An Updated Theory for the Quasi-Biennial Cycle of the Tropical Stratosphere[1]

JAMES R. HOLTON

Dept. of Atmospheric Sciences, University of Washington, Seattle

AND RICHARD S. LINDZEN

Dept. of the Geophysical Sciences, The University of Chicago

(Manuscript received 10 January 1972, in revised form 10 April 1972)

ABSTRACT

The authors' earlier theory of the quasi-biennial oscillation is reevaluated in the light of more recent observational analyses and theoretical developments. The main contention of the earlier theory, that the stratospheric oscillation is driven by the vertically propagating, short-period (5–15 days) waves excited in the upper troposphere, is now well established. However, the assumption that the waves interact with the mean flow through critical level absorption is now seen to be largely, if not entirely, incorrect. Instead it appears that the short-period waves are strongly attenuated primarily by infrared cooling. As they are attenuated, their mean momentum flux is deposited in the mean flow. It is shown that this mechanism leads to an oscillation in the mean flow as readily as the earlier critical-level mechanism, and with fewer assumptions.

1. Introduction

In an earlier theory of the quasi-biennial oscillation (Lindzen and Holton, 1968) we showed how upward-propagating, internal, equatorial, planetary-scale waves could drive the quasi-biennial oscillation in zonal wind observed in the tropical stratosphere. It was assumed that the vertical flux of momentum due to the waves was deposited in the mean flow at critical levels (levels where the waves' Doppler-shifted frequency went to zero) where the waves were absorbed by the mean flow. The theory required a spectrum of waves whose speeds ranged from the maximum easterly to the maximum westerly velocities found in the quasi-biennial oscillation of the mean flow. It is the purpose of this paper to reexamine this theory in the light of more recent observational and theoretical developments.

2. Observations

It is now quite certain that the quasi-biennial oscillation is indeed forced by waves propagating vertically from below. Wallace and Kousky (1968a) have shown that the divergence of the Reynolds' stress due to the waves agrees closely with the acceleration of the mean flow. The precise nature of the waves, themselves, has not been definitely established. However, the evidence to date (Kousky and Wallace, 1971; Maruyama, 1969) suggests that the waves consist primarily in a single westerly Kelvin wave during the descending westerly phase of the quasi-biennial oscillation and in a single easterly mixed gravity-Rossby wave during the descending easterly phase. Moreover, the Kelvin wave, at least, appears to have a phase speed of about 30 m sec^{-1} which is in excess of the maximum westerly flow in the quasi-biennial oscillation. Thus, at least below 30 km, it seems that the Kelvin wave does not encounter a critical level. The above facts, in general, suggest very strongly that critical-level absorption does not dominate the interaction of the short-period waves with the mean flow.

3. Theoretical developments

In numerical studies Holton (1970) and Lindzen (1970) showed that equatorial planetary-scale waves did indeed undergo critical-level absorption. However, in Lindzen (1970) and more explicitly in a later analytical study of equatorial waves (Lindzen, 1971), it was shown that even when critical levels are present, equatorial waves are primarily absorbed by existing dissipation—mostly due to infrared cooling, i.e., dissipation increases as the Doppler-shifted frequency becomes smaller so that the waves are damped out below the critical level. This absorption leads quite obviously to a divergence of the total momentum flux [which as shown by Jones (1967) and Bretherton (1969) consists in both the usual Reynolds stress $\rho_0 uw$ and an additional term due to Coriolis torques] and a consequent acceleration of the mean flow. It was suggested by Lindzen (1971) that this was probably the most important source of mean flow-wave interaction in the quasi-

[1] Contribution No. 253, Department of Atmospheric Sciences, University of Washington.

biennial cycle. However, no explicit description of how such an interaction could lead to an oscillation in the mean flow was presented. In subsequent numerical studies we have found that regardless of the precise distribution of the dissipation, and the precise nature of the waves and their phase speeds (as long as there are both easterly and westerly waves each with a sufficiently large phase speed), there is, indeed, no difficulty in generating a long-period oscillation in the mean flow. The remainder of this paper will describe one such study involving conditions which seem to us most likely on the basis of present knowledge.

4. Prototype calculation

The obviously correct approach to studying the effect of wave absorption on the mean flow is to integrate the full appropriate equations of motion (conservation of momentum, mass and energy, plus the gas law) for the long-period oscillation, with the short-period wave absorption appearing in forcing terms. This was the approach described in Section 6 of Lindzen and Holton (1968), and although entirely feasible, it is computationally lengthy. More significantly, it is an approach which yields more complexity and detail than we are really interested in at this stage. Fortunately, we found, in the earlier study, that the basic mechanism of the wave-mean flow interaction is clearly and adequately revealed in one-dimensional prototype calculations wherein the local zonal acceleration of the mean flow is balanced by the vertical convergence of the momentum flux due to the short-period waves. Such prototype models can also include vertical diffusion of momentum as well as other sources of forcing. It is such a model that we will examine in this note. Our equation is

$$\frac{\partial \bar{u}}{\partial t}=-\frac{1}{\rho_0}\frac{\partial}{\partial z}\bar{F}_{MW}+K\frac{\partial^2 \bar{u}}{\partial z^2}+G, \qquad (1)$$

where:

- $\bar{u}$ mean zonal flow (positive to the east)
- $\rho_0(z)$ mean density
- $\bar{F}_{MW}$ average vertical flux of mean zonal momentum due to synoptic-scale equatorial waves
- K an eddy diffusion coefficient
- G semiannual forcing
- t time
- z altitude

The presence of G simulates the observed presence of a semiannual oscillation in the zonal wind above 28 km in the tropics (Reed, 1966). More explicitly, we take

$$\left.\begin{aligned} G&=0, \quad \text{for } z\leqslant 28\text{ km}\\ G&=\omega_{sa}u_{sa}, \quad \text{for } z>28\text{ km}\\ u_{sa}&=2(z-28\text{ km})\text{ m sec}^{-1}\text{ km}^{-1}\sin(\omega_{sa}t)\\ \omega_{sa}&=(2\pi/180)\text{ day}^{-1}\approx 4\times10^{-7}\text{ sec}^{-1}\end{aligned}\right\}. \qquad (2)$$

In the course of these simplified calculations some diffusion was found necessary. However, our choice of

$$K=3\times10^3\text{ cm}^2\text{ sec}^{-1} \qquad (3)$$

for an eddy coefficient is small compared to common estimates, and is probably not an unduly unrealistic representation of effects we have omitted. We shall return to this more specifically in our discussion of results. The specification of $\bar{F}_{MW}$ is, of course, the crucial element in the construction of the prototype model. We shall here consider the mean of F_{MW} at each level. As shown in Lindzen (1971), away from regions of excitation, $\bar{F}_{MW}$ will be independent of z except at critical levels as long as there are no damping mechanisms operating. However, there is important dissipation—primarily in the form of infrared cooling. This, in turn, leads to the decay of wave amplitudes and hence $\bar{F}_{MW}$ with increasing height. The precise rate of decay depends on the damping rate, the wave's type, the wave's phase speed and frequency, and the mean flow. The detailed analysis of this behavior is given in Lindzen (1971).

In general, we write

$$\bar{F}_{MW}=\sum_i A_i \exp\left\{-2\int_{17\text{ km}}^{z} g_i(z)dz\right\}, \qquad (4)$$

where 17 km is taken to be the height at which the waves originate. The index i refers to the individual waves. For this study we assume the presence of only two waves, an easterly mixed gravity-Rossby wave and a westerly Kelvin wave; these appear to be dominant in current observational analyses. Moreover, there appears to be no way that the inclusion of other modes would significantly alter our results. Lindzen (1971) derived expressions for g_i when there are both Rayleigh friction and Newtonian cooling. Re-deriving these expressions when only Newtonian cooling is present yields the following approximate formulae for g_i:

$i=0$, Kelvin wave

$$g_0\approx\frac{1}{2}\frac{\sqrt{gS}}{(c-\bar{u})^2}\frac{\alpha}{k} \qquad (5)$$

$i=1$, mixed gravity-Rossby wave

$$g_1\approx\frac{1}{2}\frac{\sqrt{gS}}{(c-\bar{u})^3}\frac{\beta}{k^2}\frac{\alpha}{k}\left[1-\frac{k^2}{\beta}(\bar{u}-c)\right] \qquad (6)$$

where:

$$S=\frac{1}{T_0}\left(\frac{dT_0}{dz}+\frac{g}{C_p}\right)$$

- g acceleration
- T_0 mean temperature
- C wave's phase speed

β $2\Omega/a$
Ω earth's rotation rate
a earth's radius
k zonal wavenumber of wave
α rate coefficient for dissipation

We have not explicitly dealt with critical levels, and Eqs. (5) and (6) indeed assume $|\alpha| \ll |k(\bar{u}-c)|$; however, (5) and (6) formally lead to infinite absorption at critical levels as long as α is finite, and hence, qualitatively simulate a major feature of critical levels. More important, they imply continuous absorption in the absence of critical levels.

For both the Kelvin and the mixed wave we have chosen

$$S = 4.75 \times 10^{-5}\ \text{m}^{-1} \qquad (7)$$

corresponding to an isothermal atmosphere with a scale height

$$H = 6\ \text{km}. \qquad (8)$$

We assume that α is due to infrared cooling to space. We have adopted the distribution

$$\left.\begin{aligned} \alpha &= \frac{1}{21}\text{day}^{-1} + \left(\frac{1}{7}-\frac{1}{21}\right)\text{day}^{-1}\frac{(z-17\ \text{km})}{13\ \text{km}}, \\ &\qquad\qquad \text{for } 17\ \text{km} \leqslant z \leqslant 30\ \text{km} \\ \alpha &= \frac{1}{7}\text{day}^{-1}, \quad \text{for} \quad z \geqslant 30\ \text{km} \end{aligned}\right\}. \qquad (9)$$

The increase of α between 20 and 30 km will prove of considerable importance in the present prototype calculation. Our choice of α at 30 km is larger than is usually used, though the overall variation is within the range of possibilities described by Dickinson (1968). The scale of variation for α is sufficiently large compared with $(1/2\pi)$(vertical wavelength) to permit the use of formulae derived on the basis of constant α.

For the Kelvin wave we have chosen

$$\left.\begin{aligned} c &= 30\ \text{m sec}^{-1} \\ k &= (2\pi/40{,}000)\ \text{km}^{-1} \end{aligned}\right\}, \qquad (10)$$

values taken from Wallace and Kousky (1968b).

For the mixed wave we have chosen

$$\left.\begin{aligned} c &= -30\ \text{m sec}^{-1} \\ k &= (2\pi/10{,}000)\ \text{km}^{-1} \end{aligned}\right\}. \qquad (11)$$

The wavenumber corresponds to that found by Maruyama (1967). The phase speed is a bit greater than that reported by Maruyama, his value being -23 m sec^{-1}. We chose the larger value primarily to avoid a critical level in these calculations. Our point is simply to see whether an oscillation can be set up without critical level absorption. However, it ought to be added that -30 m sec^{-1} is well within the range of observational uncertainty.

For the A_i's in Eq. (4) we have chosen

$$\left.\begin{aligned} A_0 &= 4\times 10^{-3}\ \text{m}^2\ \text{sec}^{-2} \times \rho_0(17\ \text{km}) \\ A_1 &= -4\times 10^{-3}\ \text{m}^2\ \text{sec}^{-2} \times \rho_0(17\ \text{km}) \end{aligned}\right\}, \qquad (12)$$

again compatible with present observational analyses.

For our boundary conditions we have taken

$$\left.\begin{aligned} \bar{u} &= 0 \quad \text{at} \quad z = 17\ \text{km} \\ \bar{u} &= u_{SD} \quad \text{at} \quad z = 35\ \text{km} \end{aligned}\right\}. \qquad (13)$$

The above completely specifies our prototype model. Eq. (1) was integrated using a numerical scheme due to Matsuno (1966). Our resolution was

$$\Delta z = 250\ \text{m}$$
$$\Delta t = 24\ \text{hr}.$$

The integration was carried on for 2160 time steps (i.e., $\sim$6 years).

5. Results

The results of the above calculation are shown in Fig. 1, where a time-height cross section of the zonal wind is presented. A semiannual oscillation is clearly present above 30 km (as we might expect from the forcing). However, below 30 km the flow is clearly dominated by a longer period oscillation, which for our choice of the A_i's happens to have an average period of $\sim$26.5 months. As with the observed oscillation, phase propagates downward at the rate of about 1 km per month. Indeed, the only significant difference between Fig. 1 and the observed cycle is that easterly shears in Fig. 1 are too sharp. However, as shown in Lindzen and Holton (1968) nonlinear advective terms (which are not included in the prototype calculations) tend to produce westerly accelerations at the equator, which might, in a more realistic calculation, eliminate this discrepancy.

6. Discussion of results

In this section we will describe qualitatively the mechanisms involved in producing the results displayed in Fig. 1. First we shall show how the damping of waves by Newtonian cooling can produce a downward moving shear zone. Essential to the mechanism are the following features:

(i) From Eq. (12) we see that the wave whose phase speed is easterly relative to the mean flow carries easterly momentum upward while the westerly wave carries westerly momentum upward.

(ii) Due to the presence of Newtonian cooling the waves are attenuated as they propagate upward, their momentum flux being deposited in the mean flow.

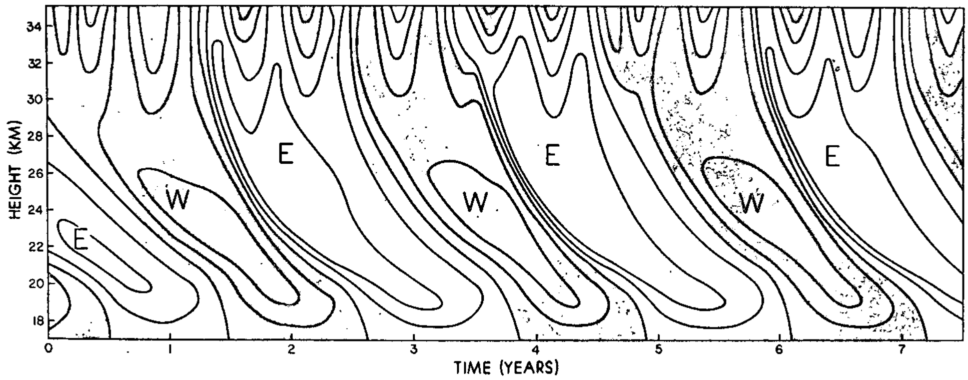

FIG. 1. Time-height cross section of mean zonal wind. Contours have been drawn at 10 m sec^{-1} intervals. Regions of westerly flow have been shaded.

(iii) From Eqs. (5) and (6) we see that the rate of attenuation for the waves increases as α increases and as $|\bar{u}-c|$ decreases.

(iv) The dissipation rate coefficient α increases with altitude through the region 20–30 km.

How the above features can cause a shear zone to descend is readily seen in the following simple example. Consider an atmosphere at rest with no semiannual forcing. Now, at some time t_0, begin generating a westerly wave at 17 km. The wave will propagate upward with relatively gradual attenuation until it gets near 30 km where because of increasing α the attenuation will become rapid, leading to a pronounced westerly acceleration of the mean flow. The westerly flow will cause a diminution of $|\bar{u}-c|$ which in turn will lead to increased absorption of the wave momentum flux at a lower level. The wave flux absorbed at the lower levels will no longer be available to accelerate the upper levels, and the acceleration of the upper levels will decrease. However, there will be increased acceleration at lower levels leading to decreasing $|\bar{u}-c|$ at these levels further lowering the level at which most of the momentum flux is absorbed. In this manner, a westerly shear zone will descend to the lower boundary where eventually a balance will be established between the convergence of $\bar{F}_{MW}$ and downward diffusion of momentum. Now, when we have both an easterly and a westerly wave propagating upward, the above mechanism leads to an oscillation in the mean wind. To see this, imagine a westerly shear zone approaching the lower boundary. Westerly waves will essentially be prevented from reaching the upper levels. However, easterly waves will propagate upward producing an easterly shear zone which will follow the westerly shear zone. The presence of a semiannual oscillation is not essential to this mechanism, but it does help matters by periodically reducing $|\bar{u}-c|$ and in doing so it synchronizes the longer period oscillation in a manner already described in Lindzen and Holton (1968). In attempting to visualize the above mechanism, the reader may encounter difficulty at the lower boundary. Instead of obtaining a long-period oscillation, we might instead obtain, for example, an extremely thin westerly jet just above 17 km bounded by a regime of easterly flow above the jet. The presence of sufficient eddy viscosity prevents this by effectively diffusing the jet and permitting the easterly regime to propagate to the bottom while simultaneously allowing the westerly waves to again propagate upward. Eddy diffusion (plus the lower boundary condition, $\bar{u}=0$ at $z=17$ km) in the present model replaces the switching mechanism used in Lindzen and Holton (1968). While neither mechanism is likely to be realistic, both serve to parameterize the very plausible assumption that when a shear zone descends to the level of excitation, it will cease to serve any longer as a barrier to the upward propagation of waves. Parenthetically, we might add, that in the absence of eddy diffusion the above described mechanism would produce unstable shear zones which in turn would spawn the eddies we had initially omitted. Since the presence of eddy diffusion in our model proved important only near the lower boundary, it might indeed prove a more plausible description of reality than we presently expect.

7. Conclusions

We have shown that the primary hypothesis of our earlier model of the quasi-biennial oscillation, namely that the long-period oscillation is mechanically forced by shorter period equatorial waves propagating upward from the troposphere, is physically consistent *independent* of particular mechanistic assumptions concerning the details of the interaction of these waves with the mean flow. More specifically, we have replaced our earlier critical-level absorption mechanism with a new (and observationally more realistic) mechanism wherein the damping of the short-period waves by infrared cooling produces the convergence of wave momentum flux which in turn accelerates the mean flow.[2] In reality, the presence of some critical-level absorption (which from current observations seems likely for the easterly waves) would not significantly affect our present model since it would merely enhance the effects of diminishing $|\bar{u}-c|$ which are already incorporated in the model. More important, the present interaction mechanism permits us to eliminate the more restrictive assumptions in our earlier model; namely:

(i) We no longer need a full spectrum (in phase speed) of short-period equatorial waves.

(ii) The mesospheric semiannual oscillation, while important, is no longer absolutely essential to the overall theory.

While further observational and theoretical findings may be expected to modify our present picture of the quasi-biennial oscillation, we feel that our broad conceptual understanding of its origins is now reasonably certain.

Acknowledgments. The support of the Atmospheric Sciences Section, National Science Foundation, under Grants NSF GA 25904 (R. S. Lindzen) and NSF GA 23488 (J. R. Holton) is gratefully acknowledged. R. S. Lindzen also wishes to thank the Alfred P. Sloan Foundation whose funds made possible a visit to the University of Washington.

[2] In calculations not presented here, we found that the inclusion of Rayleigh friction with a coefficient equal to our Newtonian cooling rate coefficient modified our results only slightly—leading primarily to a somewhat shorter quasi-biennial cycle (26 months instead of 26.5 months). We also investigated what would happen if we were to use height-independent coefficients for Rayleigh friction and Newtonian cooling. A "quasi-biennial" oscillation resulted, but its height distribution did not simulate the observed distribution particularly well.

REFERENCES

Bretherton, E. P., 1969: Momentum transport by gravity waves. *Quart. J. Roy. Meteor. Soc.*, **95**, 213–243.

Dickinson, R. E., 1968: On the excitation and propagation of zonal winds in an atmosphere with Newtonian cooling. *J. Atmos. Sci.*, **25**, 269–279.

Holton, J. R., 1970: The influence of mean wind shear on the propagation of Kelvin waves. *Tellus*, **22**, 186–193.

Jones, W. L., 1967: Propagation of internal gravity waves in fluids with shear flow and rotation. *J. Fluid Mech.*, **20**, 439–448.

Kousky and Wallace, 1971: On the interaction between Kelvin waves and the mean zonal flow. *J. Atmos. Sci.*, **28**, 162–169.

Lindzen, R. S., 1970: Internal equatorial planetary-scale waves in shear flow. *J. Atmos. Sci.*, **27**, 394–407.

——, 1971: Equatorial planetary waves in shear, I. *J. Atmos. Sci.*, **28**, 609–622.

——, and J. R. Holton, 1968: A theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 1095–1107.

Maruyama, T., 1967: Large-scale disturbances in the equatorial lower stratosphere. *J. Meteor. Soc. Japan*, **45**, 391–408.

——, 1969: Long term behavior of Kelvin waves and mixed Rossby-gravity waves. *J. Meteor. Soc. Japan*, 44, 25–43.

Matsuno, T., 1966: Numerical integration of the primitive equations by a simulated backward differencing method. *J. Meteor. Soc. Japan*, **44**, 76–84.

Reed, R. J., 1966: Zonal wind behavior in the equatorial stratosphere and lower mesosphere. *J. Geophys. Res.*, **71**, 4223–4233.

Wallace, J. M., and V. E. Kousky, 1968a: On the relation between Kelvin waves and the quasi-biennial oscillation. *J. Meteor. Soc. Japan*, **47**, 496–502.

——, and——, 1968b: Observational evidence of Kelvin waves in the tropical stratosphere. *J. Atmos. Sci.*, **25**, 900–907.