# A Theory of the Quasi-Biennial Oscillation[1]

RICHARD S. LINDZEN

*Dept. of the Geophysical Sciences, University of Chicago*

AND JAMES R. HOLTON

*Dept. of Atmospheric Sciences, University of Washington, Seattle*

(Manuscript received 12 April 1968, in revised form 29 July 1968)

ABSTRACT

A theory is presented which indicates that the quasi-biennial oscillation of the zonal wind in the tropical stratosphere is a result of the interaction of long-period, vertically propagating gravity waves with the zonal wind. We discuss the theoretical basis and observational evidence for the existence of long-period gravity waves near the equator, and the mechanism of their interaction with the zonal wind, and present some simple numerical results which show how the absorption of the momentum of these waves by the mean flow leads to a downward propagating zonal wind profile. It is shown that the interaction of these gravity waves with the observed semiannual zonal wind oscillation above 40 km will produce a downward propagating quasi-biennial oscillation. We present the results of several numerical experiments with a model of the tropical stratosphere which includes the gravity wave interaction mechanism. The quasi-biennial oscillation is simulated quite successfully. Finally, we discuss possible observational checks for our model, and some of its implications for tropical dynamics.

## 1. Introduction

The observed properties of the quasi-biennial oscillation in the zonal wind field of the tropical stratosphere have been analyzed in detail by Reed (1964) and Wallace (1967a). The observations indicate a number of peculiar features which a satisfactory theory of the oscillation must be able to explain. In particular, zonally symmetric easterly and westerly wind regimes alternate regularly with a period varying from about 24–30 months. Successive regimes first appear above 30 km, but propagate downward at an average rate of 1 km month$^{-1}$. Rocket data give evidence of the existence of the oscillation above 30 km (Reed, 1965). Its phase speed, in this region, appears to be much greater than 1 km month$^{-1}$. The downward propagation occurs without loss of amplitude between 30 and 23 km, but there is rapid attenuation below 23 km. These fluctuations are symmetric about the equator with a half-width of about 12° latitude and an amplitude of about 20 m sec$^{-1}$ at the equator.

Previous theoretical models of the quasi-biennial oscillation have invoked a variety of mechanisms in attempting to account for the origin, dynamics and structure of this phenomenon, but none have successfully explained the three most striking features of the oscillation: 1) the approximately biennial periodicity, 2) the downward propagation without loss of amplitude, and 3) the occurrence of zonally symmetric westerlies at the equator.

[1] Contribution No. 175, Dept. of Atmospheric Sciences, University of Washington, Seattle.

Most attempts to explain the periodicity have been based on an assumed biennial oscillation in either the diabatic heating rate or the horizontal eddy momentum flux divergence. The theories of Staley (1963) and Lindzen (1965, 1966) are in the former class, while the numerical model of Wallace and Holton (1968) is of the latter class. The impossibility of a radiative forcing was demonstrated by Wallace and Holton (1968), who showed that the observed amplitude of the oscillation can be simulated through radiative forcing only if the oscillation in the diabatic heating rate has a magnitude of nearly 1.5°C day$^{-1}$ at the equator. This is more than an order of magnitude greater than the response calculated by Lindzen (1965) for a 10% change in the solar ultraviolet insolation. A physical explanation for the extremely large diabatic heating rate required in a radiatively driven model has been provided by Wallace (1967b). He pointed out that the ratio of kinetic energy to available potential energy is very large in the tropical stratosphere. Thus, the conversion of available potential energy to kinetic energy in a radiatively driven model would quickly exhaust the supply of available potential energy unless strong diabatic heating was continually replenishing the supply of available potential energy. Therefore, the forcing is apparently of a dynamical rather than thermal nature. However, although some observations (Wallace and Newell, 1966) indicate a biennial periodicity in the horizontal eddy momentum fluxes in middle latitudes above 30 mb the evidence for these fluxes in tropical latitudes is not very convincing, and the amplitude of such variations is probably too small to account for the quasi-biennial oscillation.

Furthermore, models based on the above mechanisms all share a crucial defect—the inability to propagate the oscillation away from the region of forcing without substantial loss of amplitude. For example, the models of Staley (1963) and Dickinson (1968) both postulate a diffusion wave type of propagation away from a forcing region above 30 km, and in both these models the oscillation damps rapidly below 30 km. Another possible mechanism for vertical propagation which has been discussed by Tucker (1964) is vertical advection of the zonal wind by a mean subsidence throughout the tropical stratosphere. Wallace and Holton (1968) demonstrated with their diagnostic numerical model that a wide region of subsidence, as proposed by Tucker, would by continuity require a strong equatorward meridional flow, and consequent unrealistically large mean easterlies a few degrees from the equator. Wallace and Holton were able to obtain somewhat more realistic results by confining the mean subsidence to within a few degrees of the equator and allowing horizontal diffusion to spread the downward propagating oscillation away from the equator; this mechanism, however, caused rather rapid attenuation of the oscillation.

We conclude that the above models are all incorrect in their basic assumption that the quasi-biennial oscillation is the result of propagation of a disturbance away from a forcing region above 30 km. The extensive numerical experiments of Wallace and Holton have indicated beyond reasonable doubt that only a forcing mechanism which itself propagates downward can account for the observed constant amplitude propagation between 30 and 23 km. The purpose of the present paper is to demonstrate that long-period vertically propagating gravity waves are able to provide this sort of forcing, and can also explain the quasi-biennial periodicity as well as providing a source for westerly momentum generation at the equator.

In Section 2 we briefly present our theoretical reasons for expecting long-period gravity waves in the neighborhood of the equator. We also describe the present status of observational evidence on this matter. In Section 3 we outline the results of Booker and Bretherton (1967) on the interaction of a vertically propagating gravity wave and the mean flow at a critical level where the horizontal phase speed of the waves relative to the mean flow is zero. These results are extended to the problem of the interaction of waves with a continuous distribution of phase speeds with the mean flow. A simple formula for the change in the mean flow is obtained under the assumption that the mean flow changes slowly compared to the wave periods. In Section 4 some examples are given of how the interaction of gravity waves with a mean flow can cause a given velocity profile in the mean flow to propagate downward. In Section 5 we give a qualitative description of how a spectrum of gravity waves interacting with the observed semiannual oscillation in the equatorial zonal wind above 40 km can produce a quasi-biennial oscillation below 40 km. The physical meaning of the longer period is explained. In Section 6 a description is given of a simplified version of the numerical model of the equatorial stratosphere developed by Wallace and Holton (1968) which has been modified to include the gravity wave-mean flow interaction. The results of numerical simulations of the quasi-biennial oscillation with this model are given in Section 7. Finally, in Section 8, we describe some observationally verifiable predictions resulting from our model.

## 2. Equatorial gravity waves

A theory of the quasi-biennial oscillation based on the interaction of gravity waves with the mean flow, while interesting, would be academic unless we had some reason for expecting strong gravity waves to exist preferentially at the equator. The theoretical basis for this expectation is described in Lindzen (1967). Briefly, we expect that in the troposphere there will be a continuous excitation of planetary-scale disturbances by the effects of orography and the mutual adjustments of the wind and pressure fields. The question of whether such disturbances can propagate vertically was treated by Charney and Drazin (1961) who, adopting a quasi-geostrophic approximation suitable for mid-latitudes, found that disturbances could propagate only when their phase speed was westward relative to the mean zonal flow, but less than some maximum easterly speed whose value rapidly diminished with increasing horizontal wavenumber. Thus, the bulk of large-scale disturbances would appear to be vertically trapped. The quasi-geostrophic approximation assumes that the period of the waves is much longer than the pendulum day. The trapping found by Charney and Drazin is, in fact, due to the same mechanism whereby waves in a stratified, plane, rotating fluid are trapped when their frequency is less than twice the rotation rate (Eckart, 1960). Their ability to propagate under certain circumstances is due to the variation of the vertical component of rotation with latitude, i.e., the beta effect. As one approaches the equator the length of the pendulum day approaches infinity. Thus, the basis for the quasi-geostrophic approximation (as applied to oscillatory motions) breaks down, and one might expect that for a disturbance of any period and zonal wavenumber there will exist some region about the equator where the disturbance can propagate vertically. That this is so, is shown in Lindzen (1967). These disturbances propagate primarily as internal gravity waves. However, the gravest eastward propagating mode, symmetric about the equator, is a Kelvin wave (Matsuno, 1966; Holton and Lindzen, 1968), which for our purposes can be considered a particular kind of internal gravity wave. The gravest westward propagating mode, antisymmetric about the equator, propagates as a mixed internal gravity-Rossby wave (Matsuno, 1966; Maruyama, 1967; Lindzen and Matsuno, 1968) whose properties are distinctly different from those of internal gravity waves. The degree to which vertical propagation is confined to the equator is determined by the horizontal wavenumbers and the Doppler shifted frequency of a given disturbance; for meteorologically relevant parameters (period $\sim$5 days, zonal wavenumber $\sim$5) propagation is confined to within about 10°–15° of the equator. Most modes propagate vertically with very short vertical wavelengths, typically 2 km or less. Such modes could not be resolved with presently available data. Only the Kelvin mode and the gravest antisymmetric mode are expected to have significantly longer vertical wavelengths ($\sim$4–20 km), and examples of both these modes have been isolated in the data (Maruyama, 1967; Wallace and Kousky, 1968). The data may be interpreted as showing that there is nothing intrinsically unlikely about the assumed generation and propagation of the above described waves. However, the actual existence of a wide spectrum of such waves must, for the moment, be taken as conjecture.

## 3. Absorption of gravity waves at critical levels

The vertical propagation of gravity waves in a medium where there is a mean wind with shear has been studied by a number of people. The important results for our purposes are found in papers by Eliassen and Palm (1960) and Booker and Bretherton (1967).

Let the waves have a wavenumber $k$ and a real phase speed $c$ in some horizontal direction $x$. Let the basic flow $U$ be in the $x$ direction, and let $U$ be a function only of altitude $z$. Primed quantities will refer to wave fields, and overbars will refer to averages with respect to $x$ and $t$. It was shown in Eliassen and Palm that $\overline{p'w'}$ (where $p'$ and $w'$ are the pressure and vertical velocity fields associated with the wave) is the upward energy flux due to the wave, and that $\rho_0(z)\overline{u'w'}$ (where $\rho_0$ is the basic density distribution, and $u'$ is the velocity field in the $x$ direction associated with the wave) is the upward flux of momentum in the $x$ direction due to the wave. The following important relations were derived by Eliassen and Palm:

$$\overline{p'w'}=-\rho_0(U-c)\overline{u'w'}. \tag{1}$$

If $U\neq c$, then

$$\rho_0\overline{u'w'} \text{ is independent of } z. \tag{2}$$

Let $x$ represent the west-to-east direction, and let our wave source be below our region of interest. Then $\overline{p'w'}$ will be positive and (1) states that waves whose phase speeds are westerly relative to $U$ will carry westerly momentum upward, while waves whose phase speeds are easterly relative to $U$ will carry easterly momentum upward. Relation (2) states that in the absence of a critical level where $U=c$, none of this momentum will be deposited in the mean flow. Eliassen and Palm do not deal with the problem of what occurs at a critical level. However, from (1) and (2) it is clear that something striking must happen. For example, assuming $\overline{p'w'}$ remains positive as a wave goes through a critical level, then (2) certainly cannot remain true. Eq. (1) suggests that $\rho_0\overline{u'w'}$ must, in fact, change sign, implying a large exchange of momentum between the wave and the mean flow. Recently, Booker and Bretherton (1967) have solved the problem of the behavior of internal gravity waves at critical levels. Their results confirm the above picture. Let the wave source be below the critical level and let

$$\rho_0\overline{u'w'}=A \tag{3}$$

below the critical level. Booker and Bretherton show that

$$\rho_0\overline{u'w'}=-A\exp(-2\pi\sqrt{\mathrm{Ri}-\tfrac{1}{4}}) \tag{4}$$

above the critical level, where

$$\mathrm{Ri}=\frac{\dfrac{g}{T_0}\left(\dfrac{\partial T_0}{\partial z}+\dfrac{g}{c_p}\right)}{\left(\dfrac{\partial U}{\partial z}\right)^2}, \tag{5}$$

the Richardson number, is generally greater than 1; in such cases we see from (4) and (2) that the wave is almost completely absorbed by the mean field at the critical level. From (3) and (4) we see that an amount of momentum given by

$$A[1+\exp(-2\pi\sqrt{\mathrm{Ri}-\tfrac{1}{4}})]$$

is absorbed at the critical level, and as Booker and Bretherton note, the absorption of a finite momentum by a single level must lead to difficulty. The difficulty arises, as also shown by Booker and Bretherton, from the assumption that a wave with finite energy and a single phase speed $c$ will propagate for an infinite length of time through a medium with an unchanging basic velocity distribution having a critical level. There are several ways of circumventing this difficulty. The way which is relevant to this paper is to assume that the disturbances consist of a spectrum of waves with a continuous distribution of phase speeds, i.e.,

$$\rho_0\overline{u'w'}=\int_{-\infty}^{\infty}\hat{f}(c)dc, \tag{6}$$

where $\hat{f}(c)dc$ is the momentum flux due to waves of all wavenumbers $k$ having phase speeds between $c$ and $c+dc$. In writing (6) it is assumed that if $u'$ is associated with a wave of one value of $c$, and $w'$ with a wave of another value, then the average $\overline{u'w'}$ will be zero. It follows from integrating (2) with respect to $k$ that $\hat{f}(c)$

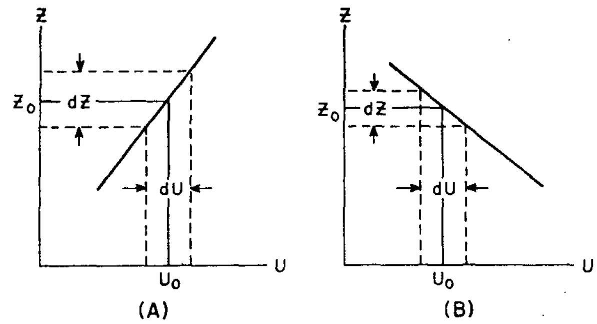


FIG. 1. Geometry of the critical layer absorption process for westerly shear (A), and easterly shear (B).

will be independent of $z$ in any region where $U\neq c$. If $\hat{f}(c)=\hat{A}$ below a critical level, it will go to $-\hat{A}\exp[-2\pi\sqrt{\mathrm{Ri}-\frac{1}{4}}]$ above, as shown by (3) and (4). However, as long as $\hat{f}(c)$ is finite, an infinitesimal layer will never absorb more than an infinitesimal amount of momentum. Unless otherwise indicated, $\hat{f}(c)$ will refer to the distribution $\rho_0\overline{u'w'}$ (i.e., Reynold's stress) at the base of the region of interest.

We now proceed to the derivation of a simple formula for the deposition of momentum in the mean flow by gravity waves. Consider first the case shown in Fig. 1a where $dU/dz$ is positive. At a height $z_0$, $U=U_0$. The amount of momentum absorbed (per unit time) in the altitude range $dz$ will be

$$\hat{f}(U_0)[1+\exp(-2\pi\sqrt{\mathrm{Ri}(z_0)-\tfrac{1}{4}})]|dU|, \tag{7}$$

and the momentum flux divergence at $z_0$ is given by

$$\frac{dF_{MW}}{dz}=\hat{f}(U_0)[1+\exp(-2\pi\sqrt{\mathrm{Ri}(z_0)-\tfrac{1}{4}})]\left|\frac{dU}{dz}\right|, \tag{8}$$

where $F_{MW}$ is the momentum flux due to waves, the result being independent of the sign of $dU/dz$. Moreover, waves whose critical level is at $z_0$ will be westerly relative to $U$ below $z_0$; hence, $\hat{f}(U_0)$ is positive. Since $dU/dz$ is also positive, (8) may be rewritten as

$$\frac{dF_{MW}}{dz}=|\hat{f}(U_0)|[1+\exp(-2\pi\sqrt{\mathrm{Ri}(z_0)-\tfrac{1}{4}})]\frac{dU}{dz}. \tag{9}$$

For the case shown in Fig. 1b, $dU/dz$ is negative. However, waves whose critical level is at $z_0$ will now be easterly relative to $U$ below $z_0$; hence, $\hat{f}(U_0)$ is also negative, and, (9) is therefore correct for this case as well. For conciseness we may rewrite (9) as

$$\frac{dF_{MW}}{dz}=f(U)\frac{dU}{dz}, \tag{10}$$

where $f(U)=|\hat{f}(U)|[1+\exp(-2\pi\sqrt{\mathrm{Ri}(z_0)-\tfrac{1}{4}})]$ and $f(U)$ is always positive. Eq. (10) is appropriate regardless of the sign of the shear, and is based on the treatment by Booker and Bretherton (1967) wherein it was assumed that $U(z)$ is independent of time. Although there are no such flows in nature (the existence of momentum flux divergence also implying that $U$ will change in time), we shall assume that the changes in $U$ will occur sufficiently slowly for (10) to remain valid at any given moment. In the present context this amounts to assuming that 26 months is sufficiently longer than a few days to make $U$ appear steady to waves of the latter period. If this is the case, the equation for mean momentum in the $x$ direction becomes

$$\rho_0\left(\frac{\partial U}{\partial t}+W\frac{\partial U}{\partial z}\right)=-f(U)\frac{\partial U}{\partial z}+\text{other terms}. \tag{11}$$

When the nonlinear term on the left-hand side of (11) and the "other terms" are neglected, (11) describes the downward propagation of a distortion in $U$. The speed of propagation is given by $-f(U)/\rho_0$. It should be noted that if $U$ is such that $U(z_1)=U(z_0)$, $z_1>z_0$, and $U(z)\neq U(z_0)$, $z_1>z>z_0$, then

$$f(U,z_1)=f(U,z_0)\exp\{-2\pi\sqrt{\mathrm{Ri}(z_0)-\tfrac{1}{4}}\}. \tag{12}$$

The above discussion is based entirely on gravity waves in a nonrotating fluid. The reader should keep in mind that in applying the above considerations to a theory of the quasi-biennial cycle we are assuming that a similar critical-level absorption mechanism exists for the rotationally influenced equatorial internal gravity waves described in Section 2. It is shown in Lindzen and Matsuno (1968) that this is, in fact, unlikely for the mixed gravity-Rossby mode. However, all the remaining modes are so similar to conventional gravity waves as to suggest that the above theory is at least approximately applicable. Some additional support for this view is provided by the work of Jones (1967) who studied critical-level absorption for gravity waves in a plane rotating fluid. His results were almost identical to those of Booker and Bretherton (1967). However, he found that Eq. (2) of the present paper was slightly modified (angular momentum replacing linear momentum) and critical-level absorption took place where $U-c=|f/k|$ (where $f$ is twice the fluid's rotation rate and $k$ is the zonal wavenumber), rather than where $U-c=0$. Such modifications are significant, but for present considerations (especially near the equator where $f=0$) we feel they are likely to produce only secondary effects.

Finally, we should mention that (10) may be more general than we have indicated since it depends only on the existence of some critical level absorption mechanism, not necessarily the one described by Booker and Bretherton (1967). Hines and Reddy (1967) show, for example, that the existence of dissipation will lead to critical level absorption, quite apart from the Booker and Bretherton mechanism.

In the next section we present some simple examples of changes in a mean flow resulting from gravity wave absorption.

## 4. Simple examples

In order to apply Eq. (11) we must know the functional form of $f(U)$. Neither observations nor theory are yet adequate for the detailed specification of this form for the equatorial gravity waves described in Section 2. However, observations do suggest that the bulk of wave energy is associated with phase speeds between about $\pm 20$ m sec$^{-1}$. The simplest choice for $f(U)$, consistent with this, is

$$\left.\begin{array}{ll} f(U)=\text{constant}, & -c_r<U<c_r \\ f(U)=0, & c_r<|U| \end{array}\right\}, \tag{13}$$

where $c_r$, a positive constant, is a high-frequency cutoff for the wave spectrum.

Our first example is for a situation where local time changes in $U$ are due solely to wave absorption. Eq. (11) becomes

$$\rho_0\frac{\partial U}{\partial t}=-f(U)\frac{\partial U}{\partial z}. \tag{14}$$

At $t=0$ we let $U=A+Bz$, where $A=-20$ m sec$^{-1}$ and $B=\frac{2}{3}$ m sec$^{-1}$ km$^{-1}$, and the vertical distribution for $\rho_0$ is taken from the equatorial standard atmosphere. From (13) with $c_r=5$ m sec$^{-1}$, we have

$$f(U)=\rho_0(27\text{ km})\cdot 1\text{ km month}^{-1},\quad -c_r<u<c_r,$$

where we have ignored the term $\exp(-2\pi\sqrt{\mathrm{Ri}-\frac{1}{4}})$ in (10), which is equivalent to assuming a very high Richardson number. This is certainly the case initially. The solution for this case, obtained numerically, is shown in Fig. 2; $t=0$ is arbitrarily identified with 1 April 1960. We note the following important features:

1) Only the region between 22.5 and 37.5 km, where the mean wind is between $-5$ and $+5$ m sec$^{-1}$, is affected by wave absorption.

2) Since $\partial U/\partial z$ is positive, wave absorption causes the addition of westerly momentum. However, the maximum velocity which can be produced is 5 m sec$^{-1}$. The mechanism ceases to operate for greater velocities.

3) Wave absorption causes the level at which $U=5$ m sec$^{-1}$ to descend at a rate given approximately by (1 km month$^{-1}$) $[\rho_0(27\text{ km})/\rho_0(z)]$. Thus, the level descends about 8.5 km from 1 April to 6 July; it descends another 3.5 km by 14 October; by 13 March it has descended only an additional 1.5 km. As long as the Richardson number $>1$, the lowest level at which $U=5$ m sec$^{-1}$ effectively shields the region above from the waves. This is independent of the distribution of $U$ in the region above.

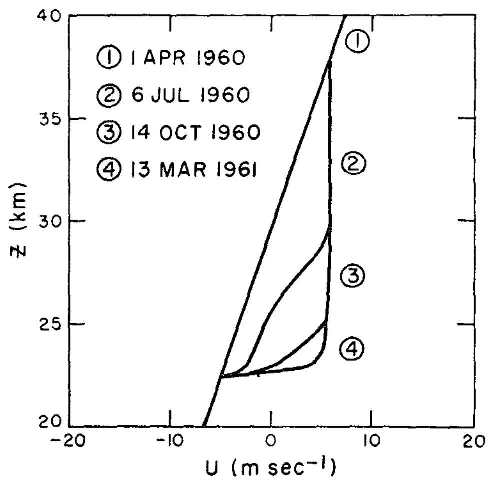


FIG. 2. Numerical results for the time variation of zonal velocity profiles in the case where local time changes in $U$ depend only on the wave absorption mechanism.

The above are the three most fundamental features of our mechanism; they are essential to all our applications. A fourth feature, peculiar to the above example, should, however, be mentioned.

4) The level at which $U=5$ m sec$^{-1}$ cannot descend below 22.5 km. Thus, in the absence of dissipative processes, a discontinuity in $U$ will form at 22.5 km, and we will, once more, have the problem, albeit an artificial one, of an infinitesmal layer absorbing a finite amount of momentum.

The shielding effect, mentioned above, is better seen in the next example where, in addition to the processes involved in our first example, we introduce a restoring force in the form of a Rayleigh friction which attempts to maintain the initial distribution of $U$. In this case Eq. (11) becomes

$$\rho_0\frac{\partial U}{\partial t}=-f(U)\frac{\partial U}{\partial z}+K[U(t=0)-U], \tag{15}$$

where $K=\rho_0(27\text{ km})\times(365\text{ days})^{-1}$, all other parameters being the same as in the first example. The numerically obtained solution is shown in Fig. 3. Again, the level at which $U=5$ m sec$^{-1}$ descends. However, now the region above, shielded from the action of the waves, relaxes to its initial form. For our particular choice of $K$ the relaxation is rather slow, but eventually we will be left with the initial profile except for a "spike" just above 22.5 km. What will happen now cannot be stated with certainty; however, the ingredients of a possible oscillation mechanism are evident. Let us assume that when the spike becomes sharp enough, it breaks down. There will then be no "shield" due to the spike, and presumably something close to the initial profile will be restored. The whole process will then begin again, forming a relaxation oscillation whose period depends on $K$, $f(U)$ and the initial $\partial U/\partial z$ (i.e., the distance

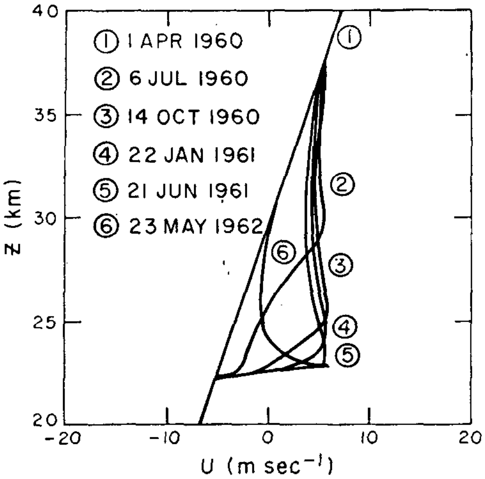


FIG. 3. Same as Fig. 2 except that a Rayleigh friction which tends to restore the initial profile in $U$ is included.

between the levels where $U=+5$ and $-5$ m sec$^{-1}$), and whose amplitude depends on the range of phase speeds over which $f$ is non-zero.

In the following sections we will show how, for the conditions obtaining in the equatorial stratosphere, an oscillation may be obtained without recourse to an arbitrary relaxation procedure.

## 5. Qualitative description of a model for the quasi-biennial oscillation

In Sections 6 and 7 we will describe the incorporation of the wave absorption mechanism into a reasonably comprehensive numerical model of the tropical stratosphere. In order to better understand the results of these rather complex calculations, it seems best to isolate beforehand those features of the model which are essential to the generation of the quasi-biennial oscillation and show, qualitatively, their roles.

The first important feature is the assumption that a spectrum of gravity waves, with horizontal phase speeds between $-20$ and $+20$ m sec$^{-1}$ propagates upward from a region below some level in the lower stratosphere. The range of phase speeds is most simply incorporated by using Eq. (13) with $c_r=20$ m sec$^{-1}$.

The second important feature is the existence in the equatorial upper stratosphere of a strong semiannual oscillation in zonal wind. This oscillation is described in detail by Reed (1966). While no complete theory has been presented for this oscillation, it is not hard to rationalize a six-month oscillation at the equator at levels where insolation absorption is important. This feature is incorporated by forcing the top level, 40 km, to oscillate with a six-month period.

The third feature is the assumption that the gravity waves are generated in the neighborhood of the tropopause. The observational basis for this assumption is meagre, primarily the finding by Yanai *et al.* (1968) that the peak energy density of equatorial waves occurs near the tropopause. We incorporate this feature into our model in the following way: As shown in Section 4 the interaction of waves with the mean flow causes a "velocity ledge" (i.e., a zone of strong wind shear) to descend. We simply shut off the shielding effect of this ledge when the middle of the ledge descends below 19 km. When the ledge is above 19 km it effectively shields the region above it from the action of waves.

How the above three features could combine to produce a quasi-biennial oscillation is shown in Figs. 4a–k. We take the constant in (13) to be such that a ledge descends from 40 to below 19 km in 15 months. This choice leads to what we shall call a synchronized oscillation. The reason for this terminology will be explained later. Let us assume at time $t=0$ that the semiannual oscillation is in its westerly phase and that the region is in an easterly phase. This is the situation in Fig. 4a. Since $\partial U/\partial z$ is positive, wave absorption will cause the descent of a westerly ledge, and this ledge shields the region above from the further action of waves. Thus, the region above will continue to execute a semiannual oscillation. This situation is depicted schematically in Figs. 4b–e. When $t=15$ months, the middle of the westerly ledge is below 19 km and we assume that the waves can again propagate upward; this is shown in Fig. 4f, the semiannual oscillation being in its easterly phase. Hence, $\partial U/\partial z$ is negative and wave absorption causes an easterly ledge to descend while the upper region continues to execute a semiannual oscillation. This is shown in Figs. 4g–4j. Finally, when $t=30$ months (Fig. 4k), the middle of the easterly ledge has descended below 19 km and the semiannual oscillation is in its westerly phase; the whole process can now begin anew. The reason for calling the above sequence a synchronized oscillation should now be clear: when the middle of a westerly ledge descends below 19 km, the semiannual oscillation is in its easterly phase; and when the middle of an easterly ledge descends below 19 km, the semiannual oscillation is in its westerly phase. Synchronization requires a "ledge" to descend from 40 km to below 19 km in $(2m+1)\times 3$ months ($m=0, 1, 2, 3\cdots$). Assuming easterly and westerly ledges can descend at different rates, we find that synchronous oscillations can take place with periods of

$$T_{syn}=(p+1)\times 6 \text{ months}, \quad p=0, 1, 2, \cdots. \tag{16}$$

The period is merely the sum of the times it takes for the wave interaction to bring a westerly and an easterly ledge from 40 km to below 19 km. The observed wave amplitudes in the data analysis of Wallace and Kousky (1968) are consistent with a momentum flux which suggests that $p$ is around 3 or 4. An average period of 26 months would result from two 24-month oscillations and one 30-month oscillation.

What happens when we do not have synchronization is shown in Fig. 5. The oscillation is merely delayed for

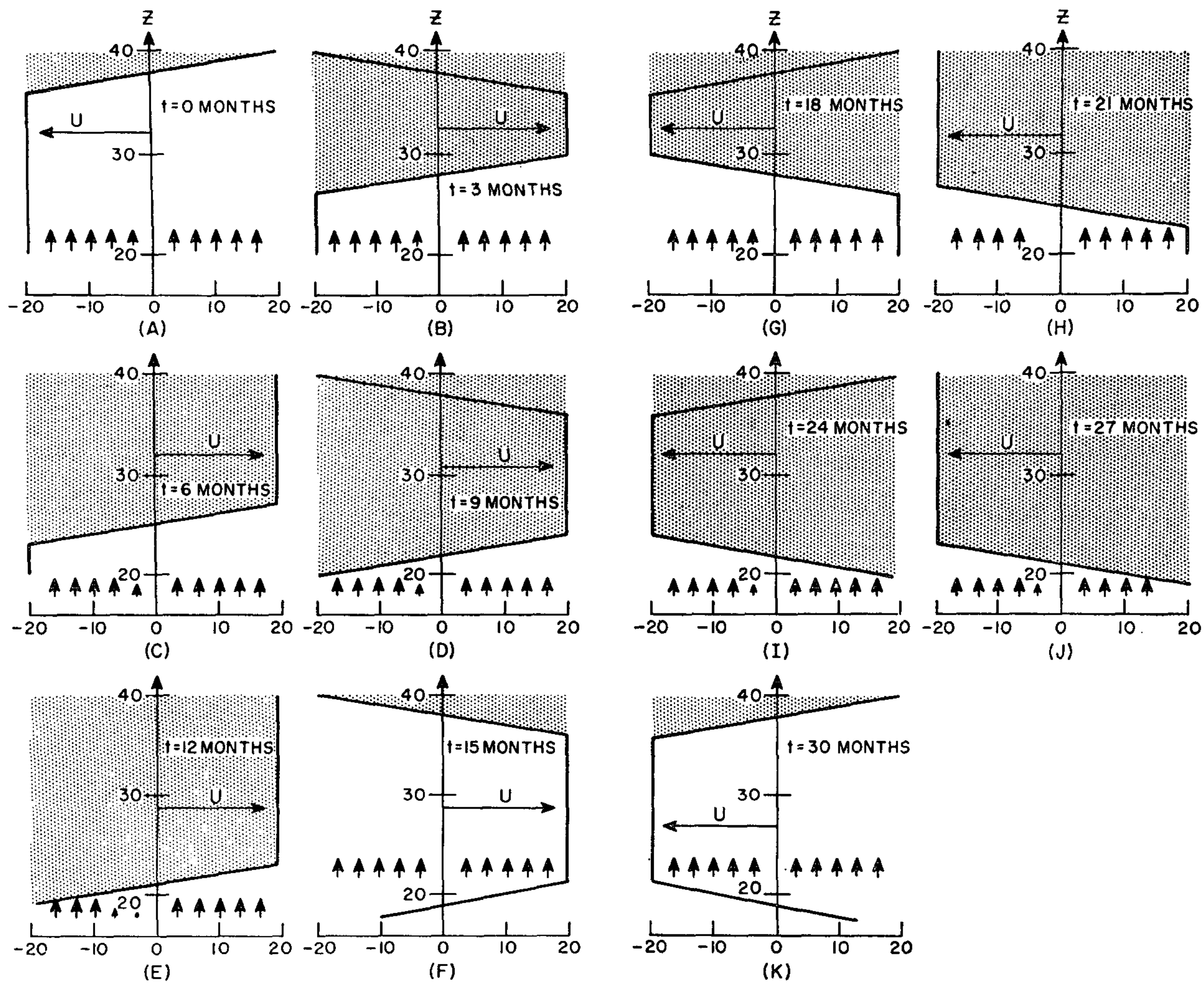


FIG. 4. A schematic time sequence illustrating the proposed theory of the quasi-biennial oscillation for the synchronous case at three-month intervals from 0–30 months, (a)–(k). Vertical arrows indicate gravity wave propagation. Shaded regions are shielded from the waves. See text for details.

a length of time up to 3 months until synchronization is re-established. In the meantime, the gravity waves simply continue to propagate upward beyond 40 km.

That this rather novel mechanism is apparently consistent with the observational evidence may be seen in the time-height section of Fig. 6. We note that the westerly regimes in the quasi-biennial oscillation are clearly synchronized with the westerly phase of the semiannual oscillation.

Before proceeding to a description of our numerical model let us briefly review the parameters which determine the characteristics of the oscillator in our model:

1) The period of the oscillation is determined by a) the upward flux of momentum due to gravity waves which can, but need not be, a statistically steady quantity; and b) the distance between the region dominated by the semiannual oscillation and the region of wave generation.

2) The amplitude of the oscillation is determined by the range of phase speeds over which we have a significant energy density of gravity waves.

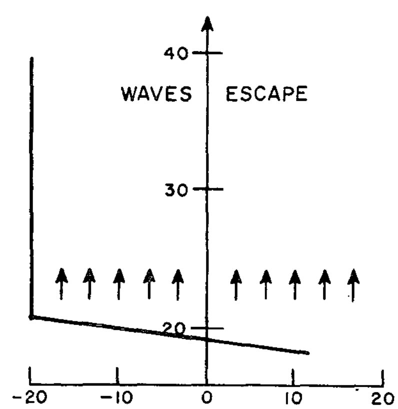


FIG. 5. Schematic illustration of gravity wave escape for non-synchronous oscillation. See text for details.

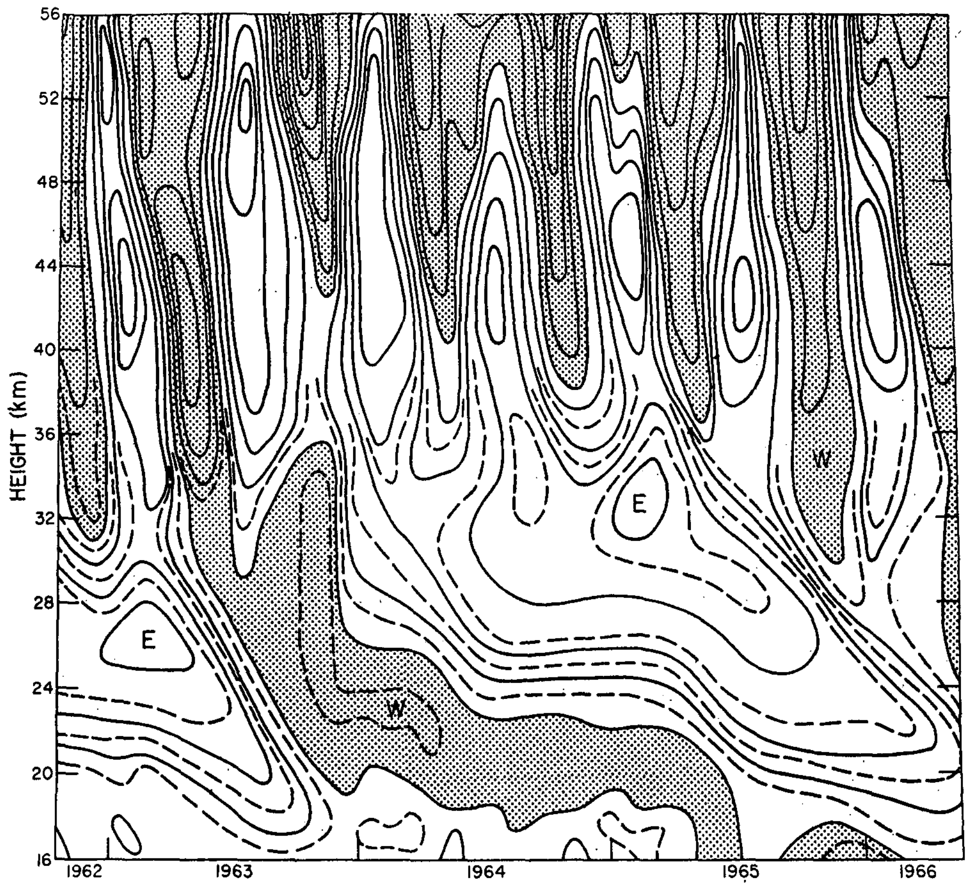


FIG. 6. Time-height section of zonal wind at 8° latitude with annual cycle removed. Solid isotachs are placed at intervals of 10 m sec$^{-1}$. Shaded areas indicate westerlies. Below 35 km monthly mean rawinsonde data for the Canal Zone (9N) and Ascension Island (8S) were averaged together to remove all fluctuations with odd symmetry about the equator. Above 34 km, this procedure could not be used because rocket data were available for Ascension Island only. At these levels the annual cycle was removed by harmonic analysis. Some minor smoothing was done to make the analyses compatable at 35 km. Figure prepared by J. M. Wallace and V. E. Kousky.

It should finally be added that the above, while containing the basic features of our mechanism, is merely a schematic description. As we shall see in the following sections, a more realistic treatment contains differences in details.

## 6. The numerical model

The basic equations of the model are a simplification of the set developed by Wallace and Holton (1968). The details of the derivation will not be repeated here. Wallace and Holton showed that in nondimensional form the zonal momentum, thermal wind, thermodynamic energy, and continuity equations could be written as follows:

$$\frac{\partial u}{\partial t}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}-fv=F-G, \tag{17}$$

$$f\frac{\partial u}{\partial z}+\frac{\partial \theta}{\partial y}=0, \tag{18}$$

$$\frac{\partial \theta}{\partial t}+v\frac{\partial \theta}{\partial y}+w\left(\frac{R\theta}{c_p}+\frac{\partial \theta}{\partial z}\right)+\frac{w}{\epsilon}=-P+Q, \tag{19}$$

$$\frac{\partial v}{\partial y}+e^{z}\frac{\partial}{\partial z}(e^{-z}w)=0. \tag{20}$$

Here, $t$ is time dimensionalized by $\tau$, the annual period; $y$ is the northward distance dimensionalized by $L$, the latitudinal half-width of the oscillation ($\sim$2000 km); $z$ is the vertical coordinate dimensionalized by $H$, the scale height; $u$ is the zonal velocity dimensionalized by $2\Omega(\sin\phi_0)L$, where $\Omega$ is the angular velocity of the earth and $\phi_0$ is a convenient reference latitude (say, 6N) chosen to make $u$ order unity; $v$ is the meridional velocity dimensionalized by $\tau/L$; $w$ is the vertical velocity dimensionalized by $\tau/H$; and $\theta$ is the deviation of temperature from its horizontal average dimensionalized by $[2\Omega(\sin\phi_0)L]^2/R$, where $R$ is the gas constant. The remaining symbols in (17)–(20) are $c_p$, the specific heat at constant pressure; $f=\sin\phi/\sin\phi_0$, where $\phi$ is the latitude; and $\epsilon$, the static stability parameter, defined by

$$\epsilon=\frac{[2\Omega(\sin\phi_0)L]^2}{gH\left(\frac{R}{c_p}+\frac{d\ln T_0}{dz}\right)},$$

where $T_0$ is the horizontally averaged temperature.

The source terms on the right-hand side in (17) and (19) are defined as follows: $F$ and $G$ are the frictional dissipation due to small scale eddies, and the zonally averaged divergence of the eddy momentum flux, respectively, dimensionalized by $[\tau/2\Omega(\sin\phi_0)L]^{-1}$; and $P$ and $Q$ are the eddy heat flux and the rate of diabatic heating, respectively, dimensionalized by $[2\Omega(\sin\phi_0)L]^2/R$.

In the stratosphere where $\epsilon\ll1$, (19) may be greatly simplified for long-period oscillations by neglecting terms of order $\epsilon$ and obtaining a simple balance between the adiabatic cooling by vertical motion and diabatic heating by radiation. Thus,

$$w=\epsilon Q. \tag{21}$$

We assume that the diabatic heating rate $Q$ may be approximated by Newtonian cooling, so that

$$Q=k_e(\theta_e-\theta),$$

where $\theta_e$ is the (nondimensional) radiative equilibrium temperature and $k_e$ an inverse radiative relaxation time dimensionalized by the annual period. We can then combine (18) and (21) to obtain

$$\frac{\partial^2\Psi}{\partial y^2}=\epsilon\kappa_e\left(\frac{\partial\theta_e}{\partial y}+f\frac{\partial u}{\partial z}\right)e^{-z}, \tag{22}$$

where $\Psi$ is a meridional mass transport streamfunction defined by the relations

$$\frac{\partial\Psi}{\partial z}=-e^{-z}v,\quad \frac{\partial\Psi}{\partial y}=e^{-z}w. \tag{23}$$

Eqs. (17) and (22) together with the definitions (23) form a complete set for prediction of the evolution of the zonal and meridional flow, provided that the forcing functions $F$, $G$ and $\theta_e$ are specified. A linearized version of this system was discussed by Dickinson (1968). In the present model $G$ represents the critical layer absorption of vertically propagating gravity waves which we parameterize in terms of the zonal wind shear as explained in Section 3. We represent dissipation by small-scale eddies as an eddy diffusion process by letting

$$F=\frac{1}{\mathrm{Re}_z}\frac{\partial^2u}{\partial z^2}+\frac{1}{\mathrm{Re}_y}\frac{\partial^2u}{\partial y^2},$$

where $\mathrm{Re}_z=H^2/\tau K_{zz}$ and $\mathrm{Re}_y=L^2/\tau K_{yy}$, with $K_{zz}$ and $K_{yy}$ the vertical and horizontal eddy diffusivities, respectively.

A finite difference analog to a linearized version of these equations has been formulated by Holton (1968). The numerical solutions described in the next section were obtained using a straightforward extension of his scheme. We compute the dependent variables $u$ and $\Psi$ at the points of a rectangular grid which extends laterally from the equator to 26N in 1° intervals, and vertically from 16–40 km in 1-km intervals.

Boundary conditions are applied such that 1) $\Psi=0$ at the equator and 26N, 2) $u=0$ at 16 km and at 26N, 3) $\partial u/\partial y=0$ at the equator, and 4) $u=\cos(4\pi t)\times\exp(-\phi^2/289)$ at 40 km, where $\phi$ is the latitude in degrees. Thus, a semiannual oscillation in the zonal wind is imposed as the upper boundary condition.

## 7. Numerical results

We have carried out several numerical integrations to test our proposed mechanism for explaining the cause and dynamics of the quasi-biennial oscillation. In each of the four numerical experiments described below the equations were integrated for 16 years using time steps of 3.75 days. In all of our experiments $\theta_e$ was assumed independent of latitude. The radiative relaxation time was allowed to vary inversely with height from 60 days at 16 km to 7.5 days at 40 km. The eddy diffusion coefficients were assigned the values $K_{zz}=10^3$ cm² sec⁻¹ and $K_{yy}=5\times10^7$ cm² sec⁻¹. The semiannual oscillation at 40 km was set at 20 m sec⁻¹ amplitude at the equator.

*a. Experiment 1.* In this experiment the momentum equation (17) was simplified by omission of the nonlinear advection terms and the Coriolis term. Therefore, the physical processes included in the model were essentially only those of the prototype model, i.e., critical layer momentum absorption and frictional dissipation. The only latitudinal coupling in the dynamics is due to the relatively small horizontal momentum diffusion. Fig. 7 illustrates a time-height section of the zonal wind at the equator for a 6-year portion of the integration. The wave spectrum assumed here is the flat-top distribution of (13) which, in terms of our nondimensional variables (with $u=1$ corresponding to 20 m sec$^{-1}$), may be written as

$$\left.\begin{array}{l}G=\exp\{z-1.4\}\partial u/\partial z,\quad \text{for } -1<u<+1\\ G=0,\quad \text{for } |u|\geqslant1, \text{ and for the entire shielded}\\ \quad\text{region above the critical layer}\end{array}\right\}, \tag{24}$$

where the exponential gives the inverse density dependence normalized to 25 km. Eq. (24) implies a propagation rate of 1 scale height per year at 25 km. With this flat-top source and no Coriolis torques or nonlinear advection terms, the alternating easterly and westerly wind regimes are quite symmetric in appearance and propagate downward at about the same rate. The averaged period in this case is 36 months.

*b. Experiment 2.* This experiment is similar to the previous one except that the magnitude of $G$ is doubled for $0<u<1$. Thus, the momentum transport of the

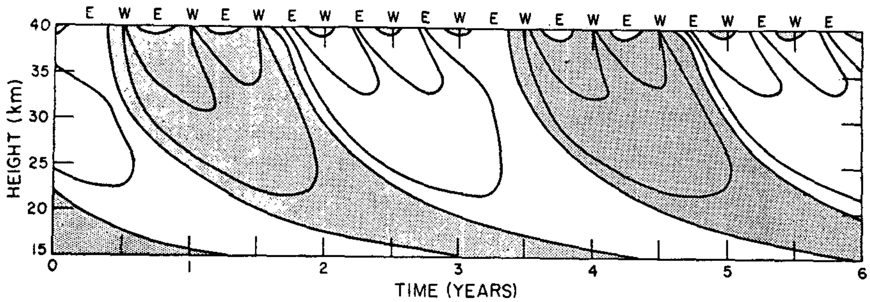


FIG. 7. Time-height section of the zonal wind velocity at the equator for Experiment 1. Solid lines are 20 m sec$^{-1}$ isopleths. Westerlies are shaded. Westerly and easterly phases of the semiannual oscillation at 40 km are denoted by W and E, respectively.

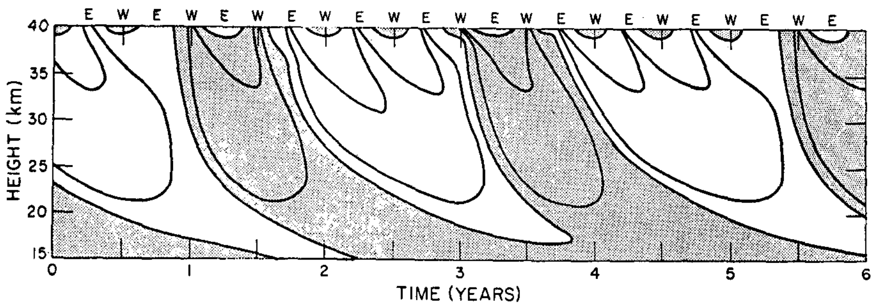


FIG. 8. Time-height section of the zonal wind at the equator for Experiment 2. (See legend for Fig. 7).

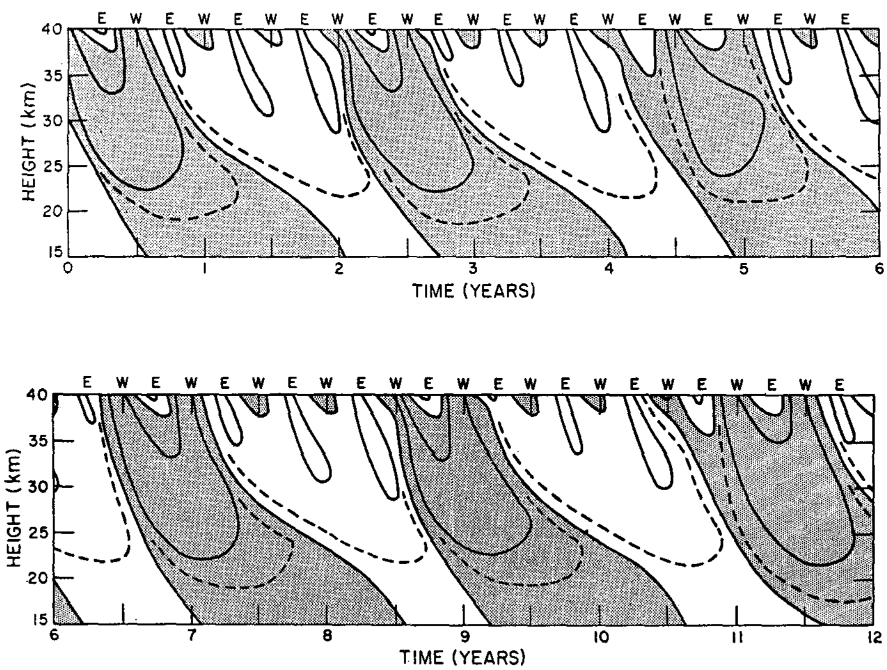


FIG. 9. Time-height section of the zonal wind at 12° latitude for Experiment 3. Solid lines are 10 m sec$^{-1}$ isopleths.

westerly waves is assumed to be double that of the easterly waves. A six-year portion of the results at the equator is shown in Fig. 8. The time-height section indicates, not surprisingly, that the difference in descent rates between the westerlies and the easterlies can be accounted for by assuming greater energy in the westerly waves than in the easterly waves. It is not necessary to depend on differences in the rate of vertical advection between the two regimes as proposed by Wallace (1967a). However, as will be shown in Experiment 4, vertical advection does play a significant role in the momentum balance. Due to the faster descent rate of the westerlies, the period of the oscillation in this experiment was less than in Experiment 1. In this case the 16-year run indicated alternating 24- and 30-month oscillations, i.e., an average period of 27 months. This variation in period results from the fact that for the conditions of this model the "natural" period of the long-term oscillation is not an exact multiple of six months, so that it is not locked in phase with the semiannual oscillation. Thus, as is indicated in Fig. 8, when an easterly regime reaches the cutoff altitude of 19 km at a time when the zonal velocity is westerly at 40 km, a new westerly regime will start downward immediately. But if the easterlies reach 19 km at a time when the zonal velocity is easterly at 40 km, there will be no critical layer for the westerly waves to encounter. Thus, formation of a new westerly regime can not begin until the semiannual oscillation returns to its westerly phase.

*c. Experiment 3.* This experiment differs from the above two in that the Coriolis acceleration due to the meridional motion associated with the oscillation is included. The results of this run indicate that the Coriolis effect has a negligible influence near the equator, but at 12° latitide it reduces the amplitude of the downward propagating regimes as is readily apparent from Fig. 9. Physically, this effect occurs because the Newtonian cooling in the model is basically a damping mechanism which through its coupling with the meridional circulation acts to diffuse zonal momentum (Holton, 1968). Since the easterly regimes propagate more slowly than the westerly regimes, the radiative damping has a longer time to dissipate the easterly momentum. Hence, the amplitude of the easterlies decays away from the source region more rapidly than that of the westerlies. Note also that Newtonian cooling causes the semiannual oscillation to diffuse downward a few kilometers below its source level.

We also note that the 12-year run shown in Fig. 9 indicates that the length of the cycle is somewhat variable. The mean period for this sample run is about 26 months. This mean period results from an irregular alternation between 24- and 30-month oscillations. Thus, the term "quasi-biennial" indeed appears to be an appropriate name for this oscillation.

*d. Experiment 4.* In this final experiment we have integrated the full momentum equation (17). The results are shown in Fig. 10 as time-height sections for a six-year portion of the run. Results for both the

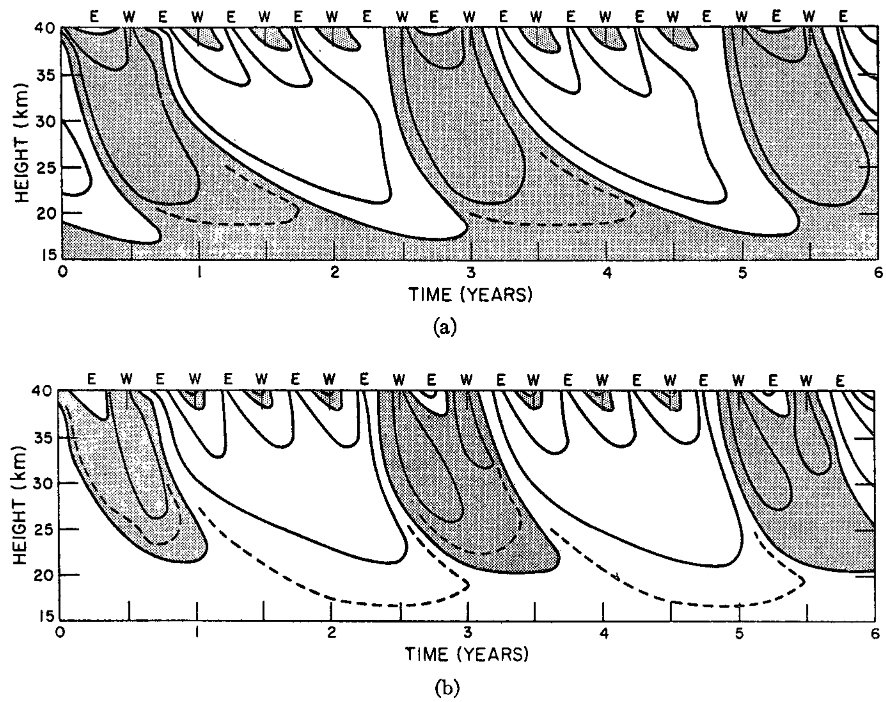


FIG. 10. Time-height section of the zonal wind at the equator (a) and 12° latitude (b) for Experiment 4. Solid lines are 20 m $sec^{-1}$ isopleths in (a) and 10 m $sec^{-1}$ isopleths in (b).

equator and 12° latitude are displayed in order to illustrate the latitudinal variation of the nonlinear terms. Horizontal momentum advection actually plays a minor role, but is included for completeness. Vertical advection, on the other hand, has an important moderating influence on the oscillation. In the long term mean the vertical advection associated with the oscillation tends to produce westerlies in the lower stratosphere at the equator, and easterlies away from the equator. In fact it may be shown that $w\partial u/\partial z$ is negative definite near the equator so that vertical advection will tend to always produce westerlies at the equator. Note that $w$ is here the vertical velocity induced by the oscillation itself. That $w\partial u/\partial z$ is always negative at the equator may be seen as follows: If $\partial u/\partial z>0$, then $\partial\theta/\partial y<0$ from the thermal wind relationship. Therefore, $\theta>\theta_e$ and there is radiative cooling which must be balanced by adiabatic warming ($w<0$). Conversely, when $\partial u/\partial z<0$ at the equator, then $\partial\theta/\partial y>0$, $\theta<\theta_e$, and $w>0$. In either case $w\partial u/\partial z<0$ so that vertical advection produces westerly momentum. This continuous westerly acceleration must, in the long term, be balanced by frictional dissipation. In fact, some numerical experiments (not shown here) have demonstrated that there is a rather sensitive balance between the eddy viscosity values assumed in the model and the mean westerly momentum obtained near the tropopause at the equator. In latitudes more than about 10° from the equator, the sign of the vertical velocity induced by the oscillation is opposite to its equatorial value. Thus, in these regions $w\partial u/\partial z>0$. The 12° latitude time-height section in Fig. 9 indicates that vertical advection is sufficiently powerful to produce mean easterlies in the lower region despite the opposing tendency of the Coriolis term. This model is therefore able to at least partly account for the observed mean easterly shear of the zonal wind with latitude between 0°–12° without the necessity of invoking horizontal eddy momentum flux processes. We note that vertical advection also propagates the westerly phase of the semiannual oscillation downward a few kilometers near the equator, while propagation of the easterly phase is suppressed. Furthermore, the effective reduction in propagation velocity of the easterly quasi-biennial regime near the equator by vertical advection results in an average period of 30 months in this experiment compared to the 26-month average period of the previous experiment. This may imply that the wave amplitude assumed in our model is a slight underestimate of the actual amplitude of the waves in the tropical stratosphere. However, the arbitrary nature of the 19-km cutoff in this model makes a closer estimate of the wave amplitude difficult.

## 8. Conclusions and suggestions

We have presented a mechanism, involving the interaction of internal equatorial gravity waves with the equatorial stratospheric zonal wind, which explains the following main features of the quasi-biennial oscillation: the period, the constant amplitude above 23 km, and the presence of zonally averaged westerly momentum. In our model, the period [Eq. (16)] is determined by the upward flux of gravity waves from the troposphere (a quantity which may be statistically steady). However, the period must be some multiple of six months. This period may change from cycle to cycle. Study of the data in Reed (1967) and Wallace and Holton (1968) indicates that the time interval between the appearance of successive westerly regimes at 30 km tends to be a multiple of six months. (See also Fig. 6.) This is an important observation, since the prediction that the periods are multiples of six months is independent of most of the assumptions in our model. An average period of 26 months would, as pointed out in Section 5, result from two 24-month cycles and one 30-month cycle.

There are a number of observations which could check various aspects of our model:

1) The rate at which stratospheric easterlies and westerlies propagate downward should be proportional to the upward flux of gravity waves at the bottom of the equatorial stratosphere. It should be possible to obtain observations of $\overline{w'p'}$ for equatorial waves and see whether there is a correlation between this quantity and the rate at which westerlies or easterlies are descending.

2) It should be possible to relate observations of short-period equatorial waves to the length of a given quasi-biennial cycle. This suggestion is, of course, related to our first suggestion.

3) As noted in Sections 5 and 7, there will be periods when the mesospheric semiannual oscillation and the stratospheric quasi-biennial oscillation are not synchronized. During such periods, equatorial gravity waves should escape into the upper mesosphere where they may, perhaps, be observed by either rocket measurements of wind or by radio observations of the *D* region.

4) A detailed study of the gravity waves described in Section 2 is necessary. In particular, there is a need for wind profiles with high vertical resolution in order to see whether waves other than those described by Maruyama (1967) and Wallace and Kousky (1968) are present.

Our theory, itself, needs to be improved and extended in some of its details: thus, 1) the interaction of gravity waves with the mean flow should be studied without assuming separation of scales; 2) the interaction of gravity waves with the mean flow should be explicitly studied for gravity waves modified by the earth's rotation; 3) the means whereby equatorial gravity waves are excited requires much more study; and finally 4) it would be satisfying to replace our *ad hoc* assumption concerning the cutoff which occurs when a velocity ledge reaches the region of wave generation with explicit calculations. This, of course, requires some progress on item 3).

It is evident that our theory is far from complete. It does, however, provide a useful framework for further study.

*Acknowledgments*. R. S. Lindzen wishes to thank NCAR for its support. Much of the work reported here was done while he was an employee of that organization. R. S. Lindzen also wishes to thank Prof. R. J. Reed for his hospitality at the University of Washington. Prof. Reed's continuing encouragement has been important to both authors. Thanks are also due to Mr. Grant Gray for help with numerical analyses and programming, and to Prof. J. M. Wallace for observational advice and stimulating discussion. We also wish to thank Prof. Wallace and Mr. Vern Kousky for permission to reproduce their unpublished data analysis (Fig. 10). This research was supported in part by the Atmospheric Sciences Section, National Science Foundation, under Grants GA-450 and GA-629x.

## REFERENCES

Booker, J. R., and F. P. Bretherton, 1967: The critical layer for internal gravity waves in a shear flow. *J. Fluid Mech.*, **27**, 513–519.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–110.

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

Eckart, C., 1960: *Hydrodynamics of Oceans and Atmospheres*. New York, Pergamon Press, 290 pp.

Eliassen, A., and E. Palm, 1960: On the transfer of energy in stationary mountain waves. *Geofys. Publikasjoner*, **22**, No. **3**, 1–23.

Hines, C. O., and C. A. Reddy, 1967: On the propagation of atmospheric gravity waves through regions of wind shear. *J. Geophys. Res.*, **72**, 1015–1034.

Holton, J. R., 1968: A note on the propagation of the biennial oscillation. *J. Atmos. Sci.*, **25**, 519–521.

——, and R. S. Lindzen, 1968: A note on 'Kelvin' waves in the atmosphere. *Mon. Wea. Rev.*, **96**, 385–386.

Jones, Walter L., 1967: Propagation of internal gravity waves in fluids with shear flow and rotation. *J. Fluid Mech.*, **30**, 439.

Lindzen, R. S., 1965: The radiative-photochemical response of the mesosphere to fluctuations in radiation. *J. Atmos. Sci.*, **22**, 469–478.

——, 1966: Radiative and photochemical processes in mesospheric dynamics: Part II. Vertical propagation of long period disturbances near the equator. *J. Atmos. Sci.*, **23**, 334–343.

——, 1967: Planetary waves on beta-planes. *Mon. Wea. Rev.*, **95**, 441–541.

——, and T. Matsuno, 1968: On the nature of large wave scale disturbances in the equatorial lower stratosphere. *J. Meteor. Soc. Japan*, **46** (in press).

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

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–43.

Reed, R. J., 1964: A tentative model of the 26-month oscillation in tropical latitudes. *Quart. J. Roy. Meteor. Soc.*, **90**, 441–466.

——, 1965: The quasi-biennial oscillation of the atmosphere between 30 and 50 km over Ascension Island. *J. Atmos. Sci.*, **22**, 331–333.

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

——, 1967: The structure and dynamics of the 26-month oscillation. *Proc. Intern. Symp. Dynamics of Large Scale Processes in the Atmosphere*, Moscow, 376–387.

Staley, D. O., 1963: A partial theory of the 26-month oscillation of the zonal wind in the equatorial stratosphere. *J. Atmos. Sci.*, **20**, 506–515.

Tucker, G. B., 1964: Zonal winds over the equator. *Quart. J. Roy. Meteor. Soc.*, **90**, 405–423.

Wallace, J. M., 1967a: On the role of mean meridional circulations in the biennial oscillation. *Quart. J. Roy. Meteor. Soc.*, **93**, 176–185.

——, 1967b: A note on the role of radiation in the biennial oscillation. *J. Atmos. Sci.*, **24**, 598–599.

——, and J. R. Holton, 1968: A diagnostic numerical model of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 280–292.

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

——, and R. E. Newell, 1966: Eddy fluxes and the biennial stratospheric oscillation. *Quart. J. Roy. Meteor. Soc.*, **92**, 481–489.

Yanai, M., T. Maruyama, Tsuyoshi Nitta and Y. Hayashi, 1968: Power spectra of large scale disturbances over the tropical Pacific. *J. Meteor. Soc. Japan*, **46**, (in press).