(\mu,t))$, $\varphi_t = c_{obs}(\mu,t)$, then (45) is exact. Here v is an arbitrary smooth function and its amplitude need not be small. When multiple waves are present, that is, when ξ is also a function of λ, (45) defines a weighted average phase speed of all wave components. To demonstrate the accuracy of (45), we show in Fig.6 the phase migration of the meridional velocity v at 37° and 20° N as functions of longitude and time (Hovmöller diagram): shown on the left are the model output of v normalized by their instantaneous rms value across longitudes, and on the right are the reconstructions by way of translating a sine function with the phase speed calculated from (45). In the model output wavenumber 6 remains dominant throughout the simulation, although wavenumber 12 also emerges at 20° N toward the end. At 37° N there is a kink in the phase lines around day 8.5, and at 20° N the phase lines are discontinuous around day 6.5 and 7.5. These irregularities are due to breakup and reconnection of phase lines, to be discussed more in detail in the next section. The reconstructions reproduce the salient features of the model output, smoothly connecting the gaps in the phase lines. Even at the end of the 15-day period the placement of the phase remains nearly identical with the model output, justifying the use of (45) as a surrogate for the phase speed. (We have also used the eddy component of relative vorticity in place of v to evaluate (45) and obtained virtually identical results). Finally, we will also compare the phase speed estimated from (43) with the linear theory [based on (30) with the small amplitude approximations to A and E (29) and the instantaneous values for the zonal-mean quantities].

Figure 7 shows the values of phase speed based on (43) (thick dots), (45) (thin solid curve), and (30) (thin dots) as functions of time. At latitudes 40° N and higher, the observed phase speed c_{obs} (solid curve) starts at around 20-25 ms^{-1} and slowly increases to about 30 ms^{-1} at the end of the seven day period. At lower latitudes the initial phase speed is much slower—in fact negative (westward) at 15° and 20° N— but quickly increases to about 25 ms^{-1} in 3-4 days. This initial latitudinal gradient in the phase speed is consistent with the development of phase tilt between 15° and 40° N in absolute vorticity (Fig.1b). The large fluctuation in the observed phase speed at 15° N is due to rapid breakup and reconnection of the phase lines across a critical line (see section 4 and Fig.11). At this latitude the theoretical estimate based on (43) matches the observed values only for a short period (day 1-3) and deviate significantly at other times, but the agreement improves as one moves higher in latitude (20° - 35° N). Overall the theoretical estimate based on (43) agrees with the observed values better than the linear theory, particularly at 15° and 60° N and during the early stage of the simulation — a somewhat surprising result given the small wave amplitude in these circumstances. A possible reason for this is that the linear dispersion relation only quantifies changes in wave geometry, while initially the wave amplitude is also changing rapidly: only the finite-amplitude theory takes the effect of changing amplitude into account. On the other hand, at 45° and 50° N the linear theory is much closer to the observation—another surprise considering the large initial wave amplitude here. At 45° N the phase speed predicted by (43) is highly variable and at 50° N it is out of the frame most of the time. A likely reason for these big discrepancies is that eddy momentum flux $\overline{u'v'}$ [or equivalently, c_A : see (27)] changes sign around these latitudes (Fig.2c). Since (43) may be rewritten as [with (41), (42), (20)]

$$\frac{c\overline{u'v'}}{\overline{u'v'}} = -\frac{\partial E^\dagger / \partial t}{\partial A^\dagger / \partial t}, \qquad (46)$$

c is indefinite where $\overline{u'v'}$ vanishes and likely subject to large uncertainties as it changes sign. In this case the local formula (44) likely produces more accurate result. Apart from these caveats, it is encouraging that (43) is capable of predicting the observed phase speeds of finite-amplitude Rossby waves in parallel shear flows, particularly considering that the slowly varying assumption is not necessarily satisfied in the simulation. That said, we do not necessarily advocate using (43) as a diagnostic method for c given that (45) is accurate and much easier to compute. True utility of the generalized dispersion relation lies in the fact that it can describe the effects of nonlinearity on wave dynamics as demonstrated in the next section.

4. Phase speed structure and critical lines

The dispersion relation (24) is built on several assumptions: (i) dynamics is conservative; (ii) the wave is near plane and consists of a single zonal harmonic; (iii) wave amplitude, phase speed, and the zonal-mean state all vary slowly in time and latitude compared with the phase. Given these restrictions, it is expected that the predicted phase speed becomes less accurate as one or more of these conditions are violated.

In linear wave theory, WKB assumption breaks down at turning latitudes ($l = 0$) and critical lines ($c = \bar{u} / \cos\phi$, CLs hereafter). At CL the linear wave theory itself breaks down due to singularity. The linear WKB dispersion relation for the Rossby wave on the beta plane reads [cf. (30)]

$$\bar{u}(y) - c = \frac{\bar{e}}{A} = \frac{\beta_{eff}(y)}{k^2 + l^2(y)}, \tag{47}$$

where β_{eff} is the local meridional gradient of absolute vorticity. As long as β_{eff} is nonzero, (47) implies $l^2 \to \infty$ as the wave approaches a CL ($\bar{u} - c = 0$). This implies that both the meridional phase and group velocities vanish at the CL:

$$c^y = \frac{k}{l} c \to 0, \qquad c_g = \frac{\partial(kc)}{\partial l} = \frac{2kl\beta}{(k^2 + l^2)^2} \to 0 \quad \text{as} \quad c \to \bar{u}, \tag{48}$$ [6]

i.e., an incident wave cannot reach the CL in finite time.

[6] Although (47) is formally invalid at the CL, these limits still hold.

Given that the dispersion relation (24) is based on finite-amplitude theory, can we expect a different behavior at CLs? To gain some insight on this, we shall examine the spatio-temporal structure of the CLs observed in the simulation described in section 3. Since in our simulation both $\bar{u}$ and c are changing in time and space, the locations of CLs can change. Figures 8a-8d show c_{obs} obtained from (45) (thick solid curve) and $\bar{u}/\cos\phi$ (thin solid curve) as functions of latitude for day 0, 3, 5 and 7, respectively. Crossing of the two curves defines the location of a CL. Initially the two curves are well separated ($c_{obs} < \bar{u}/\cos\phi$ everywhere) so there is no CL (Fig.8a). However, because of rapid initial transformation of the wave the profile of c_{obs} changes quickly, and by day 3 the two curves cross on both flanks of the jet (Fig.8b). In fact, there are two crossings on each flank due to a sharp falloff of c_{obs}, giving rise to two CLs sandwiching a region in which $c_{obs} > \bar{u}/\cos\phi$ (shaded, Fig.8b). For each shaded region, we call the CL facing the jet axis 'the *inner CL*' and the one facing away from it 'the *outer CL*.' The structure of c_{obs} remains relatively flat inside the jet from day 3 and 5, but in low latitudes it is highly transient: on day 5 two additional CLs appear to the south of the existing ones but they disappear by day 7 (Figs.8cd). The two persistent CLs on the southern flank of the jet slowly migrate northward, from $(13^\circ, 19^\circ\mathrm{N})$ on day 3 to $(20^\circ, 24^\circ\mathrm{N})$ on day 7, whereas the two CLs on the northern flank remain nearly stationary at $\sim(54^\circ, 63^\circ\mathrm{N})$. Although the initial formation of CLs is clearly due to the rapid change in the profile of c_{obs} rather than the modification of $\bar{u}/\cos\phi$, the magnitudes of $c_{obs} - \bar{u}/\cos\phi$ in the shaded regions remain comparable to the modification of $\bar{u}/\cos\phi$ (see Fig.3b), suggesting that for the maintenance of mature CLs the wave-mean flow interaction is important. In fact, the northward migration of the subtropical CLs is in step with the deceleration of the mean flow in this region.

Figures 8e-8h show the corresponding evolution of $A/\cos\phi$ and c_A. Initially the wave has no meridional tilt (Fig.1a) so $\overline{u'v'} = c_A = 0$ everywhere (Fig.8e), while the peak of $A/\cos\phi$ is located at $\sim 45^{\circ}$N. The latter separates into equatorward and poleward moving packets, of which the former moves faster because of a greater (negative) c_A (Figs.8f-8h, see also Fig.2a). These packets become increasingly more focused because c_A decreases toward the front end of the packets. Although both $A/\cos\phi$ and c_A tend to vanish in the far field, there remain significant wave amplitudes at the CLs, particularly the inner CLs. On day 3 c_A is also clearly finite at three of the four CLs. This is a significant departure from linear theory, which predicts that the shaded regions are inaccessible to the Rossby wave packet. On the contrary, the wave packet amplitudes are actually *greatest* in the shaded regions on day 7 (Fig.8h). It appears as though CLs are no longer singular and capable of (at least partially) transmitting the Rossby wave through them. The only thing that appears singular is the near discontinuous variation of c_{obs} across the outer CLs (Fig.8c,d).

As an attempt to build conceptual understanding of the CL behavior in the nonlinear simulation, let us utilize the generalized dispersion relation developed above with the beta-plane approximation. We also assume that the phase speed c is constant. (In doing so we are tacitly removing the amplitude-dependence of phase speed, so that wave-mean flow interaction is the sole nonlinear effect to be considered.) In this case the integral form of (36) is

$$E + cA = \overline{e} + \frac{A^2}{2} - (U_{REF} - c)A = \overline{e} - \frac{A^2}{2} - (\overline{u} - c)A = 0\,.^{7} \tag{49}$$

The above equation is defined *for each latitude* y and time t. Rearranging,

$$\overline{u} - c = \frac{\overline{e}}{A} - \frac{A}{2} = \frac{k^2 + l^2}{2kl} c_A - \frac{A}{2}, \tag{50}$$

where the second equality uses $c_A = kl\alpha^2 / A$ [(27)] and $\overline{e} = (k^2 + l^2)\alpha^2 / 2$. Equation (50) replaces (47) in linear theory. From (50) at the CL ($\overline{u} = c$) we have

$$c_A = \frac{klA}{k^2 + l^2}. \tag{51}$$

Thus the transport velocity c_A at the CL becomes a function of wave amplitude. Assuming l is finite, a wave with nonzero A will have nonzero c_A at a CL. This is consistent with the fact that both $A / \cos\phi$ and c_A are finite at many of the CLs in Fig.8. The implication is that under the dispersion relation (49), an incident wave packet can reach the CL in finite time and may be transmitted through it. Unlike linear theory, a CL does not require that l^2 diverge. In addition to the expression for c_A, (50) provides a constraint on the wave amplitude at the CL

$$\overline{e} = \frac{A^2}{2}, \tag{52}$$

which represents a local maximum of $\overline{e}$ with respect to A since [from the A-derivatives of (49)]

---

[7] Since we fixed c, (49) and (50) should be viewed as equations for A, given c, U_{REF}, and $\overline{e}$, rather than the formulae for c.

$$\frac{\partial \overline{e}}{\partial A} = \overline{u} - c = 0\,,$$

$$\frac{\partial^2 \overline{e}}{\partial A^2} = \frac{\partial}{\partial A}(\overline{u} - c) = \frac{\partial}{\partial A}(U_{REF} - A - c) = -1 < 0\,. \tag{53}$$

Note that (52) cannot be obtained from the linear (small-amplitude) theory because the right-hand side would be two orders of magnitude smaller in wave amplitude than the left-hand side.

To illustrate how the wave solution varies across CLs, Fig.9 schematically depicts the left-hand side of (49) as a function of A at a fixed time for three different latitudes y_1, y_2, y_3 in decreasing order. For simplicity, we assume that U_{REF} monotonically decreases from y_1 to y_3. Curve 1 corresponds to the highest latitude y_1. The intersections of the curve with the abscissa provide two possible roots for A but let's say the smaller root $A = A_1$ materializes here. This is close to the linear solution [ignoring $A^2/2$ in (49)] represented by the intersection of the dash-dotted line and the abscissa. Note that the horizontal coordinate of the vertex is $A = U_{REF}(y_1) - c$ so $A_1 < U_{REF}(y_1) - c$ or equivalently $c < \overline{u}(y_1)$. Thus, at this latitude the wave travels westward relative to the zonal mean flow. At the second latitude y_2 the wave amplitude is higher so the curve is shifted upward but also to the left, since $U_{REF}(y_2) < U_{REF}(y_1)$. The curve 2 touches the abscissa at the vertex, so A has a double root $A_2 = U_{REF}(y_2) - c$ or equivalently $c = \overline{u}(y_2)$. In another words $y = y_2$ is a CL. Now U_{REF} continues to decrease toward y_3 at which the left-hand side of (49) is described by curve 3. This curve again intersects the abscissa at two points (since A must be real the vertex of the curve cannot move above the abscissa), but this time the larger root $A = A_3$ is chosen so that $A_3 > U_{REF}(y_3) - c$, or $c > \overline{u}(y_3)$, namely at this latitude the wave travels *eastward* relative to the mean flow. This is a solution not possible in the linear theory. Because of finite A, c_A remains nonzero from y_1 to y_3, supporting the meridional transmission of pseudomomentum through the CL. A wave solution is possible in the region $\overline{u} - c < 0$ because of wave-mean flow interaction: finite-amplitude wave decelerates the mean flow to such an extent that the wave propagates faster than the flow. Since the CL corresponds to a double root of quadratic dispersion relation (49), it involves no mathematical singularity. The picture that wave activity may be transmitted through CL at finite amplitude fits the inner CLs (at $\sim 24^\circ$ and 53°N) in Figs.8b-d particularly well, where both A and c_A are finite (Figs.8f-h) although c_A at 53°N is very small.

If CLs are nonsingular in (49), is there any singularity anywhere? One might consider $U_{REF} - c = 0$ in (49) analogous to linear CL but it is not realizable at finite amplitude because it requires $\overline{e} + A^2/2 = 0$. In Figs.8a-8d, $U_{REF}/\cos\phi$ indicated by the dashed curve indeed stays above c_{obs} most of the time, except in the subtropics on day 3 and in low latitudes on day 5 ($c_{obs} > U_{REF}/\cos\phi$ implies that the wave is evanescent in latitude). Although (49) is nonsingular at $\overline{u} = c$ and $U_{REF} = c$, from (50) it is clear that $c = \overline{u} + A/2$ has the same mathematical characteristics as the linear CL, including a vanishing c_A:

$$c_A \to 0 \quad \text{as} \quad (\overline{u} + A/2) - c = \frac{\overline{e}}{A} \to 0. \tag{54}$$

Note that $\overline{u} < \overline{u} + A/2 < U_{REF}$. In Figs.8c and 8d c_{obs} starts to decrease sharply once it reaches halfway between $\overline{u}$ and U_{REF} at the flanks of the jet, and c_A nearly vanishes there (on day 5 this occurs at $\sim 19^{\circ}$ and 63°N; Figs.8c and 8g). It suggests the presence of singularity, and the sharp drop in c may be viewed as the wave's attempt to avoid it. It also gives rise to another CL ($c = \overline{u}$), the outer CL, at about the same latitude. Even though the CL itself is nonsingular, the sharp decrease of c makes its location practically indistinguishable from the latitude of singularity (54). Further transmission of wave activity through this CL may be limited given the small c_A there. In the meantime, the decreasing c_A towards the outer CL causes accumulation of A, which explains the large wave activities in the vicinity of CLs (Figs.8g and 8h).

Since the CL geometry discussed above departs significantly from the traditional view based on the linear theory, it is summarized schematically in Fig.10. In the classical linear theory the wave solution is possible where $c < \overline{u}$ yet the incident Rossby wave cannot reach the CL in finite time because a diverging l causes the meridional group and phase velocities to vanish at the CL (Fig.10a). In finite-amplitude theory, a nonzero A allows the group velocity to remain finite at the inner CL, so the wave activity may be transmitted through it into the region in which $c > \overline{u}$ (Fig.10b). However, the wave soon encounters singularity as c approaches halfway between $\overline{u}$ and U_{REF}, where c decreases sharply and c_A nearly vanishes. It creates the outer CL at the same location (Fig.10b), which transmits little wave activity through it.

When the wave energy at the CL is large, streamlines develop closed contours ('cat's eye') with substantial meridional width to an observer co-moving with the wave, a favorable condition for wave breaking (Haynes and McIntyre 1987). This is the case with the inner CLs at $\sim 24^\circ$ and 53° N in Figs.8cd. Particularly around 24° N, where there was little wave activity initially, a rapid wave breaking and stirring of absolute vorticity occurs (Figs.1d-1f). The well-stirred region between $\sim 20^\circ$ and 30° N is identified as the nonlinear *critical layer* (or 'surf zone', McIntyre and Palmer 1983). The inner CL is located at the center of the critical layer, whereas the outer CL is at the southern edge of the critical layer. Thus, the separation between the inner and outer CLs, d, is roughly (half) the width of the nonlinear critical layer. Figure 10 suggests that d is approximately written as

$$d \approx \frac{A_{crit}}{\left|\partial \overline{u} / \partial y\right|_{crit}} \approx \frac{\left(\overline{e}_{crit}\right)^{1/2}}{\left|\partial \overline{u} / \partial y\right|_{crit}}, \tag{55}$$

where the subscript *crit* indicates characteristic values in the critical layer, and we used (52) to derive the second expression (constant factors are dropped). Thus, more energetic wave and/or weaker horizontal shear make the critical layer wider. Killworth and McIntyre (1985) show that the upper bound of d is $\approx \beta^{-1}\left|\partial \overline{u} / \partial y\right|_{crit}$.

Once the wave reaches the outer CL, it encounters an abrupt drop in c. Since this violates the assumption of slow variation, the dispersion relation breaks down there (even if singularity is averted). This occurs relatively early in the simulation at low latitudes; this is why the agreement between theory and observation in Fig.7 is short-lived at 15° and 20° N. Rapid change in c implies rapid change in l, or equivalently, the tilt of the wave. Figure 11 shows the phase structure (longitude-latitude) of v-velocity between day 7.8 and 8.8 of the simulation. There are visually discernible discontinuities in the meridional tilt at $\sim 15^\circ$, 20°, 35°, and 62° N. These all coincide with gaps in c (not shown), and except at 35° N they coincide with the outer CLs. Since discontinuity in l means discontinuity in refractive index, it implies that the incident wave is partially reflected and partially refracted by the CLs. (Wave reflection is something that the WKB-like solution cannot describe.) Thus, in Figs.8c and 8d, the region between the two outer CLs at 20° and 62° N is shaping up to be a waveguide, hosting multiply reflected waves.

Figure 11 shows not only discontinuous tilt but also break-up and reconnection of phase lines at the outer CLs. Since the wave is traveling at distinct phase speeds across the CL, the phase lines inevitably deform and eventually break, and then reconnect with the next phase lines on the other side of the CL. The jump in phase lines is evident in 15° - 20° N (see also Fig.4 of Haynes and McIntyre 1987) and also near 35° N. The phase jump occurs rapidly and creates anomalous transient behaviors in phase speed as we observed in Fig.7 at 15° N and in Fig.6 at 20° N (phase discontinuities). The phase jump at 35° N around day 8.5 in Fig.11 is preceded by a gap in c (Fig.8d) and it causes a reversal of the meridional tilt, hence $\overline{u'v'}$ and c_A, as we have seen in Fig.2c. Apparently this midlatitude jump is caused by interference of two Rossby modes in the waveguide traveling at distinct speeds due to different meridional structures. Clearly, description of these rapid phase jump behaviors is beyond the ability of the dispersion relation [(24)].

5. Summary

We have extended the dispersion relation for the Rossby wave in barotropic shear flow to finite-amplitude regime using the exact conservation laws for pseudomomentum and pseudoenergy densities (Nakamura and Zhu 2010) and the well-known method based on the phase-averaged Lagrangian density for slowly modulated, near plane waves (Whitham 1965, Bretherton and Garrett 1968). The phase speed is expressed as a functional derivative of pseudoenergy density with respect to pseudomomentum density. Despite obvious limitations (conservative dynamics, single zonal harmonic, slow modulation, near plane waveform, etc.), writing the dispersion relation in terms of pseudomomentum and pseudoenergy densities enhances the versatility of theory since their conservation does not depend on specific wave geometry or amplitude, and the effects of wave-mean flow interaction and the amplitude dependence of the phase speed may be incorporated seamlessly. Furthermore, since pseudomomentum and pseudoenergy densities are readily evaluable, the theory is testable with numerical simulations.

In addition to the dispersion relation, we have developed a method to estimate the phase speed of the wave directly from instantaneous data [(45)]. Data sampled from the numerical simulation of nonlinear barotropic decay on a sphere (HP87) with an initial zonal wavenumber 6 demonstrate that the finite-amplitude theory reproduces the observed phase speed better than the linear theory as long as one is away from regions in which meridional eddy momentum flux changes signs and/or the phase speed is discontinuous. It is also found that the linear theory does not necessarily perform better at smaller amplitude when the wave amplitude is growing quickly. Consistent with previous studies (e.g., Randel and Held 1991) CLs are generally located on the flanks of the jet, but their geometry differs significantly from the standard linear theory: there are multiple CLs on each side of the jet axis. It is shown using the generalized dispersion relation that nonlinearity removes singularity from CLs and the meridional group velocity at the CL remains finite when the wave amplitude is finite. As a result, the Rossby wave may be transmitted through the CLs even into the region where $c > \bar{u}$.

In future studies we will extend the analysis to atmospheric data and examine the CL geometry in the upper troposphere and its relationship to wave breaking with the aid of the generalized dispersion relation.

Acknowledgments

This work has been supported by NSF Grant AGS-1151790.

Appendix A Derivation of (39) from (37) and (2)

With the plane waveform (37) and the beta-plane approximation, absolute vorticity q is

$$\begin{aligned} q(\theta, y) &= f_0 + \beta y + q'(\theta), \\ q'(\theta) &= -(k^2 + l^2) B \cos\theta, \qquad \theta = k(x - ct) + l\,y. \end{aligned} \tag{A1}$$

Pseudomomentum density A is defined by the beta-plane version of (2)

$$\begin{aligned} A(y) &= \frac{1}{2\pi}\left(\iint_{q \geq Q} q(\theta, y')\, dy' d\theta - \iint_{y' \geq y} q(\theta, y')\, dy'\, d\theta \right) \\ &= -\overline{\int_y^{y+\xi(\theta,y)} q\, dy'} = -\overline{\left[q(\theta, y)\xi(\theta, y) + \frac{1}{2}\beta \xi^2(\theta, y) \right]}, \end{aligned} \tag{A2}$$

where overbar denotes phase average (= zonal average) and $\xi(\theta, y)$ is the meridional displacement of the contour $q = Q$ relative to y along phase line (fixed θ). Because the meridional gradient of q along the phase line is constant (= β),

$$q(\theta, y) = Q(y) - \beta \xi(\theta, y). \tag{A3}$$

Requiring $\overline{\xi} = 0$ (displacement is area-preserving), one sees from (A3)

$$Q(y) = f_0 + \beta y, \qquad \xi(\theta) = -q'(\theta)/\beta. \tag{A4}$$

Substituting in (A2)

$$A = -\overline{\left(-\frac{q'^2}{\beta} + \frac{1}{2}\frac{q'^2}{\beta} \right)} = \frac{1}{2}\frac{\overline{q'^2}}{\beta} = \frac{(k^2 + l^2)^2 B^2}{4\beta}\ , \tag{A5}$$

which is (39). Notice that, although (A5) is identical with what the linear theory predicts, the wave amplitude B here is not assumed small.

Appendix B Calculation of *A*

One may calculate A by evaluating the area integrals in (2) with weighted box counting method. However, this method is prone to errors when the wave amplitude is small, since in that case A becomes a small difference between two large integrals. Here we use an alternative method to calculate A.

By taking the derivative of (2) with respect to μ twice, one obtains

$$\frac{\partial^2}{\partial \mu^2}\left(A\cos\phi\right) = a\frac{\partial}{\partial \mu}\left(\overline{q}(\mu,t) - Q(\mu,t)\right), \tag{B1}$$

where $Q(\mu,t)$ is absolute vorticity in equivalent latitude [Nakamura and Zhu 2010 Eq. (19)]. From the definition of zonal-mean absolute vorticity

$$\frac{\partial^2}{\partial \mu^2}\left(\overline{u}\cos\phi\right) = a\left(2\Omega - \frac{\partial \overline{q}}{\partial \mu}\right). \tag{B2}$$

Adding (B1) and (B2) then using (3b)

$$\frac{\partial^2}{\partial \mu^2}\left(U_{REF}\cos\phi\right) = a\left(2\Omega - \frac{\partial Q}{\partial \mu}\right). \tag{B3}$$

We first calculate $Q(\mu,t)$ by inverting the area-absolute vorticity relation

$$S(Q) = 2\pi a^2(1-\mu), \tag{B4}$$

where $S(Q)$ is the area of domain in which $q \geq Q$. This is evaluated with equally spaced 1024 bins of Q between $Q_{\max}$ and $Q_{\min}$, the maximum and minimum values of absolute vorticity. $S(Q)$ is then inverted for $Q(\mu)$ on equally spaced μ (1024 latitudes), using linear interpolation. Then we calculate the gradient of Q with a finite difference method and evaluate the right hand side of (B3) on each μ. Finally we solve for $U_{REF}\cos\phi$ using a tridiagonal solver with the boundary conditions

$U_{REF}\cos\phi = 0$ at $\mu = \pm 1$.

Once $U_{REF}\cos\phi$ is obtained, we interpolate the result onto Gaussian latitudes on which the model's variables are evaluated, and subtract $\overline{u}\cos\phi$ to obtain $A\cos\phi$.

References

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.

—— and ——, 1978a: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.*, **89,** 609-646.

—— and ——, 1978b: Wave-action and its relatives. *J. Fluid Mech.*, **89,** 647-664.

Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics.* Academic Press, 489pp pp.

Boyd, J. P., 1976: The noninteraction of waves with the zonally averaged flow on a spherical earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci.*, **33,** 2285-2291.

Bretherton, F. P., and C. J. R. Garrett, 1968: Wavetrains in inhomogeneous moving media. *Proc. Roy. Soc. London Ser. A*, **302,** 529-554.

Buchwald, V. T., 1972: Energy and energy flux in planetary waves. *Proc. Roy. Soc. London Ser. A*, **328,** 37-48.

Bühler, O., 2009: *Wave and mean flows.* Cambridge Univerty Press, 341pp.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66,** 83-110.

Dickinson, R. E., 1969: Theory of planetary wave-zonal flow interaction. *J. Atmos. Sci.*, **26,** 73-81.

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

Grimshaw, R., 1984: Wave action and wave-mean flow interaction, with application to stratified shear flows. *Ann. Rev. Fluid Mech.*, **16,** 11-44.

Haurwitz, B., 1940: The motion of atmospheric disturbances on the spherical earth. *J. Mar. Res.*, **3,** 254-267.

Hayes, W. D., 1970: Conservation of action and modal wave action. *Proc. Roy. Soc. London Ser. A*, **320,** 187-208.

Haynes, P. H., 1988: Forced, dissipative generalizations of finite-amplitude wave-activity conservation relations for zonal and nonzonal basic flows. *J. Atmos. Sci.*, **45,** 2352-2362.

Haynes, P. H., and M. E. McIntyre, 1987: On the representation of rossby-wave critical layers and wave breaking in zonally truncated models. *J. Atmos. Sci.*, **44,** 2359-2382.

Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: theory. *Large-scale dynamical processes in the atmopshere*, B. J. Hoskins, and R. P. Pearce, Eds., Academic Press, 127-167.

——, 1985: Pseudomomentum and the orthogonality of modes in shear flows. *J. Atmos. Sci.*, **42,** 2280-2288.

Held, I. M., and P. J. Phillips, 1987: Linear and nonlinear barotropic decay on the sphere. *J. Atmos. Sci.*, **44,** 200-207.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38,** 1179-1196.

Killworth, P. D., and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect, or over-reflect? *J. Fluid Mech.*, **161,** 449-492.

Longuet_Higgins, M. S., 1964: Planetary waves on a rotating sphere. *Proc. Roy. Soc. London Ser. A*, **279,** 446-473.

——, 1965: Planetary waves on a rotating sphere. II. *Proc. Roy. Soc. London Ser. A*, **284,** 40-68.

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

McIntyre, M. E., and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305,** 593-600.

McIntyre, M. E., and T. G. Shepherd, 1987: An exact local conservation theorem for finite-amplitude disturbances to nonparallel shear flows, with remarks on Hamiltonian-structure and on Arnold stability theorems. *J. Fluid Mech.*, **181,** 527-565.

Nakamura, N., and A. Solomon, 2010: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part I: Quasigeostrophic theory and analysis. *J. Atmos. Sci.*, **67,** 3967-3983.

Nakamura, N., and D. Zhu, 2010: Finite-amplitude wave activity and diffusive flux of potential vorticity in eddy-mean flow interaction. *J. Atmos. Sci.*, **67,** 2701-2716.

Platzman, G. W., 1968: The Rossby wave. *Quart. J. Roy. Meteor. Soc.*, **94,** 225-248.

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.

Rossby, C. G., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semipermanent centers of action. *J. Mar. Res.*, **2,** 38-55.

Salmon, R., 2013: An alternative view of generalized Lagrangian mean theory. *J. Fluid Mech.*, **719,** 165-182.

Seliger, R. L., and G. B. Whitham, 1968: Variational principles in continuum mechanics. *Proc. Roy. Soc. London Ser. A*, **305,** 1-25.

Solomon, A., and N. Nakamura, 2012: An exact Lagrangian-mean wave activity for finite-amplitude disturbances to barotropic flow on the sphere. *J. Fluid Mech.*, **693,** 69-92.

Sturrock, P. A., 1961: Energy-momentum tensor for plane waves. *Phys. Rev.*, **121,** 18-19.

Whitham, G. B., 1965: A general approach to linear and non-linear dispersive waves using a Lagrangian. *J. Fluid Mech.*, **22,** 273-283.

——, 1970: Two-timing, variational principles and waves. *J. Fluid Mech.*, **44,** 373-395.

——, 1974: *Linear and nonlinear waves*. New York: Wiley-Interscience. 656pp.

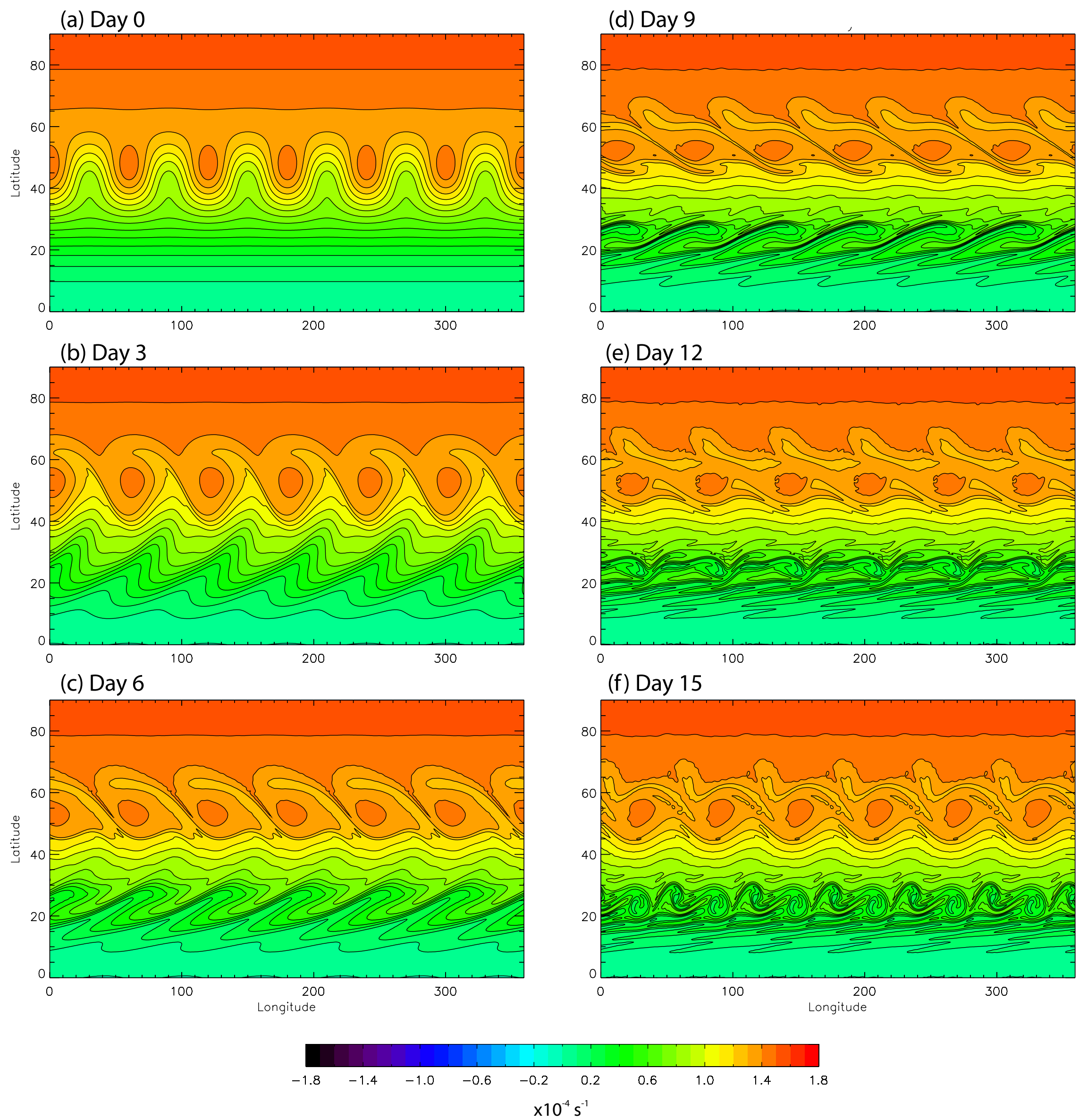

Figure 1 Evolution of absolute vorticity in the Northern Hemisphere during nonlinear barotropic decay described in section 3. Abscissa is longitude and ordinate is latitude. (a) intial condition, (b) day 3, (c) day 6, (d) day 9, (e) day 12, (f) day 15.

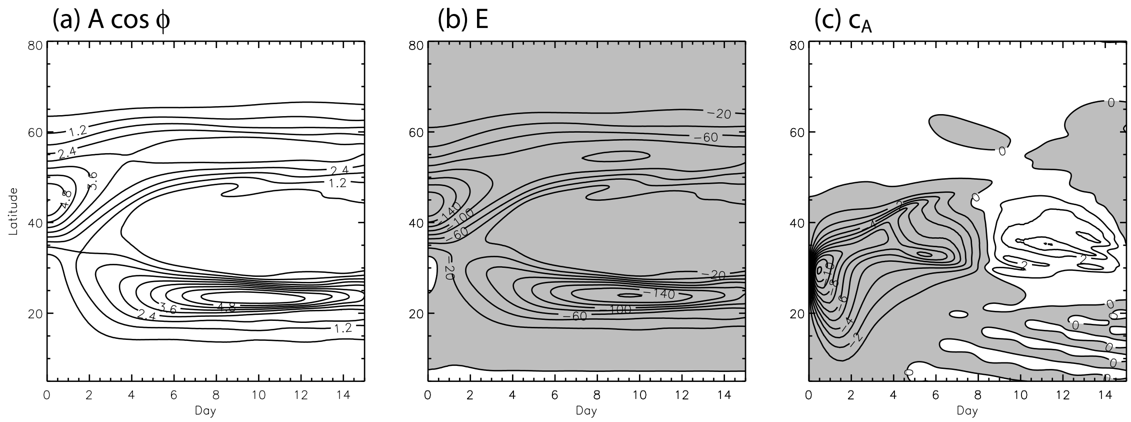

Figure 2 Wave properties as functions of time and latitude during the same numerical simulation as in Fig.1. (a) Angular pseudomomentum density $A\cos\phi$, contour interval = 0.6 ms^{-1} (b) pseudoenergy density E, contour interval = 20 m^2s^{-2} (c) effective transport velocity c_A (= group velocity in small-amplitude limit), contour interval = 1 ms^{-1}. In (b) and (c), negative values are shaded. See text for the definitions of variables.

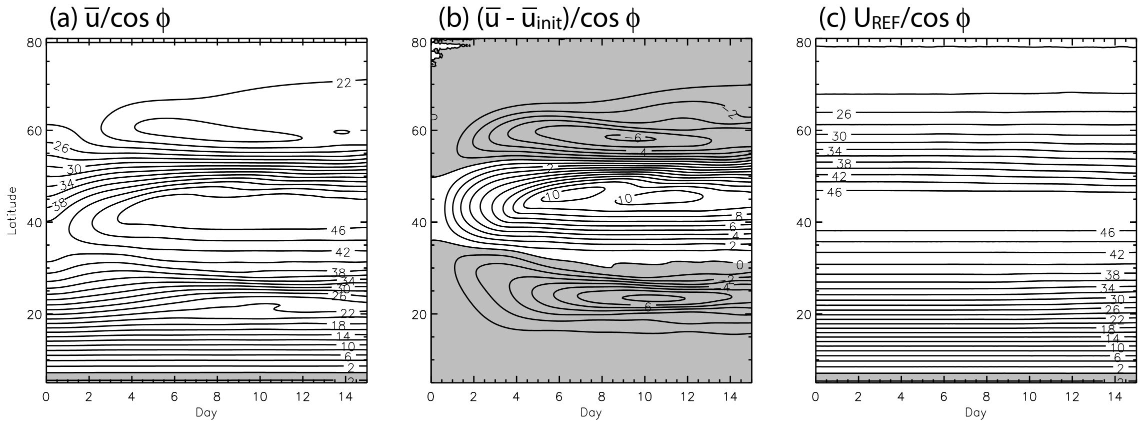

Figure 3 Same as Fig.2 but for the mean flow properties. (a) Zonal-mean angular velocity $\bar{u}/\cos\phi$, contour interval = 2 ms^{-1} (b) $\bar{u}/\cos\phi$ minus its initial value, contour interval = 1 ms^{-1} (c) Angular velocity of the reference state flow $U_{REF}/\cos\phi$, contour interval = 2 ms^{-1}. Negative values are shaded.

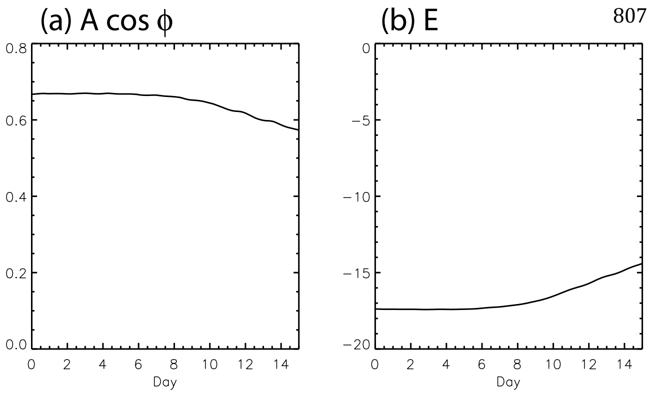

Figure 4 (a) Domain-averaged $A\cos\phi$ as a function of time during the above simulation. Unit: ms^{-1}. (b) Same as (a) but for domain-averaged E. Unit: m^2s^{-2}.

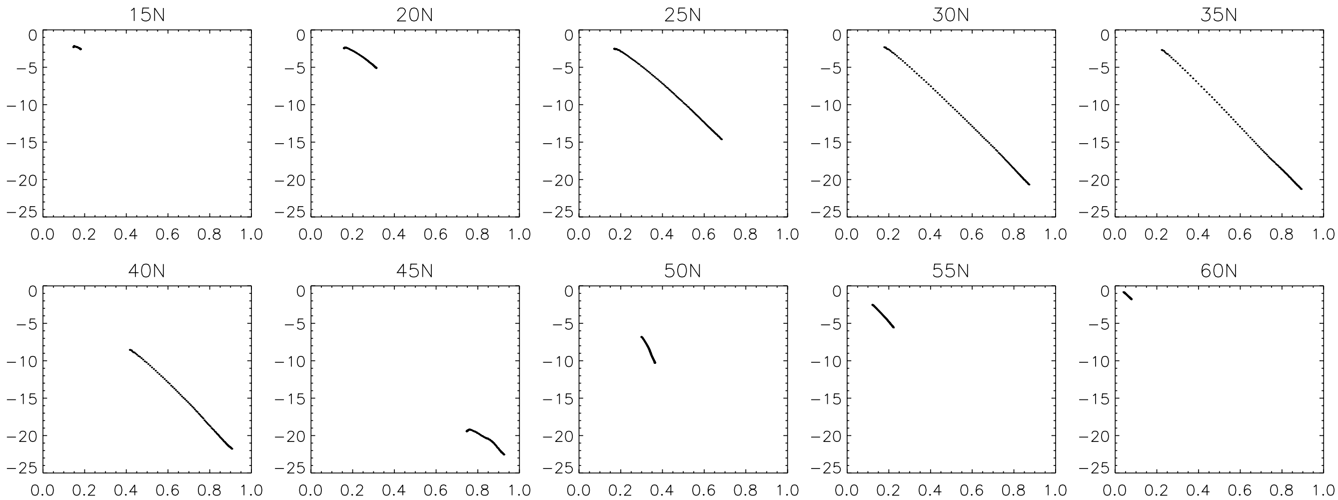

Figure 5 Scatter plots of $A^{\dagger}$ (abscissa, ms^{-1}) versus $E^{\dagger}$ (ordinate, m^2s^{-2}) at 10 different latitudes, each constructed from the first 7 days of simulation. Each frame show 101 data points, sampled every 6000 s. The wave amplitude is small in the upper left and large in the lower right. For 50°, 55°, and 60° N, $A^{\dagger}$ and $E^{\dagger}$ were evaluated by southward integration from the North Pole. See Eq. (43) and text for details.

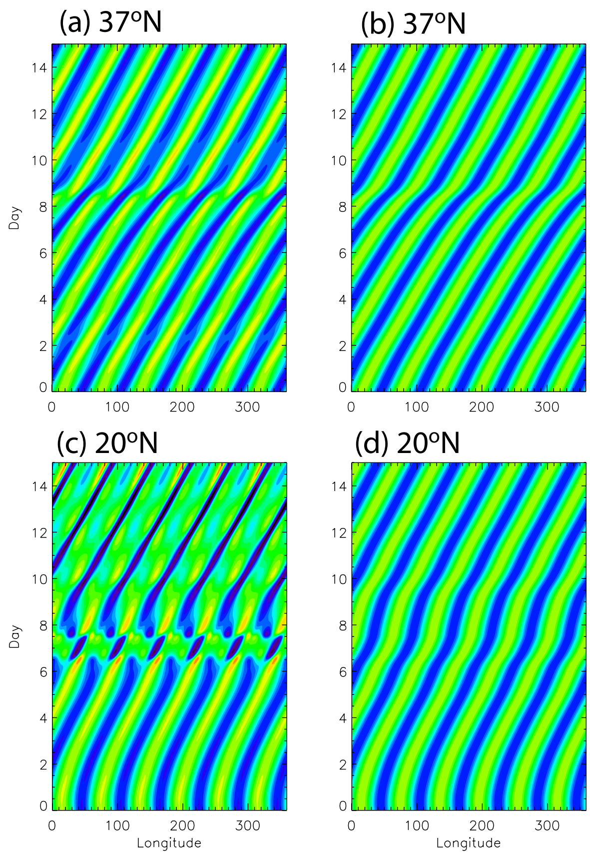

Figure 6 Left: longitude-time plots (Hovmöller diagrams) of the rms-normalized v velocity for the simulation, at 37° N (top) and 20° N (bottom). Right: Reconstruction of phase by integrating the 'observed' phase speed calculated from Eq. (45). See text for details.

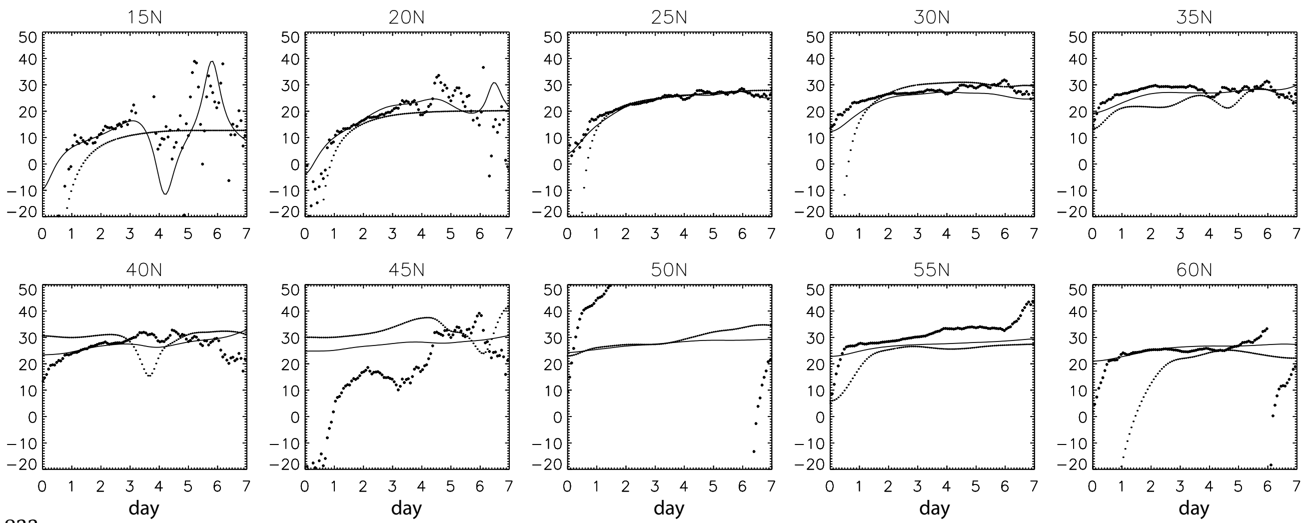

Figure 7 Phase speed of the Rossby wave c during the first 7 days of simulation at 10 different latitudes. Thin solid curve: 'observed' phase speed based on Eq. (45). Thick dots: theoretical estimates based on (43) and the $A^{\dagger}$-$E^{\dagger}$ relations in Fig.5. Thin dots: theoretical estimates based on linear theory [(30) and (29)]. The unit of the ordinate is ms^{-1}. See text for details.

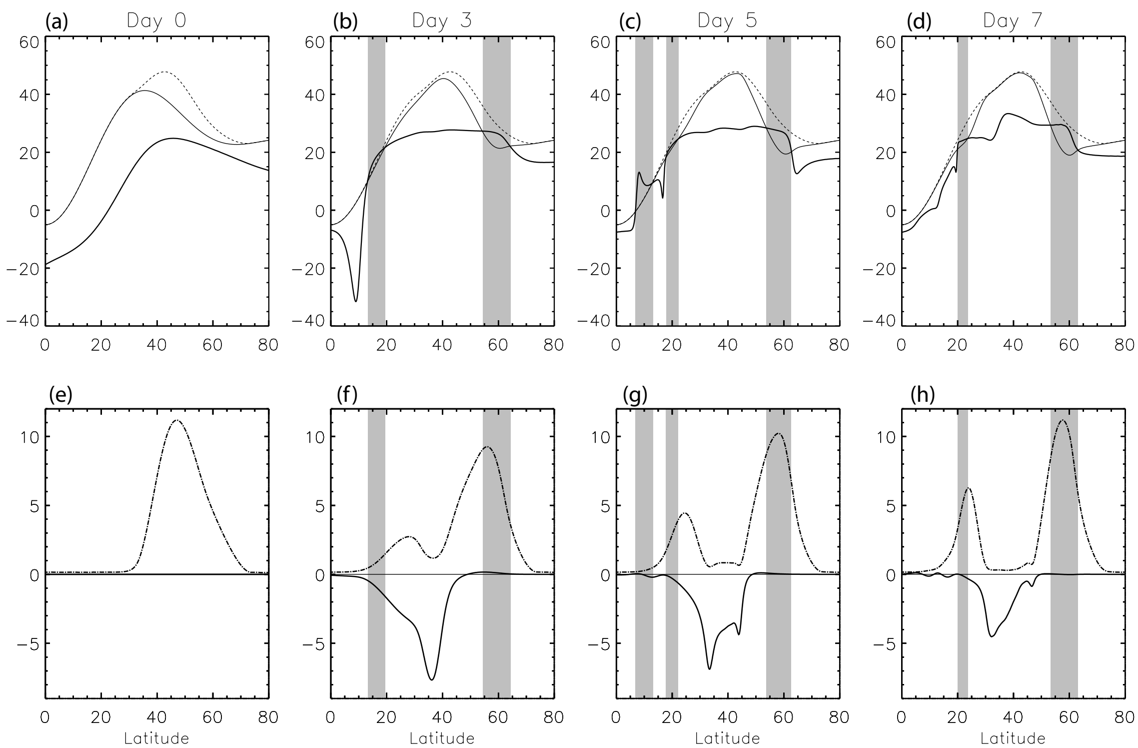

Figure 8 Top row: c_{obs} based on (45) (thick solid curve) and $\bar{u}/\cos\phi$ (thin solid curve) and $U_{REF}/\cos\phi$ (dashed curve) as functions of latitude at four different instants during the numerical simulation of nonlinear barotropic decay. (a) Day 0 (b) Day 3 (c) Day 5 (d) Day 7. In the shaded region $c_{obs} \geq \bar{u}/\cos\phi$: the boundaries of these regions mark the critical lines. Bottom row: same as the top row but for c_A based on (27) (thick solid curve) and $A/\cos\phi$ (dot-dashed curve). Zero line is also added. (e) Day 0 (f) Day 3 (g) Day 5 (h) Day 7. The unit of the ordinate is ms^{-1} in all plots.

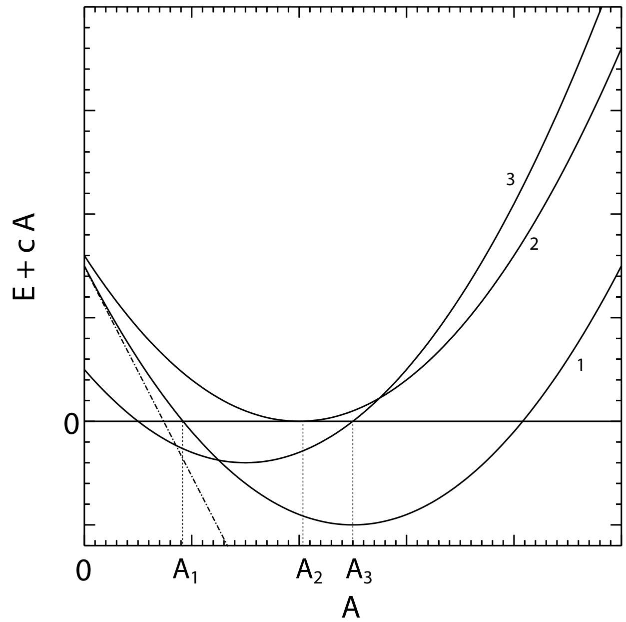

Figure 9 $E+cA$ as quadratic functions of A. Curves 1-3 correspond to three different latitudes, y_1, y_2, y_3. The dash-dotted line in indicates linear dispersion relation at y_1: it is tangent to curve 1 at $A=0$. The intercept of each curve with the ordinate equals $\overline{e}$ at that latitude and the horizontal coordinate of the vertex of each curve is $U_{REF}-c$. See text for details.

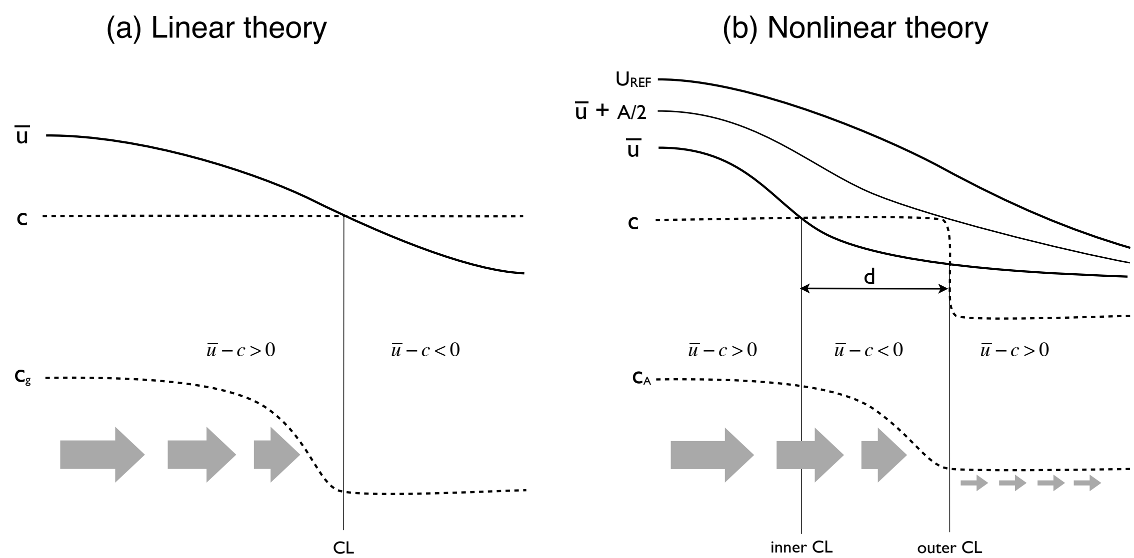

Figure 10 Schematics of mature critical line geometry (horizontal axis is latitude). Critical lines are defined by the intersections of the $\bar{u}(y)$ and $c(y)$ curves. (a) Classical linear theory. Phase speed c is constant across latitude whereas the meridional group velocity (indicated by gray arrows) drops to zero at the critical line. (b) Nonlinear theory. The inner critical line is nonsingular and transmits the wave to the right. c is constant up to $c \approx \bar{u} + A/2$, where it drops abruptly. c_A also nearly vanishes here. The outer critical line coincides with the latitude of sharp falloff of c. The distance between the two critical lines is denoted by d. See text for details.

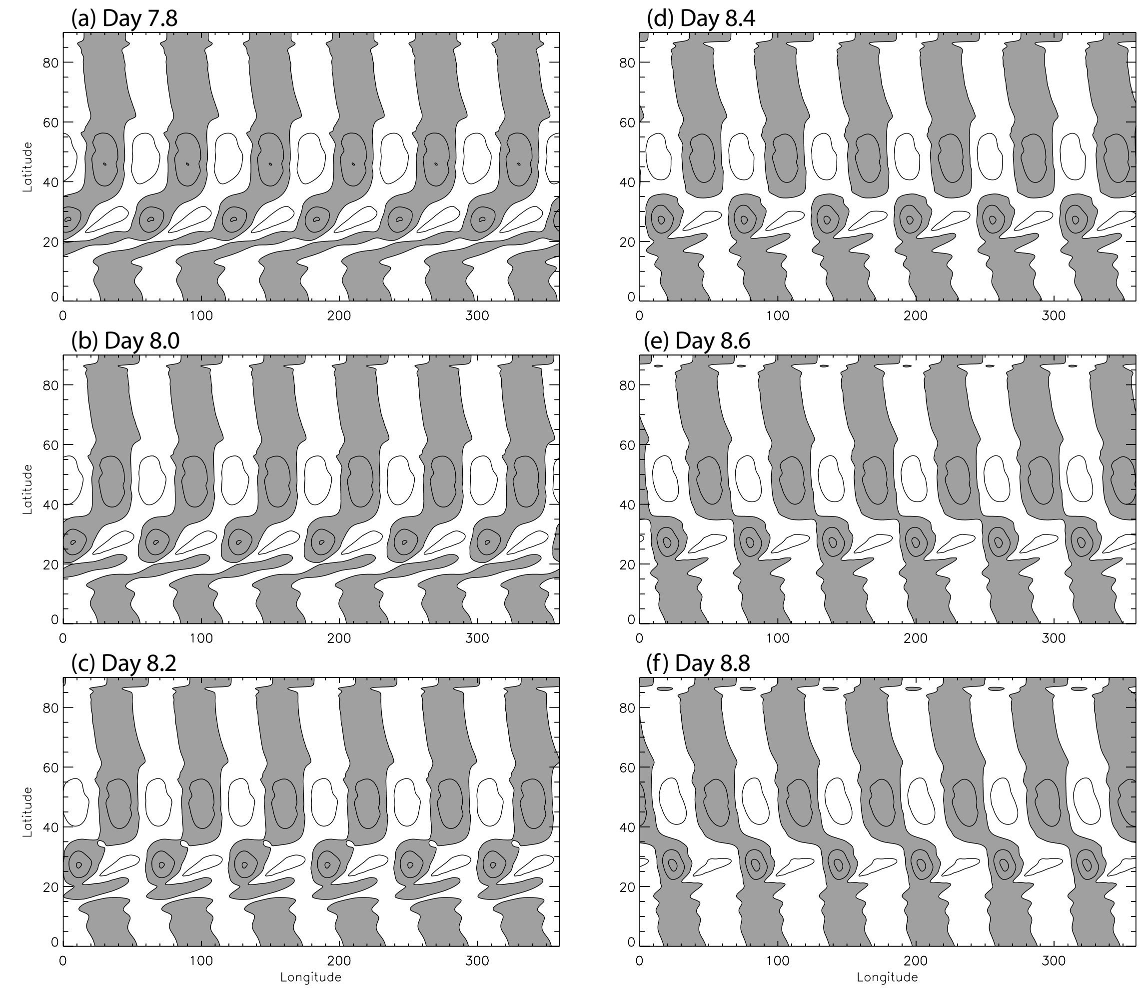

Figure 11 Reconfiguration of phase lines in v-velocity from day 7.8 to day 8.8. Notice the breakup and reconnection of phase lines around 15° - 20° N and 35° N. Contour interval is 1 ms^{-1} and negative values are shaded.