Nonlinear Forcing of Planetary Scale Waves by Amplifying Unstable Baroclinic Eddies Generated in the Troposphere

RICHARD E. YOUNG

Space Science Division, NASA-Ames Research Center, Moffett Field, CA 94035

GARY L. VILLERE

Informatics Professional Services, Inc., Palo Alto, CA 94303

(Manuscript received 25 October 1984, in final form 2 April 1985)

ABSTRACT

A global distribution of amplifying intermediate scale baroclinic modes can cause growth of planetary scale modes through wave–wave coupling, such that the growth of the long waves is comparable to the growth of the most unstable baroclinic modes. A global, spectral, primitive equation circulation model is used to study this phenomenon for two initial mean zonal wind fields having idealized meridional distributions, but realistic vertical variations. For the cases treated here, both the direct transfer of kinetic energy from intermediate scales of motion, and baroclinic conversion of available potential energy to kinetic energy within a given planetary wave, are important. It appears that direct kinetic energy transfer to the long waves can be inhibited by heat transport. Planetary wave amplitudes are considerably enhanced by wave–wave coupling over what they would be due to baroclinic instability of the zonally averaged state to planetary wave disturbances. Nonlinear forcing in the troposphere may be an important mechanism for generating eastward propagating planetary waves observed in the Southern Hemisphere. The planetary wave forcing terms have relative maxima near the tropopause. Such nonlinear forcing near the tropopause may explain the lack of coherence between the troposphere and stratosphere observed in eastward traveling waves with zonal wavenumbers 1 and 2.

1. Introduction

The present study is principally concerned with isolating the effects of wave–wave coupling on the growth of planetary-scale baroclinic modes of zonal wavenumber 1–3. Using primitive equation models, Gall et al. (1979a,b) and Frederiksen (1981c) have numerically investigated baroclinic scale selection processes and energy spectra for realistic Northern and Southern Hemisphere mean zonal states, respectively. However, only in Gall et al. (1979b; referred to hereafter as GBS), was there presented detailed analysis of the dynamics of the planetary, or ultralong, baroclinic waves for the particular mean zonal state considered. There it was argued that the ultralong waves were forced principally by planetary scale variations in the meridional heat flux convergence of the higher wavenumber modes, producing a positive correlation between planetary wave upward motion and temperature. The principal kinetic energy source for the planetary waves was the conversion of eddy available potential energy to eddy kinetic energy associated with this positive correlation.

Here we address the question of whether or not nonlinear forcing may be more important in producing eastward traveling planetary waves in the upper troposphere and lower stratosphere than baroclinic instability of zonally averaged flows to planetary scale baroclinic disturbances. Several linear growth studies have indicated that baroclinic instability in the troposphere can produce planetary scale modes which extend well into the stratosphere (Straus, 1981; Hartmann, 1979; Simmons and Hoskins, 1977). Such modes have been suggested as an explanation of eastward traveling wave features of zonal wavenumber 1–3 observed in the stratosphere from satellites, principally in the Southern Hemisphere (Leovy and Webster, 1976; Hartmann, 1976, 1979; Mechoso and Hartmann, 1982). Typical growth times for zonal wavenumber 1 disturbances are 10–20 days, decreasing to 4–8 days at zonal wavenumber 3. Hartmann (1983) has shown that barotropic instability of the southern hemisphere winter polar jet to planetary scale perturbations may produce eastward traveling planetary modes in the Southern Hemisphere stratosphere having growth times of the order of several days. However this mechanism would not explain the eastward traveling planetary modes in the troposphere discussed by Mechoso and Hartmann (1982). Under certain conditions Frederiksen (1982a,b) has shown

© 1985 American Meteorological Society
that eastward propagating modes can arise due to instability of nonaxisymmetric three-dimensional states. While such linear growth rate studies indicate growth of planetary scale modes due to various forms of instability, the question remains as to how planetary modes might be forced by intermediate scale baroclinic unstable modes occurring in the troposphere which have growth times of 2–3 days. The computations of GBS indicate that nonlinear wave–wave coupling can induce growth of planetary scale baroclinic modes, such that the growth rates achieved are comparable to the linear growth rates of the most unstable intermediate-scale baroclinic modes. Such an effect is also apparent in the results of Frederiksen (1981c). Our results indicate that as far as amplification of planetary waves is concerned, nonlinear forcing of the planetary modes is much more important than baroclinic instability of the zonally averaged state to planetary scale perturbations. Such nonlinear forcing may partially explain the observational result from Mechoso and Hartmann (1982) that Southern Hemisphere eastward moving waves with zonal wavenumber 1 and 2 have small coherence between the troposphere and stratosphere.

At first glance it might be expected that if wave–wave coupling is the primary forcing mechanism of the planetary waves, the planetary waves might grow at a rate significantly larger than the most unstable baroclinic modes. For example, consider the situation where the linear growth rate spectrum peaks near zonal wavenumber 8 (which is usually not too far from the case). Modes having zonal wavenumbers 6 and 8 will force a wave 2 disturbance, and similarly for modes with wavenumbers 5 and 7, 7 and 9, and so on. Assuming wave–wave coupling produces the dominant terms governing the growth of wave 2, one might expect the amplitude of wave 2 to grow at a rate equal to approximately twice the growth rate typical of waves having a zonal wavenumber near 8. The fact that the computed growth rate is about half this suggests that there is significant cancellation when the nonlinear terms in the equations are summed over all waves.

In order to easily isolate the effects of varying jet structure, we have chosen to consider initial mean zonal states very similar to the 45 and 30° jet cases considered by Simmons and Hoskins (1977, 1978). These states have realistic vertical distributions but idealized meridional variations. As will be shown when results are discussed, there are significant differences between the two jet cases in the growth of the planetary modes. In a future paper we shall treat climatological basic states and make more direct comparisons between characteristics of computed baroclinic modes and observed characteristics of eastward traveling planetary scale features.

Details of the numerical model are given in Section 2, the basic states are described in Section 3, and results are presented in Sections 4–7.

2. Numerical model

a. Model characteristics

The numerical model is based on the hydrostatic primitive equations on a sphere. Neither diabatic heating, moist processes, nor topography are included. The vertical coordinate is taken as \( z = -H \ln(p/p_0) \), where \( p \) is pressure, \( p_0 \) is a fixed globally averaged surface pressure, and \( H \) is a constant equal to the scale height defined with the initial globally averaged surface temperature. When referring to altitude in the text or figures we shall be referring to the quantity \( z \). The appropriate basic equations are given in Holton [1975, Eqs. (2.1)–(2.5)], and need not be repeated here, the only note being that the model equations are formulated in terms of colatitude, rather than latitude. The following notation is used:

- \( t \) time
- \( \theta \) colatitude
- \( \phi \) longitude
- \( a \) earth radius
- \( u \) zonal velocity
- \( v \) meridional velocity
- \( w \) \( dz/dt \)
- \( T \) temperature
- \( \Phi \) geopotential
- \( m \) zonal wavenumber

\[
\nabla H^2 = \frac{1}{a^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{a^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}.
\]

We define the horizontal velocities in terms of streamfunction \( \Psi \) and velocity potential \( S \), such that

\[
\begin{align*}
    u &= -\frac{\partial \Psi}{\partial \theta} + \frac{1}{\sin \theta} \frac{\partial S}{\partial \phi} \quad \text{and} \\
    v &= \frac{1}{\sin \theta} \frac{\partial \Psi}{\partial \phi} + \frac{\partial S}{\partial \theta}
\end{align*}
\]

The basic equations, when reformulated in terms of \( \Psi \) and \( S \), yield equations for the vorticity \( \zeta \) and horizontal divergence \( D \), respectively (cf. Hoskins and Simmons, 1975).

All scalar variables \( \Psi, S, T, w, \) and \( \Phi \), are represented in terms of spherical harmonics, such that

\[
F(\theta, \phi, z, t) = \sum_{m=-M}^{M} \sum_{n=|m|}^{N(m)} \sum_{l=0}^{L(m,n)} F_{m,n}(z, t) Y_{n}^{m}(\theta, \phi).
\]

In (2.2) \( Y_{n}^{m}(\theta, \phi) \) is a spherical harmonic normalized so that, denoting the complex conjugate of \( Y_{n}^{m} \) by \( Y_{n}^{m*} \),

\[
\int_{0}^{2\pi} \int_{0}^{\pi} Y_{n}^{m*} Y_{n}^{m} \sin \theta d\theta d\phi = 2\pi.
\]

In the vertical direction a grid is established and finite differences are used to evaluate derivatives with respect to \( z \). The level spacing is arbitrary, except that even
levels are placed midway between odd levels. The variables $\Phi$ and $w$ are defined at the odd levels, all other variables are kept at the even levels. The vertical differencing scheme is based on that described in Arakawa and Lamb (1977); we have chosen the scheme so that the quadratic functions $(u^2 + v^2)$ and $T^2$ are conserved with respect to a mass weighted global integral over the entire domain.

At the upper boundary $w$ is set to zero. At the lower boundary $z = 0$, the quantity

$$\frac{d\Phi}{dt} = \frac{\partial \Phi}{\partial t} + \frac{u}{a} \frac{\partial \Phi}{\partial \theta} + \frac{v}{a} \frac{\partial \Phi}{\partial \phi} + w \frac{\partial \Phi}{\partial z}, \quad (2.4)$$

which represents the geometrical vertical velocity times gravity, is set to zero (cf. Holton, 1975, p. 57). Any quantities defined at the even levels which are required to evaluate terms at the lower boundary are linearly extrapolated from the interior. The existence of a rigid upper boundary introduces some wave reflection of energy back into the region of interest. However, the upper boundary is placed high enough so that although the modal amplitude itself near the upper boundary may be comparable to values at much lower levels, the energy density is negligible. Hence any reflection occurring at the upper boundary will not significantly affect the regions of interest, namely the troposphere and lower stratosphere.

The time integration scheme used here is similar to the semi-implicit gravity wave schemes described by Robert et al. (1972), Hoskins and Simmons (1975), and numerous others. The essential aspect of the scheme is that the gravity wave terms in the temperature and divergence equations are appropriately separated from the remaining terms in the equations, and treated implicitly. All other terms are time integrated using an explicit centered time difference. A time filter suggested by Robert (1966) and used by Simmons and Hoskins (1976) was used to eliminate the computational mode from the solutions. Time steps of 1 hour were used, except when advective time scales required use of a smaller time step, in which case the time step was set to \( \frac{1}{2} \) hour.

b. Spatial resolution

In order to adequately resolve the vertical structure of the growing baroclinic modes, both in the troposphere near the surface and in the stratosphere, we chose to space odd levels of $z$ according to

$$z_k = \left(\frac{a - b}{V_b}\right) \tan^{-1} \frac{x_k}{V_b} + x_k, \quad (2.5)$$

where $x_k = 0.5 (k - 1) \Delta x$, $k = 1, 3, 5, \ldots, a = 25 \times 10^6 m^2$, $b = 4 \times 10^8 m^2$, and $\Delta x = 3.5 \times 10^7 m$.

Table 1 presents the values of $z$ associated with the odd $k$ levels. At the surface $\Delta z \approx 0.25 km$, while in the stratosphere $\Delta z \approx 3 km$. The maximum horizontal resolution was $M = 20$, $N(m) = 40$.

c. Dissipation

Some form of dissipation is required in order that nonlinear interactions do not destroy the computations by accumulating energy in the smallest resolved scales. However, care must be taken to ensure that nonlinear interactions are still accurately represented for the scales of motion that are of interest. In the present calculations we have considered scale selective damping formulations which act only on horizontal scales of motion, and which are linear in the field variables. It was found that in order to obtain solutions for which the amplitudes, etc. of the flow variables were insensitive to the amount of damping, the resolution had to be significantly larger than that required to adequately resolve the linear baroclinic eigenfunctions. Initially we tried a damping of the form $\nu \nabla^4_H$, where $\nu$ is a constant, with the damping applied in the vorticity, divergence, and temperature equations. Such a formulation has been used by Simmons and Hoskins (1978) with $\nu = 2.34 \times 10^{16} m^4 s^{-1}$; others have also used a $\nabla^4_H$ formulation. It was found that when modes of all zonal wavenumber were allowed to interact simultaneously in the calculations, the above value of $\nu$ was such that the solutions were significantly affected by the dissipation. Simmons and Hoskins noted in the cases they considered, for which $N(m) = 42$, that the above value of $\nu$ appeared to be near the minimum which maintained coherent fields of vorticity and divergence. Since $N(m)$ used here is comparable to theirs, we therefore chose a more scale-selective damping of the form $\nu \nabla^3_H$, with $\nu = 4.0 \times 10^{37} m^8 s^{-1}$ in an attempt to preserve accuracy at large and intermediate scales, while at the same time sufficiently damping the smaller scales. This

<table>
<thead>
<tr>
<th>Level</th>
<th>$z$ (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.252</td>
</tr>
<tr>
<td>5</td>
<td>0.687</td>
</tr>
<tr>
<td>7</td>
<td>1.435</td>
</tr>
<tr>
<td>9</td>
<td>2.549</td>
</tr>
<tr>
<td>11</td>
<td>4.022</td>
</tr>
<tr>
<td>13</td>
<td>5.817</td>
</tr>
<tr>
<td>15</td>
<td>7.884</td>
</tr>
<tr>
<td>17</td>
<td>10.177</td>
</tr>
<tr>
<td>19</td>
<td>12.654</td>
</tr>
<tr>
<td>21</td>
<td>15.282</td>
</tr>
<tr>
<td>23</td>
<td>18.031</td>
</tr>
<tr>
<td>25</td>
<td>20.880</td>
</tr>
<tr>
<td>27</td>
<td>23.813</td>
</tr>
<tr>
<td>29</td>
<td>26.814</td>
</tr>
<tr>
<td>31</td>
<td>29.872</td>
</tr>
<tr>
<td>33</td>
<td>32.979</td>
</tr>
</tbody>
</table>
approach proved successful. The attained kinetic energy for any particular zonal wavenumber mode increased in many instances by over a factor of 2 when the $\nabla_H^8$ dissipation formulation was used instead of the $\nabla_H^6$. However, the solutions were relatively insensitive to variations of $\nu$ about the value $4 \times 10^{37}$ m$^8$ s$^{-1}$. Reduction of $\nu$ by a factor of 2 for the $\nabla_H^8$ formulation showed $< 30\%$, but typically $< 15\%$, variation in the kinetic energy spectrum below zonal wavenumber 10 during a 12-d comparison integration. For example, the maximum kinetic energies attained by the $m = 8$ and $m = 9$ modes changed by 2% and 15%, respectively. Thus we were able to successfully represent nonlinear transfer of energy and momentum between intermediate and planetary, or large, scales and dissipate energy at the small scales. Subsequent to completing the above calculations we became aware of the work of MacVean (1983), who considered the effects of various linear scale selective damping formulations on the growth and finite amplitude behavior of intermediate scale baroclinic modes. Our conclusions regarding the relative merit of the $\nabla_H^6$ and $\nabla_H^8$ formulations, and the value of $\nu$ required, are very similar to MacVean’s (cf. also Frederiksen, 1981b). It should be noted that when $N(m) = 30$ in our calculations, the value of $\nu$ required for stable calculations was between $4.8 \times 10^{16}$ and $9.6 \times 10^{16}$ m$^8$ s$^{-1}$ for the $\nabla_H^6$ damping, and between $2 \times 10^{38}$ and $4 \times 10^{38}$ m$^8$ s$^{-1}$ for the $\nabla_H^8$ damping. In each of these cases the amount of damping was such that the solutions were considerably affected by the dissipation. The impaction is that in order to achieve an accurate representation of large and intermediate scales of motion, the necessary spatial resolution and required damping are interrelated and problem dependent, a conclusion consistent with MacVean (1983) and Williamson (1978) among others.

The effects of the $\nabla_H^8$ dissipation on the linear growth rate spectrum of the 45° jet basic state are illustrated in Fig. 1 for $\nu = 4 \times 10^{37}$ m$^8$ s$^{-1}$. The details of the 45° jet structure are given in the next section. Growth rates were computed using a primitive equation eigenvalue stability model developed by H. Houben, which has exactly the same vertical resolution as that in the nonlinear calculations, but uses rhomboidal truncation with truncation wave number 18, i.e., $N(m) = m + 18$. It is clear that for $m \leq 12$ the dissipation has little effect on growth rate. For comparison the growth rate curve for a $\nabla_H^6$ dissipation using $\nu = 2.4 \times 10^{16}$ m$^8$ s$^{-1}$ is also shown.

3. Initial conditions and linear growth rates

Two initial mean zonal states are considered, each having a mean zonal wind of the form $\bar{u} = f(z) g(\theta)$, where an overbar denotes a zonal average. The two forms of $g(\theta)$ chosen are identical to the 30 and 45° jets treated in Simmons and Hoskins (1977, 1978), namely $g(\theta) = \sin^2 \pi \mu$ and $g(\theta) = \sin^2 \pi \mu^2$, respectively, where $\mu = \cos \theta$. The numerical values for the vertical function $f(z)$ were supplied to us by E. F. Danielsen (private communication, 1984), and represent a typical Dec–Feb average at 30°N. Figure 2 illustrates the wind and temperature fields for the two jets. The global average thermal structure was taken to be that given in the U.S. Standard Atmosphere, 1976. The horizontally varying part of the mean temperature was determined by thermal wind balance with each of the above zonal wind fields. As will be discussed later, there are significant differences between the 30° jet and 45° jet cases as far as the development of the planetary-scale modes is concerned, and it was in anticipation of such a result that we chose to consider meridional variations in the mean state that could easily be identified and varied. For the 45° jet two types of solutions were obtained: one in which all wave–wave interactions were accounted for, and one in which the growing baroclinic eddies were allowed to modify only the basic state.

Each integration had as initial conditions one of the above zonal mean states perturbed by small amplitude disturbances at each zonal wavenumber. At a particular zonal wavenumber, the field variables were set to the eigenfunction corresponding to the most unstable baroclinic mode of the 30° jet at that zonal wavenumber, suitably scaled so that the maximum amplitude of the geopotential at the surface for a given value of zonal wavenumber was approximately 2.5 m. The eigenfunctions used for the initial conditions were computed with a linearized version of the present numerical model, in a manner exactly analogous to that used by Simmons and Hoskins (1976). The phases of the modes were taken from the end of this linear integration, and were therefore set to whatever they happened to be at that time. The initial conditions for the nonlinear calculations imply a global distribution of eddies. GBS pointed out that eddy development may be an essentially local process, and the forcing and
maintenance of the planetary eddies could therefore appear unduly complicated when represented in a spectral model. The justification for using "global" initial conditions here comes from the fact that eastward traveling planetary waves are clearly observed, at least in the Southern Hemisphere, implying global scale processes for their forcing and maintenance. Furthermore, it is not uncommon to observe global distributions of eastward traveling waves of higher wavenumber at low levels (Palmen and Newton, 1969). Thus, it should be appropriate in the present simulations to use a spectral model to represent energy transfer mechanisms. The question is whether global forcing is due principally to growth of the wave through baroclinic instability of the zonal mean state, or whether wave–wave forcing is important. The results obtained here indicate that wave–wave forcing plays a dominant role in producing rapid growth of the planetary modes.

For reference linear growth rates and phase speeds as a function of zonal wavenumber are presented in Fig. 3 for each of the jet profiles. The effects of dissipation are included in the curves. The essential feature of the growth rate spectrum as far as the present work is concerned is the small growth rate of wavenumbers 1–3 compared with that at higher wavenumbers. This implies that in general the planetary modes evolve in the presence of higher zonal wavenumber modes whose amplitude at some point is likely to be considerably larger than that of waves with $m = 1-3$. The length of time over which the simulations were carried out was determined so as to include several linear growth times of the planetary modes.

4. Energetics

The temporal behavior of the kinetic energy for the different cases considered is illustrated in Figs. 4–6. The quantity shown is the kinetic energy per unit mass averaged over the entire atmosphere, denoted by KE. The most noticeable feature from the figures is the difference in time dependence of KE for the planetary waves with and without wave–wave coupling; compare Figs. 4 and 5. The results shown in Fig. 4 include the effects of all wave–wave interactions. Those given in
Fig. 3. Growth rate curves for $45^\circ$ jet and $30^\circ$ jet as function of zonal wavenumber, including $\nabla_\phi$ damping.

Fig. 4. Time variation of kinetic energy per unit mass, $m^2 s^{-2}$, for $45^\circ$ jet. All wave–wave interactions included. Solid, $m = 1$; short dashed, $m = 2$; long dashed, $m = 3$; long-short dashed, $m = 8$.

Fig. 5. As in Fig. 4 but only wave–mean field interactions are included.

more, the W–W planetary mode growth rates during the first 10 days or so are essentially the same as the growth rates of the most unstable baroclinic modes. GBS and Frederiksen (1981c) also computed rapid growth of the ultralong waves during early stages of integration. The initial large growth of the planetary waves is clearly associated with amplification of intermediate-scale waves. However, it is not just the growth of intermediate-scale waves that is important, since the growth rates of the planetary waves are considerably less than those which would exist if the amplitudes of the planetary waves were simply proportional to quadratic products of amplitudes of the intermediate-scale modes. In other words, the relative phasing of the intermediate-scale modes is important, and there is significant cancellation between the different modes in the planetary-wave forcing terms. Nevertheless, the implication of the results is that the early time behavior of the planetary modes depends in general more on the global distribution of the higher wavenumber modes than on the initial distribution of the planetary waves themselves. A global distribution of amplifying baroclinic eddies can produce rapid growth of planetary waves.

The planetary modes reach values of KE 2–5 times larger for the $45^\circ$ jet than for the $30^\circ$ jet. However, the
periods of rapid growth of the planetary modes $C(m)$ and $L(m)$ are generally comparable in magnitude, with both contributing to growth of the waves. This situation is somewhat in contrast to the results presented by GBS. Although they showed that both $C(m)$ and $L(m)$ contributed significantly to growth of the planetary waves, in their computations $C(m)$ was the dominant term. In the present calculations, however, baroclinic energy conversion is not always the primary source of kinetic energy for the planetary waves, especially in the case of the 45° jet. Periods of maximum growth are clearly associated with times when both $C(m)$ and $L(m)$ are positive, but $L(m)$ in many instances contributes to

$m = 5$ and $m = 6$ modes reach about the same amplitude in both cases. This region appears to be a crossover point in terms of maximum KE, since examination of KE as a function of wavenumber shows that lower wavenumber modes reach larger KE for the 45° jet than for the 30° jet, whereas higher wavenumber modes reach larger amplitude for the 30° jet. Such a result is qualitatively consistent with that obtained by Simmons and Hoskins (1978), who found that wave 6 reached larger KE for the 45° jet, but wave 9 reached larger KE for the 30° jet. It might be expected that all modes would reach larger amplitude for the 45° jet, which generally has larger linear growth rates. The fact that there is a crossover in maximum KE near $m = 5$ indicates that characteristics of the mean zonal wind and temperature fields, such as jet location, affect the kinetic energy attained by a mode with given zonal wavenumber.

The kinetic energy budgets for waves $m = 1–3$ for each jet profile are given in Figs. 7 and 8. The quantities plotted are $C(m)$, the conversion of available potential energy to kinetic energy within zonal wavenumber $m$; $L(m)$, the transfer of kinetic energy from all other zonal wavenumbers (not equal to zero) to the kinetic energy of wavenumber $m$; $M(m)$, the conversion of mean zonal kinetic energy to eddy kinetic energy with wavenumber $m$. It can be seen from the figures that during the initial
and so is the momentum flux. The reasons for decay of \( L(m) \) will be discussed in the next section. Conversion of zonal kinetic energy to eddy kinetic energy is generally less important for the planetary modes than for the higher wavenumber modes, and relatively larger for the 45° jet than for the 30° jet.

It is interesting to note that of the planetary modes, the \( m = 2 \) mode grows to the largest KE during the 28 days of integration for either jet (Figs. 4, 6), although the effect is much more pronounced for the 45° jet. It can be seen from Fig. 7b that the \( m = 2 \) wave is being forced between days 16 and 23 primarily by direct kinetic energy transfer from the higher wavenumber modes and barotropic kinetic energy conversion from the zonally averaged flow. Computation of the quantity

\[
\frac{\partial \tilde{q}_\theta}{\partial \theta} = 2\Omega \cos \theta - \frac{\partial}{\partial \theta} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \tilde{u}) \right]
\]

shows that by approximately day 8 of the integration for the 45° jet, \( \partial \tilde{q}_\theta/\partial \theta \) changes sign at low levels near 60° latitude. As the integration proceeds, the region of negative \( \partial \tilde{q}_\theta/\partial \theta \) expands horizontally and vertically, but by about day 25 when both \( L(2) \) and \( M(2) \) are near zero, this region has decreased in size and intensity, and is located only very close to the pole between 5 and 15 km altitude. We therefore attribute the amplification of the \( m = 2 \) mode beyond day 16 to barotropic instability involving at least the zonally averaged flow. Various wave components might also be barotropically unstable during this time, which may partially explain the rise in \( L(2) \) after days 14 and 21. Baines (1976) showed that above a critical amplitude isolated waves with given \( n \) and \( m \) are barotropically unstable to other sets of two waves, with the two waves in each set having the sum of their zonal wavenumbers equal to \( m \). A survey of the streamfunction amplitudes at day 16 showed that in many instances they exceeded the critical amplitudes given in Baines. Although the stability criteria computed by Baines may not be directly applicable in the present situation, they should give some indication as to whether barotropic instability is likely. It is notable that modes which exceeded the critical amplitudes the most were modes having zonal wavenumbers 3–5, and were the ones shown by Baines to be most unstable to waves having zonal wavenumbers 1, 2, and 4 (wave 4 was also increasing during the same time as wave 2).

5. Heat and momentum transport forcing the planetary waves

The dynamics of the wave–wave forcing of the planetary modes involves consideration of both heat and momentum transport. Clearly, if the intermediate-scale waves are to force the planetary waves, there must be planetary-scale variations in their heat and momentum fluxes. GBS presented a clear picture of how uneven
distributions of growing baroclinic eddies around the globe would result in planetary scale variations of eddy heat fluxes, thereby forcing planetary scale waves. Maximum planetary-scale temperature variations are produced where the divergence or convergence of heat flux is greatest, which occurs north and south of the developing intermediate-scale baroclinic modes. Planetary-scale upward motion is then forced where the planetary-scale temperature is enhanced, and similarly, downward motion is induced where there is cooling. This leads to positive correlation between planetary-scale vertical velocity and temperature, and hence positive values of the quantity $C(m)$. As noted previously, this was the principal mechanism by which the planetary modes in GBS were forced. However, in the present calculations $L(m)$ plays a more substantial role than was the case in GBS, and hence consideration must be given to how the growing intermediate scale baroclinic modes transport momentum, as well as to how the heat and momentum transport interact with each other.

The terms which contribute most to $L(m)$ are those involving

$$\frac{-2mp}{a \sin \theta} \text{Im}[u_m((u - \bar{u})(u - \bar{u}))^\ast_m]$$

$$- \frac{2p}{a} \text{Re}\left\{u_m \frac{1}{\sin^2 \theta} \frac{\partial}{\partial \theta} [\sin^2 \theta (u - \bar{u})(v - \bar{v})]^\ast_m\right\}$$

$$- 2 \text{Re}\left\{u_m \frac{\partial}{\partial z} [p(u - \bar{u})(w - \bar{w})]^\ast_m\right\}$$

(5.1)

(see Appendix; also cf. Saltzman, 1970). In (5.1) Re and Im stand for the real and imaginary parts, respectively; $u_m$ represents the zonal wind coefficient associated with zonal wavenumber $m$; $\{ \}$ the $m$th coefficient of the quantity $\{ \}$; and so on. That part of the planetary zonal wind driven by the net of the flux divergence terms in (5.1) will be anticorrelated with the net of the flux divergences, thereby tending to make expression (5.1) positive, and hence, cause a transfer of kinetic energy from the intermediate-scale waves to the planetary waves. Regions of increasing poleward momentum flux with latitude will tend to produce easterlies, regions of decreasing poleward flux will produce westerlies. Likewise, regions of increasing upward momentum flux with altitude will tend to induce easterlies, and vice-versa. That part of the planetary-scale zonal wind driven by planetary-scale horizontal temperature gradients may or may not be correlated with the intermediate-scale momentum flux divergences given in (5.1). In other words, the heat flux due to the intermediate-scale modes, which induces planetary-scale variations in temperature, may act to decrease the direct kinetic energy transfer from intermediate to large scales by indirectly producing a zonal wind which is out of phase with the momentum flux divergences. Consider a given longitude $\phi_0$, and a particular planetary wave having zonal wavenumber $m_0$. If at $\phi_0$ there is a net warming on the poleward side of the developing intermediate scale modes for the wave with $m = m_0$, and a corresponding cooling on the equatorward side, there will tend to be a planetary scale rising motion on the poleward side and descending motion on the equatorward side. This will tend to induce a wave $m_0$ poleward meridional motion at levels near or below the level of maximum heating/cooling, but equatorward motion above. Consequently, westerly zonal motion will tend to be induced at lower levels, with easterly zonal motion at higher levels. As $\phi$ varies, the above configuration will vary also, the sense of the circulation reversing at some longitude because it is associated with a wave with $m = m_0$. Whether or not the planetary scale zonal wind generated in this way is such as to produce a transfer of kinetic energy to wave $m_0$ via (5.1) depends on the phase of the zonal wind relative to the momentum fluxes, as well as latitude locations of the different quantities. Our results indicate, at least during the first 5–10 days of integration, that as the heat flux penetrates into regions where only momentum flux had been significant before, $L(m)$ subsequently decreases.

The discussion in the rest of this section refers to the 45° jet unless explicitly noted otherwise. Qualitatively similar results apply to the 30° jet. Figure 9 presents latitude–height cross sections of the $m = 3$ component of horizontal momentum and heat flux at days 4, 6, and 8. The corresponding results for both the $m = 1$ and $m = 2$ components are similar. It can be seen that up to day 6 the heat flux is confined to the lowest levels, while the momentum flux has relatively large amplitudes at higher levels. However, by day 8 both the momentum and heat fluxes have relatively large upper-level amplitudes. Such time behavior is qualitatively consistent with the zonally averaged heat and momentum fluxes computed by Simmons and Hoskins (1977, 1978) and Frederiksen (1981a,b,c). In these cases the linear modal eigenfunctions tended to have heat flux restricted to low levels, while momentum flux occurred at higher levels (Gall's, 1976, linear results differ in this respect). However all studies showed that once the modes had developed so that nonlinear effects were significant, heat flux exhibited upper-level maxima as did the momentum flux, in qualitative agreement with observation (Oort, 1983). Figure 7 shows that between days 5–8, the magnitude of $L(3)$ normalized by KE is decreasing. However, analysis shows that (see Fig. 4) between days 5 and 6 this decrease is due entirely to the amplification of the kinetic energy, so that the actual value of $L(3)$ is still slightly increasing during this period. However by day 8, $L(3)$ has become negative. Thus, $L(3)$ has rapidly decreased between days 6 and 8, a time during which the amplitudes of both the heat and momentum fluxes are increasing. Comparison of
Figs. 4 and 5 shows that it is principally wave-wave coupling which is forcing the planetary modes during this time; changes in the mean state, which occur mostly near the surface below the levels where most of the zonally averaged and wave component of momentum transport occur, are not the primary cause of the decrease in $L(3)$. The key point is that between days 6 and 8 the heat flux has developed at upper levels. The result of the upper level heat flux is apparently to destroy the phasing between the planetary scale zonal
wind and the net of the momentum flux divergences, so that by (5.1), \( L(3) \) decreases and even becomes negative.

As was mentioned previously, the planetary modes amplify at a rate similar to the rate of the most unstable baroclinic modes during the first 10 days or so. Since they are clearly being forced during this time by quadratic wave–wave interactions, the fact that they are not amplifying at a rate larger than that of the most unstable modes deserves attention. During the time that the kinetic energies of modes 4–10 increase by over an order of magnitude, between days 2 and 8, the planetary wave components of the heat and momentum fluxes increase by a factor \( \leq 5 \). Beyond day 2, after the waves are more developed, the increase of the fluxes relative to the kinetic energies is even less. Thus, the individual contributions to the wave fluxes do not add up in a completely coherent manner. This can be understood in terms of the structures of the intermediate-scale modes and how they evolve relative to each other. For the purposes of illustration, consider the meridional wave heat transport assuming the quasi-geostrophic approximation. A typical product term in the wave \( s \) component of the heat flux (\( vT \)) then is proportional to

\[
2 \text{Re} \left( \frac{im}{\sin \theta} \Psi_m e^{i\alpha_m \phi} e^{i\phi_m} \right) \cdot 2 \text{Re} \left[ \frac{\partial \Psi_{m+1}}{\partial z} e^{i\lambda_{m+1}} e^{i(m+1)\phi} \right],
\]

(5.2)

where \( \Psi_m \) is the streamfunction coefficient in the Fourier representation of \( \Psi \) in longitude, and \( \alpha_m \) is the phase of \( \Psi_m \). Expression (5.2) can be combined to give

\[
\frac{2m |\Psi_m|}{\sin \theta} \frac{\partial |\Psi_{m+1}|}{\partial z} \sin [s \phi + (\alpha_{m+1} - \alpha_m)] + \frac{2m}{\sin \theta} \times |\Psi_m| \frac{\partial \alpha_{m+1}}{\partial z} \cos [s \phi + (\alpha_{m+1} - \alpha_m)].
\]

(5.3)

Note that if \( s = 0 \), i.e., corresponding to the zonal average, the well known fact that there must be a phase tilt with height to obtain a nonzero heat flux is evident. For the wave components \( (s \neq 0) \) there is no such restriction, and it can be seen that there is a nonzero contribution to the heat flux due to the variation of amplitude with height. Expression (5.3) must be summed over \( m \) to obtain the \( s \)th component of the heat flux. It is clear from (5.3) that the relative phases and amplitude structures of the individual intermediate scale modes are important in determining the net flux. This is in contrast to the expression giving the kinetic energy of a particular mode, which involves only the amplitude and phase of the mode itself [a situation analogous to the case \( s = 0 \) in (5.3)]. As the intermediate scale modes grow and evolve with time, there is significant cancellation among the terms in the sum of (5.3) over \( m \), so that the \( s \)th component of the heat flux grows at a slower rate than the kinetic energy of the intermediate scale modes. The fact that once the waves are somewhat developed the flux terms increase relatively slowly, means that the amplitudes of the planetary modes increase according to a more or less fixed amount of forcing, rather than a forcing which grows with the amplitudes of the intermediate scale modes.

6. Structure

Figures 10 and 11 present representative temperature amplitudes of the planetary waves at days 6 and 26 for both jet profiles. Only the mode with \( m = 2 \) is shown for the purposes of illustration. The quantity plotted is the amplitude associated with zonal wavenumber 2 in an expansion of the form (2.2). The real physical amplitude would be twice as large because of the sum over positive and negative values of \( m \). Several general comments can be made. The amplitude of the 45° jet modes is typically larger than that for the 30° jet modes. At day 6 the waves are confined mainly to the troposphere, but by day 10 (not shown) they have significant relative amplitude in the stratosphere. Amplitudes continue to increase with time beyond day 10 in the stratosphere, but not a great deal at the lower levels. At low levels the waves spread towards high and low latitudes, but in the stratosphere they tend to be confined to latitudes above 30°. In the levels near the surface the double maximum structure discussed by GBS is usually apparent, although this structure can become somewhat more complicated as Figs. 10(b) and 11(b) indicate. With regard to multiple maxima, it should be noted that a double maxima structure can occur without wave–wave coupling. Hartmann (1979) computed such structures for the linear planetary wave eigenfunctions using a zonal wind field appropriate to the Southern Hemisphere. Multiple maxima are also apparent in the temperature structures of the M–W (no wave–wave coupling) solution for waves 1–3. Furthermore, the linear calculations of Hartmann, and those for the M–W solution, indicate that the phase shift is about 180° between the maxima. Thus multiple maxima at low levels shifted in phase by about 180° are not by themselves indications of wave–wave forcing, although the W–W solutions (which include wave–wave coupling) do exhibit such characteristics.

Even though the growth of kinetic energy is enhanced by wave–wave coupling, it is not obvious that the amplitude of the planetary modes should be significantly enhanced in the stratosphere, since higher wave number modes are confined to the troposphere. However, as Fig. 9 illustrated, by day 8 the planetary wave forcing terms have relative maxima near the tropopause, as is observed for the zonally averaged heat and momentum fluxes. Thus, the planetary waves
7. Comparison with observations

Because of the rather idealized nature of the initial zonal wind fields we have considered, together with the fact that we have not included latent heat release or topography, no concerted attempt is made here to quantitatively compare values of heat flux, etc., with observation. However some qualitative comparisons can be made. For the following discussion observational results are taken from Oort (1983). Figure 12 illustrates the total zonally averaged meridional heat transport at day 26 for the 45° jet. Only the contributions from modes with \( m > 0 \) are included. It can be seen that an upper-level local maximum has developed which is comparable to the low level heat trans-
port. The existence of an upper-level maximum is qualitatively consistent with observations of the northward heat transport by transient eddies, however the computed magnitudes at low levels are smaller than observed. Also the location of the upper maximum is somewhat higher than observed by day 26, being located near 160–180 mb rather than 200 mb. Figure 13 shows the computed zonally averaged total momentum flux, again including only modes with $m > 0$. Here the maximum momentum transport occurs at upper levels, in qualitative agreement with observation, although the magnitude of the transport once the waves are fully developed appears somewhat low. It is interesting to note that for both heat and momentum fluxes the maximum amplitudes do not diminish as the intermediate scale modes go through the decay phase of their life cycle. Thus the planetary waves, together with waves 4 and 5, make the major contributions to the fluxes in the later stages of integration, and maintain the fluxes at or somewhat above the earlier values.

The low-level double maximum structure in the planetary wave temperature fields, discussed previously, is seen in observations of eastward traveling planetary waves (Mechoso and Hartmann, 1982). From the observations alone it is not possible to determine whether or not wave–wave coupling is important in producing this structure, since as previously mentioned, linear eigenmodes can also exhibit a similar pattern. A set of calculations using realistic Southern Hemisphere mean zonal winds is needed so that detailed comparisons can be made between linear and nonlinear computed planetary wave structures and the observations.

Mechoso and Hartmann point out that there is little coherence between the troposphere and upper stratosphere for eastward propagating waves 1 and 2. They suggest several possible explanations. One explanation suggested by the result that the computed planetary wave forcing terms have relative maxima near the tropopause is that waves 1 and 2 are nonlinearly forced in the upper troposphere and lower stratosphere. Under these circumstances one would expect little correlation of the structures of waves 1 and 2 between the lower and middle troposphere and the upper stratosphere (see also the following discussion concerning Fig. 14). Mechoso and Hartmann used a pressure level of 300 mb for their reference point in computing the coherency. As mentioned above, the computed nonlinear forcing terms for waves 1 and 2 have local maxima above this level, consistent with the requirement that whatever causes the lack of coherence occurs between 300 mb and the upper stratosphere. If nonlinear forcing explains the lack of coherence of waves 1 and 2, one needs to explain why waves 3 and 4 are observed to be coherent from the surface to the upper stratosphere. The reason may be that the linear growth rate is substantial enough that nonlinear forcing is not so important for waves 3 and 4. Hartmann's (1979) results indicate that, depending on the particulars of a climatological Southern Hemisphere mean zonal wind profile, wave 3 has about twice the growth rate of wave 2; wave 4 should have a still higher growth rate. In the present calculations the ratio of wave 3 to wave 2 linear growth rate is near 1.5 (see Fig. 3), and nonlinear forcing of wave 3 has been significant. However, it is true that for $m = 3$ the computed structures of the W–W and M–W solutions at later stages in the integration appear more similar than those for waves 1 or 2.

If observations of eastward traveling planetary waves in the stratosphere refer to baroclinic modes forced mainly by wave–wave interactions, such baroclinic modes must have a consistent eastward phase progression. While it is expected that a linear baroclinic wave would have a regular eastward phase speed, it is not clear that a mode forced mostly by wave–wave interactions would have any such regularity. It turns out that the planetary waves computed here generally do exhibit an eastward phase progression. Although it is difficult to define a precise measure of the phase change with time for a nonlinear mode, the overall pattern seems to shift in an approximately consistent manner at a given altitude. It should be noted, however, that there are instances, for example the $m = 2$ mode of the 30° jet at low levels, which at times do not exhibit any apparent phase progression. Figure 14 shows the horizontal temperature pattern at $z = 16.5$ km and $z = 2$ km for the $m = 2$ mode of the 45° jet, spaced at a 6 day interval starting at day 21. It can be seen from Fig. 14 that there is an eastward progression of the wave pattern, such that over the 6 days covered by the figure the pattern has shifted by roughly 60°. The implied phase speed of the pattern progression varies with position, but roughly speaking the pattern shift in longitude, given the uncertainty of determination, is ap-

![Fig. 13. Zonally averaged momentum flux (wu), due to all waves with $m > 0$; 45° jet. In m² s⁻². Negative values correspond to poleward flux. Day 26.](image-url)
approximately consistent with phase speeds of the linear eigenmodes. The phase shift appears larger at the lower altitude in Fig. 14, and the locations in latitude and longitude of positive and negative values have changed significantly relative to the upper altitude. Such a result is consistent with significant nonlinear forcing being present between the two levels. As mentioned above, the 30° jet offers an even more pronounced example of the lack of correlation between lower and upper layers.

8. Conclusions and discussion

It has been shown that a global distribution of amplifying intermediate-scale baroclinic modes can cause rapid growth of planetary-scale modes. In fact, the growth rates of the planetary modes are comparable to the growth rates of the most unstable baroclinic modes during the first 5–10 days. Thereafter the planetary waves continue to grow (albeit at a slower rate and not necessarily monotonically), and do not exhibit
a decay phase as pronounced as that which is characteristic of the intermediate-scale modes. It is clear that during the initial period of rapid growth, wave–wave coupling is principally responsible for the amplification, with intermediate-scale modes feeding energy into the planetary waves. Such results are similar to those obtained by GBS, and confirm the view that wave–wave interactions must be accounted for when considering the energy sources for planetary waves in the troposphere and lower stratosphere. The total kinetic energy of the planetary waves is between one and two orders of magnitude less, after 28 days of integration, if wave–wave coupling is not accounted for, i.e., if the baroclinic modes are allowed to feedback only on the zonally averaged fields. As to why the planetary waves do not amplify at an even faster rate than they do, since they are being forced by quadratic nonlinear terms, has to do with the relative structure, both amplitude and phase, of the different intermediate-scale modes. The evolution of the different waves is such as to cause a lack of coherence in so far as the forcing of the planetary waves is concerned. This is evidenced by the fact that the planetary wave components of the heat and momentum fluxes produced by intermediate-scale modes do not amplify at a rate as large as the rates at which the kinetic energies of the intermediate scale modes amplify.

There are two ways in which the intermediate-scale baroclinic modes transfer kinetic energy into the planetary modes: direct kinetic energy transfer from the intermediate-scale modes, and available potential energy production, in which heat transport by the intermediate-scale modes causes planetary scale temperature variations, and a subsequent conversion of planetary available potential energy to kinetic energy. In the present calculations both processes can be of comparable importance for generating the planetary-scale waves. This is not always the case, however, and there are instances where baroclinic conversion is dominant, a situation consistent with that found by GBS. It is not clear at the moment what circumstances determine the relative importance of each type of energy transfer. In the present case it was found that during the time when the heat transport was confined to levels near the surface, but the momentum transport occurred at higher levels, the transfer of kinetic energy to the planetary waves from the intermediate-scale modes was relatively large. However, once the heat transport also had significant values at the same levels as the momentum transport, direct kinetic energy transfer to the planetary waves rapidly decreased, and even reversed sign. The wind fields indirectly generated by the heat flux divergences were evidently not in phase with the momentum flux divergences, hence the decrease in direct kinetic energy transfer.

That amplifying intermediate-scale baroclinic eddies should produce rapid growth of planetary waves is a result which appears to be independent of the particular form assumed for the initial zonally averaged basic state or the detailed spatial dependence of small initial disturbances. The present study, GBS, and Frederiksen (1981c) used different basic states and initial small disturbances, yet each calculated growth of planetary waves well in excess of the growth rate corresponding to linear baroclinic instability of the basic state to planetary scale perturbations. However, since each of the studies scaled the initial disturbances so that they were of small amplitude, a question which deserves future attention is the sensitivity of nonlinear forcing of planetary modes to initial perturbations of differing but substantial amplitude.

Amplitudes of the computed planetary modes are considerably enhanced in the upper troposphere and in the stratosphere over what they would be in the absence of wave–wave coupling, i.e. enhanced over what they would be if the growing unstable baroclinic modes interacted only with the zonally averaged state. The zonally averaged zonal wind fields considered here are barotropically stable everywhere, at least initially. Therefore propagation of the planetary modes from the regions in which they are forced, the lower troposphere and near the tropopause, accounts for the amplitude enhancement in the stratosphere. The computed planetary wave components of heat and momentum fluxes, due to all other waves, exhibit relative maxima near the tropopause, a situation also observed for the zonally averaged heat and momentum fluxes of transient eddies (cf. Oort, 1983). Although nonlinear forcing is dominant in generating waves 1 and 2 in the present computations, these waves tend to exhibit, but not always, an eastward phase progression which is approximately that of the linear baroclinic eigenmodes. The lack of coherence of eastward traveling waves 1 and 2 between the stratosphere and troposphere reported by Mechoso and Hartmann (1982) may be related to nonlinear forcing of these waves near the tropopause. Future work involving realistic Southern Hemisphere mean states is required in order to better assess this possibility, and at the same time investigate the importance of the instabilities discussed by Hartmann (1983) and Frederiksen (1982b).

Acknowledgments. We would like to thank H. Houben for providing his linear eigenvalue stability model to compute linear growth rates. Thanks are also due to L. Pfister and J. R. Barnes for many valuable discussions.

APPENDIX

Expression for Kinetic Energy Conversion

The total expression for the transfer of kinetic energy from all wavenumbers greater than 0 to wavenumber $m(\neq 0)$ is given by the sum of $L_1$ and $L_2$, where
\[ L_1(m) = \int_V \left\{ \frac{2m}{\alpha \sin \theta} \text{Im}\{u_m[(u - \bar{u})(v - \bar{v})]^*_m \right. + v_m[(u - \bar{u})(v - \bar{v})]^*_m \right. \\
+ \frac{2}{a} \text{Re}\left\{ \frac{u_m}{\sin^2 \theta} \frac{\partial}{\partial \theta} [\sin^2 \theta (u - \bar{u})(v - \bar{v})]^*_m \right\} \\
+ \frac{2}{a} \text{Re}\left\{ v_m \frac{\partial}{\partial \theta} [(v - \bar{v})^2]^*_m \right\} \\
+ \frac{\beta}{a} \cot \theta \text{Re}\left\{ u_m [(u - \bar{u})^2 - (u - \bar{w})^2]^*_m \right\} \right\} dV, \\
\]

\[ L_2(m) = \int_V \left\{ 2 \text{Re}\left\{ u_m \frac{\partial}{\partial z} [p(u - \bar{u})(w - \bar{w})]^*_m \right\} \\
+ v_m \frac{\partial}{\partial z} [p(v - \bar{v})(w - \bar{w})]^*_m \right\} dV, \\
\]

\[ dV = \sin \theta d\theta d\phi dz. \]

Here \( L_1 \) and \( L_2 \) are basically the same quantities given in Saltzman (1970).

Scaling arguments based on the continuity equation suggest that zonal velocity for the planetary waves should be larger than meridional velocity, since meridional length scales are usually shorter than zonal length scales at latitudes not too near the poles. In the calculations the planetary wave meridional velocity amplitude was typically less than or about one-third that of the zonal velocity, being less for \( m = 1 \) but somewhat greater for \( m = 3 \). This was the basis for neglecting the meridional velocity terms in (5.1).

REFERENCES


