

The Evolution of a Stratospheric Wave Packet

NILI HARNIK

Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 18 August 2000, in final form 11 July 2001)

ABSTRACT

This work examines the extent to which a few basic concepts that apply to plane waves, for example, the refraction of waves up the gradient of the index of refraction, apply to stratospheric planetary waves. This is done by studying the relation between group velocity (C_g) and the *wave activity velocity*, which is defined as the Eliassen–Palm flux divided by the wave activity density ($\mathbf{V}_a = F/A$). It is shown that although in the limit of plane waves \mathbf{V}_a equals C_g , the two velocities are not equal for stratospheric waves, because of reflection, tunneling, and superposition. The use of conservation of wave activity to understand the spatial variations of wave structure is explored. This is done by defining a wave activity packet as part of the wave that moves with \mathbf{V}_a . Integral lines of \mathbf{V}_a are then used to keep track of the wave packet location and volume. In the idealized case of an almost-plane wave, conservation of wave activity leads to variations in the amplitude of the wave when it is refracted by the slowly varying basic state. This effect is related to changes in wave packet volume. The wave activity packet framework is used to examine the importance of the “volume effect” for explaining the spatial variations of stratospheric waves.

The wave packet formulation is also used to study the evolution of a wave propagating from the troposphere to the stratosphere. It is shown that the consequence of the polar night jet being a leaky waveguide is that perturbations initially concentrate up into the waveguide and only later leak out to the equatorial region. This can explain the observed stratospheric wave life cycle of baroclinic growth followed by a barotropic stage. Finally, integral lines of \mathbf{V}_a are used to estimate vertical propagation timescales of an observed wave, and it is shown that this estimate is consistent with linear wave dynamics.

1. Motivation

Midlatitude stratospheric planetary waves are generally considered to be vertically propagating Rossby waves (Charney and Drazin 1961), and one of the basic questions is how their structure depends on the basic state and damping fields. Linear wave propagation theory and wave activity conservation have both been used to address this question. Karoly and Hoskins (1982, referred to as KH) used linear wave propagation theory, in particular, ray tracing, to show that plane waves on a slowly varying medium (in the WKB sense) refract up the gradient of the index of refraction. Refraction can cause the group velocity to converge or diverge, which will result in spatial variations of wave amplitude, due to conservation of wave activity (Lighthill 1978). Karoly and Hoskins (1982) used this to obtain an analytic expression for the variation of wave amplitude along a wave ray. The main limitation of KH’s result is how to apply it to realistic cases, which are not plane waves.

The use of wave activity conservation, on the other

hand, is not limited to almost-plane waves. A convergence or divergence of a wave activity velocity, defined as the flux of wave activity divided by the wave activity density (denoted by \mathbf{V}_a), is accompanied by changes in wave amplitude. This has been used to understand variations in wave amplitude, both in very idealized cases of one-dimensional wave propagation (e.g., Lighthill 1978; Swanson et al. 1997) and in more realistic cases of horizontally propagating tropospheric waves (e.g., Hoskins and Karoly 1981; Sobel and Bretherton 1999, and references therein). The question is how does \mathbf{V}_a relate to the basic state. In the limit of plane waves on a slowly varying medium, Edmon et al. (1980) showed that \mathbf{V}_a reduces to the group velocity (denoted by C_g). Given this limit, it is tempting to extend KH’s results to stratospheric waves in general. Indeed, there is a general assumption in the literature that Eliassen–Palm (EP) fluxes should refract up the gradient of index of refraction, and \mathbf{V}_a is often referred to as the “group velocity.” There are a few reasons, however, why this may be a problem. First, \mathbf{V}_a and C_g are equal only for an almost-plane wave, and not when we have a superposition of plane waves. In the stratosphere, we generally have a superposition of northward- and southward-propagating waves, and sometimes, of upward-propagating and downward-reflected waves (Harnik and Lindzen 2001,

Corresponding author address: Nili Harnik, Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10964.
E-mail: nili@ldeo.columbia.edu

hereafter HL). Second, damping affects the direction of EP fluxes whereas ray tracing does not take into account the effects of damping on wave propagation. For example, it is often observed that the EP flux is directed equatorward at middle latitudes, where the local index of refraction increases poleward—an indication in this case of leakage from the stratospheric waveguide to the tropical critical surface.

One of the goals of this study is to understand the relation between \mathbf{V}_a and \mathbf{C}_g , and to better understand how \mathbf{V}_a is affected by the index of refraction. The condition that the wave activity flux \mathbf{F} reduce to the product of group velocity and wave activity density in the limit of almost-plane waves propagating on a slowly varying medium is considered one of the advantages of formulations of wave activity conservation. It is in fact used as a justification for the specific choice of \mathbf{F} , which is not defined uniquely from wave activity conservation (e.g., Edmon et al. 1980; Plumb 1986, and references therein¹). The applicability of this limit to realistic waves has not been explicitly tested before. While most studies mention that this limit might be violated because WKB conditions do not strictly hold, there is little discussion of the violation of the plane wave limit. This is important since violation of WKB conditions may still allow \mathbf{V}_a to behave qualitatively like a group velocity,² but violation of the plane wave limit causes \mathbf{V}_a to behave qualitatively different from \mathbf{C}_g (even where WKB conditions hold).

Another goal is to determine whether the contribution of spatial variations of \mathbf{V}_a to changes in wave amplitude are at all important for stratospheric waves, and what processes or features of the basic state cause the spatial variations in \mathbf{V}_a . We also compare with nonconservative processes that affect wave activity, like damping and time variations in wave source. Keeping track of the variations of wave activity requires defining a “wave packet” that carries it, in analogy to fluid particles. We do this by calculating integral lines of \mathbf{V}_a , and using them to define a coordinate system in which grid boxes are the wave packets. The contribution of convergence or divergence of \mathbf{V}_a to wave activity density changes is very intuitive in this framework because it is directly related to changes in wave packet volume. We will discuss how the basic processes of refraction, tunneling, reflection, and superposition affect \mathbf{V}_a and its divergence.

Finally, one of the most direct links between the troposphere and the stratosphere is through the vertical

propagation of Rossby waves. The correlation between the two is strongly affected by how the waves propagate, and how long it takes them. Existing methods that estimate vertical propagation timescales are either statistical, for example, time lag correlations (Randel et al. 1987; Randel 1987), or are based on highly idealized calculations, like ray tracing (KH), for which the relation to an actual wave event is ambiguous. Tracking wave packets using our method allows us to estimate the vertical propagation time for a given wave event.

We will start by presenting the basic wave activity formulation (section 2). We then define our wave packets using a new coordinate system, and study the effect of \mathbf{V}_a convergence on wave structure in a steady state (section 3) and a time-dependent (section 4) wave field. In section 5 we discuss the relation between our diagnostic and the ray tracing technique, applied to the stratosphere by KH. In section 6 we use \mathbf{V}_a to estimate vertical propagation timescales from observations. We summarize and discuss our results in section 7.

2. Formulation—A definition of wave packets based on the wave activity velocity

In the following section we present the mathematical formulation of our diagnostic. The derivation is based on the conservation of wave activity. We use the linear quasigeostrophic (QG) formulation, but much of what follows can be done for more general forms of wave activity provided they obey a conservation law. We start from the conservation equation. Our notation follows Andrews et al. (1987), to which the reader is referred for a derivation

$$\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{F} = D, \quad (1)$$

where \mathbf{F} is the EP flux, A is a wave activity density, which is defined as follows on a β plane:

$$A \equiv \frac{\overline{\rho q'^2}}{2\overline{q}_y}, \quad (2)$$

where \overline{q}_y is the zonal mean potential vorticity (PV) gradient, q' is the PV perturbation, ρ is density, D is damping of wave activity, and an overbar denotes zonal averaging. Note that $\overline{q'^2}$ is related to the wave geopotential height variance ($\overline{\varphi'^2}$), hence variations in A imply variations in the wave amplitude of geopotential height. The relation between A and $\overline{\varphi'^2}$ is not in generally straightforward, but a sense of it can be obtained by looking at the simple case of a stationary, inviscid wave. From PV conservation (see, e.g., Andrews et al. 1987), we get $q' \approx -(\overline{q}_y \varphi' / \overline{U})$, which implies $A \approx \rho \overline{q}_y \overline{\varphi'^2} / \overline{U}^2$. In the middle of the stratospheric waveguide, where \overline{q}_y and the zonal mean wind do not vary significantly,³

¹ It should be noted, however, that the condition that \mathbf{F} be parallel to \mathbf{C}_g in the plane wave limit does not uniquely define the flux of wave activity in more generalized formulations when the basic state is nonzonal (Plumb 1985).

² Group velocity does not strictly exist when WKB is violated; however, if WKB is only weakly violated, for example, because the scale over which the basic state varies is slightly smaller than the wavelength, we can still calculate a group velocity assuming that WKB does hold.

³ The spatial pattern of \overline{q}_y is qualitatively like the index of refraction squared (dashed lines in Fig. 6). The \overline{q}_y field for this model run is shown in Fig. 1b of HL.

variations in A are roughly proportional to variations in $\rho|\varphi|^2$, but this is not necessarily the case in regions where \bar{q}_y and \bar{U} vary significantly.

We define a *wave activity velocity*, which we denote by \mathbf{V}_a , as follows (e.g., Palmer 1982):

$$\mathbf{V}_a \equiv \frac{\mathbf{F}}{A}, \quad (3)$$

where \mathbf{V}_a is the velocity at which wave activity density propagates along the wave field. Plugging Eq. (3) into Eq. (1) gives an equation for the variation of wave activity density following \mathbf{V}_a :

$$\frac{\partial A}{\partial t} + \nabla \cdot (\mathbf{V}_a \cdot A) = \frac{\partial A}{\partial t} + \mathbf{V}_a \cdot \nabla(A) + A \nabla \cdot \mathbf{V}_a = D, \quad (4)$$

$$\frac{dA}{dt} = D - A \nabla \cdot \mathbf{V}_a, \quad (5)$$

where the material derivative is defined following the wave activity velocity ($d/dt \equiv \partial/\partial t + \mathbf{V}_a \cdot \nabla$). We see that following \mathbf{V}_a , wave activity density increases when lines of \mathbf{V}_a converge, and vice versa.

It is convenient to define a coordinate system that follows integral lines of \mathbf{V}_a , and define wave packets as grid boxes in this coordinate. This way a wave packet is a part of the wave that moves with \mathbf{V}_a , in analogy to material fluid parcels moving with the flow. Assuming the wave packets initially have a unit volume, the Jacobian of the transformation (J) equals the volume of the wave packets as they move along. It then follows that the fractional change of J following the wave packet equals the divergence of \mathbf{V}_a [see Aris (1962), or any other appropriate fluid dynamics text book for a derivation]:

$$\frac{1}{J} \frac{dJ}{dt} = \nabla \cdot \mathbf{V}_a. \quad (6)$$

Plugging in Eq. (4), we get the following conservation equation:

$$\frac{d}{dt}(AJ) = DJ. \quad (7)$$

This formulation allows a very intuitive description of the evolution of wave activity: AJ is the total amount of wave activity in the packet (A is wave activity density and J is the volume of the packet) and DJ is the volume-integrated damping of wave activity. Here AJ is conserved as it moves along \mathbf{V}_a , unless there is dissipation. The wave activity density does change, however, because the volume of the wave packet changes due to divergence of \mathbf{V}_a . We are interested in understanding what processes affect \mathbf{V}_a , because they affect A , and hence wave amplitude.

In the case of an almost-plane Rossby wave, \mathbf{V}_a equals the group velocity (Edmon et al. 1980). The condition that \mathbf{F} reduce to the product of wave activity density

and group velocity in the limit of a plane wave is in fact necessary to uniquely define it (Edmon et al. 1980). The need for an additional condition stems from the fact that any nondivergent vector \mathbf{G} can be added to \mathbf{F} , for wave activity conservation [Eq. (1)] to hold.

In general, we do not have an almost-plane wave, rather, we have a superposition of almost-plane waves, because we have reflection. This has a large effect on the relation between \mathbf{V}_a and group velocity (\mathbf{C}_g). As is shown in appendix A, in the case of an almost-plane Rossby wave, $\mathbf{V}_a = \mathbf{C}_g$. When we have reflection the wave solution is composed of a superposition of almost-plane waves propagating in different directions. In this case \mathbf{V}_a equals the superposition of the group velocities of the different components. For example, a superposition of poleward and equatorward propagating waves, $\phi = (\alpha_+ e^{i \int l dy} + \alpha_- e^{-i \int l dy}) e^{i(kx - \omega t + m dz)}$, under WKB conditions will result in

$$\mathbf{V}_a = \frac{\mathbf{C}_{g+} |\alpha_+|^2 + \mathbf{C}_{g-} |\alpha_-|^2}{\left| \alpha_+ \exp\left(i \int l dy\right) + \alpha_- \exp\left(-i \int l dy\right) \right|^2}. \quad (8)$$

The meridional component of \mathbf{V}_a can vary considerably, depending on the relative magnitudes of the wave components, and in the case of a standing wave (i.e., $\alpha_+ = \alpha_-$) it will vanish, since $C_{g_{y+}} = -C_{g_{y-}}$ (the subscript y denotes the meridional component). Nonzero \mathbf{V}_a implies a net propagation of wave activity, which, in steady state requires the existence of a wave activity sink.

3. The evolution of wave packets and the volume effect in a simple model

In this section we examine how conservation of wave activity directly affects the structure of stratospheric waves in a simple linear QG β -plane model of the stratosphere. We restrict our discussion to waves of a single wavenumber. We force the model at the tropopause with a stationary wave 1 perturbation. The height in all model runs is nondimensionalized by the density scale height, which is 7 km.

Figure 1 shows the basic-state wind, the sponge layer, and the steady-state wave 1 geopotential height field along with the shape of the forcing at the bottom. In all figures, the pole is on the left-hand side (small y). The basic state is characteristic of an early winter Southern Hemisphere stratospheric jet. The jet tilts equatorward and widens with height. The basic-state Brunt-Väisälä frequency varies only with height in our model (not shown), and is specified to look like a standard mid-latitude winter profile. The PV gradient (not shown) and index of refraction (see Fig. 6) have similar features, with a ridge that follows the jet and negative regions on both sides of the jet. We have a sponge layer in the form of linear damping on temperature and momentum at the top and at low latitudes, to approximate, respec-

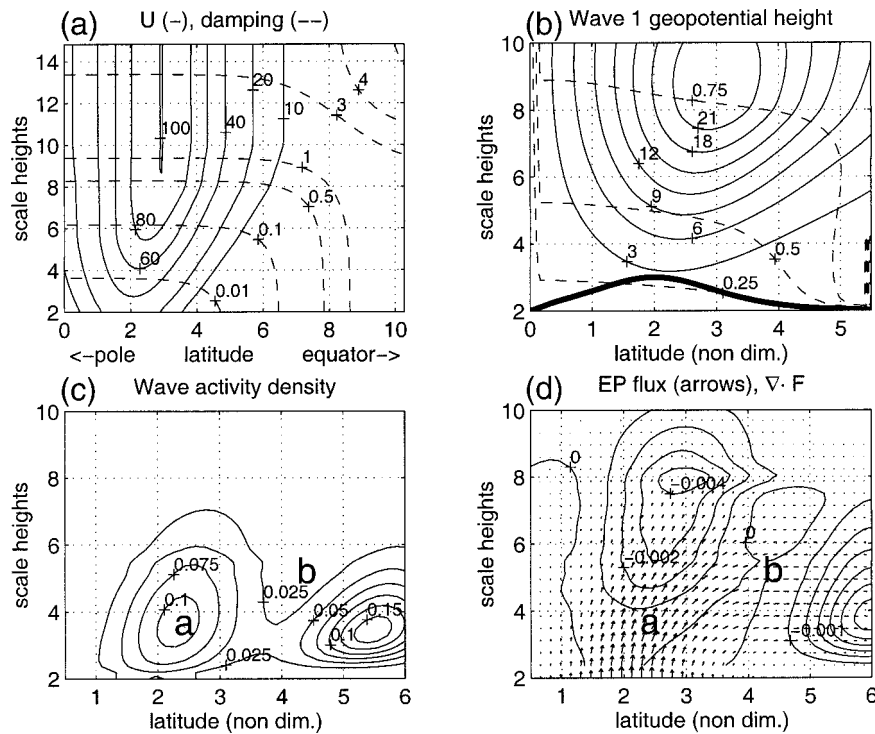


FIG. 1. Characteristics of the β -plane model run. (a) Zonal mean wind (solid, m s^{-1}) and the damping coefficient (dashed, day^{-1}). Note: to show the sponge layer, the domain of this plot is larger than the others. (b) Wave 1 stationary geopotential height amplitude (solid, arbitrary units), phase (dashed, in units of π), and the forcing at the bottom (thick line, magnitude varies between zero and one, with no phase variation with y). (c) Wave activity density (A , arbitrary units). (d) The EP flux (arrows) and EP flux divergence (arbitrary units). Height is in scale-heights (7 km) and latitude is in radii deformation ($L_d = 1190$ km). The **a** and **b** in the bottom two plots also appear in Figs. 2 and 3.

tively, a radiation condition and the effect of either absorption at an equatorial critical surface or radiation to the other hemisphere.

The same model is used in HL, to which the reader is referred for more details. From the various model runs we have done, with different basic states and forcings, the overall picture we get is that the stratospheric jet acts to guide wave activity from the troposphere upward along its axis. Correspondingly, the wave geopotential height peaks in the middle of the waveguide, and in this specific run, the sponge layer damping causes the amplitude to decrease above nine scale heights. The equatorial boundary of the waveguide is leaky because of the wave sink at the equator (represented by the sponge layer). We will show later some consequences of a leaky waveguide configuration.

Figure 1 also shows the wave activity density (A), and the EP flux and its divergence, calculated using Eq. (2) and the middle expressions of Eq. (A4), respectively. We see two distinct maxima of wave activity density, one in the middle of the wave guide, and the other equatorward of it. The EP flux vectors are vertical at lower levels and tilt toward the equatorial sponge layer higher up. Here $\nabla \cdot \mathbf{F}$ is large at the equatorial sponge

layer where there is a maximum of wave activity density, and in the upper stratosphere in the middle of the waveguide.

The corresponding wave activity velocity \mathbf{V}_a is shown in Fig. 2 (arrows). We define a coordinate system that follows \mathbf{V}_a , shown in Fig. 2 by the solid lines. One coordinate, denoted by s , is obtained by calculating integral lines of \mathbf{V}_a .⁴ We refer to s lines as *wave packet paths*, and the value of s represents the time it takes a parcel to reach its location. The other coordinate, denoted by r , represents the latitude at which a wave packet enters the stratosphere at the bottom. The use of this coordinate system for the steady state case is mostly illustrative, but it will prove very useful for keeping track of the wave activity budget in the time-varying case. Wave packets are defined as grid boxes in $s - r$ space. The wave packets move along s lines, and change their volume both because the spacing between s lines

⁴ Since the definitions of wave activity density and wave activity velocity are ambiguous in regions of zero and negative PV gradients, and since the regions of negative PV gradients are small, we have artificially set \mathbf{V}_a to zero there. This explains why wave activity paths end in a point in the middle of the domain.

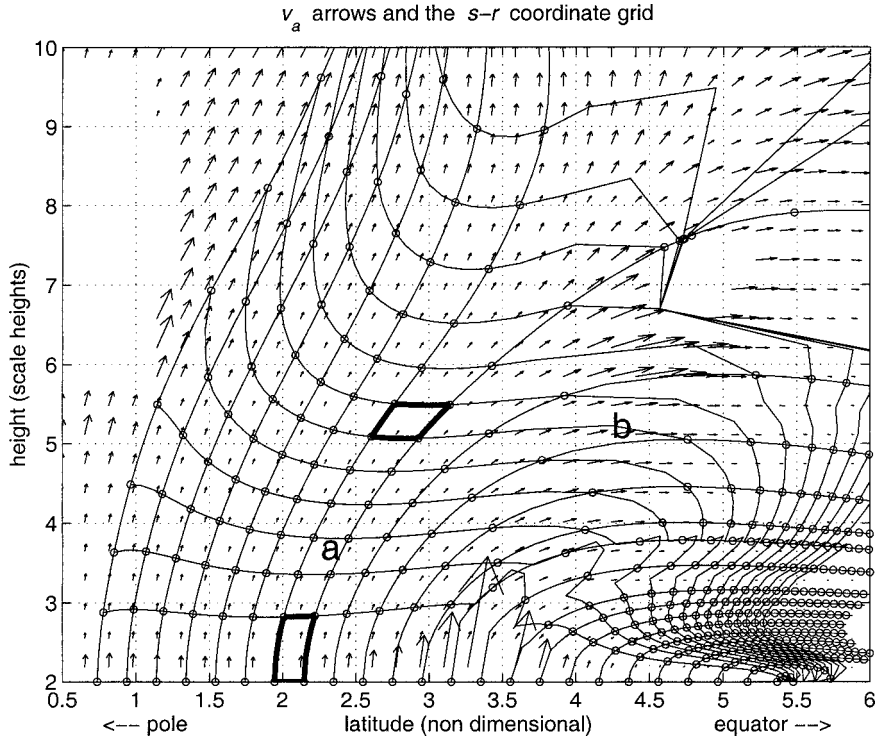


FIG. 2. The wave-based coordinate, plotted in geometric space (latitude–height). The s coordinate is tangent to \mathbf{V}_a (arrows). The circles are spaced 1 day apart on s lines. The a and b are also marked in Figs. 1 and 3.

changes and because the magnitude of \mathbf{V}_a changes along the packet path. The variation of the volume of a wave packet in $y - z$ space is clearly illustrated by the two thick grid boxes marked on Fig. 2. Since the total wave activity in a wave packet (AJ) is conserved (damping

is very small in the region of the two marked grid boxes), A has to change when the volume (J) does. Technical details of these calculations are given in appendix B. Using this framework, in which we can keep track of the wave activity budget of a wave packet, it is tempting to explain the variations in A as *caused* by the variations in wave packet volume, which are induced by local effects of the basic state on \mathbf{V}_a as the wave packet propagates. Indeed, this has been done in various past studies (e.g., Swanson et al. 1997; Sobel and Bretherton 1999; Hoskins and Karoly 1981). As we will show, however, there are some processes, most notable reflection, which affect \mathbf{V}_a nonlocally, and as a result such causality cannot always be assumed.

Figure 3 shows the volume contribution to changes in A [the second term on the rhs of Eq. (5)]. This should be compared to $\nabla \cdot \mathbf{F}$ (Fig. 1d), which equals the damping contribution to changes in A when $\partial A / \partial t = 0$ [first term on the rhs of Eq. (5)]. We see that variations of volume have a significant effect on wave activity density, larger in magnitude than the effect of damping (although the regions of influence are different and the upper-stratospheric damping acts where A is very small).

There are a few mechanisms that affect wave packet volume. One mechanism, of course, is the refraction of the wave by spatial variations of the basic state. This is essentially the mechanism shown by KH of refraction of the group velocity up the gradient of n_{ref}^2 . To get some

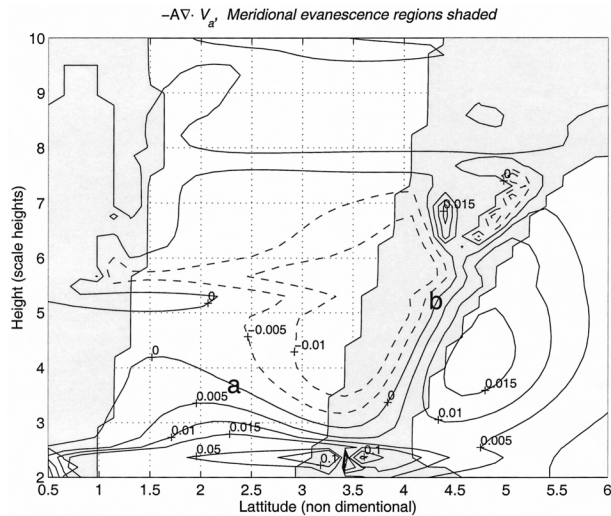


FIG. 3. The variations of wave activity density due to changes in packet volume [second term on the rhs of Eq. (5), contours, negative values dashed], and regions of evanescence in the meridional direction (shaded). The a and b , are also marked in Figs. 1 and 2.

sense of the contribution of refraction, however, we need to better understand the relation between \mathbf{C}_g and \mathbf{V}_a . We discuss this in detail in section 5. Two other processes that affect the volume of wave packets are wave tunneling and reflection. Both processes cause \mathbf{V}_a to differ from \mathbf{C}_g . Wave tunneling is seen on the equatorial side of the waveguide, where the PV gradients are small and n_{ref}^2 is small or negative (point b, and the region with large \mathbf{V}_a at the bottom between $y = 3 - 4$, see Fig. 6). The shaded regions in Fig. 3 are where the meridional wavenumber squared (as defined in HL) is negative, meaning these are wave evanescence regions through which the wave tunnels. There are a few points to note about tunneling. First, \mathbf{V}_a is very large in the tunneling regions, compared to propagation regions. This has a large effect on the volume of wave packets—the volume decreases as the wave packets enter the tunneling region (poleward of point b) and increases as they leave it on the equatorial side. Correspondingly, there is a minimum of wave activity density (point b, see Fig. 1c) in the evanescent region, with maxima on both sides. It is interesting that wave propagation (as represented by \mathbf{V}_a) is faster in evanescent regions than in propagation regions; however, this is true only after the tunneling is established. As we show in section 4, when wave forcing is turned on, the wave does not initially propagate through the evanescent region, rather, it takes time for tunneling to develop. Consequently, the propagation across the tunneling region of “information,” as represented by the leading edge of the forced wave, is not actually faster than regular wave propagation. The second point to note is that while \mathbf{V}_a shows tunneling through evanescence regions, the group velocity is reflected away from these regions (hence the distinction from propagation regions). The phenomena of tunneling, which occurs when the wavelength is on the order of the width of the evanescent region, is not captured by the group velocity or by wave rays. This limitation of the wave ray formulation is well known for light waves. In section 5 we discuss this limitation for stratospheric Rossby waves.

The effect of reflection is less obvious. We see that the volume effect increases wave activity density in the waveguide ($y = 1 - 3$) in the lowest third of the domain and decreases it above, and the peak in wave activity density (point a) falls exactly on the zero line. These volume variations are the result of downward reflection from a turning surface that exists in the upper stratosphere in this model run.⁵ The EP flux field changes very smoothly in space, and its spatial pattern is not affected much by reflection (rather, reflection reduces it everywhere). At the same time, the superposition of the incident and downward reflected components of the wave causes its amplitude and wave activity density

patterns to vary in space quite a lot. As a result, \mathbf{V}_a maximizes where A is minimum and vice versa [Eq. (3)]. This leads to a convergence “upstream” (assuming “stream lines” follow \mathbf{V}_a) of the maximum in A and a divergence “downstream” from it, which is consistent with the volume-induced variations in A . The volume effect in this case cannot be seen as causing the variations in wave activity density, rather, it is consistent with them. This effect of reflection is true not only for stratospheric waves but for any waves for which wave activity conservation applies.

Finally, we note that damping can also affect the volume of wave packets by affecting \mathbf{V}_a . This is a separate effect from the direct decrease of wave activity. From this and other runs (see HL), we find that damping affects \mathbf{V}_a mostly by indirectly affecting the amount of reflection from upper-stratospheric reflecting surfaces. We find a noticeable local effect on \mathbf{V}_a only in the sponge layer, where the damping becomes quite large. Our sponge layer, which only serves as an energy sink, is most likely not a very realistic description of the effects of damping on the waves. Since the magnitude of damping in the stratosphere, and in particular, the extent to which a linear formulation is applicable, are not very well known, we leave the discussion of damping effects for future work.

4. The transient evolution of wave activity density and \mathbf{V}_a

Observed stratospheric waves occur in episodes (Hirota and Sato 1969; Randel 1987) and are continually varying in time. When the wave field varies with time, the wave activity density, the EP flux, and hence \mathbf{V}_a and our wave-based coordinate vary with time as well. Time variations in the tropospheric source of wave activity are advected into the stratosphere, resulting in a contribution to the spatial variations of A . To understand how the various mechanisms affect wave activity density in this case, we study the evolution of a stratospheric wave in a time-dependent version of our model (described in HL). Our basic state, which is constant with time, is the same as in the steady-state model (Fig. 1). We start with no perturbation, turn on the forcing over a period of 12 days and decrease it to 0.6 of its maximum on days 14–19. The latitudinal shape of the forcing is the same as in Fig. 1. The technical details of the calculations presented in this section are described in appendix B.

Figure 4 shows the wave activity density, and the integral lines of the corresponding instantaneous \mathbf{V}_a fields, for 3 days of the model run. We find a two-stage evolution, with the perturbation initially propagating up the waveguide and later on “spreading sideways.” This evolution is due to the existence of an evanescent region separating the midlatitude waveguide from the large equatorial n_{ref}^2 region. Initially, the wave is “not aware” of the large index of refraction at the equator, and it

⁵ HL calculated the vertical wavenumber for this run and found a reflecting surface for vertical propagation at around eight scale heights.

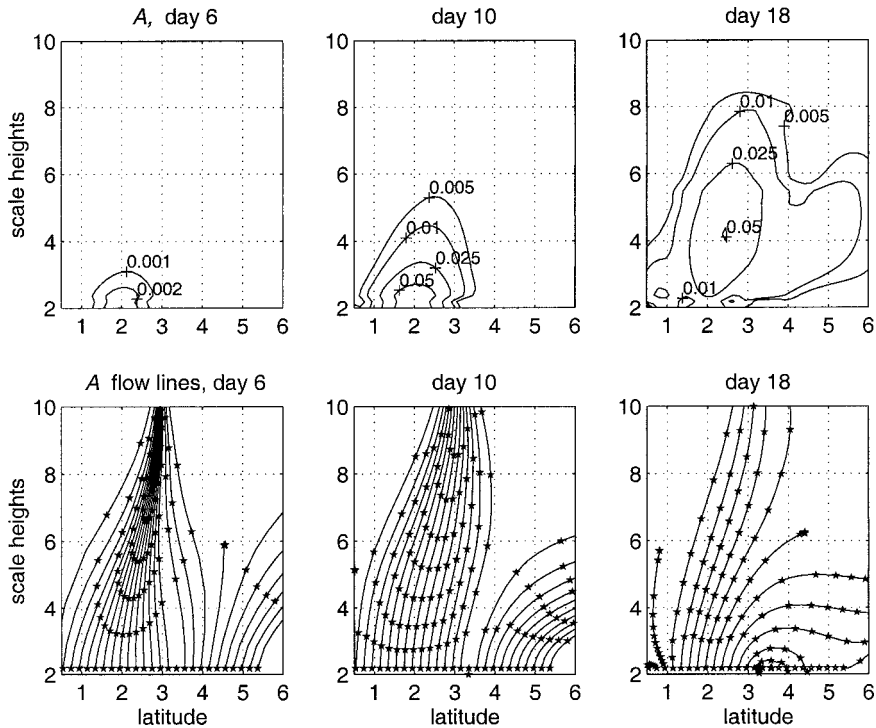


FIG. 4. Three days (6, 10, 18) of (top) wave activity density and (bottom) integral lines of \mathbf{V}_a (stars mark day intervals) for the model run described in the text. Note that contour intervals are not the same for all days.

travels up the local n_{ref}^2 gradient into the middle of the waveguide. Since the width of the evanescent region is on the order of the wavelength, some wave activity eventually leaks through to the equator, where it is absorbed in the sponge layer, and tunneling develops. Correspondingly, \mathbf{V}_a lines tilt equatorward. This behavior is found in all the runs we did in which a tropospheric source is turned on. Randel et al. (1987) in a study of life cycles of stratospheric planetary waves find an initial baroclinic stage (EP fluxes point upward), followed by a barotropic stage (EP fluxes point equatorward), suggestive of the wave evolution shown here. Viewed as an evolution on a leaky waveguide, it is clear why the “barotropic stage” (meridional propagation) occurs only later on in the wave life cycle—it takes time for the wave to “see” the damping region in the equator and tunnel to it. Note that tropospheric waves have a similar life cycle of upward followed by equatorward EP fluxes (e.g., Edmon et al. 1980), and this picture of initial propagation into the waveguide, followed by eventual leakage to the equator might apply there too.

The time variations of the wave field complicate the interpretation of wave activity conservation. The advection of a time-dependent wave source introduces spatial changes in wave activity density. We can get rid of this effect by following a single wave packet. This requires a different calculation of \mathbf{V}_a lines because those shown in Fig. 4 do not represent the propagation of a single wave packet (analogous to the distinction be-

tween streamlines and trajectories in fluid flow). We follow the set of wave packets that leave the tropopause at a given day, taking into account their time varying \mathbf{V}_a . For clarity, we denote the integral lines of the velocity field *following wave packets by wave packet paths*. Correspondingly, the fractional change in wave packet volume as it propagates is given by $\nabla \cdot \mathbf{V}_a$ with \mathbf{V}_a following a wave packet (and not by the divergence of the instantaneous \mathbf{V}_a field). Wave packet paths are analogous to trajectories in fluid flow, while the integral lines of the instantaneous \mathbf{V}_a field are analogous to streamlines.

Figure 5 shows the wave packet paths for a set of wave packets that leave the bottom of the model on day 16. We mark intervals of 1 model day by circles, so, for example, the lowest set of circles (at around three scale heights) are the location of the wave packets on day 17. To illustrate the conservation of wave activity of this set of wave packets we need to keep track of their wave activity density. This is most convenient to do if we define the wave packet paths as our $s - r$ coordinates. We then determine A of the packet from its correct day and location, by transforming the A field of each model output time to the $s - r$ coordinates,⁶ and stacking the A values from the appropriate s of each time. For example, a plot of this composited wave ac-

⁶ The transformation of a scalar field is simply done by unskewing it so that s lines are vertical (see appendix B).

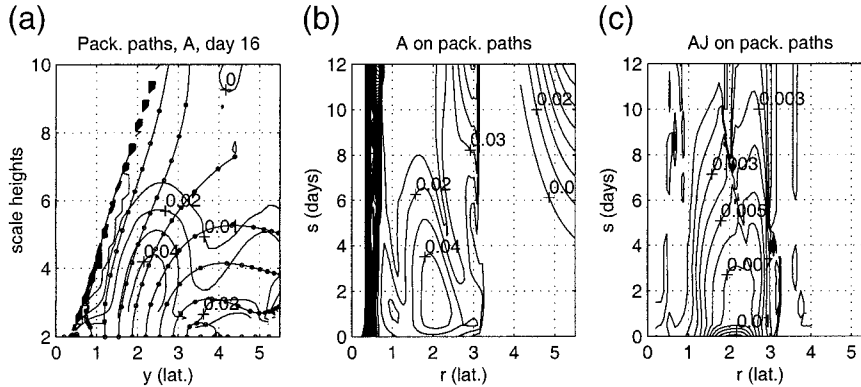


FIG. 5. Wave activity budget for packets that leave the bottom on day 16. (a) Wave packet paths and the corresponding wave activity density that the packets have. (b) A of the wave packets, plotted on $s - r$ coordinates. (c) The AJ of the wave packets, plotted on $s - r$ coordinates. The pole is at small y/r .

tivity density field, plotted in the $s - r$ coordinates, is shown in Fig. 5b. The values of A that are plotted on the $s = i$ row are taken from the $s = i$ row of the A field of day 16 + i , where $i = 0:0.5: \dots 12$ (our model output time step is half a day). The “vertical” axis in this plot is the direction of propagation of each of the wave packets, and the variations of A along the vertical direction indicate it changes as the wave packets move along. This is due both to variations in packet volume and to damping. Since we are tracking given wave packets, however, we do not have a contribution from the time variation of the forcing. For reference, we transform the composited A field back to $y - z$ space and plot it on the wave packet paths (Fig. 5a). The packets that leave the bottom on day 16 travel along these paths, while wave activity density varies according to the contours.

To illustrate that indeed we are following wave packets that conserve wave activity in regions where there is no damping, we multiply A by the corresponding packet volume (which is the Jacobian of the transformation J), and plot in $s - r$ coordinates (Fig. 5c). Again, since s is the direction of wave packet propagation, vertical contours mean a constant value with time. We see that indeed AJ is quite constant for the first 3 days, roughly the time it takes the packets to traverse four scale heights and reach the sponge layer. The strong decrease in AJ near the surface is probably not real, because the Jacobian, which involves derivatives along the packet paths, is not very accurate at the bottom.

To summarize, in steady state, we can calculate the damping and the volume effect in Cartesian coordinates, using $\nabla \cdot \mathbf{V}_a$ and $\nabla \cdot \mathbf{F}$ directly. In the time-varying case, in order to separate out the variations of wave activity due to advection of time variations of the wave source, we need to follow a wave packet. The damping that a specific wave packet feels, or the changes of A due to the volume effect, equal $\partial A / \partial t + \nabla \cdot \mathbf{F}$ and $A \nabla \cdot \mathbf{V}_a$ [Eqs. (1) and (5)], calculated with keeping track of how A ,

\mathbf{V}_a , and \mathbf{F} vary as the wave packets move along. The most practical way to do this is using the $s - r$ coordinates. The amount of damping the wave packets feel is the vertical derivative of the $s - r$ plot of AJ . The volume contribution can be calculated from a similar derivative of J . On the other hand, the more commonly used quantity $\partial A / \partial t + \nabla \cdot \mathbf{F}$, calculated using instantaneous fields, gives the local variation of wave activity density due to damping for all wave activity packets that compose the wave field at that time.

5. The relation to ray tracing

In section 3 we discussed various processes that affect wave packet volume, and only mentioned wave refraction. Wave refraction, by definition, affects the group velocity. In this section we want to understand how wave refraction affects \mathbf{V}_a . We do this by comparing wave activity paths to the wave rays calculated by KH. Wave rays are the integral lines of \mathbf{C}_g . The integration of $dy/dt = C_{gy}$ and $dz/dt = C_{gz}$ is done along with an integration of the equations for the variation of the vertical and meridional wavenumbers (m and l , respectively) along the wave rays, subject to a set of initial conditions: y_o, z_o, l_o, m_o , and assuming an almost-plane wave with a given zonal wavenumber and phase speed. The wave rays essentially show where wave activity will propagate if a *point source* (y_o, z_o) is put into the medium, for a given initial wave vector (l_o, m_o). Wave activity velocity paths, on the other hand, are diagnostic in nature, meaning the path of wave activity is diagnosed for a given wave field. The calculation requires a knowledge of the wave field and the basic state, but no wavenumbers have to be specified. Karoly and Hoskins (1982) calculated rays for a range of initial propagation angles, on specified stratospheric basic states and various locations of the point source. They used the rays as an indication for where waves will propagate, in order to gain insight into their structure. They showed that

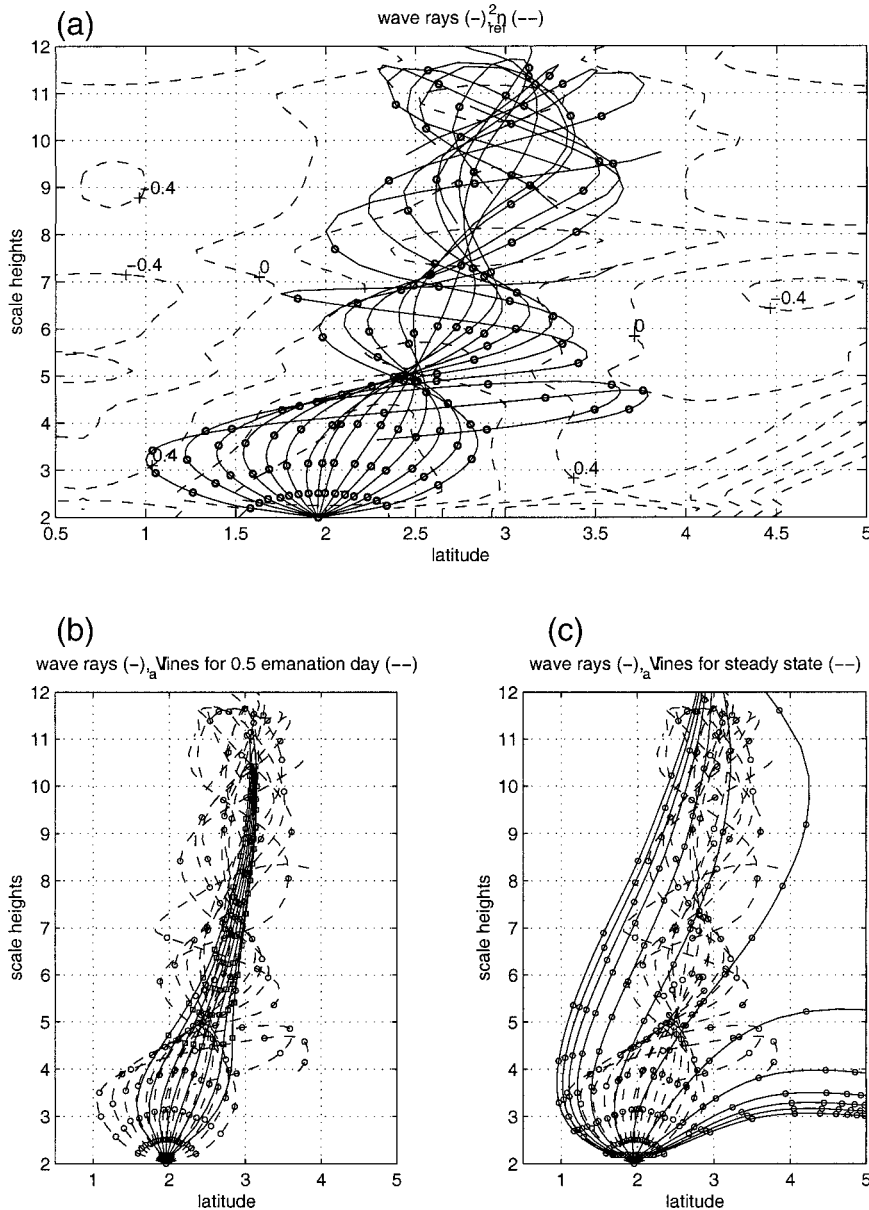


FIG. 6. (a) Wave rays (solid), superposed on the index of refraction squared (dashed). (b) and (c) A comparison between wave rays (dashed) and wave packet paths (solid), for (b) wave packets that leave the bottom at day 0.1 and (c) for steady state, for a model run with point forcing turned on at day 0, over 4 days. Half-day intervals are marked on all rays and packet paths. See text for details.

wave rays are refracted up the local gradient of n_{ref} (in spherical coordinates rays also refract equatorward). The relation of wave rays to a given wave field and its EP fluxes, was not discussed in detail in KH.

Figure 6 shows the wave rays along with the wave packet paths for a stationary wavenumber 1 point source that is turned on at the bottom of the waveguide at $t = 0$, both for wave packets that leave the bottom during the initial stages of wave development (0.1 days) and for the steady-state solution. The basic state is the same as used in previous sections. The circles/squares on both

mark half-day intervals. We also plot the wave rays superposed on n_{ref}^2 , to show that the rays are reflected back and forth in the latitudinal direction along the n_{ref}^2 waveguide. When a point source is turned on, wave activity propagates in all directions, and the wave rays tell us how it propagates in *one* such direction. When the wave activity reaches the sides of the waveguide, it reflects back and superposes. The actual wave field is the superposition of wave activity propagating from all directions. At later stages, the wave tunnels to adjacent propagation regions and is dissipated in regions with

damping. In the bottom 1.5 scale heights, before the wave field reaches the sides of the waveguide, the wave packet paths follow the wave rays quite well. Above the level at which meridional reflection occurs, wave packet paths initially concentrate into the middle of the waveguide, while the wave rays oscillate around them. At later stages the wave rays do not change but wave packet paths tilt to the equator.

We see significant differences between \mathbf{V}_a and C_g that arise from the fact that wave packet paths are a diagnostic of the waves while wave rays characterize the *basic state* for a given zonal wavenumber. While wave packet paths show the flow of wave activity in the *total* wave field, which is affected nonlocally by processes like tunneling, reflection, and damping, wave rays reflect *local* wave propagation properties. They do not “tunnel” through evanescent regions and are not affected by damping. A more subtle effect is the fact that the stratospheric waveguide, through reflection in the meridional direction, sets the structure (i.e., the meridional wavenumber) of the waves. Harnik and Lindzen (2001) show that this results in a reflecting surface forming at around eight scale heights in this model, because $n_{\text{ref}}^2 - l^2 < 0$. Harnik and Lindzen (2001) also show that this reflecting surface has a very large effect on the wave structure and evolution. Ray tracing, on the other hand, does not capture its effect, because the meridional wavenumber is set locally, and by the initial wave vector, and there is a wide range of initial l_o values for which wave rays propagate beyond the reflecting surface. The difference between the evanescent region for vertical propagation and the evanescent region on the equatorial side of the waveguide is that in the latter, n_{ref}^2 is negative or very small. The $n_{\text{ref}}^2 < 0$ region, which is often observed in the real stratosphere, is a result of the small meridional PV gradients there. Another source of difference between \mathbf{V}_a and C_g is the fact that ray tracing applies to point sources that have a pure sinusoidal oscillation in time (a single frequency), hence it is not an accurate description at the initial stages when the point source is turned on. This is probably why we see such a poor correspondence between the timescales of the wave rays and \mathbf{V}_a lines (the distances between the circles or squares) in the bottom scale height of the 0.1 day plot and quite a good correspondence in the steady-state case. Finally, in the general case of non-localized forcing the wave field is the superposition of the response to many point sources, and the relation between wave packet paths and wave rays becomes even more complicated.

To summarize, meridional reflection off of the sides of the waveguide, and later on, tunneling to the equator and downward reflection, cause \mathbf{V}_a and C_g to differ greatly from each other. Essentially, ray tracing cannot capture these phenomena because they occur when the characteristic wavelength is on the order of the meridional scales of the waveguide and the evanescent regions. This limitation of ray tracing is not the same as

the limitations to WKB conditions being valid. In fact in our model run, while WKB conditions are violated near reflecting surfaces, they do hold locally quite well in large parts of the domain (see HL, appendix B). In addition, C_g represents the propagation from a point source, while \mathbf{V}_a is diagnosed from the superposition of the wave field from many point sources. An important implication for the stratosphere is that we do not expect \mathbf{F} to refract up the local gradient of the index of refraction just because C_g does. The general tendency of \mathbf{F} to refract toward the large equatorial index of refraction values is more likely because there is a critical surface there, which absorbs wave activity.

6. Diagnosing vertical propagation timescales and the application to observations

In this section we use wave packet paths as a diagnostic of vertical propagation time scales of the waves. By doing this on observed waves, we also examine whether our framework of a wave being composed of many small wave activity packets makes sense for real waves, and whether some of the features we find in our model apply to the real world.

Since vertical wave propagation is one of the main ways in which the stratosphere and troposphere are linked, the vertical propagation timescale is one of the fundamental timescales of the coupled troposphere–stratosphere system. As mentioned in the introduction, the existing methods for estimating vertical propagation timescales are based on ray tracing (KH) or on time lag correlations (Randel et al. 1987; Randel 1987). Ray tracing, as we saw in section 5, is a highly idealized calculation, and the relation of it to the actual propagation of a wave is complicated. The time–lag correlation method is statistical, meaning it gives an averaged timescale over many wave events. Tracking wave packets using our method allows us to estimate the vertical propagation time for a given wave event, and to see how much it varies between different wave episodes. By tracking different parts of the wave (individual wave activity packets), we get a sense of how characteristic a single propagation timescale is of the entire wave field. This is useful for understanding how the choice of reference point in the time–lag correlation calculations might affect estimated propagation times.

To apply to observations, we first interpolate the observed geopotential height, zonal mean wind, and temperature onto a high-resolution grid, and then calculate our wave activity diagnostics. The observational data we use, as well as the technical details of applying our diagnostic to observations are described in appendix C. The vertical coordinate in the figures is log-pressure height, using a 7-km scale height.

We should note that Randel’s timescale estimates are based on correlations of geopotential height. Harnik and Lindzen (2001) studied how reflection affects the geopotential height structure of the waves. This allowed

them to obtain an independent estimate of a vertical propagation timescale. A comparison with our method can therefore be used to increase our confidence in both estimates. We use an observed wave 1 event from 18 July to 19 August 1996, in the Southern Hemisphere, which was studied in detail by HL. They showed that during most of this period the basic state allowed vertical propagation throughout the stratosphere. The wave underwent the following life cycle. It propagated from the troposphere to the stratosphere, and after a few days started decelerating the mean flow in the upper stratosphere. Eventually, the deceleration resulted in the formation of a reflecting surface. The wave responded by reflecting downward, at which time its source decreased. Downward reflection was accompanied by acceleration of the zonal mean wind, which caused the reflecting surface to disappear. This cycle repeated itself twice. Reflecting surfaces formed during 29 July–1 August and 8–10 August. We choose this wave event because it allows us to test the effects of downward reflection on our estimates. As we saw in section 3, downward reflection considerably affects \mathbf{V}_a , therefore a test of our timescale estimates should include a comparison using reflected and nonreflected waves. Also, a comparison between the two life cycles will give us a sense of the variability in propagation times.

The top plot of Fig. 7 shows a height–time plot of the latitudinal averaged wave 1 geopotential height amplitude during this wave event. The reader is referred to HL for plots of the evolution of other wave and basic-state properties. Starting 18 July, we calculate the wave packet paths for each emanation day, and use them to calculate the time it takes various wave packets to reach a certain height, and correspondingly, the height wave packets reach after a certain time. The most striking property of these calculations is how variable they are, both as a function of the emanation latitude, and as a function of time. Figure 7b shows the maximum height reached after a certain time, by wave packets that start at 16 km on 18 July–19 August [calculated by integrating Eqs. (B1)–(B4)]. The maximum height is chosen from the different emanation latitudes. The bottom plot shows the height reached by packets after 3 days of propagation, for all latitudes and emanation days. The line corresponding to 3 days in the middle plot is constructed by plotting the daily maximum value from the bottom plot, and the bottom plot gives an indication of how representative the maximum height is of the entire wave field. We see that propagation is fastest when the wave is growing in the stratosphere (due to propagation from the troposphere), and it is smallest during downward reflection periods (28 July–2 August and 10–15 August). This decrease due to reflection is because the wave activity velocity represents the total wave field, which is a superposition of upward- and downward-propagating waves [e.g., the vertical version of Eq. (8)]. To get a real propagation timescale, we therefore need to look at the initial stages of a wave event, when the

wave is only propagating upward. Since the wave source decreased during our wave event, we have two such periods (before 24 July and 3–6 August). The estimated propagation time from 16 to 45 km is 3–4 days for the second growth period (using 4–5 August) and roughly 5 days for the first growth period (18 July gives 4–5 days and 24 July gives 5–6 days). The difference between the two periods is probably due to differences in the basic state (the zonal mean winds were weaker by more than 10 m s^{-1} during the second period). Randel (1987) estimated 4 days for wave 1 to propagate from midtroposphere to midstratosphere. While the two calculations are not of exactly the same thing, it is encouraging that they give similar numbers.

A different test we can do is to check whether our diagnosed times are consistent with the observed wave dynamics. Downward reflection causes the wave phase tilt with height to change, such that the original upward-propagating structure of a westward phase tilt with height changes to a standing wave structure (vertically aligned phase lines) when the wave is fully reflected. If the source of the wave is constant with time, this happens when the wave propagates from the turning surface to the tropopause, but if the source decreases with time during downward reflection (as is the case here), it should happen in half the time.⁷ Harnik and Lindzen (2001) showed that it takes the wave roughly 2 days to tilt to a vertical position from the time of the formation of the reflecting surface (on 10 August), meaning the vertical propagation time should be around 4 days. This is indeed consistent with the timescale estimated from the wave activity packets. Since our estimate uses the early stages of the wave event, while vertical phase tilting occurred later on, the two calculations are independent.

In previous sections we showed that the effect of the leaky waveguide on wave packet paths is to cause them to initially concentrate up the waveguide and only later on to spread sideways and leak to the equator. This is also found in observations when we look at the evolution of the wave through the location of wave packets. Figure 8 shows wave packet locations for 21, 26, 28, and 31 July, superposed on the corresponding outline of regions where $n_{\text{ref}}^2 < 0$ (shaded regions). Each symbol represents a wave packet. Packets are marked by the day they emanate at the bottom (16.1 km in this case), and the color and symbol are kept the same in all plots for a given emanation day. For example, on 21 July we see the locations of packets that left the bottom on 18 July (magenta o's), 19 July (yellow o's), 20 July (black squares), and 21 July (the purple line at the bottom). These plots are created by compositing the locations of wave packets from the wave packet

⁷ This estimate is for an idealized case. In reality the formation of a turning surface is gradual, hence the time it takes the structure to change is not exactly the vertical group propagation time, but it should be close to it.

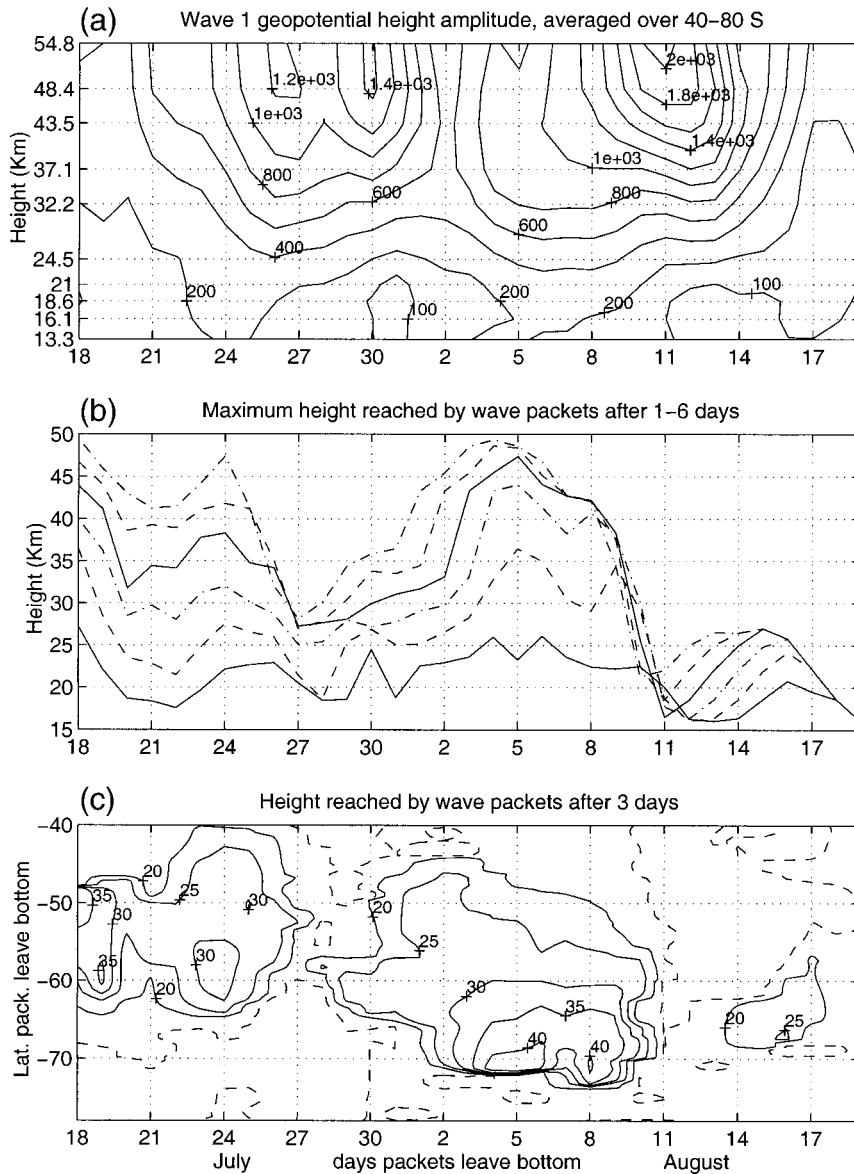


FIG. 7. Height–time sections (18 Jul–19 Aug 1996) of (a) wave 1 geopotential height amplitude (m), averaged over 40°–80°S. (b) The maximum height reached by wave packets 1, 2, 3, 4, 5, and 6 days after leaving the bottom (16.1 km), as a function of emanation day (in km). The curves are stacked in order on 18 Jul (1 day is the lowest). (c) The height (in km) reached by wave packets 3 days after they leave the bottom (16.1 km), as a function of latitude and day of emanation. Dashed lines indicate wave packets that propagate downward at 16.1 km.

paths of all emanation days prior to the day in which the packet locations are shown. We see that it takes the packets of 18 July 3 days to ascend from 16.1 to 40 km (which is consistent with the estimates from Fig. 7). We also see the slowing down of packet propagation later on. The wave packets show a clear evolution of an initial propagation up into the middle of the waveguide followed by a spreading out and leaking out of the sides, mostly toward the equator.

A note is needed as to the quality of these diagnostics. First, in our calculations, we set the wave activity ve-

locity to zero in regions of negative PV gradients, hence the bunching of packets at equatorward vertical lines. The magnitude (not the direction) of \mathbf{V}_a depends strongly on \bar{q}_y ($\mathbf{V}_a = \mathbf{F}/A \propto \mathbf{F}/(1/\bar{q}_y) \propto \bar{q}_y$). The direction of the propagation, on the other hand, depends on the phase structure of the wave. Plots of successive wave packet locations are useful in showing us general propagation patterns, but the exact location of a given wave packet on a given day is not very accurate. An idea of the uncertainty in packet locations can be drawn from the large differences between calculations of $\nabla \cdot \mathbf{F}$ of the

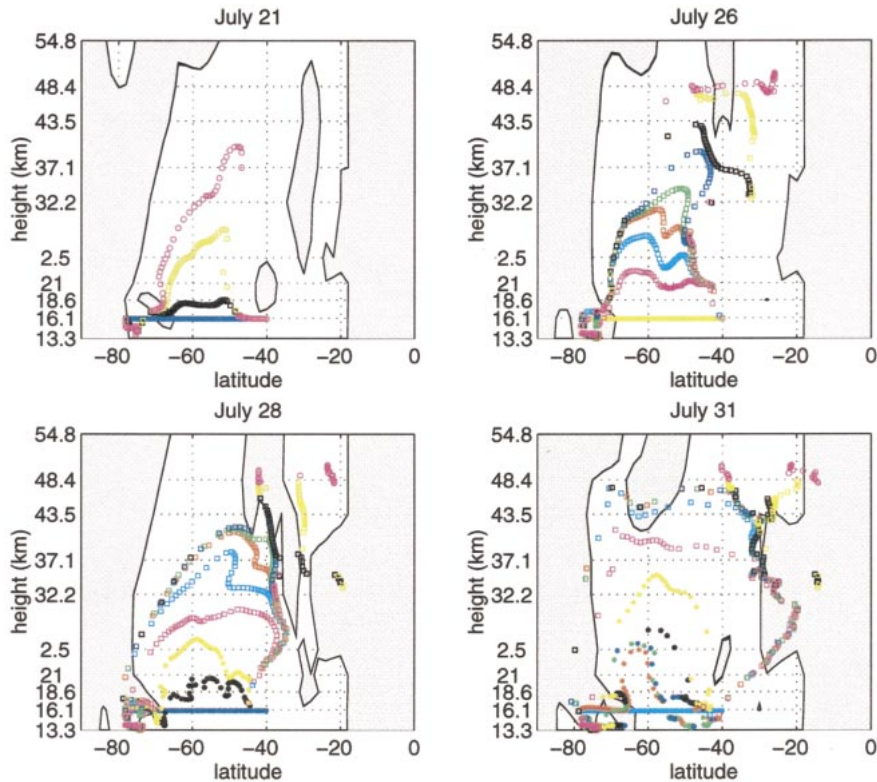


FIG. 8. Latitude–height plots of wave packet locations, for 21, 26, 28, and 31 Jul 1996, superposed on the shading of regions of negative n_{ref}^2 of the same day. Each symbol denotes a wave packet. Packets are plotted for each day, and the colors and symbols mark the day of emanation at the bottom. These are consistent within the different plots; e.g., magenta circles denote packets that emanated on 18 Jul.

various analyses products, especially for the Southern Hemisphere (see Grose and O'Neill 1989).

7. Summary and discussion

In this paper we use the conservation of wave activity to study how a few very basic wave processes that relate to the group velocity of the waves apply to stratospheric waves. More specifically, KH showed that the group velocity refracts up the gradient of the index of refraction (for Rossby waves), and they combined this with conservation of wave activity to obtain an analytic expression of the variation of wave amplitude along a wave ray, for almost-plane waves propagating on a slowly varying basic state. This provides the mechanism by which spatial variations in the basic state cause the spatial variation in wave amplitude.

We ask whether we can obtain similar understanding of wave structure for more realistic waves. To do this, we need to understand how the wave activity velocity $\mathbf{V}_a = \mathbf{F}/A$, which equals the group velocity \mathbf{C}_g in the plane wave limit, relates to \mathbf{C}_g in the case of stratospheric waves. We do this by comparing the integral lines of \mathbf{V}_a to wave rays (which are integral lines of \mathbf{C}_g). We find that meridional reflection off of the sides

of the waveguide, and later on, tunneling to the equator and downward reflection, cause \mathbf{V}_a and \mathbf{C}_g to differ greatly from each other. In addition, \mathbf{C}_g represents the propagation from a point source, while \mathbf{V}_a is diagnosed from the superposition of the wave field from many point sources. One of the things this means is that contrary to what is often assumed in the literature, the EP flux is not refracted up the local gradient of the index of refraction. It is true that the EP flux tends to refract to the equator, where the index of refraction is very large, but this is due to the existence of a critical surface there.

We also use the integral lines of \mathbf{V}_a to define a coordinate system. We then define grid boxes of this coordinate as *wave activity packets* that are analogous to compressible fluid parcels. Each wave packet contains a certain amount of wave activity, which is conserved unless there is damping. The convergence of \mathbf{V}_a will result in changes of wave packet volume, which are accompanied by changes in wave activity density, and hence wave amplitude. This volume effect, which was shown by KH for the almost-plane wave limit, is very basic in wave theory (not only for the stratosphere)—it provides a physical explanation of why the first-order WKB solution of an almost plane wave has a term

inversely proportional to the square root of the wave-number (see KH). Also, if changes in \mathbf{V}_a are caused *locally* by the variations in the basic state (as is the case with \mathbf{C}_g), the corresponding variations in wave activity density can be viewed as *caused* locally by the spatial variations in the basic state, through the volume effect. It would be useful to have such an understanding for stratospheric waves, and if that is not possible, it is interesting to understand why. We use our wave activity coordinate, as well as our understanding of the relation between \mathbf{V}_a and \mathbf{C}_g , to show how refraction, tunneling, and reflection affect $\nabla \cdot \mathbf{V}_a$ and the volume effect in a simple steady-state model. Since these processes are characteristic of a leaky waveguide, we expect them to occur in the real stratosphere. One of the important things we show is that causality (in the sense discussed above) is lost when we have reflection, because \mathbf{V}_a is no longer related to the basic state in a local way. Rather, the structure of the wave is determined by the superposition of the incident and reflected waves, and the corresponding variations in \mathbf{V}_a and wave packet volume, are merely *consistent* with the wave structure. Note that this effect of reflection is a general property of waves, and is not specific to stratospheric Rossby waves.

We also use \mathbf{V}_a as a diagnostic of the vertical propagation timescale for the waves. This timescale is important for the coupled troposphere stratosphere system because vertical wave propagation is one of the main links between the two. We test our method by applying it to an observed wave event. We choose a wave event that occurred during a time when the basic state had no reflecting surfaces, but during its life cycle reflecting surfaces formed (also analyzed by HL). This allows a comparison of the timescale estimates with and without reflection. We also compare with an independent estimate that HL obtained based on their understanding of how wave structure changes when a wave reflects down. We find that both estimates give a timescale of 4 days for wavenumber 1 to traverse the stratosphere. The fact that these estimates, one based on \mathbf{V}_a and wave activity, and the other on the geopotential height structure, are consistent, strengthens our confidence in our dynamical understanding of the wave life cycle and in the wave activity framework presented here. Note that our definition of a wave packet, as being a part of the wave field, which is much smaller than a wavelength, is different from the more traditional view that a wave packet contains a few wavelengths.

Finally, we identify a characteristic evolution of a wave propagating up a leaky waveguide, of initial propagation up into the middle of the waveguide, and a later tunneling out to the equator. The leaky waveguide picture was suggested quite early on in studies of stratospheric waves (Dickinson 1968; Matsuno 1970), but the consequences of such a configuration, in particular, on the life cycle of stratospheric waves, have not been demonstrated in detail. We show that tunneling has a large

effect on the wave activity density, through the volume effect. This may have some implications for the structure of waves in the tropical region, because we expect the factors affecting the magnitude of wave activity density to be different in the case of direct propagation into the tropical region (if the evanescence region did not exist, or as a result of direct vertical propagation from the tropical tropopause). Also, the evolution we find corresponds nicely to the baroclinic growth–barotropic decay life cycle that was observed for stratospheric waves by Randel et al. (1987). It explains why the barotropic stage, of meridional propagation, comes later—it takes time for the wave to feel the wave activity sink at the equator and tunnel out to it. While this propagation picture is obviously relevant to the stratosphere, it is unclear to what extent it is relevant to life cycles of tropospheric waves, where we also expect a contribution to the EP fluxes from baroclinic growth due to instability, and barotropic decay due to shearing of the perturbation by the basic state (see Edmon et al. 1980, and references therein).

Acknowledgments. This work is part of my Ph.D. dissertation. I thank my advisor, Richard Lindzen for his inspiring guidance. I thank Alan Plumb and Adam Sobel for helpful discussions, and, along with two anonymous reviewers, for helpful comments on this manuscript. I wish to thank Dr. A. J. Miller of the Climate Prediction Center at NCEP, and Dr. Paul A. Newman of the Atmospheric Chemistry and Dynamics Branch (code 916) at the Goddard Space Flight Center for producing these data, and to thank Paul Newman, Larry Coy, and Eric Nash for providing assistance and easy access to the data. I also wish to thank the Goddard DAAC for distributing the data as part of NASA's Mission to Planet Earth program. This research was supported by Grants ATM9813795 from the National Science Foundation, and NAG5-5147 from National Aeronautics and Space Administration.

APPENDIX A

The Relation between \mathbf{V}_a and Group Velocity

The following derivation is quite standard, but since we want to emphasize a result that is not always appreciated in the literature [Eq. (8)], we bring it here for completeness. We use standard notation and a full derivation can be found in Andrews et al. (1987), or in HL.

We assume linear quasigeostrophic Rossby waves, that satisfy WKB conditions, and an almost-plane Rossby wave of the form:

$$\phi \propto \exp \left[i \left(kx - \omega t + \int l dy + \int m dz \right) \right]. \quad (\text{A1})$$

The dispersion relation is

$$\omega = kc = kU - \frac{k\bar{q}_y}{k^2 + l^2 + \frac{m^2}{N^2} - F(N^2)}, \quad (\text{A2})$$

where $F(N^2) \equiv f_o^2(e^{z/2h}/N)(\partial/\partial z)[(e^{-z/h}/N^2)(\partial/\partial z)(e^{z/2h}N)]$, which is obtained when the equations are transformed into canonical form (e.g., see HL). The density is assumed to vary exponentially with a scale height $h = 7$ km.

The group velocity of this wave equals

$$C_{gy} = \frac{\partial\omega}{\partial l} = \frac{2\bar{q}_y kl}{\left[k^2 + l^2 + \frac{m^2}{N^2} - F(N^2)\right]^2}$$

$$C_{gz} = \frac{\partial\omega}{\partial m} = \frac{2\bar{q}_y km}{N^2 \left[k^2 + l^2 + \frac{m^2}{N^2} - F(N^2)\right]^2} \quad (\text{A3})$$

For a wave of the form (A1), the EP flux and the wave activity density equal

$$F_y = \rho \frac{k}{2} |\phi|^2 \text{Im} \left(\frac{\phi_y}{\phi} \right) = \rho \frac{kl}{2} |\phi|^2$$

$$F_z = \rho \frac{k}{2N^2} |\phi|^2 \text{Im} \left(\frac{\phi_z}{\phi} \right) = \rho \frac{km}{2N^2} |\phi|^2, \quad (\text{A4})$$

$$A = \left[k^2 + l^2 + \frac{m^2}{N^2} - F(N^2) \right]^2 \frac{\rho |\phi|^2}{4\bar{q}_y}. \quad (\text{A5})$$

We see that $C_g = \mathbf{F}/A$. In general, however, Eq. (A1) does not hold, rather, we have a superposition of waves, for example,

$$\phi = \left[\alpha_+ \exp \left(i \int l dy \right) + \alpha_- \exp \left(-i \int l dy \right) \right] \times \exp \left[i \left(ky - \omega t + \int m dz \right) \right], \quad (\text{A6})$$

where α_+ and α_- are complex coefficients of the two wave components. It is easy to show that under WKB conditions Eq. (A6) implies

$$\text{Im} \left(\frac{\phi_y}{\phi} \right) = l \frac{|\alpha_+|^2 - |\alpha_-|^2}{\left| \alpha_+ \exp \left(i \int l dy \right) + \alpha_- \exp \left(-i \int l dy \right) \right|^2}. \quad (\text{A7})$$

A similar relation can be derived for a superposition of waves in the vertical. Finally, using Eqs. (3) and (A3)–(A7), we get Eq. (8).

APPENDIX B

Technical Details of the Wave-Based Coordinate Diagnostics

The setup of the $s - r$ coordinate is described in section 3 and shown in Fig. 2. We denote the location of $s - r$ coordinate grid points in $y - z$ space by $[y_s(y, z), z_s(y, z)]$. To obtain $[y_s(y, z), z_s(y, z)]$, we integrate the following set of equations for a set of equally spaced emanation latitudes (y_0):

$$\frac{dy_s}{dt} = V_{ay} \quad (\text{B1})$$

$$\frac{dz_s}{dt} = V_{az} \quad (\text{B2})$$

$$\frac{ds}{dt} = 1 \quad (\text{B3})$$

$$\frac{dr}{dt} = 0 \quad (\text{B4})$$

subject to the following initial conditions at $t = 0$:

$$\begin{aligned} z_s(t = 0) &= 2, & y_s(t = 0) &= y_0, \\ s(t = 0) &= 0, & r(t = 0) &= y_0. \end{aligned} \quad (\text{B5})$$

After each time step we need to calculate \mathbf{V}_a at the new location (y_s, z_s) , which we do by interpolation from adjacent grid points. Here r equals the latitude at which the \mathbf{V}_a integral line emanates at the bottom ($z = 2$ scale heights), and s is the time it takes to reach the given location $[y_s(y, z), z_s(y, z)]$ from the initial location $(y_0, 2)$. In the new coordinate system \mathbf{V}_a is nondivergent (rather, the physical volume of a grid box varies along the coordinate), and constant s lines are spaced proportionally to $|\mathbf{V}_a|$ in geometric space.

The Jacobian of the transformation is very simple to calculate in $s - r$ space:

$$J = \frac{\partial y_s}{\partial r} \frac{\partial z_s}{\partial s} - \frac{\partial y_s}{\partial s} \frac{\partial z_s}{\partial r}. \quad (\text{B6})$$

In the time-dependent model run we calculate the $s - r$ coordinate in the same way, taking into account that \mathbf{V}_a is now a function of time. Since our model integration time step is larger than the coordinate integration time step, we first interpolate the \mathbf{V}_a field in time.

The wave activity budget calculations involve transforming A to the new coordinate. The transformation of a scalar field onto the new $s - r$ grid is simply an interpolation of these fields from the regular y, z model grid onto the irregular y_s, z_s grid.

APPENDIX C

Application to Observations

The observations we use are from the National Aeronautics and Space Administration Goddard Space Flight

Center stratospheric dataset, which is based on satellite retrievals in the stratosphere and radiosonde data in the troposphere. There are 18 levels between 1000 and 0.4 mb, 9 of them at or above 100 mb. The latitudinal resolution is 2° . Since the observations analyzed in this paper are the same as in HL, the reader is referred there for more details.

Harnik and Lindzen (2001, manuscript submitted to *Ann. Geophys.*) verified that the coarse vertical resolution of the satellite observations are capable of reasonably resolving the vertical structure of stratospheric planetary waves, by simulating the retrieval process in our model. Also, to check the effect of low resolution of the observations on the present calculations, we sample the geopotential height and the basic state of our model at a resolution roughly similar to the observational product (9 levels between 2 and 8 scale heights, equally spaced in our sample but not in observations), and interpolate back to high resolution for our calculations. Comparing $A(y, z)$ and $AJ(s - r)$ to the original high-resolution version, we find that the sampled-interpolated version captures the fact that AJ is constant, except near nodes in A , where the high-resolution calculations are problematic also. The effect of sampling is to spread the node in the vertical, over 1–2 observational grid points. This can have a large effect on \mathbf{V}_a especially if we have a node in midstratosphere, where the vertical resolution of the observations is quite low.

For the observations we use a spherical coordinate version of the calculations of appendix B. The wave activity equation in spherical coordinates is derived the same way as in the β plane (see section 2), and is exactly like Eq. (1), only the definitions of A , \mathbf{F} , $\nabla \cdot \mathbf{F}$, and D are different

$$A = \frac{a^2 \rho \cos \varphi}{2 \bar{q}_\varphi} q'^2, \quad (\text{C1})$$

$$F_\varphi = -a \rho \cos \varphi \overline{v'v'}, \quad (\text{C2})$$

$$F_z = -a \rho \cos \varphi \frac{gf}{T_o N^2} \overline{v'T'}, \quad (\text{C3})$$

$$\nabla \cdot \mathbf{F} = \frac{\partial F_z}{\partial z} + \frac{1}{a \cos \varphi} \frac{\partial (F_\varphi \cos \varphi)}{\partial \varphi}, \quad (\text{C4})$$

where φ is latitude.

To calculate the $s - r$ coordinate, we define the wave activity velocity as in the β plane [Eq. (3)], and integrate (C5) along with Eqs. (B2)–(B4):

$$\frac{dy_s}{dt} = \frac{V_{ay}}{a} \quad (\text{C5})$$

subject to the same initial conditions (B5) where $y_o = a\varphi_o$.

REFERENCES

- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Aris, R., 1962: *Vectors, Tensors, and the Basic Equations of Fluid Mechanics*. Dover, 286 pp.
- 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., 1968: Planetary waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci.*, **25**, 984–1002.
- Edmon, H. J., Jr., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen–Palm cross-sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616; corrigendum: **38**, 1115.
- Grose, W. L., and A. O'Neill, 1989: Comparison of data and derived quantities for the middle atmosphere of the Southern Hemisphere. *PAGEOPH*, **130**, 195–212.
- Harnik, N., and R. S. Lindzen, 2001: The effect of reflecting surfaces on the vertical structure and variability of stratospheric planetary waves. *J. Atmos. Sci.*, **58**, 2872–2894.
- Hirota, I., and Y. Sato, 1969: Periodic variation of the winter circulation and intermittent vertical propagation of planetary waves. *J. Meteor. Soc. Japan*, **47**, 390–402.
- 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.
- Karoly, D. J., and B. J. Hoskins, 1982: Three dimensional propagation of planetary waves. *J. Meteor. Soc. Japan*, **60**, 109–123.
- Lighthill, M. J., 1978: *Waves in Fluids*. Cambridge University Press, 504 pp.
- Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.*, **27**, 871–883.
- Palmer, T. N., 1982: Properties of the Eliassen–Palm flux for planetary scale motions. *J. Atmos. Sci.*, **39**, 992–997.
- Plumb, R. A., 1985: An alternative form of Andrews' conservation law for quasi-geostrophic waves on a steady, nonuniform flow. *J. Atmos. Sci.*, **42**, 298–300.
- , 1986: Three-dimensional propagation of transient quasi-geostrophic eddies and its relationship with the eddy forcing of the time-mean flow. *J. Atmos. Sci.*, **43**, 1657–1678.
- Randel, W. J., 1987: A study of planetary waves in the southern winter troposphere and stratosphere. Part I: Wave structure and vertical propagation. *J. Atmos. Sci.*, **44**, 917–935.
- , D. E. Stevens, and J. L. Stanford, 1987: A study of planetary waves in the southern winter troposphere and stratosphere. Part II: Life cycles. *J. Atmos. Sci.*, **44**, 936–949.
- Sobel, A. H., and C. S. Bretherton, 1999: Development of synoptic-scale disturbances over the summertime tropical northwest Pacific. *J. Atmos. Sci.*, **56**, 3106–3127.
- Swanson, K. L., P. J. Kushner, and I. M. Held, 1997: Dynamics of barotropic storm tracks. *J. Atmos. Sci.*, **54**, 791–810.