The Effect of Reflecting Surfaces on the Vertical Structure and Variability of Stratospheric Planetary Waves

NILI HARNIK AND RICHARD S. LINDZEN

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

ABSTRACT

The effects of an upper-stratospheric reflecting surface on the vertical structure of stratospheric planetary waves are considered. A diagnostic of the basic-state wave propagation characteristics, which is particularly useful for determining the existence and location of turning surfaces for meridional and vertical propagation, is developed. The diagnostic used is a more accurate indicator of wave propagation regions than the index of refraction because it diagnoses meridional and vertical propagation separately.

The diagnostic is tested on a series of simple models, both steady state and time dependent. It is found that the stratospheric waveguide sets the meridional wavenumber of the waves, regardless of the characteristics of their tropospheric forcing, making it easier to understand the effects of damping and turning surfaces on the vertical structure of the waves. The diagnostic is then applied to observations of the Southern Hemisphere winter of 1996. It is shown that the differences in vertical wave structure between middle and late winter can be explained as a linear response to the seasonal evolution of the basic state, which involves a formation of a reflecting surface in late winter. It is also shown that on daily timescales wave–mean flow interactions cause significant changes in the basic-state propagation characteristics for periods of a few days. These changes, along with the time variations in the forcing of the waves, are responsible for the observed daily timescale variations in wave structure. The fact that the observed evolution of the waves and the basic state are consistent with linear or quasi-linear wave theory (depending on the timescale looked at) supports the applicability of the theory, as well as the validity of the observations.

1. Introduction and motivation

In this work we study how the basic-state wave propagation characteristics, and in particular the geometry of reflecting surfaces, affect the variability in the vertical structure of stratospheric planetary waves in a simple model and in observations of the Southern Hemisphere winter.

The most common theory for stratospheric planetary waves is linear wave theory. Charney and Drazin (CD; 1961) were the first to treat stratospheric perturbations as vertically propagating Rossby waves forced from below. Charney and Drazin showed that the waves can propagate vertically if the basic-state zonal mean wind is westerly, and below a critical value that depends on the zonal wavenumber. Their formulation essentially divides the stratosphere into wave propagation and wave evanescence regions, separated by a *turning surface,* associated with a critical wind value.

The formation of a CD turning surface has been invoked to explain observed minima in wave activity during winter in the Southern Hemisphere, which occurs when the jet is strongest (Plumb 1989), and the episodic

minima in wave activity during Northern Hemisphere winter (e.g., Hirota and Sato 1969). A turning surface will inhibit propagation into the stratosphere if it forms close enough to the tropospheric source; however, if the turning surface forms in the upper stratosphere, reflection will change the wave's structure rather than inhibit it. Sato (1974) pointed out that observed waves in the Northern Hemisphere often have a node in the geopotential height structure at 100 mb; they attributed this to downward reflection of the waves from a CD turning surface. Matsuno (1970) summarized the results of his model study by drawing a picture of a stratospheric *wave cavity,* where waves that are guided up the stratospheric jet axis are bounded from above by a CD turning surface. He then suggested this can result in stratospheric wave amplitudes peaking in the stratosphere. In this study we seek to determine whether reflecting surfaces do indeed affect wave structure and evolution in the stratosphere, by specifically and explicitly diagnosing reflecting surfaces and their effect on wave structure. We focus in this study on the Southern Hemisphere, because the dynamics are considered to be more linear, compared to the Northern Hemisphere, and the analyses correspondingly more straightforward.

Theoretically, we expect a reflecting surface to have

Corresponding author address: Nili Harnik, Room 54-1626 Massachusetts Institute of Technology, Cambridge, MA 02139.
E-mail: nilih@cirrus.mit.edu

a large effect on wave structure, both the steady-state and transient evolution. The direction of wave propagation is directly related to the tilt of the wave phase lines in the vertical–zonal plane (Charney and Drazin 1961). Also, strong downward reflection can cause wave amplitudes to peak in the stratosphere, which will otherwise grow exponentially with height due to the density effect.

There is observational evidence that the vertical structure of the waves changes, and the observed changes are suggestive of downward reflection. We will focus on two kinds of variations of particular interest: a daily timescale variation in the phase tilt of the waves and a seasonal timescale change in the amplitude structure. Since wave activity is observed to be episodic, with episodes lasting a few weeks (e.g., Hirota and Sato 1969; Hartmann 1976), daily phase structure changes are observed within a given episode, while seasonal amplitude changes are in fact a variation in wave structure from one episode to the other.

On seasonal timescales, the peak of wave geopotential height amplitude is observed to shift poleward and downward toward the end of winter (Hartmann 1976, 1977; Randel 1992). The peak of the Southern Hemisphere polar night jet is also observed to move downward and poleward during the Southern Hemisphere winter (e.g., Harwood 1975; Hartmann 1976, 1977; Mechoso et al. 1985; Shiotani and Hirota 1985; Randel 1988). Hartmann (1977) first noted this correlation between the seasonal evolution of wave structure and the jet, and attributed it to a linear response of the waves to the changes in the basic state. We want to test this more explicitly. As mentioned above, downward reflection can cause wave amplitudes to peak in the stratosphere. Damping, however, can also cause wave amplitudes to peak where they do, and it is interesting to determine which mechanism is more relevant. Estimates of the radiative component of thermal damping, which is relatively well known, suggests damping rates are strongest in the upper stratosphere (Dickinson 1969), and may vary during the season (Pawson et al. 1992). It should be noted, however, that other forms of damping, which are poorly known, may also be important.

On daily timescales, we observe a variability that involves changes in the wave phase tilt with height. For example, Hartmann (1976) analyzed satellite data from one Southern Hemisphere winter and found a case where the phase lines of wavenumber 1 went from a westward to an eastward phase tilt with height over a period of a few days, during the decay stage of the wave life cycle. Randel et al. (1987) also found a case in observations when an upward propagating wave shifted its propagation direction downward for a few days. Figure 7, shows another example of such a tilting of phase lines over a few days, during a Southern Hemispheric wave event that lasted from 18 July to 19 August 1996. This kind of change also occurred earlier during the wave event, and a couple of times during September 1996.

As with the seasonal timescale variations, there are a few processes that could explain this variability. If a reflecting surface exists, time variations in the source of the wave can cause variations in the phase tilt of the wave, because the relative magnitudes of the upward and downward components of the wave will change with time. Given the highly transient nature of the forcing, such structure changes could occur quite often. In the absence of a reflecting surface, however, wave packets will only propagate upward, and the wave will have a westward phase tilt with height at all times. Another possibility is that the basic state varies, and reflecting surfaces form temporarily, resulting in temporary downward reflection.

Given that the basic state undergoes episodic deceleration and acceleration in the upper stratosphere (e.g., Hartmann et al. 1984; Shiotani and Hirota 1985), this mechanism is plausible. We therefore want to determine the location and time evolution of reflecting surfaces.

To determine whether reflecting surfaces exist for a given basic state, we develop a new diagnostic. The need for a new diagnostic arises because the traditional one for wave propagation, namely, the index of refraction (Charney and Drazin 1961; Matsuno 1970), is not exactly what we want. In order to deduce the location of turning surfaces for vertical propagation from the index of refraction, we need to subtract the meridional wavenumber contribution. This was pointed out in a number of papers (e.g., Dickinson 1968; Matsuno 1970; Simmons 1974), but without a general method of doing so. The difficulty lies in the basic state being nonseparable in latitude and height. While the index of refraction is a function of the basic state and a prescribed zonal wavenumber and phase speed, there is no clear way to prescribe the meridional wavenumber.

In section 2 we present the relevant linear wave propagation theory and introduce our diagnostic. We then test it and demonstrate its robustness on a series of steady-state model runs (section 3). In section 4 we apply our diagnostic to observations, during periods when the waves exhibit variations in their vertical structure. Section 4a examines the seasonal timescale, while sections 4b and 4c examine daily timescales. We also use a time-dependent model run to test our findings with observations and describe these results when necessary. We summarize and discuss our results in section 5. The observations we use are described in appendix A and the formulation of our diagnostic in spherical coordinates is described in appendix C. We test the validity of the WKB approximations in appendix B.

2. The wavenumber diagnostic

In this section we describe a diagnostic of the wave propagation characteristics of a two-dimensional basic state. We start by reviewing conventional theory in order to explain the need for a new diagnostic and its relation to the index of refraction diagnostic.

We use the quasigeostrophic (QG) equations in log-pressure coordinates, linearized around a zonal mean basic state. We start with a QG β-plane model and extend our results to spherical coordinates for use with observations. The derivation and notation are standard, and can be found in Andrews et al. (1987). Much of the theory was developed by Charney and Drazin (1961).

Assuming a normal mode structure in longitude for the geopotential height perturbation ϕ:

$$\phi(y, z) = \varphi(y, z)\, e^{ik(x-ct)}, \tag{1}$$

where k and c are the zonal wavenumber and phase speed of the perturbation, the pseudo–potential vorticity (PV) conservation equation becomes

$$\frac{f_o^2}{\rho}\frac{\partial}{\partial z}\left(\frac{\rho}{N^2}\frac{\partial\varphi}{\partial z}\right) + \frac{\partial^2\varphi}{\partial y^2} + \left(\frac{\overline{q}_y}{U-c} - k^2\right)\varphi$$

$$= \frac{i}{k(U-c)}\left[\frac{f_o^2}{\rho}\frac{\partial}{\partial z}\left(\frac{\rho\alpha}{N^2}\frac{\partial\varphi}{\partial z}\right) + r\nabla^2\varphi + \frac{\partial r}{\partial y}\frac{\partial\varphi}{\partial y}\right], \tag{2}$$

where x, y, z, t are the zonal, meridional, height, and time coordinates; ρ is density; f_o is the Coriolis parameter; and $U, N^2, \overline{q}_y$ are the zonal mean basic-state wind, Brunt–Väisälä frequency, and meridional PV gradient, respectively. Here N^2 and ρ are functions of height only. We use Newtonian damping: $\mathcal{H} = -\alpha(\partial\varphi/\partial z)$, and Rayleigh friction: $(\nabla \times \mathcal{F}) \cdot \hat{\mathbf{k}} = -(r/f_o)\nabla^2\varphi - (1/f_o)(\partial r/\partial y)(\partial\varphi/\partial y)$.

Equation 2 constitutes our linear wave equation, and we concentrate on forced perturbations for which k and c are specified and c is a real number.

The discussion in terms of wave geometry is best illustrated for the simple case where we have no meridional variations, and no damping. Setting $\partial/\partial y$, α, and r to zero and transforming φ to a new weighted geopotential height variable ψ, we get

$$\varphi = e^{z/2h}\sqrt{N^2}\,\psi, \tag{3}$$

$$\psi_{zz} + \left[\frac{\overline{q}_y}{U-c} - k^2 + F(N^2)\right]\frac{N^2}{f_o^2}\psi = 0, \tag{4}$$

where $\rho = \rho_o e^{-z/h}$, h is the density-scale height, and

$$F(N^2) \equiv f_o^2\frac{e^{z/2h}}{N}\frac{\partial}{\partial z}\left[\frac{e^{-z/h}}{N^2}\frac{\partial}{\partial z}(e^{z/2h}N)\right]. \tag{5}$$

Equation (4) is a wave equation, and the index of refraction for vertical propagation is

$$n_{\mathrm{ref}}^2 = \frac{N^2}{f_o^2}\frac{\overline{q}_y}{U-c} - k^2\frac{N^2}{f_o^2} + F(N^2)\frac{N^2}{f_o^2}. \tag{6}$$

Under WKB conditions,[1] the solution is of the form

$e^{\pm i\int mdz}$, where m is the vertical wavenumber that satisfies $m^2 = n_{\mathrm{ref}}^2$. In regions where m^2 (hence n_{ref}^2) is positive we have wave propagation, and in regions where it is negative we have exponential behavior (wave evanescence). The wave geometry framework separates the basic state into wave propagation and wave evanescence regions, and in the stratosphere, these regions are separated by a turning surface (where $m^2 = 0$). Under WKB conditions, the full solution (in wave propagation regions) can be written in terms of a superposition of an upward- and a downward-propagating wave:

$$\phi \propto \left[\frac{A}{\sqrt{m(z)}}e^{i\int m(z)\,dz} + \frac{B}{\sqrt{m(z)}}e^{-i\int m(z)\,dz}\right]e^{ik(x-ct)}e^{z/2h}, \tag{7}$$

where A and B are integration constants multiplying the upward- and downward-propagating waves, respectively, and the factor $m^{-1/2}$ is necessary to satisfy conservation of wave activity.

It is easy to show from (7) (and assuming WKB) that if the amplitude of the upward-propagating wave is larger (meaning we have a net upward flux of wave activity), the phase lines will tilt westward with height, and for a given basic state the tilt will increase the larger the net upward flux is. A downward flux will be associated with an eastward phase tilt with height, and a standing wave (where the upward and downward waves are equal and there is no net vertical flux) will have vertical phase lines. Note that if there is no downward reflection ($B = 0$), the amplitude of the wave varies as $m^{-1/2}$, but if A and B are nonzero, the vertical structure of the wave (characterized, e.g., by the amplitude and phase of ϕ) is not set entirely by m because the relative magnitudes of the upward- and downward-propagating components (A, B) also matter. Also, under WKB conditions, which imply that the wavenumber and the basic state change more slowly than the wavelength divided by 2π, the vertical wavenumber can be calculated from the solution as follows:

$$\mathrm{Re}\left(\frac{\psi_{zz}}{\psi}\right) = -m^2. \tag{8}$$

Note that if A and B are nonzero, $\mathrm{Im}(\psi_z/\psi)$ does not equal m.

The picture is more complicated when the basic state varies both in latitude and height in a nonseparable way. Equation (4) is now

$$\frac{\partial^2\psi}{\partial z^2} + \frac{N^2}{f_o^2}\frac{\partial^2\psi}{\partial y^2} + n_{\mathrm{ref}}^2\psi = 0, \tag{9}$$

where an additional meridional propagation term has been added. In analogy to the one-dimensional case, an approximate WKB solution of the following form can be assumed:

---

[1] WKB is valid if the wavelength of the solution is smaller than the scale of variations of the medium, allowing us to make the separation between a wave and a slowly varying basic state. See appendix B for a quantitative discussion.

$$\phi \propto [(A_1 e^{i\int ldy} + A_2 e^{-i\int ldy})e^{i\int mdz}$$
$$+ (B_1 e^{i\int ldy} + B_2 e^{-i\int ldy})e^{-i\int mdz}]e^{ik(x-ct)}, \quad (10)$$

where l and m can either be real or imaginary (in which case we have an exponential rather than a wave behavior). The amplitude functions A_1, A_2, B_1, B_2, which are functions of y and z, have to satisfy conservation of wave activity locally, but unlike the one-dimensional case, this condition does not determine these coefficients uniquely. In addition, the meridional and vertical wavenumbers (l and m) satisfy the following dispersion relation:

$$m^2 + \frac{N^2}{f_o^2}l^2 = \frac{N^2}{f_o^2}\left[\frac{\overline{q}_y}{U - c_r} - k^2 + F(N^2)\right] \equiv n_{\text{ref}}^2. \quad (11)$$

The main problem now is how to separate the index of refraction into vertical and meridional propagation. Past approaches include specifying the zonal mean wind in a way that renders the equation exactly separable (Dickinson 1968) or approximately separable (Simmons 1974), or ignoring shear terms in the wave equation, which also makes it separable (Schoeberl and Geller 1977). Note that Dickinson's (1968) approach resulted in waves being guided up weak westerlies rather than up the jet axis (Matsuno 1970). We take a different approach to separating n_{ref}^2, which is diagnostic. Rather than making approximations, we solve the equations using a model, and *diagnose* the wavenumbers, as follows:

$$\text{Re}\left(\frac{\psi_{zz}}{\psi}\right) = -m^2, \quad (12)$$

$$\text{Re}\left(\frac{\psi_{yy}}{\psi}\right) = -l^2. \quad (13)$$

Equation (10) and the derivation of relations (12)–(13) from it depend on WKB being valid, which is not obvious for the stratosphere. In appendix B we show that WKB is not too bad an assumption. Also, if we have no damping and a pure real phase speed, and we use Eqs. (12) and (13) as a *definition* of l^2 and m^2, Eq. (11) is exactly satisfied regardless of whether WKB holds or not. This can be seen by dividing Eq. (9) by ψ and equating the real parts. In the next section we test our diagnostic using a series of β-plane model runs. We will show that the wavenumbers diagnosed from the steady-state wave solution to a given basic state using Eqs. (12) and (13) are meaningful in terms of wave propagation characteristics, at least for stratospheric basic states.

3. Model results

The analysis of the previous section suggests that the wavenumbers calculated using Eqs. (12)–(13) show the wave geometry of the *basic state*. For this to be true, the wavenumbers need to be insensitive to things like damping, or the latitudinal shape of the forcing. In this part we test this by conducting a series of steady-state model runs with various parameter values to see how the meridional and vertical wavenumbers change. We also use these runs to understand what determines how n_{ref}^2 is divided into meridional and vertical wavenumbers [Eq. (11)].

The model we use is described in detail in Harnik (2000), and we present the basic features here. Our model is quasigeostrophic, linear, and on a β-plane centered at 55°S, and it spans 10.3 deformation radii (about 12 000 km) in latitude, and 15 scale heights (105 km) in the vertical. There are 64×71 grid points, latitude by height. We solve Eq. (2) for φ, given the basic state, damping, and a bottom forcing. We specify a basic-state wind U to be a function of latitude and height, and a basic-state temperature that varies only with height. We have a sponge layer at the top, to approximate a radiation condition, and a sponge layer at low latitudes to include the effect of either absorption at a tropical critical surface or radiation to the other hemisphere. The sponge layers are a combination of Newtonian cooling and Rayleigh damping with equal coefficients, which we specify such that waves are essentially absorbed before they reach the model boundaries, where we set the perturbation to zero. We force the model by specifying the zonal wavenumber and phase speed, and φ at the bottom, which is at two scale heights (14 km). The forcing is constant with time, except for a zonal propagation with the prescribed phase speed. In the runs we display here, the phase speed is taken to be zero, and the zonal mean wind is positive everywhere, hence we do not have a critical surface. In runs where we do have a critical surface, we add a small constant damping of perturbation PV, to ensure numerical convergence.

The model is not computationally demanding and can be run many times. The results we will show, unless otherwise stated, apply to all the runs we did. We find a large variety of vertical structures, similar to the observed, and this variability is related to the existence of turning surfaces in the stratosphere. The overall picture we get is that the stratospheric jet acts to guide wave activity from the troposphere upward. The equatorial boundary of the waveguide is leaky because of the wave sink at the equator (represented by the sponge layer). This picture of a leaky waveguide was suggested quite early on in studies of stratospheric waves (Dickinson 1968; Matsuno 1970), but the consequences of such a configuration have not been demonstrated in detail. We will show that the waveguide sets the meridional wavenumber, defined by Eq. (13). As a result, the meridional wavenumber is insensitive to the forcing characteristics.

Figure 1 shows the basic state and damping distribution we use. All β-plane model results are presented on a nondimensional grid, where height is in scale heights ($h = 7$ km) and latitude is in deformation radii [$L_d = (N_o h/f_o) = 1190$ km, $N_0^2 = 4.1 \times 10^{-4}$ s^{-2}]. The basic state, which is specified analytically, is characteristic of early Southern Hemisphere winter. The jet tilts equatorward

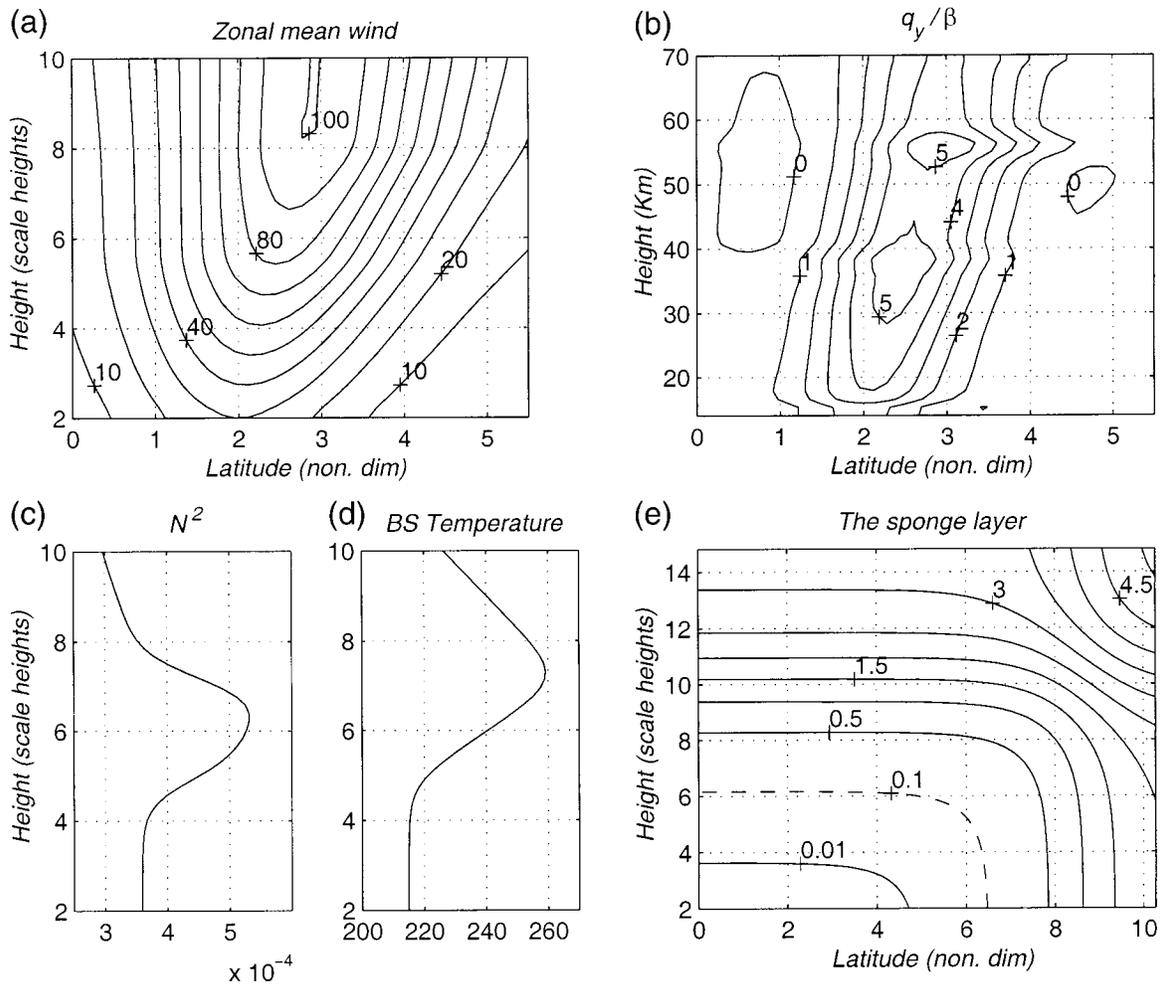FIG. 1. The basic state used in the β-plane model run. (a) Basic-state zonal mean wind (m s^{-1}). (b) Meridional PV gradients, in units of $\beta = 1.3 \times 10^{-11}$ (s m)$^{-1}$. (c) Basic state N^2 in s^{-2}. (d) Basic-state temperature (K). (e) The damping coefficient used (day^{-1}) for both Newtonian cooling and Rayleigh damping. For reference, the vertical coordinate of the $\overline{q}_y$ plot is in log-pressure height (km) while the rest are in scale heights. Note also that the sponge layer plot shows a larger domain than the other three plots. The latitude is in units of radii deformation ($L_d = 1190$ km).

and widens with height, with maximum winds of about 100 m s^{-1}. Winds are constant with height above 10 scale heights. The PV gradient field has a ridge that follows the jet and negative regions on both sides of the jet. The basic-state temperature is specified to look like a standard midlatitude winter profile in the stratosphere, and it changes smoothly to a constant value above 12 scale heights. The Brunt–Väisälä frequency is calculated from temperature. We use Newtonian damping and Rayleigh friction with equal coefficients, which increase from zero at the bottom at high latitudes (small y) to a value of 2– 5 day^{-1} at the top and equatorial boundaries, with most of the increase above 42 km ($z = 6$) and equatorward of latitude $y = 6.5$.

Figures 2 and 3 show the geopotential height perturbations, n_{ref}^2, l, and m [Eqs. (6) and (12)–(13)] for stationary wavenumbers one and two. Wavenumber values

are only shown for propagation regions. The shape of the forcing is also shown in Fig. 2 (top-left plot). We find that the meridional wavenumber is similar for waves one and two. This clearly shows a waveguide that is oriented vertically with an equatorward tilt, following the jet axis. Correspondingly, larger n_{ref}^2 values for wave 1 compared to wave 2 result in larger m^2 values (i.e., a larger propagation region). We see a horizontal reflecting surface ($m^2 = 0$) for both waves, but at a much lower altitude for wave 2. As a result, wave 2 has a much smaller amplitude (the geopotential height peaks roughly at the height of the reflecting surface). Also, the phase lines of wave 2 are vertical, while wave 1 exhibits some upward propagation (westward phase tilt with height). Wave 1 also has turning surfaces at six and seven scale heights, resulting in a region of evanescence between six and seven scale heights. Since this region is much smaller
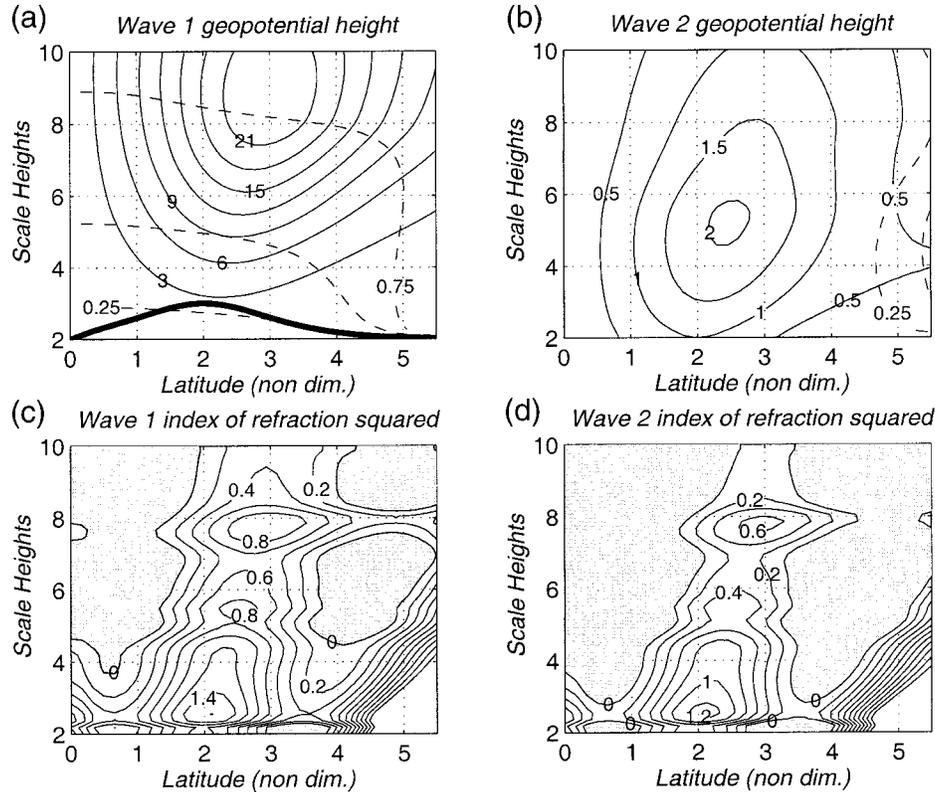
FIG. 2. (a) Wave 1 and (b) two stationary geopotential height amplitude (solid, arbitrary units) and phase (dashed, in units of π), for the basic state of Fig. 1. Also shown in (a) is the amplitude of forcing at the bottom (thick line). The magnitude is zero at the sides and one in the middle and there is no phase variation with y. The index of refraction [Eq. (11)], for (c) wave 1 and (d) 2, in units of $(7 \text{ km})^{-2}$. Only positive values are contoured, negative regions are shaded.

than a vertical wavelength, the wave tunnels through it easily. Note that the n_{ref}^2 fields do not reveal the existence of a reflecting surface, and the explicit diagnosis of the vertical wavenumber is necessary to show it. This is an important point since n_{ref}^2 is often used in the literature as an indicator of vertical wave propagation. While in observations we generally find that negative m^2 values occur because of small or negative n_{ref}^2, the two fields often exhibit important differences. We will discuss this further in section 4.

The insensitivity of the meridional wavenumber to the zonal wavenumber is interesting. This suggests that the structure of the waveguide is set by the basic state, in this case by the PV gradient ridge, and that the waveguide, in turn, sets the meridional wavenumber regardless of the forcing characteristics. We test this result further by varying the forcing characteristics. Varying the zonal phase speed also results in very similar meridional wavenumbers, as long as the phase speed is not large enough for the critical surfaces to be inside of the waveguide. Also, we force the model with a forcing constant with latitude as well as a point source, and find that the meridional wavenumber above one scale height

from the bottom is similar to Fig. 3. While the magnitude of the response changes for the different forcings, its latitudinal shape above one scale height does not. Within one scale height, the perturbation spreads out to the sides of the waveguide and reflects back, losing the information about the shape of the forcing at the bottom. The conclusions from this and many other runs, show that the stratospheric waveguide acts to determine the meridional wavenumber of the perturbation and that Eqs. (12)–(13) are indeed a robust diagnostic of these wavenumbers. While the existence of the waveguide is not surprising based on past studies (Matsuno 1970; Karoly and Hoskins 1982), its robustness (as seen from the meridional wavenumber), and our ability to diagnose vertical propagation so clearly, is striking.

An important test of our diagnostic as a diagnostic of the *basic-state* propagation characteristics is its sensitivity to damping. We expect damping to affect the amplitude of the waves, and the relative magnitude of the upward and downward propagating components (i.e., the phase tilt of the waves with height), but not to affect the location of reflecting surfaces or the shape of the waveguide. Testing this is especially important since
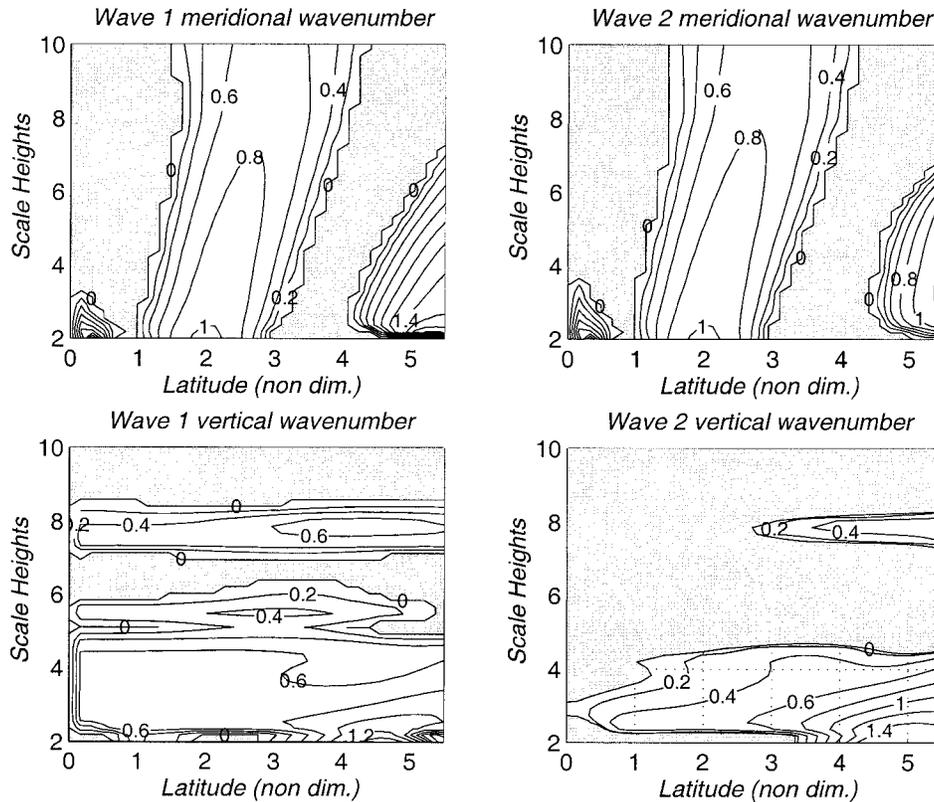
FIG. 3. (top) Meridional and (bottom) vertical wavenumbers [as defined by Eqs. (13) and (12)] for (left) the stationary wave 1 and (right) two perturbations of Fig. 2. Only propagation regions are contoured, and evanescence regions are shaded. Here m is scaled by $(7 \text{ km})^{-1}$ and l by $(1190 \text{ km})^{-1}$.

damping in the real atmosphere is not well known. We specify a few forms of damping, and vary the shape of the sponge layers, keeping the basic state and forcing constant, and test the sensitivity of our diagnostics. We find that as long as the damping is not too large or too rapidly changing,[2] it does not affect the wavenumbers of the solution, even when the vertical structure changes considerably. For example, when we lift the sponge layer and the lid up by five scale heights, and force the model with the same stationary wave 1 forcing, we get wavenumbers that are very similar to Fig. 3, with the $m^2 = 0$ reflecting surface remaining at eight scale heights. At the same time, downward reflection is much larger and the phase lines are much more vertical than in the lower sponge run.

One exception is when the basic state has no $m^2 = 0$ reflecting surface. In this case the sponge layer limits the amplitude of the wave and causes it to peak, and the wavenumber diagnostic shows a spurious reflecting surface at the edge of the sponge layer. It is easy to distinguish this from a real reflecting surface, because

it moves up when the sponge layer is raised. It is important to note that in the model run presented here (Fig. 3), the reflecting surface and the sponge layer are roughly at the same height for wave 1. Since the diagnosed reflecting surface does not move when the sponge layer is raised, but the wave amplitude peaks higher (because of the density effect), we conclude that the peak in wave amplitude is due to damping and not to reflection from the turning surface. We will seek to determine whether in observations waves peak because of downward reflection or damping.

To summarize, we show in this section that the meridional wavenumber is a robust feature of the basic state, while the vertical wavenumber is also a function of the zonal wavenumber and phase speed of the forcing (through the index of refraction). While damping affects the vertical structure, it hardly affects the vertical and meridional wavenumbers. The practical implication of these results is that we can use the wavenumbers of the steady-state wave solution to diagnose propagation characteristics of the *basic state*. Moreover, the wavenumbers cannot tell us much about actual wave structures, which depend not only on the propagation characteristics but on the damping and the time evolution of the forcing as well.

[2] By too quickly, we mean rapidly enough for WKB to be violated. Regions of sharp variations in damping will partially reflect the waves.

4. Observations

In this section we use our diagnostic to show that reflecting surfaces exist in the upper stratosphere, and we test whether observed variability in wave structure, both on seasonal and daily timescales, is consistent with downward reflection from such surfaces. For a given observed basic state, we diagnose the propagation characteristics and the location of reflecting surfaces using the steady-state linear wave solution. It is important to stress that even though we use a steady-state wave model, we can use it to diagnose the propagation characteristics of an *instantaneous* basic state. This is because our diagnostic is of the basic-state characteristics, not of the waves. For the diagnosed propagation characteristics to be meaningful for understanding the evolution of wave structure, we do, however, assume that the time variations of the basic state are slow enough for the response of the wave to be linear.

We use a spherical coordinate model with a corresponding version of the wavenumber diagnostics (described in appendix C). All other quantities we show are calculated from the observational data without the use of the model (e.g., zonal mean winds, wave geopotential height and temperature fields, n_{ref}^2, and the meridional PV gradients). Note that the wavenumber diagnostic [Eqs. (12) and (13)] is only meaningful if it is calculated from a wave in steady state, hence we do not apply it directly to observed waves. We also use a time-dependent version of our β-plane model with a specified basic state and forcing to test some of our results in an idealized setting. When needed, we present results from these runs.

We analyze observations from one randomly chosen Southern Hemisphere winter (1996). Since our goal is to establish whether reflecting surfaces play a role in stratospheric wave dynamics, as well as to see if our diagnostic approach works on observations, one winter suffices. However, a more comprehensive study of more years is needed to determine whether our results are characteristic of the winter stratosphere in general. We use the National Aeronautics and Space Administration Goddard Space Flight Center (NASA GSFC) 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°. The dataset is described in more detail in appendix A. All the figures based on observations are presented on log-pressure height, where a scale height of 7 km is used. The vertical grid is the same as for the observations.

The model we use for the wavenumber diagnostic spans the Southern Hemisphere, and for simplicity, we use the latitudinal resolution of the operational observational product. The vertical domain extends from 14–105 km (2–15 scale heights). We use the observed basic state, interpolated in the vertical to the model grid, and kept constant above 0.4 mb (in order to apply a sponge

layer). For simplicity, we specify a temperature that varies only with height, by taking an average of the observed temperature over 40°–70° latitude.[3] We use similar damping as in the β-plane model, meaning we have a sponge layer at the top and at the equator, with the equatorial sponge reaching roughly $-20°$S. We also test the sensitivity of our results to the model damping by repeating many of our runs with the lid and sponge layer raised by 35 km. The results we show are not sensitive to sponge-layer height.

To verify that our diagnostic works in spherical coordinates, we repeat the runs of section 3, with analytically specified basic states and different forcing and damping distributions. The most important difference from a β plane is that in spherical coordinates the index of refraction is infinite at the equator, which causes the waveguide to be much less separated from the equatorial propagation region compared to Cartesian coordinates. This causes waves to refract equatorward (Karoly and Hoskins 1982; Matsuno 1970), resulting in smaller wave amplitudes than in the β-plane model. Even taking this into consideration, we find that the meridional wavenumber remains insensitive to the parameters used for forcing and damping.

a. The differences between midwinter and later winter wave structure

As mentioned in the previous section, we analyze observations from the randomly chosen winter of 1996. We believe, however, that our findings are relevant to other years, because the seasonal evolution of the basic state and wave structure during 1996 are similar to the evolution during the Southern Hemisphere winter of other years. There were two major wavenumber 1 events during this winter, one during 18 July–19 August (referred to as the midwinter wave) and the other in September (referred to as the late winter wave). Figure 4 shows the latitude–height amplitude and phase structures of the time mean waves in these two periods (both geopotential height and temperature).[4] The temperature amplitude of the midwinter wave has a single peak in the stratosphere while the late winter wave has two. Correspondingly, the geopotential height amplitude peaks at least 10 km lower in September.[5] Also, the phase tilt with height is smaller in late winter. The down-

---

[3] A comparison with runs using the full two-dimensional temperature field gives very similar results (similar enough given the uncertainties in the observations).

[4] The phase of the wave θ is defined as follows. The wave 1 components of the geopotential height and temperature fields, at each latitude and height, are expressed as $A \cos(\lambda + \theta)$, where λ is the longitude in radians, A the amplitude, and θ the phase. Note that $-\theta$ is the longitude of the ridge. Since the waves shown are stationary, their phase does not average out to zero over time.

[5] During most days in August, the geopotential height increases with height throughout the stratosphere. The time mean geopotential height peaks at 48 km because of a few days in which the wave changed structure—see section 4b.
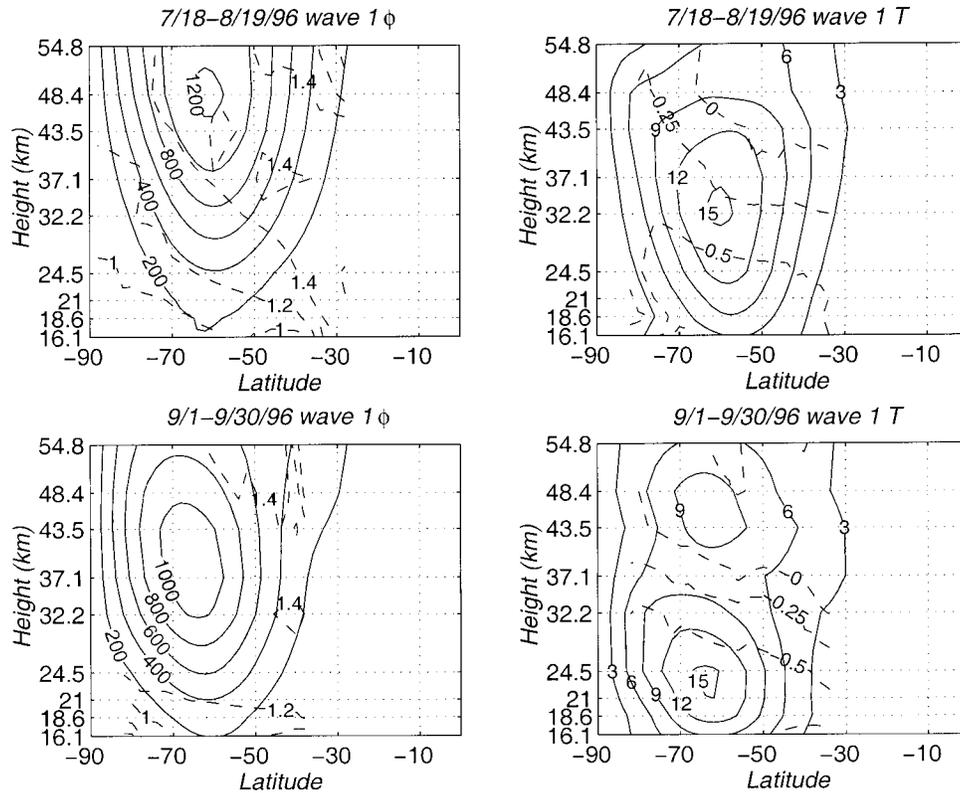
FIG. 4. (left) Time-averaged wave 1 geopotential height and (right) temperature amplitude (solid) and phase (dashed), for (top) 18 Jul–19 Aug 1996 and (bottom) 1–30 Sep 1996. Geopotential height amplitude is in meters, temperature amplitude in degrees kelvin, and phase in units of π. Phase is plotted only in latitudes where the waves are noticeable. The vertical grid is the observation grid in log-pressure height (km), using a scale height of 7 km. Time averaging was done on the amplitude and phase separately.

ward shift in geopotential height maximum was pointed out by Hartmann (1976, 1977), and is also seen in the stratospheric climatology of Randel (1992). The corresponding formation of a double-peaked structure in temperature toward the end of winter is also found in the references above.

Figure 5 shows the observed zonal mean wind averaged over each of the two wave events. The polar night jet peak varied from a magnitude of 70 m s^{-1} at around 40 km, 50°S in July–August to 50 m s^{-1} at 35 km, 60°S in September. A similar weakening along with a downward and poleward movement is a commonly observed feature of the Southern Hemisphere winter (e.g., Hartmann 1976; Shiotani and Hirota 1985; Randel 1988).

As discussed in the introduction, this seasonal evolution of wave structure has been attributed to changes in the basic state (Hartmann 1977), but without explicit evaluation. In particular, it is not obvious whether the peaks of the observed waves occur where they do because of damping in the upper stratosphere or because of downward reflection from a turning surface. In the latter case we should observe a turning surface that moves downward during winter, to account for the

downward shift in the maximum of the waves. We test this by applying our diagnostic to the time-averaged basic states of the two wave events of 1996.

The meridional and vertical wavenumbers of the corresponding steady-state solutions obtained using a wave 1 stationary forcing that is constant with latitude (geopotential height of 100 m) are shown in Fig. 5. The wavenumber diagnostics reveal a large difference in the propagation characteristics. In August there is vertical propagation in most of the domain (evanescent regions, where $m^2 < 0$, are shaded), while in September we see a turning surface at around 38.5 km in midlatitudes, where the geopotential height peaks. This suggests that as the winter progresses, the basic state changes with a turning surface forming in the upper stratosphere. Correspondingly, the vertical structure of the waves changes, and the geopotential height amplitude peaks lower down during September. Note that even though there is only a hint (at 60°S) of a reflecting surface in the upper stratosphere in July–August, the time mean waves during this period peak in the upper stratosphere. A look at the daily wave geopotential height fields shows that during most of this period waves peak at the top or above the domain of observations, but on certain days
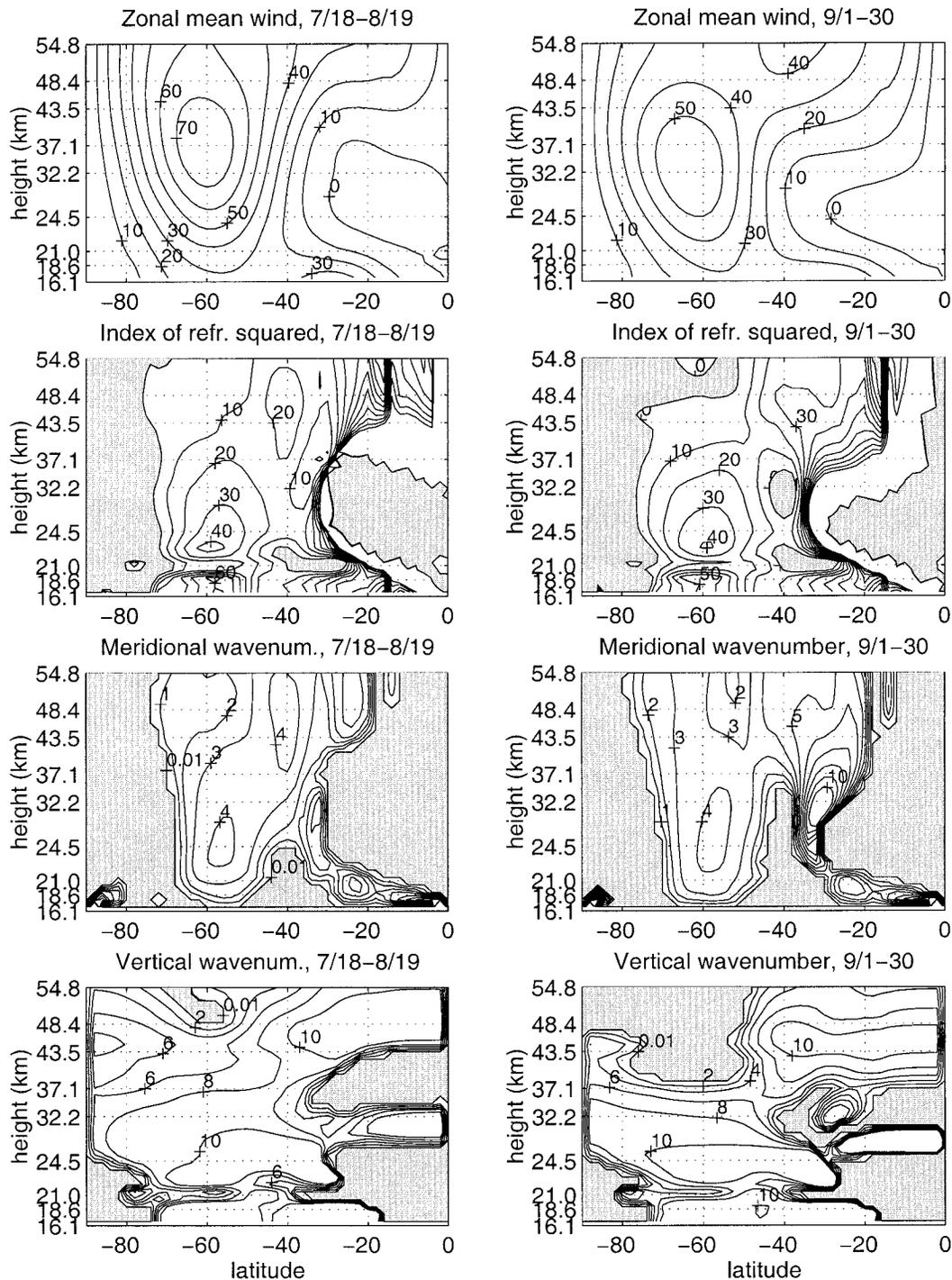
FIG. 5. (top to bottom) Observed time mean of zonal mean wind, the index of refraction squared [normalized by N^2, the term in brackets, Eq. (C2)], and the meridional and vertical wavenumbers calculated from the steady-state model solution, for (left) 18 Jul–19 Aug and (right) 1–30 Sep 1996. Wind in meters per second meridional wavenumber in inverse radians and vertical wavenumber in 10^{-5} m^{-1}. Negative values in the lower three rows are shaded. See text for details.
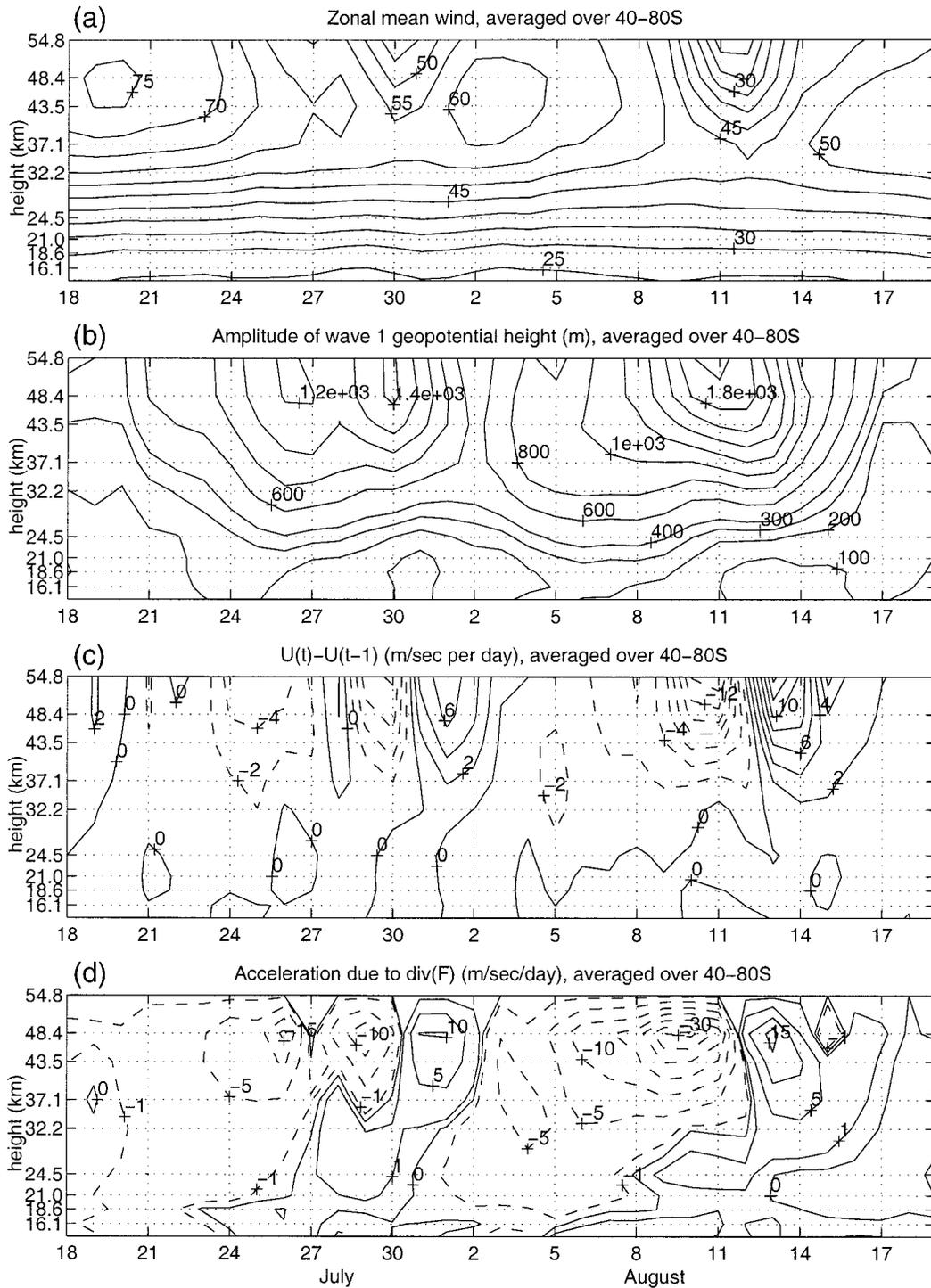
FIG. 6. Height–time sections (18 Jul–19 Aug 1996) of the 40°–80°S average of (a) zonal mean wind (contour interval of 5 m s^{-1}). (b) Wave 1 geopotential height amplitude (contours at 0, 100, and 200–2000, in jumps of 200 m). (c) The change in zonal mean wind over 1 day [$U(t) - U(t - 1)$]. Contour interval is 2 m s^{-1}, negative values dashed. (d) The acceleration due to wave driving: $(\nabla \cdot \mathbf{F})/a_e \rho \cos\varphi$. Contours at ± 0, ± 1, and ± 5–30 in jumps of 5 m s^{-1} day^{-1}, negative values dashed. All quantities, except the wave geopotential height amplitude are volume averaged over latitude (weighted by $\cos\varphi$).
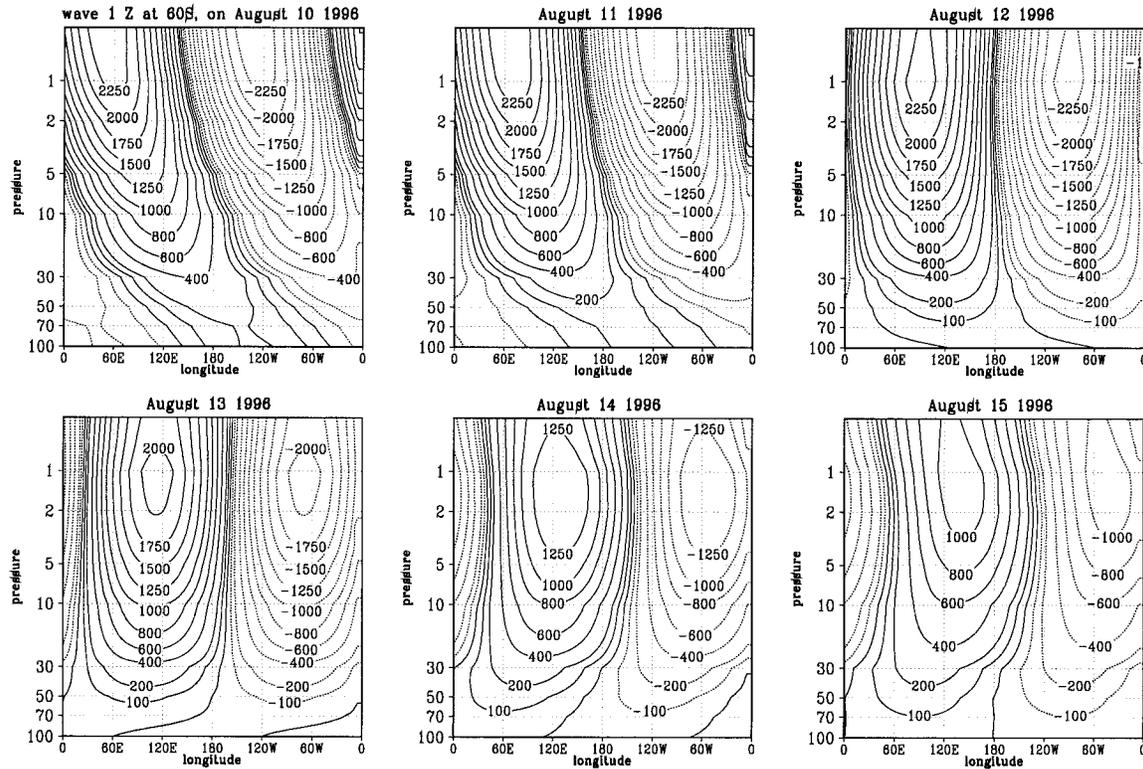
FIG. 7. Daily longitude–height cross sections at 60°S of wave 1 geopotential height for 10–15 Aug 1996. Contour intervals are at 0, ±100, ±200, ±400, ±600, ±800, and ±1000–2500, in jumps of 250 mb. Negative values are dashed. The vertical grid is the observational grid in millibars (100–0.4 mb).

the waves peak in the upper stratosphere. We will show in the next section that during those days a reflection surface actually forms. We therefore need to be careful in the interpretation of the time mean structure.

It is interesting to compare m^2 with the index of refraction squared, calculated from the time mean basic states (also shown in Fig. 5). While in July–August both m^2 and n^2_{ref} are positive throughout the stratosphere, in September $m^2 < 0$ above 35 km, while $n^2_{\text{ref}} < 0$ above 48 km. This difference, while only quantitative, is important since m^2 is clearly more consistent with the vertical structure of the observed time mean waves. Looking at n^2_{ref} it is hard to explain why the peak in geopotential height and the node in temperature are at 40 km, but it is very clear from the vertical wavenumber.

b. Daily timescale variations: July–August

As we discussed in the introduction, there is observational evidence of daily timescale changes in wave structure that look like downward reflection. We expect such changes to occur if there is a time varying forcing with a reflecting surface. Using a time-dependent version of our β-plane model, we do indeed observe such structure changes (a tilting of phase lines to a vertical, then eastward tilt with height) under these conditions.

In the following two sections we will show a few observed cases from the winter of 1996. As was shown in section 4a, in 1996 there is no time-mean reflecting surface in July–August, and there is one in September. This suggests the source of variability in wave structure may be different during these two periods. We therefore discuss them separately, starting with July–August.

Figure 6 shows the time series of wave 1 geopotential height amplitude, zonal mean wind, and its acceleration, and the EP flux divergence term in the zonal momentum equation [Eq. (3.5.2.a) in Andrews et al. 1987], for the period of 18 July–19 August 1996, averaged over the latitudes 40°–80°S. There are strong decelerations of the zonal mean wind at the end of July and in mid-August. A comparison of the EP flux divergence and the observed acceleration shows a strong relation between the two, with the former being much larger and preceding the latter by a day or two. Such episodic changes in the zonal mean wind, with corresponding $\nabla \cdot \mathbf{F}$ patterns has been observed before (e.g., Hartmann et al. 1984; Shiotani and Hirota 1985). During both deceleration periods the amplitude of the wave starts decreasing roughly when minimum winds are reached. Figure 7 shows the evolution of the vertical structure of wave 1 geopotential height at 60°S during the second deceleration period. A similar structure change is observed during the earlier
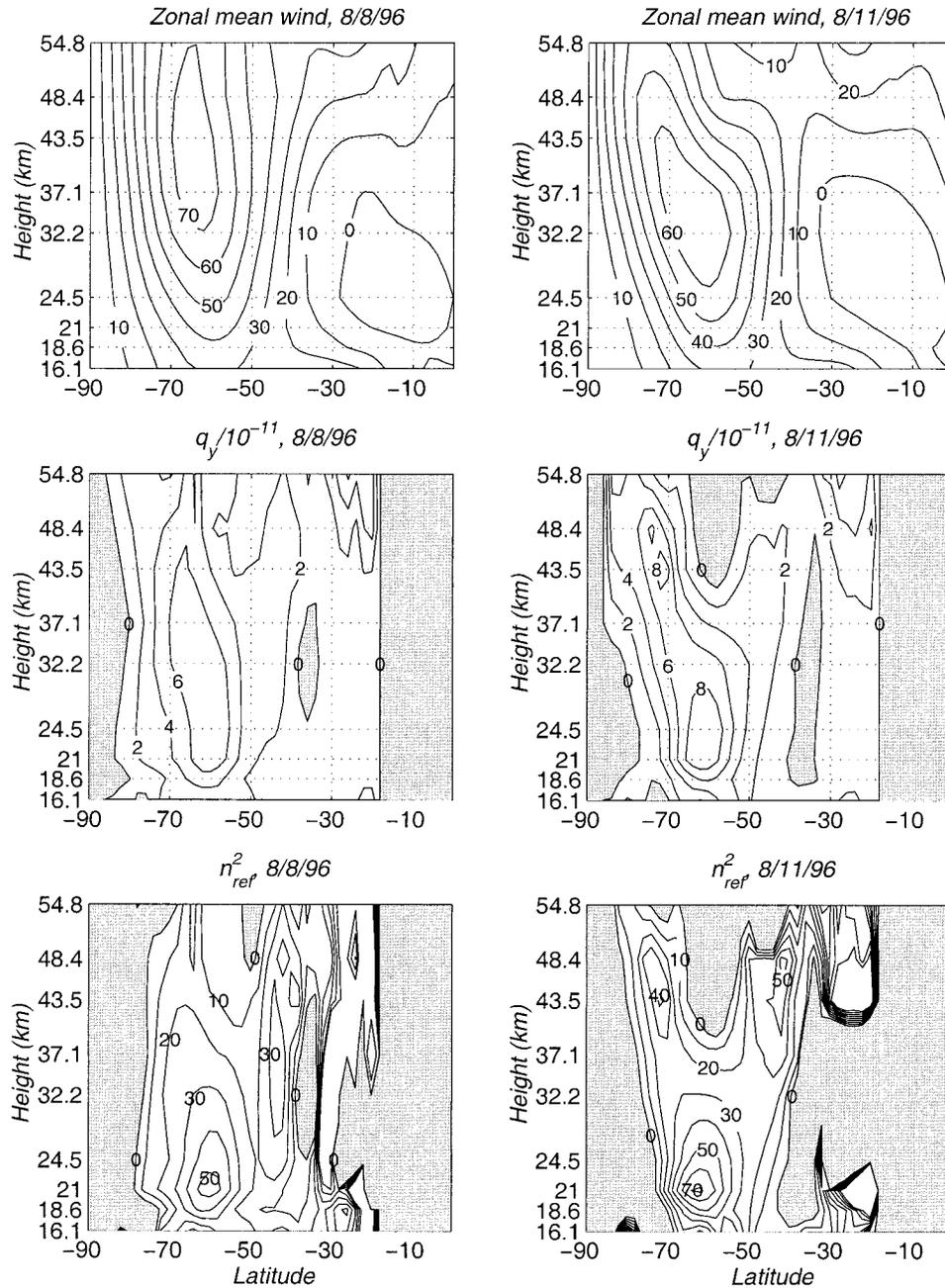
FIG. 8. (top) Zonal mean wind (contour interval of 10 m s^{-1}). (middle) Meridional PV gradient (units of 10^{-11} s^{-1} m^{-1}, contours at 0–8 in jumps of 2, negative regions are shaded). (bottom) Index of refraction squared (normalized by N^2, the term in brackets, Eq. (C2), contour interval is 10, negative regions are shaded), on (left) 8 Aug and (right) 11 Aug 1996.

deceleration as well. We see that along with the decrease in amplitude there is a change in the vertical structure of the wave. The wave has the structure of an upward-propagating wave (westward phase tilt with height) during most of the wave episode (before 29 July, 5–11 August). At the time of maximum deceleration and a few days afterward (31 July–2 August, 12–16 August),

the phase of the wave tilts into the vertical (characteristic of a standing wave in the vertical), and eventually tilts eastward with height (characteristic of a downward-propagating wave), suggestive of a turning surface forming.

Figure 8 shows the observed zonal mean wind, meridional PV gradient [$\overline{q}_y$, Eq. (C5)] and index of re-
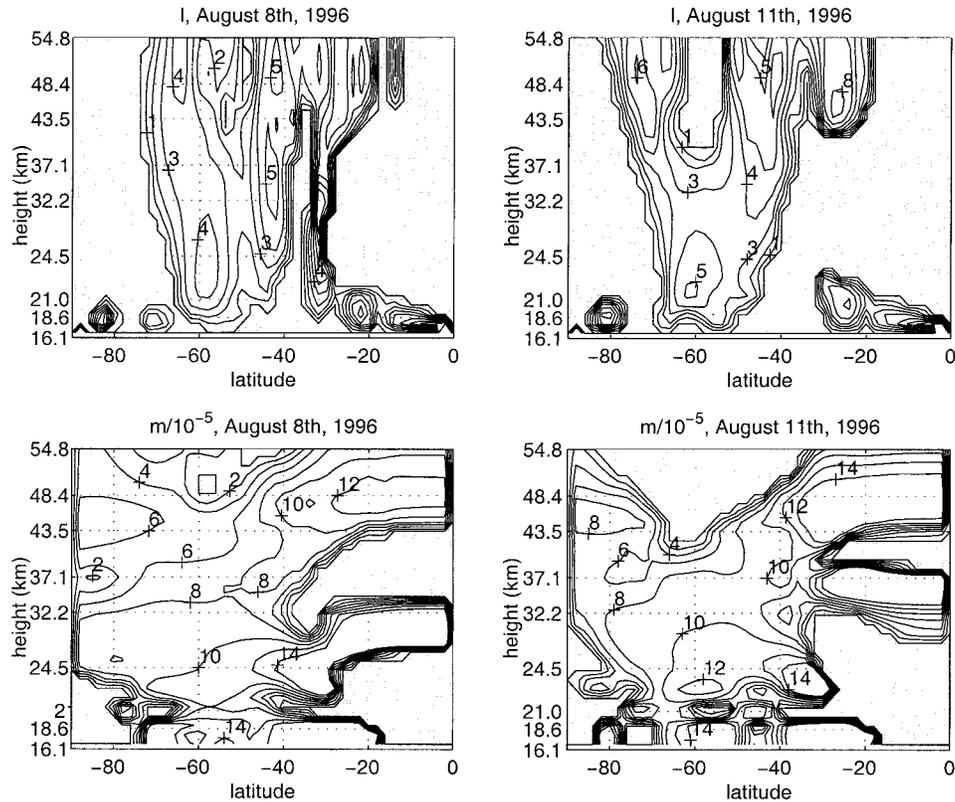
FIG. 9. Latitude–height sections of (top) meridional (contours at 0.01, and 1–8 in jumps of 1 rad^{-1}) and (bottom) vertical (contours at 0.01, and 2–14 in jumps of 2, 10^{-5} m^{-1}) wavenumbers, for the observed basic states on (left) 8 Aug and (right) 11 Aug 1996. Regions of evanescence (negative l^2, m^2) are shaded.

fraction squared [n_{ref}^2 Eq. (C2)] for stationary wave 1, on 8 and 11 August. The strong deceleration on 8–12 August results in the formation of a region of negative $\overline{q}_y$, and n_{ref}^2 in the upper stratosphere, between 55° and 65°S, on 11–12 August. In a one-dimensional model, the formation of a region of negative n_{ref}^2 would lead to downward reflection. The present case is more complicated, since the PV gradients become negative only in a midlatitude region, and essentially the waveguide in the upper stratosphere splits into poleward and equatorward branches (the latter is more pronounced). It is unclear if this would cause the wave to reflect downward, or to bypass the negative $\overline{q}_y$ region and propagate up one or both of the branches of the split waveguide.

Figure 9 shows the meridional and vertical wavenumbers calculated from the steady-state response to a steady forcing that is constant with latitude, for the basic states of Fig. 8. Regions of evanescence are shaded. The interesting result is that on 11 August, the vertical wavenumber is negative in a large region in the upper stratosphere (above 45 km in midlatitudes) while on 8 August it is positive almost everywhere in the domain of observations. This indicates clearly that a reflecting surface ($m^2 = 0$ surface) formed sometime between 8 and 11 August. At the same time, the meridional wavenumber

on 11 August has a very clear split waveguide structure, while on 8 August it looks like a waveguide about to split in two in the upper stratosphere. The emerging picture is that the observed time evolution is a wave–mean flow interaction, where the waves grow sufficiently to decelerate the mean flow and form a reflecting surface. The wave in turn responds qualitatively linearly to these basic state changes by reflecting downward.

The role of wave forcing is also important. To get an eastward phase tilt with height, which implies net downward propagation, we need the source of the wave to decrease with time while there is a turning surface. Figure 10 shows the time evolution of the vertical wavenumber (thick lines denote the turning surface), the amplitude of the wave at 150 mb (considered as the wave forcing) and the phase of the wave, averaged over a range of latitudes, for the July–August wave event. We see that during most of this period, there is no turning surface in the domain of observations. During both deceleration events a turning surface forms, along with a decrease in forcing, resulting in a tilting of the wave phase lines. The observed coincidence between the decrease in the forcing amplitude and the formation of a reflecting surface is interesting and will be discussed later on.

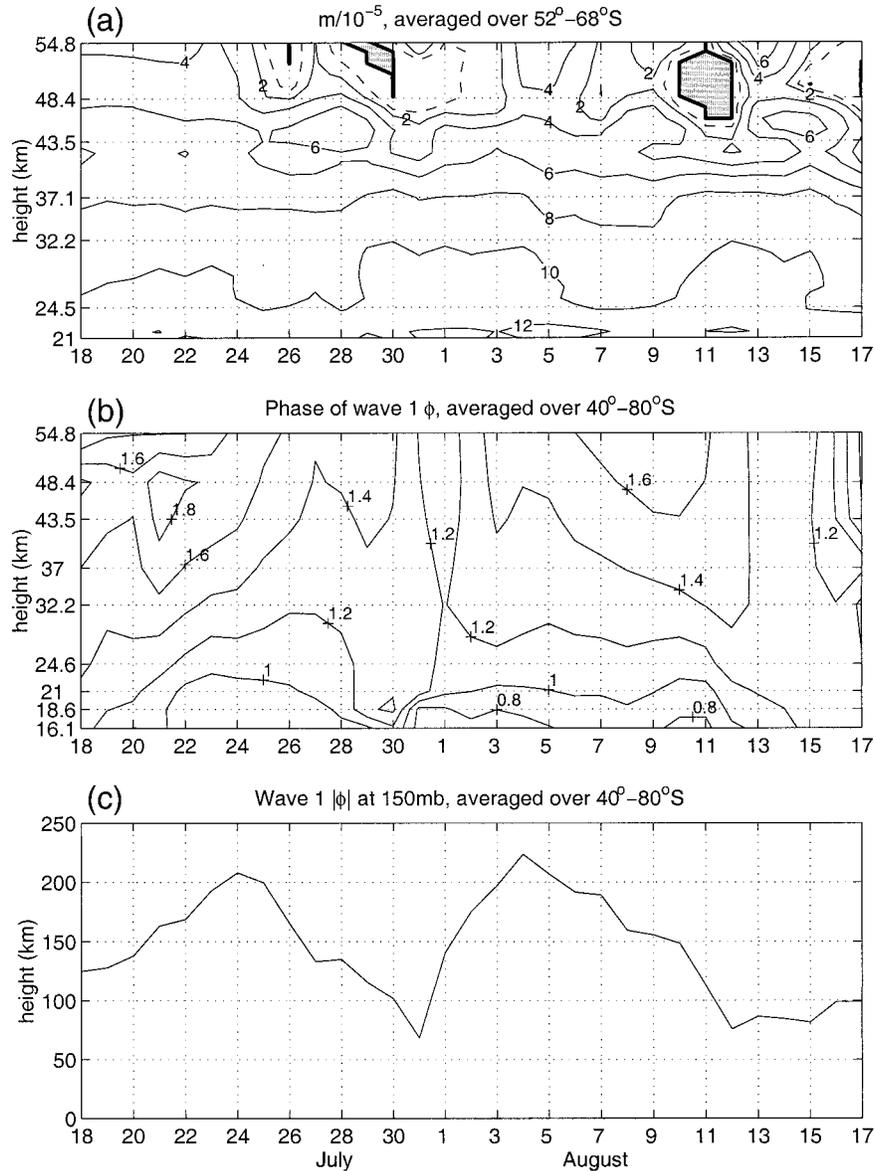The time evolution of the basic state, forcing, and wave

FIG. 10. Height–time sections (18 Jul–19 Aug 1996) of a latitudinal average of (a) vertical wavenumber m, calculated from the wave 1 steady-state solution to the daily observed basic state (10^{-5} m^{-1}). (b) The phase of wave 1 geopotential height (units of π). (c) The amplitude of wave 1 geopotential height at 150 mb (in m). Geopotential height quantities are averaged over 40°–80°S and m is averaged over 52°–68°S. Evanescent regions of m are shaded, and the thick and dashed lines denote values of 0.01 (the turning surface) and 1, respectively.

structure are so far consistent, according to quasi-linear theory. We need, however, to make sure the timescales of the observed changes are consistent with the vertical group propagation timescale. The changes in wave structure brought about by the formation of a turning surface are such that the time it takes the wave to reach a vertically aligned phase structure should equal the vertical group propagation time. When the bottom forcing decreases simultaneously with the formation of a turning surface, the relevant time is for the wave to reach an eastward phase

tilt with height (meaning the time it takes the wave to reach a vertical structure is half the vertical group propagation time). From Fig. 7 we see that it takes the wave roughly 2 days to tilt to a vertical position, meaning the vertical propagation time should be around 4 days. We estimate the vertical group propagation timescale by using integral lines of $\mathbf{F}/A$, where $\mathbf{F}$ is the Eliassen–Palm flux, and A is the QG wave activity (Edmon et al. 1980). A more detailed discussion of diagnostics that are based on integral lines of $\mathbf{F}/A$ is given in Harnik (2001). In the case
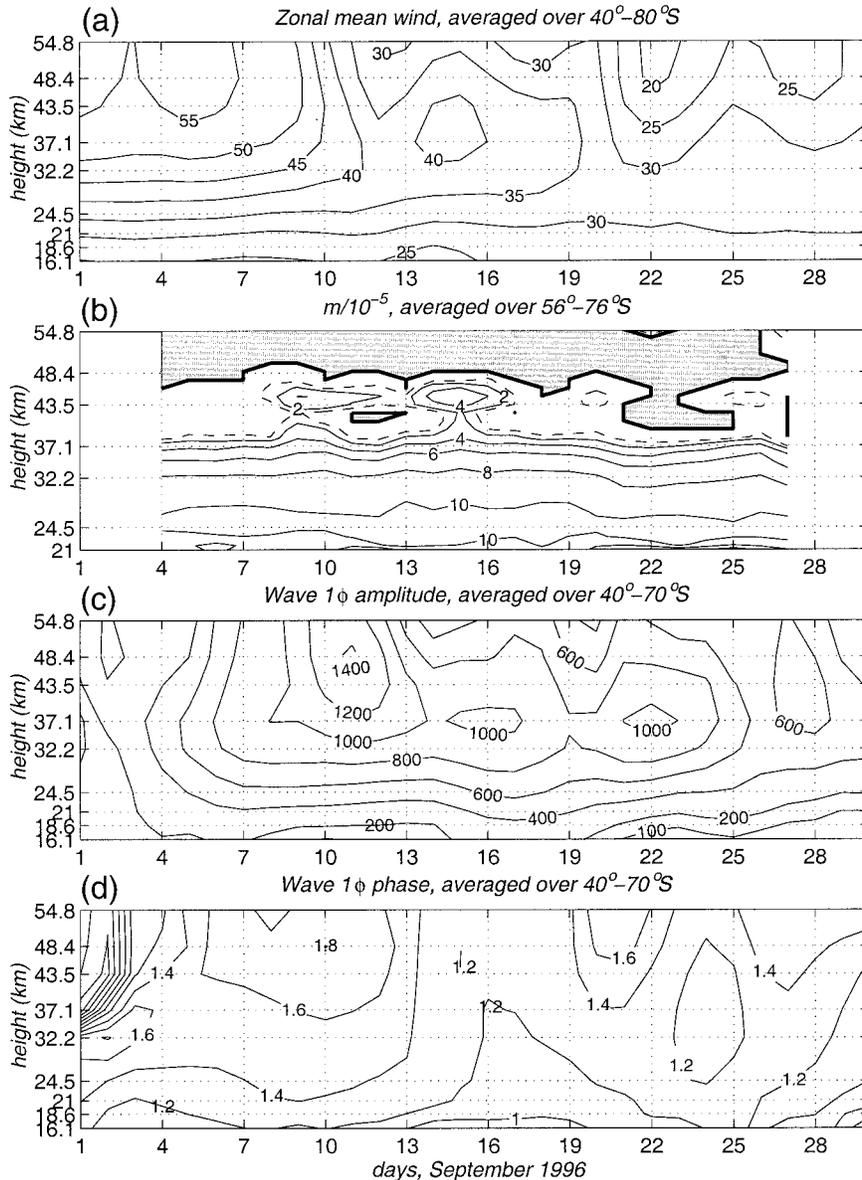
FIG. 11. Height–time sections (1–30 Sep 1996) of a latitudinal average of (a) zonal mean wind (m s^{-1}). (b) Vertical wavenumber m, calculated from the wave 1 steady-state solution to the daily observed basic state (10^{-5} m^{-1}). (c) Wave 1 geopotential height amplitude m. (d) Wave 1 geopotential height phase (units of π). Wind is volume averaged (weighted by cosφ), over 40°–80°S. Geopotential height quantities are averaged over 40°–70°S, and m is averaged over 56°–76°S. Evanescent regions of m are shaded, and the thick and dashed lines denote values of 0.01 (the turning surface) and 1, respectively. Vertical grid is the observation grid.

of a pure plane Rossby wave, this velocity equals the group velocity. Since **F** represents the net flux of wave activity, it changes when the wave reflects down. We therefore choose the period before reflection starts, when the wave is growing, to do these estimates. Starting at the tropopause on 5 August, it takes 3–4 days to reach the turning surface at around 42 km. This supports our picture that the changes in wave structure are due to downward reflection from a turning surface.

Finally, there are a few other mechanisms that may cause variability in wave structure, which we briefly examine here. One possibility is a superposition of a stationary and a propagating wave (e.g., Salby and Garcia 1987; Madden 1983; Lindzen et al. 1982). Such a superposition should result in periodic changes, with a period of the propagating wave. We test this by subtracting the time mean wave from the total wave field. We use a time mean of the entire period, or of 30 July 30–12 August,
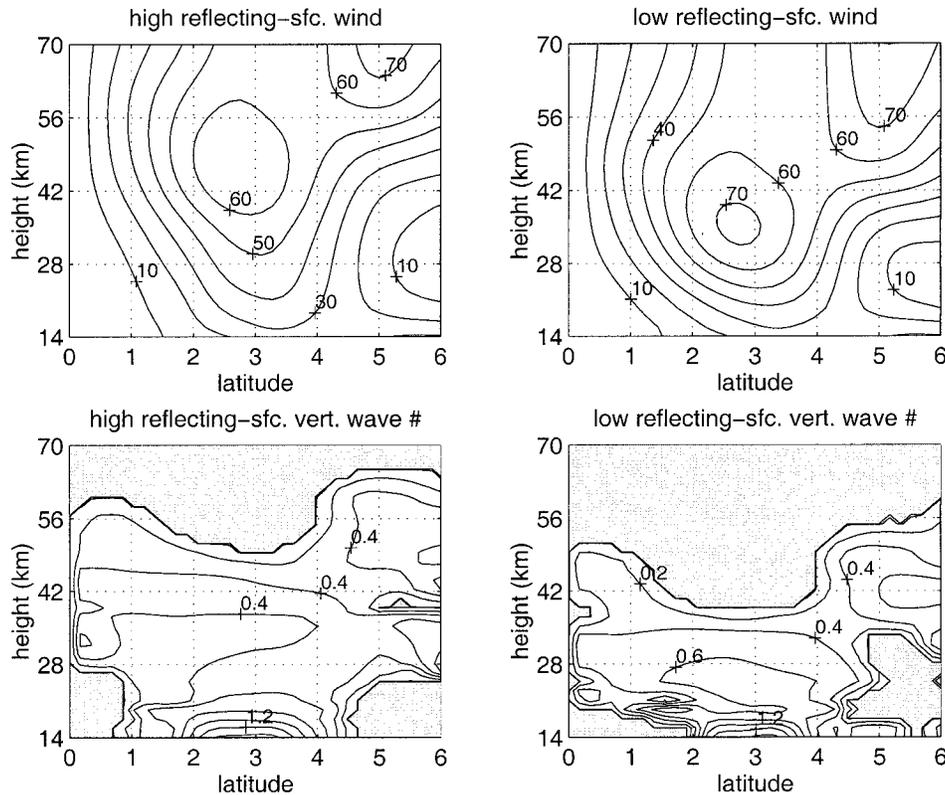
FIG. 12. Characteristics of the (left) initial and (right) final basic states of the model run described in the text. (top) Zonal mean wind. (bottom) The vertical wavenumber of the steady-state response. Only regions of propagation ($m^2 > 0$) are contoured (within the 0.01 contour line), and evanescent regions are shaded. Latitude, which is in units of 1190 km, increases equatorward (0 is the Pole).

which is the period between the 2 days on which the wave is vertical. In both cases, the remaining transient wave does not look at all like a traveling mode. The phase speed is not the same at all levels, and it changes with time. Another possibility is that the waves simply tilt with the shear as their source is shut off. Experiments with a time-dependent version of our β-plane model where we force a wave and then shut off its source show the perturbation continues to propagate vertically, and unless there is a reflecting surface, the phase tilt of the wave remains westward with height as the wave propagates into the sponge layer and decays there (see Harnik 2000, section 7.2.1 for more details). We also need to address the possibility that the observed structure changes are spurious, because they are based on satellite retrievals of temperature that have a coarse vertical resolution and are sampled asynoptically (Salby 1982). We have elsewhere established that the satellite observations are capable of resolving the observed wave structure and evolution (Harnik and Lindzen, manuscript submitted to *Ann. Geophys.*). We present more details in appendix A.

c. Daily timescale variations: September

In section 4a we showed that in September, a reflecting surface exists in the stratosphere at all times, which

means the time variations in wave forcing are sufficient to cause wave structure changes. In July–August 1996, on the other hand, a deceleration of the mean flow was necessary to get downward reflection. In this section we will show that wave–mean flow interactions do affect vertical wave structures during September 1996.

Figure 11 shows time–height plots of the observed zonal mean wind, the wave 1 vertical wavenumber calculated from the steady-state solution using the daily basic state, and the wave 1 geopotential height amplitude and phase. Wind is averaged over 40°–80°S, wave geopotential height over 40°–70°S, and the vertical wavenumber (m) over 56°–76°S (these latitudes were chosen to represent the values of m in middle–high latitudes, see, e.g., Fig. 5). There are two strong deceleration events during this period (9–12 September, 20–22), followed by acceleration (13–15 September, 19, 23–25). Wave 1 $\nabla \cdot \mathbf{F}$ (not shown) is strong enough to account for the zonal mean wind changes (although wave 2 $\nabla \cdot \mathbf{F}$, which is negligible before 12 September, is needed to account for the accelerations on 12–14 September). The wave generally has a westward phase tilt with height (an increase of the phase with height), but following both decelerations, the wave temporarily tilts to
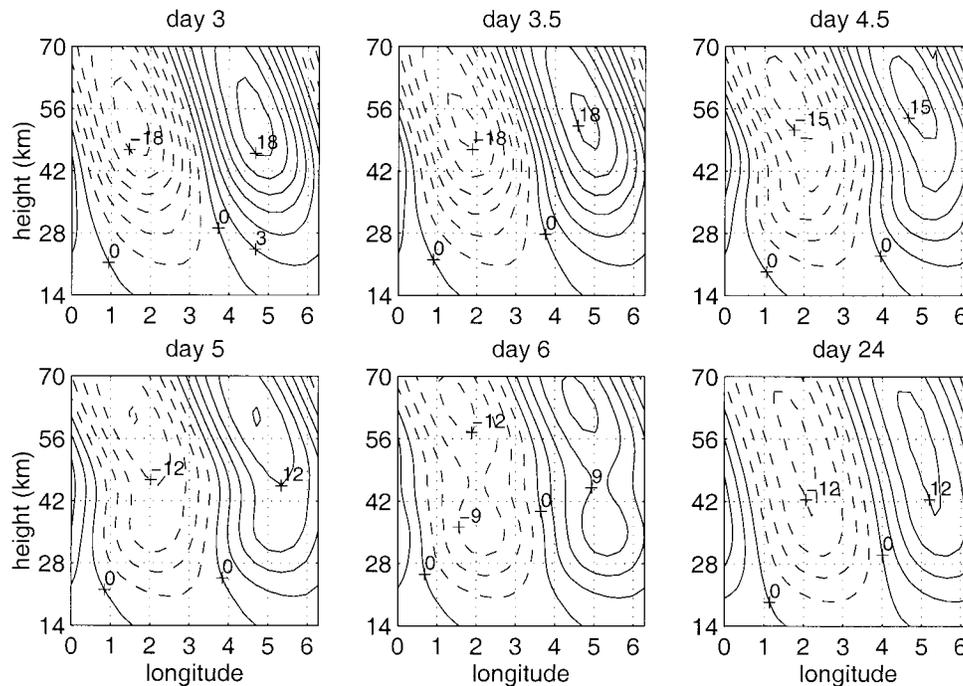
FIG. 13. Zonal–height sections of the geopotential height perturbation for days 3, 3.5, 4.5, 5, 6, and 24 in the model run of Fig. 12. Days 3 and 24 are essentially the initial and final steady-state wave fields. Sections are taken at a latitude of 2.5-radii deformation.

a vertical position (geopotential height phase becomes constant with height).

Looking at the time series of m in Fig. 11, we see that a turning surface exists within the observation domain throughout September, but that on 11–13 and 21–24 September the turning surface dips down and on 8–10 and 14–16 September it moves upward. Comparing the time evolution of m with the evolution of the wave phase tilt, we see that the wave phase tilt roughly follows m, such that a downward motion of the turning surface is followed by the wave reflecting downward and tilting vertically, and an upward motion of the turning surface is followed by the wave tilting more westward with height. Note that the observed wave forcing (amplitude near the tropopause) decreases with time within 3 days of the vertical tilting of the wave (and of the downward shift of the turning surface), therefore it could account for the observed structure changes. It is interesting, however, to see whether the observed variations in the height of the turning surface, which are not very large, also contribute.

To do this we use a time-dependent version of our β-plane model with a zonal mean wind that changes in time such that an existing turning surface shifts up or down as observed. To assure numerical stability, we add a constant damping on perturbation PV with a timescale of 25 days. Figure 12 shows the initial and final zonal mean winds, along with the vertical wavenumber calculated from the steady-state response, for one such run. The final state is the observed zonal mean wind of 6

September, which has a turning surface at 40 km. The initial state is constructed by stretching the 6 September wind field in the lowest 20 km over a region of 35 km, resulting in a basic state that has a turning surface at 50 km.[6] We vary the wind linearly between the initial and final states, and initialize the model with the steady-state response to the initial wind, while keeping the bottom forcing constant. The wind starts changing after model day 2 and reaches the final state on model day 3.5. We see that as the zonal mean wind changes, the turning surface moves down.

Figure 13 shows longitude–height sections of the wave geopotential height fields at a succession of times, as it responds to the changes in the basic state. Days 3 and 24 are quite close to the steady-state wave structure of the initial and final basic states. After the basic-state changes, the phase tilt with height decreases with time below 45 km, such that on days 4–5, the wave has vertical phase lines, after which the phase tilt increases and readjusts to the final westward-tilting steady-state solution

---

[6] We constructed the initial field this way because it results in a turning surface that is similar to the final state only higher. In particular, the wave geometry at and above the turning surface does not change. This is important since we need to make sure there is no wave propagation above the domain of observations. Note that our time-dependent model is on a β plane, hence the propagation characteristics for a given wind field can be different than in spherical coordinates. Despite the differences between the two coordinate systems, we expect the waves to behave qualitatively the same (see section 4).

(in steady-state damping at and above the turning surface will result in a small westward phase tilt with height). This supports the theory that a downward movement of the turning surface can cause a temporary downward reflection of the wave. Note that in Fig. 13 we do not get eastward phase tilt with height. This is because the forcing at the bottom is constant during this run. When we decrease the forcing at the time that the turning surface moves, we get eastward phase tilt with height.

Finally, the amount of damping affects the results, such that when the damping is very large, the downward-reflected wave decays rapidly and the phase lines become vertical only very close to the turning surface. This implies that the observed structure changes can occur only if the damping in the atmosphere is small enough. When we repeat our runs with a damping time of 6.25 days instead of 25 days, we get very little downward reflection that is evident only very close to the turning surface. Dickinson's (1969) estimates of thermal damping timescales are 20 days in the lower stratosphere and 2 days at 50 km. More careful studies, however, are needed to obtain meaningful numbers, because the form of damping matters. For example, we find that damping on momentum is much more effective in reducing wave amplitudes than damping on temperature. The momentum damping in the real atmosphere (mostly due to gravity wave drag and Rossby wave breaking) is less known than radiative damping. Another complication is that thermal damping is dependent on the vertical scale of the perturbation (Fels 1982), and this scale can vary considerably when the waves reflect downward, especially near the reflecting surface. Also, the linear form of damping is convenient for modeling but is not obviously a good parameterization of actual damping, especially on momentum.

5. Discussion and conclusions

In this paper we establish that reflecting surfaces form in the upper stratosphere, and that downward reflection from these surfaces can have a large effect on the vertical structure of the waves and its time evolution, both in models and in observations.

To diagnose the wave propagation characteristics of the basic state, and in particular, to determine the existence and geometry of turning surfaces, we use a steady-state wave model. We show that the vertical wavenumber calculated form the steady-state wave solution is a good diagnostic of the vertical propagation characteristics of the *basic state*. Our diagnostic is more suited for determining vertical propagation than the commonly used index of refraction because the latter includes a contribution from the meridional wavenumber, which our diagnostic separates out.

We analyze observations from one Southern Hemisphere winter (1996), and find that downward reflection affects the vertical structure on a seasonal timescale as well as on daily timescales. We discuss the two time-scales separately. On the seasonal timescale, we find that the evolution of the jet during the winter is accompanied by the formation of a reflecting surface. This explains the evolution of vertical wave structure, from a wave that propagates through the stratosphere to a wave that is reflected downward. These changes are most notable in the temperature field, which has one peak in the stratosphere in midwinter and two in late winter. Correspondingly, the peak in geopotential height forms above the reflecting surface and shifts downward toward the end of winter. There is evidence that this evolution is not specific to the year we analyzed. The downward shift of the jet is a well-known feature of the Southern Hemisphere winter (e.g., Harwood 1975; Hartmann 1976, 1977; Mechoso et al. 1985; Shiotani and Hirota 1985; Randel 1988). The formation of a double-peaked temperature structure toward late winter is shown in Hartmann (1976), as well as in the climatology of Randel (1992). We also looked at the daily wave fields in Southern Hemisphere winters of other years, and found the formation of a double-peaked temperature structure toward late winter in wave 2 as well as wave 1.

On daily timescales, we find the evolution to be quasi-linear, meaning the waves grow and change the mean flow, and then respond in a linear way to these changes. We show a clear case in mid-August 1996, of wave 1 decelerating the flow, forming a turning surface and reflecting downward. Similar behavior is observed a few more times during 1996, although in September 1996 the structure changes result from a vertical shifting of an existing turning surface. The wave response in both cases is reproduced by a time-dependent model.

There is evidence that this process of waves decelerating the mean flow, causing a reflecting surface to form and reflecting downward, occurs during other years. A vertical tilting of phase lines, similar to what we find in 1996, was found by Hartmann (1976) and Randel et al. (1987). We also skimmed a few years of data for signatures of a deceleration of the upper-stratospheric zonal mean winds, followed by a poleward heat flux (downward reflection). Such signatures were found in the Southern Hemisphere for wave 1 in September 1986, and 1983, and for wave 2 in September 1983, and in the Northern Hemisphere, for wave 1 in winters 1990–91 and 1995–96. Note that if downward reflection occurs on a regular basis, it may appear as a frequency peak in a space–time Fourier decomposition, because the wave structure changes associated with the reflection appear as an eastward phase propagation at some levels. This may account for at least some of the observed frequency peaks found, for example, by Mechoso and Hartmann (1982). Additional calculations are needed to determine this.

There is also evidence that this process of "self-induced downward reflection" may be fundamental to the stratosphere. Recently Giannitsis and Lindzen (2001a,b, manuscripts submitted to *J. Atmos. Sci.*) studied the mechanisms by which wave amplitudes are limited in the upper stratosphere, using a nonlinear QG model.

They found wave–mean flow interactions are the leading process by which wave amplitudes are limited. When the initial basic state zonal mean wind is strong, a reflecting surface forms and reflects the waves downward, and when the initial zonal mean wind is weak, a critical surface forms and absorbs the waves.

Whether the role of downward reflection in the dynamics of stratospheric waves is fundamental or secondary, the consistency between the time evolution of the waves and the basic state provides confidence in the relevance of quasi-linear theory to the stratosphere. It is also supportive of the quality of the observations and the wavenumber diagnostics. This is important since much of the time variation of the basic state occurs at or above 5 mb (e.g., Figs. 6 and 11), where the observations start losing reliability. The structure changes of the waves, on the other hand, are observed throughout the depth of the stratosphere.

Finally, we note the interesting possibility that downward reflected waves affect the troposphere. Recently, there has been a new interest in mechanisms by which the stratosphere can affect the troposphere, motivated by observational evidence of a link between the strength of the polar vortex and the tropospheric annular mode (Thompson and Wallace 1998; Baldwin and Dunkerton 1999). While the largest stratosphere–troposphere coupling is observed in the Northern Hemisphere during sudden warming events (Baldwin and Dunkerton 1999), downward reflection may provide a mechanism by which the stratosphere can affect tropospheric circulation in the Southern Hemisphere and during the quieter Northern Hemisphere winters. It is quite striking that a deceleration in the upper stratosphere, and a decrease of wave amplitude at the tropopause, occur within a few days of each other throughout the winter of 1996, as well as in other years we have looked at. This naturally raises the question of whether the downward reflected wave affects its source. It is possible, however, that the lifetime of the sources of the waves and the time it takes a wave to grow enough and change the wave geometry in the upper stratosphere are similar and this is only a coincidence. A more comprehensive study, in particular, of the sources of planetary-scale waves in the troposphere is needed to determine this.

APPENDIX A

The Data

We use the stratospheric analysis product compiled and distributed by the NASA Goddard Space Flight Center (GSFC) Stratospheric Chemistry and Dynamics Branch. Temperature and geopotential height fields are provided on 18 levels (1000–0.4 mb), with radiosonde data in the troposphere, and satellite retrievals in the stratosphere [above 100 (70) mb in the Southern (Northern) Hemisphere]. The horizontal resolution is $2°$ latitude by $5°$ longitude. Stratospheric temperature is retrieved from satellite measurements, while geopotential height is obtained by integrating the retrieved temperatures in the vertical, using the 100-mb tropospheric analysis as a lower boundary condition. The observations are interpolated onto a regular grid from the satellite grid using an objective interpolation (Cressman 1959). Winds, vorticity, and Ertel's potential vorticity are calculated at GSFC from the geopotential heights using a balanced wind approximation (Randel 1987). For more details see the NASA GSFC Web site. The data is also described in Harnik (2000), along with a discussion of its quality.

It is important to consider the ability of satellite retrievals to observe vertical wave structure and evolution. The main limitation is the coarse vertical resolution. We have elsewhere established the ability of these observations, to resolve the vertical structure of stratospheric planetary waves. Harnik and Lindzen (Harnik and Lindzen, manuscript submitted to *Ann. Geophys.*) tested the retrieval process by simulating it in a model (same as used in this study) and comparing the retrieved waves to the actual model waves from which they were obtained. They found that stratospheric planetary waves, under most conditions, have large enough vertical scales to be resolved quite well, below 5 mb. Above 5 mb errors start growing, with very little real observational data going into the retrievals above 1.5 mb. Note, however, that there are many other errors involved in the retrieval process, which may exist in the observational product.

Another possible source of error is aliasing due to the asynoptic sampling by the satellite (Salby 1982). Aliasing will occur for perturbations that propagate in the zonal direction with periods shorter than 1–2 days. To test this we asynoptically sample an analytically specified wave undergoing structure changes like the observed. We find a reduction in wave amplitude in the upper stratosphere of 30%, for a wave that moves about $135°$ longitude in 1 day (which is larger than observed), and hardly any distortions in the phase tilt of the waves. A more detailed description of this calculation is found in Harnik (2000, section 2.3).

APPENDIX B

Validity of the WKB Approximation

In order for Eqs. (12) and (13) to hold, with l and m being the wavenumbers of a solution of the form (10), we need to assume that the wavelength of the solution, divided by 2π, is much smaller than the length over which the amplitude of the wave and the wavenumber itself vary.

In one dimension, we can express the amplitude of the wave in terms of the wavenumber [$A \propto m^{-1/2}$, Eq. (7)], hence the above conditions can be stated in terms of the wavenumber only:
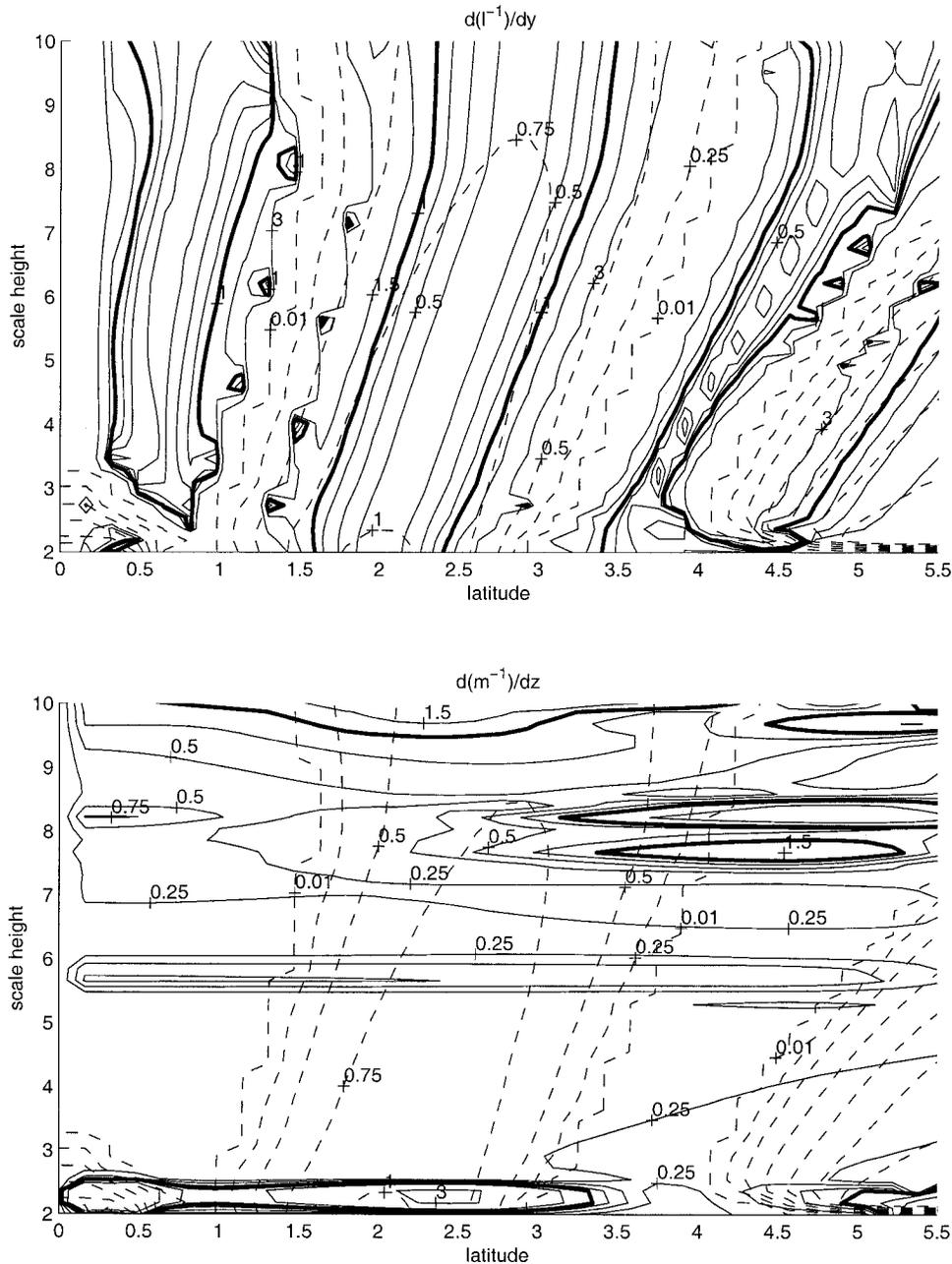

FIG. B1. The validity of the WKB approximation: Absolute values of (a) d/dy ($1/l$) and d/dz ($1/m$) [Eqs. (B1)–(B2)], which are conditions on the meridional and vertical wavenumbers, respectively. Regions larger than 0.5 are shaded. Contour values are 0.25–1 in jumps of 0.25, 1.5, and 3. The 1.0 contour is thick. Also plotted for reference is the meridional wavenumber (dashed), with contour values of 0.01, and 0.25–1 in jumps of 0.25 $(1190 \text{ km})^{-1}$. See text for details.

$$\frac{d}{dz}\left(\frac{1}{m}\right) \ll 1. \qquad (B1)$$

The extension to two dimensions is not as simple, since we cannot express the amplitude in terms of the wavenumber explicitly. The best we can do is to obtain a set of necessary conditions by applying (B1) along with a similar condition on the meridional direction:

$$\frac{d}{dy}\left(\frac{1}{l}\right) \ll 1. \qquad (B2)$$

Note that the solution may still be of wavelike nature, as long as the $\ll$ in the inequalities (B1)–(B2) is replaced by $<$. In this case the WKB form of the solution will not be accurate, but the qualitative wave features of the solution will still hold. If the left-hand sides of (B1)–(B2) are greater than 1.0, the interpretation of the solution in terms of wave structure and wave propagation is ambiguous.

Figure B1 shows the absolute value of the left-hand sides of (B1)–(B2), calculated from the high sponge layer model run of section 3 (other runs yield similar results). The shaded regions are for values greater than 0.5, which we consider as regions where WKB is violated. We also highlighted the 1.0 contours by making them thick. We see that the vertical wavenumber satisfies the WKB conditions quite well in most of the domain, and violates it near turning surfaces, and near the bottom (where we have a node, due to the downward reflection from the turning surface). The meridional wavenumber, on the other hand, satisfies WKB only in a narrow region near the center of the waveguide. Since our model runs and observational studies suggest that m^2 is a meaningful diagnostic of the propagation characteristics, the narrowness of the region in which the meridional WKB criterion is satisfied does not seem to matter.

APPENDIX C

The Formulation in Spherical Coordinates

There are two main complications in formulating the QG equations on a sphere. First, QG scaling requires the characteristic length scales of the flow to be much smaller than the radius of the earth, allowing us to ignore meridional derivatives of the geometric sphericity factors relative to meridional derivatives of the flow. Characteristic length scales of our waves (planetary) are larger than earth's radius. In spite of this, QG is used and its validity tested a posteriori. The results presented in this paper may be viewed as an assessment of the applicability of QG linear theory to stratospheric planetary waves. The second complication is that the geostrophic velocity, when defined in the traditional way, is divergent on a sphere. As a result, the EP flux divergence is not proportional to the PV flux as it is on a β plane (Palmer 1982). To get round this problem, past studies

(e.g., Matsuno 1970; Palmer 1982) have essentially redefined the vertical component of the vorticity perturbation as follows: $\zeta = f\nabla \times \mathbf{v}/f$. For our calculations we take the approach of R. A. Plumb (1999, personal communication[C1]), where instead of redefining the vorticity, we redefine the geostrophic winds by using the geopotential height scaled by the Coriolis parameter, as follows: $\Psi = \phi/f$, $v = 1/(a \cos\varphi)(\partial\Psi/\partial\lambda)$, $u = (1/a)(\partial\Psi/\partial\varphi)$, and $T = (T_o/g)(\partial\phi/\partial z) = [(fT_o)/g](\partial\Psi/\partial z)$. Here $f = 2\Omega \sin(\varphi)$ is the latitude-dependent Coriolis parameter, φ and λ are the latitude and longitude angles, z is log pressure, a the earth's radius, Ω the earth's rotation rate, and all other variables are as in the β-plane model.

Using these relations, and assuming a normal mode solution in longitude and time and the transformation 3, $\Psi = \Psi \exp\{is[\lambda - (c/a \cos\varphi)t]\} = \psi e^{z/2h} N \exp\{is[\lambda - (c/a \cos\varphi)t]\}$, we can derive the wave propagation equation:

$$\frac{1}{\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial\psi}{\partial\varphi}\right) + \frac{f^2 a^2}{N^2}\frac{\partial^2\psi}{\partial z^2}$$
$$+ \left[\frac{a\overline{q}_\varphi}{U-c} - \frac{s^2}{\cos^2\varphi} + a^2 f^2 F(N^2)\right]\psi = \text{damping}, \qquad (C1)$$

where $F(N^2)$ is defined in (5). As in the β plane, we define an index of refraction and meridional and vertical wavenumbers [Eqs. (6), (13), and (12)]:

$$n_{\text{ref}}^2 = N^2\left[\frac{a\overline{q}_\varphi}{U-c} - \frac{s^2}{\cos^2\varphi}\right.$$
$$\left. + a^2 f^2 F(N^2)\right] \qquad (C2)$$

$$\text{Re}\left[\frac{\frac{1}{\cos\varphi}\frac{\partial}{\partial\varphi}\left(\cos\varphi\frac{\partial\psi}{\partial\varphi}\right)}{\psi}\right] = \text{Re}\left(\frac{\psi_{\varphi\varphi} - \psi_\varphi \tan\varphi}{\psi}\right)$$
$$\equiv -l^2 \qquad (C3)$$

$$\text{Re}\left(\frac{\psi_{zz}}{\psi}\right) \equiv -m^2, \qquad (C4)$$

where

$$\frac{\partial\overline{q}}{\partial y} = \frac{1}{a}\frac{\partial\overline{q}}{\partial\varphi}$$

$$= \beta - \frac{1}{a^2}\frac{\partial}{\partial\varphi}\left(\frac{1}{\cos\varphi}\frac{\partial(U \cos\varphi)}{\partial\varphi}\right) - \frac{f^2}{pa^2}\frac{\partial}{\partial z}\left(\frac{p}{N^2}\frac{\partial U}{\partial z}\right)$$
$$(C5)$$

$p = p_o e^{-z/H_o}$ is pressure, and $\beta = (2\Omega \cos\varphi/a)$. All other variables are as defined in the β-plane model.

[C1] The derivation is taken from R. A. Plumb's Middle Atmosphere class notes and can be found in Harnik (2000).

Note that Plumb's derivation of the equations results in a cleaner definition of the meridional wavenumber because the meridional derivative term is the meridional component of ∇^2 in spherical coordinates. Matsuno's (1970) derivation, on the other hand, results in a different, more complicated meridional derivative term.

REFERENCES

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

Baldwin, M. P., and T. J. Dunkerton, 1999: Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *J. Geophys. Res.,* **104,** 30 937–30 946.

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.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374.

Dickinson, R. E., 1968: Planetary waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci.,* **25,** 984–1002.

——, 1969: Vertical propagation of planetary Rossby waves through and atmosphere with Newtonian cooling. *J. Geophys. Res.,* **74,** 929–938.

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.

Fels, S. B., 1982: A parameterization of scale-dependent radiative damping rates in the middle atmosphere. *J. Atmos. Sci.,* **39,** 1141–1152.

Harnik, N., 2001: The evolution of a stratospheric wave packet. *J. Atmos. Sci.,* in press.

Hartmann, D. L., 1976: The structure of the stratosphere in the Southern Hemisphere during late winter 1973 as observed by satellite. *J. Atmos. Sci.,* **33,** 1141–1154.

——, 1977: Stationary planetary waves in the Southern Hemisphere. *J. Geophys. Res.,* **82,** 4930–4934.

——, C. R. Mechoso, and K. Yamazaki, 1984: Observations of wave–mean flow interactions in the Southern Hemisphere. *J. Atmos. Sci.,* **41,** 351–362.

Harwood, R. S., 1975: The temperature structure of the Southern Hemisphere stratosphere August–October 1971. *Quart. J. Roy. Meteor. Soc.,* **101,** 75–91.

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.

Karoly, D. J., and B. J. Hoskins, 1982: Three dimensional propagation of planetary waves. *J. Meteor. Soc. Japan,* **60,** 109–123.

Lindzen, R. S., B. Farrell, and D. Jacqmin, 1982: Vacillations due to wave interference: Applications to the atmosphere and to analog experiments. *J. Atmos. Sci.,* **39,** 14–23.

Madden, R. A., 1983: The effect of interference of traveling and stationary waves on time variations of the large-scale circulation. *J. Atmos. Sci.,* **40,** 1110–1125.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.,* **27,** 871–883.

Mechoso, C. R., and D. L. Hartmann, 1982: An observational study of traveling planetary waves in the Southern Hemisphere. *J. Atmos. Sci.,* **39,** 1921–1935.

——, ——, and J. D. Farrara, 1985: Climatology and interannual variability of wave, mean-flow interaction in the Southern Hemisphere. *J. Atmos. Sci.,* **42,** 2189–2206.

Palmer, T. N., 1982: Properties of the Eliassen–Palm flux for planetary scale motions. *J. Atmos. Sci.,* **39,** 992–997.

Pawson, S., R. S. Harwood, and J. D. Haigh, 1992: A study of the radiative dissipation of planetary waves using satellite data. *J. Atmos. Sci.,* **49,** 1304–1317.

Plumb, R. A., 1989: On the seasonal cycle of stratospheric planetary waves. *Pure Appl. Geophys.,* **130,** 233–242.

Randel, W. J., 1987: The evaluation of winds from geopotential height data in the stratosphere. *J. Atmos. Sci.,* **44,** 3097–3120.

——, 1988: The seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere. *Quart. J. Roy. Meteor. Soc.,* **114,** 1385–1409.

——, 1992: Global atmospheric circulation statistics, 1000–1 mb. NCAR Tech. Note NCAR/TN-336+STR, 156 pp.

——, 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.

Salby, M. L., 1982: Sampling theory for asynoptic satellite observations. Part I: Space–time spectra, resolution, and aliasing. *J. Atmos. Sci.,* **39,** 2577–2600.

——, and R. R. Garcia, 1987: Vacillations induced by interference of stationary and traveling planetary waves. *J. Atmos. Sci.,* **44,** 2679–2711.

Sato, Y., 1974: Vertical structure of quasi-stationary planetary waves in several winters. *J. Meteor. Soc. Japan,* **52,** 272–281.

Schoeberl, M. R., and M. A. Geller, 1977: A calculation of the structure of stationary planetary waves in winter. *J. Atmos. Sci.,* **34,** 1235–1255.

Shiotani, M., and I. Hirota, 1985: Planetary wave-mean flow interaction in the stratosphere: A comparison between Northern and Southern Hemispheres. *Quart. J. Roy. Meteor. Soc.,* **111,** 309–334.

Simmons, A. J., 1974: Planetary-scale disturbances in the polar winter stratosphere. *Quart. J. Roy. Meteor. Soc.,* **100,** 76–108.

Thompson, D. J., and J. M. Wallace, 1998: The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25,** 1297–1300.