Nonlinear Saturation of Vertically Propagating Rossby Waves

CONSTANTINE GIANNITSIS* AND RICHARD S. LINDZEN

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received and in final form 10 October 2008)

ABSTRACT

The interaction between vertical Rossby wave propagation and wave breaking is studied in the idealized context of a beta-plane channel model. Considering the problem of propagation through a uniform zonal flow in an exponentially stratified fluid, where linear theory predicts exponential wave growth with height, the question is how wave growth is limited in the nonlinear flow. Using a numerical model, the authors examine the behavior of the flow as the bottom forcing increases through values bound to lead to a breakdown of the linear solution within the computational domain. Focusing on the equilibrium flow obtained for each value of the bottom forcing, an attempt is made to identify the mechanisms involved in limiting wave growth and examine in particular the importance of wave–wave interactions. The authors also examine the case in which forcing is continuously increasing with time so as to enhance effects peculiar to transiency; it does not significantly alter the main results.

Wave–mean flow interactions are found to dominate the dynamics even for strong bottom forcing values. Ultimately, it is the modification of the mean flow that is found to limit the vertical penetration of the forced wave, through either increased wave absorption or downward reflection. Linear propagation theory is found to capture the wave structure surprisingly well, even when the total flow is highly deformed. Overall, the numerical results seem to suggest that wave–wave interactions do not have a strong direct effect on the propagating disturbance. Wave–mean flow interactions limit wave growth sufficiently that a strong additional nonlinear enstrophy sink, through downscale cascade, is not necessary. Quantitatively, however, wave–wave interactions, primarily among the lowest wavenumbers, prove important so as to sufficiently accurately determine the basic state and its influence on wave propagation.

1. Introduction

The behavior of the stratosphere during the winter season, when vertical propagation of tropospherically forced stationary anomalies is permitted, has long been a subject of interest in the meteorological community. While some of the most important work was accomplished early on (Charney and Drazin 1961; Dickinson 1969; Matsuno 1971), in the past several decades interest in stratospheric dynamics has been renewed, to a large extent owing to concerns about ozone and, more recently, the role of the stratosphere in climate. Moreover, the use of satellite observations, allowing the construction of reliable detailed synoptic maps of the flow, has led to a new understanding of the dynamical interactions, with the emphasis shifting from a quasi-linear interpretation of the dynamics to a framework that focuses more on the synoptic characteristics of the flow (McIntyre 1982; Juckes and McIntyre 1987; O’Neill and Pope 1988, hereafter OP88). Particular attention has been paid to the frequent episodes of strong wave growth that are seen in observations to lead to a strongly nonlinear behavior of the flow, with pronounced deformation of the polar vortex and the stripping of filaments of potential vorticity away from its core (Juckes and McIntyre 1987; OP88).

The dynamics of such events, commonly called Rossby wave breaking events (McIntyre and Palmer 1983, 1985), have for the most part been successfully explained in terms of horizontal advection of the preexisting potential vorticity field by the velocity field of the anomalously large wave. In particular, much attention has been paid to the interaction between the propagating stratospheric waves and the tropical zero-wind line, alluding to the early work of Warn and Warn (1978, hereafter WW78) and Stewartson (1978). More recent studies have also

* Deceased 31 December 2000.

Corresponding author address: Richard S. Lindzen, 54-1720, MIT, Cambridge, MA 02139.
E-mail: rlindzen@mit.edu

DOI: 10.1175/2008JAS2978.1

© 2009 American Meteorological Society

However, although a good mechanistic understanding of how wave breaking evolves has been achieved, its effect on the propagating disturbance itself is by no means clear. One problem is the relative lack of three-dimensional numerical studies focusing on the link between wave breaking and vertical wave propagation. Much of the exploration of the wave breaking dynamics has been based on one-layer models, which use a prescribed bottom topography as forcing (Juckes and McIntyre 1987; Polvani et al. 1995). Such models, however, completely bypass the question of a possible feedback onto the propagating disturbance by the deformed interior flow. Of the few three-dimensional studies that exist (OP88; Robinson 1988; Dritschel and Saravanan 1994; Polvani and Saravanan 2000) the one that has focused most directly on the issue of how wave breaking affects the vertical propagation of a forced disturbance is that of Dritschel and Saravanan (1994). Interestingly, their study shows the forced disturbance being inhibited from penetrating to high altitudes once vigorous Rossby wave breaking occurs. However, as their results were not analyzed in further depth, the details of the process were not identified.

At the same time, however, a broad qualitative picture of the role of nonlinear wave interactions has slowly been established (McIntyre and Palmer 1983, 1985; Robinson 1988) and wave breaking has been largely associated with a downscale enstrophy cascade, leading to local dissipation of the propagating disturbance. Furthermore, attempts have been made to parameterize such an effect and obtain a qualitative description of the full dynamics (Garcia 1991; Randel and Garcia 1994) with results that seem quite successful. However, while the general argument concerning the downscale cascade of enstrophy during wave breaking is correct, the proposed parameterizations were based on assumptions that were not fully justified. Thus, it is still fair to say that a clear understanding of how the nonlinear flow reacts to episodes of strong wave growth is still incomplete.

In the present study we attempt to take a step back, approaching the question of how nonlinear dynamics influence the vertical propagation of Rossby waves in a very simplified framework. While our ultimate interest is in the dynamics of the winter stratosphere, our approach leaves aside many of the complications of the real stratospheric flow, such as the interaction of the propagating waves with the tropical zero-wind line. Thus, our focus is not on wave breaking per se—wave breaking in the stratosphere is often viewed as a consequence of the presence of a zero-wind line—but rather on the question of how the amplitudes of vertically propagating waves saturate. Given the constraint placed on wave amplitudes by potential enstrophy considerations (Lindzen and Schoeberl 1982; Schoeberl 1982, hereafter S82) and the prediction of linear theory that vertically propagating disturbances grow exponentially with height, the question of wave saturation arises naturally. Our goal is thus to examine how the breakdown of linear theory comes about and through what mechanism the nonlinear dynamics limit wave growth.

2. Model description and setup

A quasigeostrophic multilevel beta-plane channel model is used, forced at its lower boundary by a prescribed streamfunction anomaly. The specified bottom anomaly is taken to be of zonal wavenumber 1 and sinusoidal in $y$, with a meridional half-wavelength equal to the width of the numerical domain. The horizontal dimensions of the channel are chosen: $L_x = 30,000$ km and $L_y = 5,000$ km. A grid of $64 \times 32$ points is used in the horizontal and a spectral formulation is employed for the model, using Fourier sine/cosine functions as a spectral basis. The specific amplitude of the forcing is, in some sense, irrelevant. Any amplitude leads, in principle, to saturation at some altitude, given the growth of amplitude with altitude. This is discussed further in section 5.

As a basic state we use a flow with spatially uniform zonal mean winds, uniform potential vorticity gradients equal to $\beta = 1.6 \times 10^{-11}$ m$^{-1}$ s$^{-1}$, and constant Brunt–Väisälä frequency equal to $N_x = 2.0 \times 10^{-2}$ s$^{-1}$. Furthermore, the planetary vorticity is taken as $f_0 = 1.0 \times 10^{-4}$ s$^{-1}$ and the scale height is set to $H = 7.0$ km. Thus, the forced wave has characteristics broadly consistent with those of middle-latitude stratospheric planetary-scale disturbances.

Because of the spatial uniformity of our basic state, the strength of the zonal mean winds does not affect the mean potential vorticity gradients or the initial mean available potential enstrophy. However, to explore the model parameter space, numerical integrations with two different basic states are performed, $U_0$ being given the values of 10 and 25 m s$^{-1}$, respectively (we will later examine intermediate values of $U_0$); in the context of our model parameters the maximum wind speed allowing vertical propagation of the forced disturbance is 28 m s$^{-1}$.

To restore the zonal mean flow to its initial value and damp the nonzonal disturbances, spatially uniform Newtonian cooling is used, with a time scale of 28 days. Estimates of radiative cooling in the stratosphere range between 20 days for the lower stratosphere and 5 days for the stratosphere level (Dickinson 1973; Fels 1982; Randel 1990; Kiehl and Solomon 1986). However, in the
present, highly idealized study, the primary concern is not so much a realistic simulation but to ensure that damping does not become a priori dominant in the saturation process, thus maximizing the possibility for nonlinear interactions to enter the dynamics.

In addition to Newtonian cooling, a sponge layer is also used, close to the upper boundary of the domain, to absorb the upward propagating wave activity; see schematic in Fig. 1. In the sponge layer equal Newtonian cooling and Rayleigh friction coefficients are used. Given the model setup, different choices for the basic state winds produce stationary waves with different vertical scales. As the efficiency of the sponge layer depends, among other factors, on its depth compared to the vertical wavelength of the impinging wave (Yanowitch 1967; Lindzen 1968), the sponge layer is adjusted appropriately in each case, with \( \nu(z) = 4.2 e^{(z - 100 \text{ km})/5 \text{ km day}^{-1}} \) for \( U_0 = 10 \text{ m s}^{-1} \) and \( \nu(z) = 1.3 e^{(z - 100 \text{ km})/12.5 \text{ km day}^{-1}} \) for \( U_0 = 25 \text{ m s}^{-1} \). To ensure that a large portion of the numerical domain remains outside the influence of the sponge layer, the vertical extent of the model is set to 100 km. A resolution of \( dz = 1 \text{ km} \) is used in the vertical. A third-order Adams–Bashforth scheme is used for forward time stepping.

In examining the dynamics of wave saturation, two approaches are taken: a quasi-equilibrium approach and a fully transient approach. In the first approach, the bottom forcing is increased in small steps, allowing the interior flow sufficient time to equilibrate before a further increase. Initially, the forcing is too small to lead to nonlinearity anywhere in the domain. The evolution of the equilibrated flow is then examined as the bottom forcing reaches values leading to strong deviations from linear behavior in the interior of the domain. Each step in the flow “evolution” is described as a snapshot picture since the equilibrated flow is found to present relatively small amounts of time variance for all values of the bottom forcing, representing in most cases a true steady state. Roughly speaking, equilibration occurs at a particular amplitude, but as the forcing increases, this level occurs at progressively lower altitudes. In the second approach, bottom forcing is increased linearly in time to some value that leads to saturation; however, different runs use different rates of increase. The second approach does not lead to any new results but serves to clarify some issues. As a result, we will only deal with it briefly. A more complete treatment is found in Giannitsis (2001).

3. Numerical results—Quasi-equilibrated wave saturation

Since the focus of the present study is on the saturation of the amplitude of the propagating wave, an obvious quantity to examine is the wave-1 potential enstrophy, given by

\[
\psi(x,y,t) = \psi_{\text{bottom}} \sin(kx) \sin(my)
\]

Figure 2 depicts the maximum wave-1 potential enstrophy and the minimum mean available potential enstrophy obtained for each value of the bottom forcing, where both quantities represent global extrema rather than referring to a particular vertical level. One naturally observes saturation of wave-1 amplitude as the forcing increases to high values. Although linear theory predicts a quadratic dependence of the wave-1 potential enstrophy on the bottom forcing, the results of the nonlinear integrations show that beyond a certain point wave amplitudes do not follow the increase in the strength of the forcing. Interestingly, the saturation limit for the wave-1 potential enstrophy appears to depend on the choice of basic state wind speed, despite the fact that the initial mean available potential enstrophy is identical for both basic states. Moreover, the wave saturation does not appear to necessarily come as a result of depletion of the mean available potential enstrophy, as is clearly evident for the basic state \( U_0 = 10 \text{ m s}^{-1} \); this is consistent with the findings of Lindzen and Schoebert (1982). While a prediction for the maximum wave-1 amplitude based on the initial mean available potential enstrophy...
clearly fails to capture the results of the numerical integrations, one can obtain a better prediction by taking into account the effect of Newtonian cooling. As argued by S82, at steady state one has a balance between sources and sinks of potential enstrophy at each vertical level. Using a very simple representation of the flow, S82 predicted a saturation amplitude of

$$|q_1| = \frac{\beta}{\sqrt{2k_y}},$$

where $k_y$ is simply $\pi/L_y$. In the present case that estimate would give a value of $4.0 \times 10^{-11}$ s$^{-2}$ for the wave-1 potential enstrophy, matching reasonably well the values obtained in our numerical integrations.

It should be noted, however, that S82 interpreted the limit obtained by the expression above not as a saturation limit for wave amplitudes but rather as the maximum eddy amplitude for which a steady state was possible, having in mind the oscillation cycles obtained by Holton and Mass (1976). In the present numerical integrations no behavior analogous to that of Holton and Mass is found; rather, the saturation amplitude obtained represents a true upper limit on wave amplitudes. The occurrence of oscillation cycles in Holton and Mass may be attributable to the severe spectral truncation of their model, although more recent results call this into question (Scott and Haynes 2000). Thus, the difference remains essentially unresolved.

4. Mean flow modification

To understand the dynamics leading to the wavenumber-1 saturation one needs to look at the overall changes that occur in the flow as the bottom forcing reaches sufficiently high amplitudes. As a first step we focus on the changes in the zonal mean flow, approaching the dynamics from a wave–mean flow perspective. The vertical penetration and consequent growth of the forced wave are generally found to lead to a strong modification of the zonal mean flow. The convergence of Eliassen–Palm flux, due to the damping of the propagating disturbance, leads to a decrease in the zonal mean winds and a corresponding reduction of the mean potential vorticity gradients at the center of the channel (see Fig. 3). Although the changes in the mean potential vorticity gradient are not entirely straightforward, presenting an increase in the mean gradient values at the flanks of the channel, their overall reduction in the central latitudes can be easily related to the changes in the zonal mean wind field through the expression

$$\delta Q_y = -\left\{ \frac{\partial^2}{\partial y^2} + \frac{1}{\rho(z) \partial z} \left[ \frac{\rho(z) f_0^2}{N^2(z) \partial z} \delta U \right] \right\} \delta U.$$

Given the elliptical character of the Laplacian operator, it is straightforward to deduce that negative wind anomalies lead to negative anomalies in the mean potential vorticity gradients as well.

It is important to keep in mind that changes in the zonal mean flow, such as shown in Fig. 3, can significantly affect the structure of the propagating disturbance, at least to the degree that linear propagation notions still apply to the dynamics of the interior flow. The changes shown in Fig. 3 do not depend on $U_0$, but the resulting impact on the index of refraction does.

Because a modification of the mean winds and potential vorticity gradients corresponds to a change in the

---

**Fig. 2.** Minimum mean available potential enstrophy (*) and maximum wave-1 potential enstrophy (o) as functions of the bottom forcing: (a) $U_0 = 10$ and (b) $U_0 = 25$ m s$^{-1}$. 

---

**Table 1.** Summary of results for different bottom forcings.
basic state through which the wave-1 disturbance propagates, an examination of the changes in the refractive index values seems the most natural approach to understanding the evolution of the flow. In the context of quasigeostrophic theory, the index of refraction can be directly linked to the vertical and meridional structure of the forced disturbance. Taking \( C_9(x, y, z) = Y(x, y, z)e^{ik_x x} e^{2H y} \), one has

\[
\left( \frac{N_x^2}{f_0^2} \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) Y' + \nu^2 Y' = 0 \quad \text{and} \quad (1)
\]

\[
\nu^2 = \left( \frac{Q_y}{U} - k_x^2 \right) \frac{N_z^2}{f_0^2} - \frac{1}{4H^2}. \quad (2)
\]

As the bottom forcing and the geometry of the channel largely impose the meridional structure of the propagating disturbance, a change in the index of refraction reflects directly on the ability of the wave-1 disturbance to penetrate in the vertical. Given the overall decrease of the zonal mean winds, one possibility is for mean easterlies to form as the forcing increases. In the context of linear theory, a critical line, marking a singularity in the equation for \( Y'(y, z) \), acts as a perfectly absorbing layer, preventing the penetration of the propagating wave to higher levels (in the presence of damping, however small). On the other hand, it is also possible for the mean potential vorticity gradients to turn negative without the mean winds turning easterly. In such a case, a refractive index turning surface is said to form, denoting the zero contour separating positive from negative values for the square of the refractive index, and is expected to lead to downward reflection of the propagating wave, to the degree again that the predictions of linear theory are relevant to the dynamics of the nonlinear flow. In both cases, the response of the fluid is to prevent the unbounded growth of the linear vertically propagating wave by preventing the propagation itself.

Which of the two above possibilities will materialize in the numerical integrations turns out to depend on the strength of the basic-state winds \( U_0 \). For a given wind anomaly \( \delta U \), one has

\[
U_{\text{min}} = U_0 + \delta U
\]

\[
Q_{y, \text{min}} = \beta \left\{ \frac{\partial^2}{\partial y^2} + \frac{1}{\rho(z) \frac{\partial}{\partial z}} \left[ \rho(z) \frac{f_0^2}{N_z^2(z) \frac{\partial}{\partial z}} \right] \right\} \delta U.
\]

Thus, one expects weak westerly \( U_0 \) to allow the formation of a zero-wind line without a significant reduction of the mean potential vorticity gradients, whereas in the case of a strongly westerly \( U_0 \) one might expect the opposite. Indeed, the results of the nonlinear numerical integrations do confirm the above hypothesis (see Fig. 4). Given the different behavior expected for the propagating disturbance, based on the predictions of linear theory, it is important to examine the wave saturation dynamics for each basic state separately.

5. Wave-1 vertical structure

Examining the response of the wave-1 disturbance to the strengthening of the bottom forcing one observes, as
expected, an increased confinement of the wave to the lower levels of the model domain. Note that, since we are concerned with the saturation of waves that grow exponentially with altitude, increased forcing simply moves saturation to lower levels. For an unbounded, inviscid domain, any level of forcing would lead to saturation at some altitude. However, our model does have a top as well as a sponge layer. Therefore, we increase forcing and lower the saturation altitude so as to make sure that our results are robust. Interestingly, increasing the forcing and lowering the altitude of saturation causes wave-1 amplitudes at the upper levels to decrease significantly. For the basic state with $U_0 = 10$ m s$^{-1}$, the changes in the wave-1 structure are plotted in Fig. 5. To first order the behavior of the wave-1 disturbance appears to be consistent with the notion that the formation of a zero-wind line leads to absorption of the propagating disturbance, preventing a penetration of the wave to the upper levels. Our results indicate that the development of modifications in the basic flow that prevent the propagation of waves is associated with specific amplitudes for the wave. As forcing increases, a given amplitude appears at a lower level. As we see in Fig. 5, the solution for larger forcing is very similar to that for the smaller forcing except that it occurs at a lower altitude. However, small differences due to residual effects of the sponge layer near the top of the domain and restricted distance for adjustment near the bottom can be discerned.

Plots of the Eliassen–Palm fluxes reveal a pattern of strong flux convergence and wave absorption inside the region of zonal mean easterlies, consistent with the predictions of linear WKB theory (Fig. 6). This contrasts with the prediction of Killworth and McIntyre (1985, hereafter KM85) that a critical line would act in the long time limit as a perfect reflector. Of course, the results of KM85 primarily concerned the case of a barotropic flow, following the work of WW78 and Stewartson (1978), where meridional propagation of a wave through a mean flow with a zero-wind line was considered rather than the case of vertical propagation. Nonetheless, their conclusions can be extended to the problem of vertical propagation.

Not surprisingly, damping proves crucial to this difference—damping being omitted in the analytical study of KM85 as well as in the study of WW78. The nature of the difference goes well beyond the simple difference between absorption and reflection. While in WW78 strong wrapping of potential vorticity contours occurs inside the nonlinear critical layer, Fig. 7 shows a striking absence of significant zonal asymmetries in the potential vorticity field, again suggestive of linear behavior. At the same time, however, strong potential vorticity gradients appear at the sides of the channel and a large region of homogenized potential vorticity forms at the center. In the stratosphere the appearance of such regions of homogenized potential vorticity is usually linked to mixing by breaking planetary waves and the associated nonlinear behavior of the flow (McIntyre and Palmer 1984; Polvani et al. 1995). That does not seem to be the case here. In the present case, despite the significant growth of the propagating wave, the homogenization of potential vorticity cannot be viewed as a simple result of nonlinear mixing. This is clear from an

![Fig. 4](image-url)
examination of maps of a conservative tracer (Fig. 8) whose initial structure is identical to that of the potential vorticity. Clearly the tracer field shows no sign of homogenization: Rather, the tracer isopleths seem to follow the smooth undulations in the streamfunction field. Thus, the role of Newtonian cooling appears to be crucial as advection alone is unable to account for the structure in the potential vorticity field. In a way this should be no surprise: in the absence of damping, the total mass enclosed between any two potential vorticity isolines has to be conserved, implying a conservation of the area enclosed between neighboring potential vorticity contours. In other words, advection cannot on its own lead to a flattening of the potential vorticity gradients. Rather, the action of a nonconservative mechanism is required for that—Newtonian cooling playing that role in the present case.

On the other hand, the behavior of the flow seems to be rather different in the case of more strongly westerly basic states. As shown in Fig. 4, for \( U_0 = 25 \) m s\(^{-1}\) a refractive index turning surface forms when the wave reaches sufficiently high amplitude (due to either increased forcing or to growth with altitude) owing to the reduction of the mean potential vorticity gradients at the center of the channel. As the bottom forcing increases from weak to strong values, a region of negative mean potential vorticity gradients is first found to form at about level 60. Further increases of the forcing, however, lead to a downward migration of the region of negative gradients as well as a slight increase of its area. As already noted, this is primarily a matter of a given amplitude for the wave being reached at a lower altitude.

The wave-1 disturbance in turn seems to react to these changes in close agreement with the predictions of linear theory, as in the case of \( U_0 = 10 \) m s\(^{-1}\). As shown in Fig. 9, a significant downward migration of the maximum in the wave-1 streamfunction is observed, as well as an overall reduction in the wave-1 amplitudes at the upper levels of the model domain. Thus, the changes in the wave-1 structure seem to be broadly consistent with the prediction of linear WKB theory that the refractive index turning surface would reflect the propagating wave downward. Additional computations for intermediate values of \( U_0 \) show that this behavior pertains to all choices of \( U_0 \geq 13 \) m s\(^{-1}\).

While the behavior of the wave-1 disturbance appears to be qualitatively consistent with a notion of linear adjustment of the wave to changes in the zonal mean flow, a
more stringent test comes from the comparison of the nonlinear model results to the predictions of a purely linear calculation. To take into account the modification of the zonal mean flow, which represents the basic state through which the wave propagates in the linear calculation, the results of the fully nonlinear numerical integration are used, extracting the zonal mean component from the full flow. As the nonlinear flow is allowed to fully equilibrate for each value of the bottom forcing, a comparison with the steady-state linear calculations is straightforward. Differences between the linear and nonlinear results for \( \Psi \) should thus be indicative of deviations from linear dynamics and of the importance of wave–wave interactions for the saturation of the wave-1 disturbance.

Indeed, the comparison of the linear results with the nonlinear model behavior shows remarkably good agreement (see Fig. 10), not only in the overall structure of the wave-1 disturbance, but also quantitatively. Not surprisingly, this agreement is largely independent of forcing amplitude (see Fig. 11). The primary impact of strong forcing is that it causes the turning surface to occur at lower levels where there is less distance between the forcing level and the reflecting level; this places greater demands on the calculations.

The success of the linear calculations is quite surprising, given the high amplitudes reached by the propagating disturbance, although Haynes and McIntyre (1987) also noted the success of quasi-linear calculations. Particularly in the case of \( U_0 = 25 \text{ m s}^{-1} \), horizontal maps of potential vorticity reveal a very strong deformation (see Fig. 12) with structures closely resembling the Rossby wave breaking patterns described in WW78 and observed in

**FIG. 7.** Horizontal maps of (left) streamfunction and (right) potential vorticity at level 60 for \( U_0 = 10 \text{ m s}^{-1} \): (top) \( \Psi_{\text{bottom}} = 5.0 \times 10^5 \) and (bottom) \( \Psi_{\text{bottom}} = 6.0 \times 10^5 \text{ m}^2 \text{ s}^{-1} \).
the stratosphere (Juckes and McIntyre 1987). The formation of closed circulation regions, marking a dominance of nonlinear advection by the large amplitude wave over zonal advection by the mean flow, leads to strong wrapping of potential vorticity contours around the cyclonic and anticyclonic centers, respectively, with the filaments of potential vorticity becoming thinner and more elongated as the bottom forcing increases.

However, the deformation of the potential vorticity field remains relatively limited. In no case does one find violent mixing and potential vorticity homogenization inside the regions of closed circulation. Rather, contrary to the behavior seen in passive tracer maps, one observes only limited wrapping of potential vorticity contours. The presence of Newtonian cooling, which tends to restore zonal symmetry of the initial flow, appears to oppose quite effectively the tendency of nonlinear advection to homogenize the potential vorticity field, despite the fact that the damping time scale used in our model is longer than considered realistic.

A point that requires more attention regards the stability and time steadiness of the equilibrated flow, as the deformation of the potential vorticity field leads to locally reversed potential vorticity gradients. Haynes (1985) and KM85 argued that such structures would lead to the development of small-scale barotropic instabilities that would in turn wipe out the negative gradient regions. In our numerical integrations no such instabilities arise. That could perhaps be a consequence of insufficient resolution: the fine scales predicted to become unstable by Haynes (1985) and KM85 are not resolved by our model (Lander and Hoskins 1997). On the other hand, given the strong background shear and the presence of Newtonian cooling, it is not clear that such instabilities should be expected. Previous studies (e.g., Waugh and Dritschel 1991) have shown that barotropic shear can stabilize filaments of potential vorticity that are stripped off the core of the polar vortex, in line with our findings. Moreover, the semicircle theorems limit growth rates to small values.

Finally, since the two regimes described above (development of an absorbing critical surface versus development of a reflecting surface) are dynamically rather different, we have attempted to better delineate the transition between the cases and to establish whether saturation proceeds differently within particular regimes.
Calculations were therefore made that spanned the range in $U_0$ from 10 to 28 m s$^{-1}$.

Results for the saturation wave-1 potential enstrophy amplitudes obtained for each basic state are presented in Fig. 13. An analysis of the model results shows that for all basic states with $U_0 \geq 13$ m s$^{-1}$, the modification of the mean flow leads to the formation of a refractive index turning surface. Thus, with the exception of $U_0 = 10$ m s$^{-1}$, all other cases fall into a single dynamical regime. Nonetheless, as Fig. 13 reveals, no unique saturation limit is found, even though all basic states have the same initial mean available potential enstrophy. Thus, although S82’s prediction gives a reasonable first-order estimate for the wave-1 potential enstrophy saturation value ($\approx 4.0 \times 10^{-11}$ s$^{-2}$), it fails to account for the differences observed between various choices of basic states.

An explanation for this behavior could perhaps be sought in the varying significance of damping for differing basic states. Although the Newtonian cooling coefficient is kept constant in all cases, its effectiveness in damping the propagating wave depends strongly on the vertical wavelength of that disturbance. For weakly westerly basic state winds the propagating wave presents strong westward tilt of its phase lines in the vertical, whereas for basic states approaching the critical value of 28 m s$^{-1}$ the vertical wavelength becomes very long and the propagating disturbance almost barotropic. In the latter case Newtonian cooling produces only weak dissipation.

However, numerical runs using $U_0 = 23$ m s$^{-1}$ and a wide range of Newtonian cooling values do not support the above hypothesis. As we see in Fig. 14, the saturation wave-1 potential enstrophy values that are obtained do not vary much, despite changes in the Newtonian cooling by a factor of 10. Thus, no simple physical picture for what sets the saturation limit seems to emerge. It appears that details of the equilibrated flow, such as perhaps the strength of the mean shear, play a relatively important role, leading to quantitatively different results depending on the choice of basic state. In this respect a simple physical parameterization of the saturation process does not appear likely.

6. Wave-1 potential enstrophy budget

Overall, the model results seem to suggest that to first order the wave-1 disturbance is not influenced by interactions with other wavenumbers. However, no particular information has been provided on the role of nonlinear wave interactions. To understand the effect of wave breaking on the dynamics of the flow one needs to go beyond the simple comparison between linear and nonlinear models. As far as the direct effect of nonlinearities on the wave-1 component is concerned, a straightforward approach is to form a potential enstrophy budget. One then has

\[
\left\langle \frac{\partial}{\partial t} \left( \frac{q_1^2}{2} \right) \right\rangle_{\text{linear}} + \left\langle s_1 q_1^2 \right\rangle_{\text{damping}} + \left\langle q_1^1 \sum_N \sum_M -J(\Psi, q_1^M) \right\rangle_{\text{wave-wave}}.
\]  (3)
Overbars denote longitudinal averages; angle brackets correspond to averaging over latitudes. On the rhs of the expression above there is a clear separation between different physical processes and one can distinguish between the linear term representing the ability of the wave to propagate through the modified zonal mean flow, a damping term representing the dissipation by Newtonian cooling and the sponge layer, and a nonlinear term representing the effect of interactions between different wavenumbers on the wave-1 amplitudes.

In the “steady state” approach, one will always have a nearly perfect balance between the three different terms. As with every balanced budget it is not easy to think of any particular term as “forcing” a change in the wave-1 disturbance. One can, however, examine the differences in magnitude between terms and attempt to diagnose the relative importance of linear versus nonlinear dynamics in the saturation of the wave-1 disturbance.

In general, wave–wave interactions are found to act as enstrophy sinks in a channel-average sense (see Fig. 15), leading to local damping of the wave-1 potential enstrophy, in line with the usual description of the dynamics associated with Rossby wave breaking (McIntyre and Palmer 1983, 1985; Garcia 1991). What is interesting, however, is that nonlinear wave interactions seem to be rather inefficient in dissipating the “excess” wave-1 potential enstrophy. The nonlinear enstrophy “sink” appears to be at most of the same strength as the direct dissipation of the propagating wave, despite the rather weak Newtonian cooling used in our model. (Note that the relatively large role for wave–wave interactions for $U_0 = 10 \text{ m s}^{-1}$ is probably irrelevant, since for high levels of forcing, the level being looked at is well above the critical level.) This is particularly striking for the case of $U_0 = 25 \text{ m s}^{-1}$, where the horizontal deformation of the potential vorticity field was found to be quite strong.

In this respect the direct comparison of the various terms in the enstrophy budget underlines again the fact that nonlinearities associated with wave breaking do not have a strong direct effect on the wave-1 disturbance. On the other hand, given the overall a priori dominance of the linear propagation term in the enstrophy budget, one should not compare the wave–wave term to the absolute magnitude of the linear propagation term but, rather, to changes in that term as the bottom forcing increases. More specifically one needs to focus on the differences between the actual values of the linear propagation term in the nonlinear flow and the values expected for a purely linear system, where the basic state remains unchanged.

For such a system the linear term would be expected to depend quadratically on the bottom forcing strength. In the numerical integrations, however, one observes a saturation of its magnitude or even a reduction to zero as the bottom forcing strengthens. Such differences from the linear prediction give an indication of how the changes in the zonal mean flow affect the propagating disturbance. Comparing to the strength of the wave–wave term, one sees that the dynamics behind the wave-1 saturation can be explained almost entirely by referring to the modification of the mean flow propagation characteristics. It appears that, because wave–mean flow interactions are so efficient in blocking the penetration of the forced wave, more strongly nonlinear interactions never have a chance to become important.

7. Quasi-linear model

Given the above results, it is interesting to examine to what extent the nonlinear model results can be reproduced by a quasi-linear model. To address this issue the numerical integrations presented above have been repeated with a model that allows interactions between the waves and the zonal mean flow but suppresses wave–wave interactions. The overall setup of the quasi-linear model is chosen to be identical to that of the nonlinear case. In particular, the model formulation, as such, allows for the same number of degrees of freedom as in the nonlinear model, both in the zonal and the meridional direction. However, because the bottom forcing only drives a wavenumber-1 disturbance, higher zonal wavenumbers can only arise as a result of instability of the zonal mean flow.
Part of the motivation for proceeding with a quasi-linear integration comes from the fact that in the non-linear model, the Eliassen–Palm flux divergence, responsible for the reduction of the zonal mean winds, was found to be dominated to more than 80% by contributions from the wavenumber-1 disturbance. If, indeed, the main characteristics of the modified mean flow can be accounted by retaining the wave-1 fluxes only, then one would expect the quasi-linear model to perform quite well, as wave–wave interactions were found to have a rather weak direct effect on the wave-1 disturbance.

However, the actual comparison between the non-linear and quasi-linear model results only shows a zero order agreement in the evolution of the flow. Significant differences are found both in the overall structure and in the amplitudes of the wave-1 disturbance (see Figs. 16 and 17). For the case $U_0 = 10 \text{ m s}^{-1}$ the quasi-linear flow does show a reduction in the penetration of the wave-1 disturbance to the upper levels (Fig. 16) and also has an overall structure consistent with the formation of a zero-wind line. However, it is also clear that the wave 1 in the quasi-linear model refracts much more strongly from the flanks of the channel, managing to penetrate to higher altitudes. Thus, the wave-1 disturbance reaches an amplitude about twice as large as in the fully non-linear integration.

A similar description applies to the case of $U_0 = 25 \text{ m s}^{-1}$. Both the downward migration of the wave-1 streamfunction maximum, from 70 to 50 km of altitude, and the decrease in the upper-level wave-1 amplitudes compare well to the nonlinear model results. Again, one seems to have a downward reflection of the wave-

![Fig. 12. Horizontal maps of (left) streamfunction and (right) potential vorticity fields at level 50 for $U_0 = 25 \text{ m s}^{-1}$: (top) $\Psi_{\text{bottom}} = (top) 1.8 \times 10^6$ and (bottom)$\Psi_{\text{bottom}} 2.2 \times 10^6 \text{ m}^2 \text{ s}^{-1}$.](image_url)
1 disturbance from a refractive index turning surface. However, as in the case of $U_0 = 10 \text{ m s}^{-1}$, the differences from the nonlinear model results are quite significant. The downward reflection of the forced wave is clearly much weaker in the quasi-linear model, leading to stronger penetration to high altitudes.

This disagreement between the nonlinear and quasi-linear model results is obviously related to differences in the respective zonal mean wind fields as, in the nonlinear model, the wave-1 disturbance was shown to adjust almost linearly to changes in the zonal mean flow. Interestingly, however, a comparison of the mean flow fields obtained with the two models shows rather small differences. In the case $U_0 = 10 \text{ m s}^{-1}$, for example, the quasi-linear model produces a zero-wind line at the right altitude (see Fig. 18). At the same time, however, the width of the region of easterlies is considerably smaller than in the nonlinear model. Thus, part of the wave activity of the propagating disturbance can more readily refract from the sides, as shown in Fig. 16.

Similarly, in the case of $U_0 = 25 \text{ m s}^{-1}$, one has a clear reduction of the refractive index square values in the center of the channel, albeit slightly weaker than in the nonlinear model (see Fig. 19). An examination of the zonal mean potential vorticity field shows only a weak decrease of the mean potential vorticity gradients in the center of the channel in the quasi-linear integration, with the gradients remaining always positive, contrary to the nonlinear model results. Nonetheless, because of the meridional scale of the propagating disturbance, even weakly positive refractive index square values lead to vertical trapping and downward reflection. The decay, however, of the wave-1 amplitudes above the reflecting surface is relatively weak, leading to somewhat increased penetration with height.

Altogether the quasi-linear model results suggest an indirect role for nonlinearities in the saturation process. While wave–mean flow interactions are sufficient to first order to account for the saturation of the wave-1 disturbance, the results are quite sensitive to the detailed basic state. Neglecting wave–wave interactions leads to an incorrect prediction for the modified zonal mean flow, which produces an equilibrated state significantly different from that in the fully nonlinear model.

8. Transient runs

It is conceivable that our quasi-equilibrium approach did not permit breaking to play as big a role as it might. Observations of the forcing at the tropopause level suggest variations occurring over a time scale of only a few days. Therefore, we also considered a transient case in which forcing at the bottom is allowed to linearly increase to a value associated with breaking over varying time periods ranging from 200 days to 10 days. The results did not differ essentially from those obtained in quasi-equilibrium runs. Saturation was again associated with the development of closed Eulerian streamlines and, for the case in which wave–mean flow interactions lead to the development of a reflecting surface, PV contours are again found to “wrap,” indicating wave breaking (viz. Fig. 20, where we see the marked development of breaking shortly after closed streamlines form). These results did not depend on the turn-on time, although the shorter times did produce more pronounced vorticity wrapping. However, even for the short turn-on scales, results could be reasonably replicated in
linear calculations using the nonlinearly obtained basic flow. Nevertheless, the transient runs permitted us to focus on some features that were not altogether evident in the quasi-equilibrium runs. In the transient runs, the flow was examined for the first 220 days. This was sufficient for saturation to occur within the domain; further increases only caused saturation to appear at a lower altitude. One of the more interesting results was the appearance of a shorter time scale when saturation occurred. This was clearest when we used the long turn-on times; for short turn-on times the new scale was not so readily distinguishable. Figure 21 (where $\tau_{\text{turn-on}} = 200$ days) shows this clearly. As saturation occurs with the formation of closed Eulerian streamlines, a shorter time scale associated with eddy advection of vorticity enters (around day 160). The enstrophy budget was examined for this case as well. Once again, the wave–wave interactions do not act as local damping but, rather, act to reinforce wave-1 amplitudes in the center while weakening wave-1 amplitudes at the sides (much as found by Robinson 1988).

In the above transient calculations, it was found that the wave–wave term in the enstrophy budget was dominated by interactions involving wavenumbers 1 and 2. We therefore examined how well a low-resolution model could reproduce the results obtained with full resolution. For this purpose a numerical integration with a resolution of 16 grid points in both the zonal and meridional
direction, down from an original resolution of 64 × 32, was used.

The results showed the inadequacies of the low-resolution model for some purposes. On day 190, when in the “control” run the wave breaking event had fully developed, no sign of strong potential vorticity deformation was found, despite the appearance of closed contours in the streamfunction field. Naturally, the coarse horizontal resolution of the model would not allow for finescale filaments to develop. However, one would expect to observe a structure resembling qualitatively, albeit with much coarser features, the wave breaking deformation in the control run. Indeed, at very long times a wave breaking pattern does develop in the low-resolution model. However, the complete absence of such a signature from the potential vorticity maps on day 190 seems to signal the inability of the low-resolution model to properly capture the dynamics and the time evolution of the flow.

However, a closer examination of the flow revealed a significantly different picture. A comparison between low- and high-resolution model results for the time evolution of the mean potential vorticity gradient at level 60 and for the response of the wave-1 component to the changes in the zonal mean flow shows extremely good agreement (see Fig. 22). Although the low-resolution model does not show the same abrupt reduction in the mean potential vorticity gradient values as the control run does when wave breaking initiates, the overall agreement remains remarkably good. Moreover, the low-resolution model captures almost perfectly the time evolution of the wave-1 disturbance found in the high-resolution run.

**FIG. 17.** As in Fig. 16 but for $U_0 = 25$ m s$^{-1}$.

**FIG. 18.** Modified zonal mean wind field as observed in (a) the nonlinear numerical model and (b) the quasi-linear version for $U_0 = 10$ m s$^{-1}$, $\Psi_{\text{bottom}} = 7.0 \times 10^5$ m$^2$ s$^{-1}$.
The conclusion thus seems to be that the details of the small-scale deformation during wave breaking are not important for the large-scale dynamics. While small scales do cascade enstrophy downscale, their role being absolutely crucial in an undamped system, in the present case they seem to be of no particular importance to the large-scale flow. To the degree, of course, that the role of the nonlinear wave interactions is simply to cascade enstrophy downscale and to dissipate the primary wave (McIntyre and Palmer 1985; Garcia 1991) the numerical truncation and filtering that is applied in the low-resolution model should lead to qualitatively similar results. However, it should be kept in mind that the analysis of the flow revealed a much more complex behavior. Thus, the success of the low-resolution model suggests that although high-resolution models may be necessary if one is interested in the exact horizontal structure of potential vorticity or a chemical tracer, low-resolution models appear to be perfectly adequate to describe the basic dynamical interactions that affect the large-scale flow itself.

It is possible that wave–wave interactions are underestimated in the present study because only wavenumber 1 is forced at the bottom of our domain. Observationally, there are usually several wavenumbers present in the stratosphere: mostly wavenumbers 1 and 2. To check the possible influence of wavenumber-2 forcing on the saturation of wavenumber 1, we conducted runs in which wavenumber-2 forcing was permitted to equilibrate in the sense of the earlier sections of this paper, and then wavenumber-1 forcing was turned on in the time-dependent fashion introduced in this section.

The forcing of the wave-2 disturbance is such that the maximum wave-2 amplitude is on the order of 25% of the maximum amplitude attained by the wave-1 disturbance in the control run. In the stratosphere the wave-2 disturbance is generally observed to be significantly weaker than the wave-1 component (possibly because of refraction toward the equator, which would prevent it from penetrating to high altitudes; however, as we will show, the saturation of wavenumber 1 also acts to diminish wavenumber 2 at even lower altitudes). In any event, our choice for the strength of the wave-2 disturbance in the basic state is such as to leave space for a flow evolution sufficiently different from the control run while at the same time making sure one is not led to a totally unrealistic configuration. As it turns out, except for the presence of the wave-2 disturbance, the basic state is almost identical to that of the control run, making the comparison with the results of the control run relatively straightforward. In point of fact, the presence of the wave-2 component in the basic state did not seem to lead to any significant differences in the evolution of wavenumber 1.

Isolating and looking at the time evolution of the wavenumber-2 component at the levels where wave breaking occurs shows why there is so little difference. The wave-2 streamfunction amplitude is found to decrease significantly once the wave-1 forcing is turned on, despite the fact that the forcing of the wave-2 component remains constant throughout the integration. By day 170 when wave breaking begins, the wave-2 amplitude is found to have already been reduced by a factor of 6. With such low amplitudes it is only natural that the evolution of the flow is identical to the case of the purely zonal basic state.

At least one reason for the reduction of the wavenumber-2 component is clear from Eq. (2). The changes in the zonally averaged state that act to trap and reflect wavenumber-1 will be more effective for higher wavenumbers—and lead to their trapping at lower levels. To
check this, we compared the nonlinear model results with a linear run, again obtaining the zonal mean flow from the nonlinear integration and imposing it as a time-dependent basic state. In the linear calculation both the wave-1 and -2 components are allowed to adjust to the changes in the zonal mean flow; however, no interaction between them is allowed. The results of this comparison showed that the rate of decrease of the wave-2 amplitudes in the nonlinear model is at least twice as fast as in the purely linear integration. Thus, interactions with wavenumber 1 are important for the reduction of wavenumber 2. On the other hand, the linear calculation gives excellent results for the wave-1 disturbance, producing almost perfect agreement with the nonlinear model. The picture that arises from the synoptic maps thus seems to be confirmed—in that the presence of the wave-2 disturbance in the initial flow does not seem to significantly affect the development of either wavenumber 1 or the zonally averaged flow. Our results, moreover, offer a plausible explanation for the observed anticorrelation between the wave-1 and the wave-2 amplitudes, with episodes of strong wave-1 growth coinciding with weak wave-2 amplitudes and vice versa (OP88).

9. Summary and discussion

The results of the present study clearly suggest that one can understand the saturation of vertically propagating Rossby waves through a wave–mean flow framework. In the stratospheric literature, wave–mean flow interactions are sometimes dismissed as inadequate in
limiting the penetration of planetary-scale anomalies. In our model, however, while wave breaking seems to be intimately related to saturation of the propagating anomaly, a wave–mean flow point of view seems to allow for a fuller physical understanding of the dynamics and the interactions that affect the evolution of the flow.

In other words, wave breaking appears to be an unavoidable consequence of the penetration and growth of the propagating disturbance. At the same time, however, the nonlinear wave interactions, associated with it, do not seem to have a strong direct effect on the large-scale disturbance. Rather, it is the changes in the zonal mean component of the flow that inhibit the penetration of the wave in the vertical and ultimately lead to its saturation. Thus, the evolution of the flow seems to follow a simple cycle, whereby the penetration of the forced wave leads to a decrease in the zonal mean winds and potential vorticity gradients, which in turn reduce (and even eliminate) its ability to propagate in the vertical.

In view of these results it is important to examine in more detail why wave–mean flow theory has often been disfavored in the recent stratospheric dynamics literature. A first point that is raised regards the separation of the flow into Fourier components. Such a separation, as is justly argued, is at times rather unphysical, particularly when the deformation of the potential vorticity field produces localized features. It is further argued however that, even when dealing with large-scale patterns such as those arising from the growth of a planetary-scale anomaly, one should not reason in terms of a wave–mean flow system but rather consider the deformed flow as an entity (OP88). Thus, the usual practice is to describe the flow in terms of a displaced polar vortex rather than referring to the superposition of a zonal mean flow with a propagating wave, because it is argued that the propagation of the eddy disturbance cannot be well captured by a theory that relies on an Eulerian mean representation of the basic state (McIntyre 1982; Palmer and Hsu 1983).

However, while the above arguments warrant attention, our results suggest that a wave–mean flow type of description possesses strong interpretive powers. In this respect the validation for the use of such an approach comes from the results themselves. As has been shown, despite the strong horizontal deformation of the potential vorticity field, the behavior of the forced disturbance can be explained quite successfully through linear propagation notions. The strong deviation from the small wave amplitude requirement for linear theory to be valid does not appear to hinder a meaningful separation of the flow into a zonal mean and an eddy component.
It should be noted, however, that the present study, while having identified the mechanism through which saturation is achieved, has not led to a quantitative estimate of a saturation limit for wave amplitudes. Attempts to derive such a limit and to parameterize the saturation dynamics have of course been made in the past (Lindzen and Schoeberl 1982; S82; Garcia 1991). Schoeberl’s study was particularly interesting because it attempted to derive an upper limit on wave amplitudes from first principles, without any particular assumptions about the dynamics leading to saturation. Schoeberl’s derivation simply relied on the notion that at steady state a balance between enstrophy sources and sinks should exist at all vertical levels. Assuming that the interactions are primarily quasi-linear, S82 calculated the saturation limit to be

\[ |q'| = \frac{\beta}{\sqrt{2k_y}}. \]

Interestingly, this limit was found to be independent of damping. Strictly speaking, that conclusion is true only for the particular formulation of damping used in S82 in which equal Rayleigh friction and Newtonian cooling coefficients were used. It was argued, however, that the results were not strongly dependent on that assumption. In contrast, our study finds no unique saturation limit.

A point that needs further attention concerns the relevance of our results to the real stratosphere. The absence of a zero-wind line from our basic state configuration, in conjunction with other idealizations of our setup, limits our ability to directly relate our results to observations. In the present study, wave breaking is solely a result of the vertical penetration of the forced anomaly. In the stratosphere it is argued that the presence of the zero-wind line is essential for the breaking of the waves. In this respect it is not obvious a priori that the question of wave saturation (i.e., of a dynamical mechanism limiting the vertical penetration of a forced anomaly) applies to the observed nonlinearities of the stratosphere at all. On a conceptual level one has to reconcile the description viewing wave breaking as a consequence of “excessive” wave growth with a WW78 type of description, where concerns about the amplitude of the propagating wave are peripheral, only affecting the width of the critical layer that is predicted to form.

However, there is evidence that our model results might have counterparts in the stratosphere. We have already mentioned the observed anticorrelation between wavenumbers 1 and 2 (OP88). In addition, Harnik and Lindzen (2001) present results, based on analyzed data of the Southern Hemisphere, showing a strong modification of the zonal mean flow at the upper stratosphere during episodes of strong wave-1 growth. Their results show a good correlation between the wave-1 induced Eliassen–Palm flux divergence and the observed mean wind deceleration, thus attributing the modification of the mean flow to the penetration of the wave-1 anomaly. Calculating the change in the index of refraction from the modified zonal mean wind field, Harnik and Lindzen found a refractive index turning surface forming as a result of a reduction in the mean potential vorticity gradients. At the same time the observations showed a downward reflection of the wave-1 disturbance and a reduction of its amplitude at the upper levels of the stratosphere. Thus, the patterns described in Harnik and Lindzen seem to be directly analogous to the flow evolution in our idealized numerical integrations. That said, more recent results (Harnik 2009) seem to indicate that realistic simulations depend on the finiteness of the stationary wave packet, and our results do not consider a trailing edge for the vertically propagating wave. This, however, does not detract from our main goal, which was to investigate how vertically propagating waves that grow inversely with mean pressure in their linear regime saturate nonlinearly.

Although the question of the relevance of our results to the observed stratosphere is not tackled directly, the present results suggest that the traditional wave–mean flow approach deserves more attention than it has received in recent years, with the added caution that such results may depend significantly on details of the mean flow that, in turn, depend on higher-order nonlinearities.

Acknowledgments. The authors thank Alan Plumb for many useful discussions and suggestions on this work, as well as Kirill Semeniuk and Nili Harnik for comments on the manuscript. Support for this research came from Grants ATM-9813795 from the National Science Foundation, NAG5-6304 from the National Aeronautics and Space Administration, and DEFG02-93ER61673 of the Department of Energy (DOE-CCPP).

REFERENCES


