The zonally symmetric flow in the tropical lower stratosphere exhibits three types of variability: (1) an annual cycle that has odd symmetry about the equator with easterlies in the summer hemisphere and westerlies in the winter hemisphere, (2) a semiannual oscillation with a rather complex distribution of amplitude and phase, and (3) irregular long-period fluctuations with an average period of somewhat >2 years and even symmetry about the equator, successive easterly and westerly wind regimes of this so-called quasi-biennial oscillation being observed to propagate downward at a rate of ~1 km month⁻¹. Superposed on the zonally symmetric circulation are two types of planetary scale wave disturbances which propagate both zonally and vertically with periods of 10-15 days and 4-5 days, respectively. These have been identified with the gravest modes of the family of forced zonally propagating waves on a rotating sphere. The first type is an atmospheric ‘Kelvin wave,’ which bears some similarity to the shallow water wave that propagates along coastlines. The second type is a ‘mixed Rossby-gravity wave.’ Both waves transport energy upward which suggests a tropospheric energy source. The waves are capable of modifying the zonally symmetric flow in the stratosphere through the vertical transport of zonal momentum. Consideration of the interactions between the waves and the zonal flow suggests a plausible mechanism for producing the observed quasi-biennial oscillation in zonal wind.
1. INTRODUCTION

The tropical stratosphere encompasses the region within ~20° of the equator and extends from the tropopause at ~17 km upward to the stratopause near the 50-km level. It is a layer of very strong static stability, temperature increasing with height throughout much of its depth. For the discussion of the dynamics it is convenient to divide this region into an upper layer extending from 35 to 50 km and a lower layer extending from 17 to 35 km. The upper layer is marked by large tidal motions and a strong semiannual oscillation in the mean zonal wind. Owing to the sparsity of data, very little is known about the wave disturbances, other than tides, in the upper stratosphere. The circulation of the lower stratosphere is marked by the presence of large-scale vertically propagating wave disturbances with periods ranging from a few days to a few weeks. These are superposed on a mean zonal flow pattern that exhibits a curious low-frequency nonseasonal oscillation with a characteristic period of slightly >2 years. There is the distinct possibility that an analogous oscillation might exist in the tropical regions of other atmospheres. For example, Owen and Staley [1963] have shown evidence of variability in the strength of the zonal winds in the equatorial atmosphere of Jupiter.

2. ZONAL MEAN CIRCULATION

The spread of the dust cloud produced by the 1883 Krakatoa eruption gave scientists their first view of the wind circulations of the tropical stratosphere. 'Apart from offshoots toward Japan and South Africa immediately after the explosion, the main body of the cloud moved from east to west at an average speed of 73 miles per hour, completing at least two circuits of the earth in equatorial latitudes' [Wexler, 1951]. Berson's upper air studies over Africa in 1908–1909 with balloons unexpectedly showed evidence of a westerly (west to east) circulation in the same region. From that time until the late 1950's the belief in the coexistence of the two opposing flow regimes prevailed. Palmer [1954] described Berson's westerlies as 'a narrow "thread" of steady winds whose axis lies at about 2°N and whose base lies near 20 km. The upper transition to the Krakatoa winds varies from month to month and year to year.' With the advent of more regular soundings, evidence on the variability of the wind structure increased. Korshover [1954] showed that the transition level between westerlies and easterlies varied between 21 and 27 km in a series of observations spanning several years. McCreary [1959] followed the downward movement of this transition level until the upper regime eventually replaced the lower. As a result he suggested that a dynamic view of the tropical stratospheric circulation fitted the observations better than the traditional steady state description did. Within the next year, Reed [1960] in the United States and Veryard and Ebdon in England independently discovered what has come to be called the 'quasi-biennial' or '26-month' oscillation in the wind structure of the tropical stratosphere. Subsequent investigations have shown that it is possible to explain virtually all the variance of the mean (zonally averaged) zonal wind in the tropical stratosphere on the basis of a superposition of the
annual and semiannual cycles and this quasi-biennial oscillation. The numerous observational studies that provided the documentation for the quasi-biennial oscillation and led to the discovery of the semiannual oscillation at higher levels are summarized in an earlier review by Reed [1965a].

a. Observations. To illustrate the mean zonal wind structure in the tropical stratosphere, I will begin by showing a number of time-height sections based on monthly mean data for various stations and combinations of stations. Figure 1 shows sections for a group of stations near 8°N, for Canton Island (3°S) and for Ascension Island (8°S). The stations are grouped without any reference to longitude because these low-frequency zonal wind fluctuations show a very high degree of zonal symmetry. (This is in contrast to the troposphere, where the monsoon circulation produces large deviations from zonal symmetry.) The presence of the quasi-biennial oscillation is evident in all three sections, but it is most clearly seen in the 3°S section as a succession of downward propagating easterly and westerly wind regimes. In the 8° sections the quasi-biennial oscillation is superimposed on an annual cycle that is marked by easterly winds that increase with height in the summer season and westerly

Fig. 1. Time-height sections of mean zonal wind at the latitudes indicated. Solid lines are placed at increments of 10 m sec⁻¹. Shaded areas are westerlies. Stations used in preparing the section are indicated in Wallace [1967].
winds in the winter season. Composite sections for 14°, 20°, and 32°N are shown in Figure 2. The amplitude of the annual cycle is seen to increase with latitude, whereas the quasi-biennial oscillation decays until it is just barely detectable at 20°N.

The annual cycle is well understood as a response to seasonal changes in the latitudinal distribution of incoming solar energy in the ozone layer. Radiative heating in the summer hemisphere and cooling in the polar night region in the winter hemisphere drive a pole to pole mean meridional circulation cell. Because of the rotation of the earth this gives rise to a strong zonal flow with odd symmetry about the equator, easterlies in the summer hemisphere, and westerlies in the winter hemisphere [Leovy, 1964]. Since the annual cycle is mainly an extratropical phenomenon and it does not interact strongly with the other circulation systems in the tropics, I will not discuss it here in any further detail.

Because of the odd symmetry of the annual cycle with respect to the equator it is possible to eliminate its effects by averaging time-height sections for stations located at comparable latitudes in the northern and southern hemispheres. This has been done in Figure 3 for 8° and 20° latitude. The Canton

---

**Fig. 2.** Time-height sections of mean zonal wind at the latitudes indicated. Solid lines are placed at increments of 10 m sec⁻¹. Shaded areas are westerlies. Stations used in preparing the section are indicated in Wallace [1967].
Island (3°S) section is repeated for comparison. The similarity between the 8° section and the 3°S section even in some of the more subtle details is quite remarkable when we note that these sections are based on data from widely different longitudes. This similarity is evidence of the longitudinal uniformity of these low-frequency zonal wind oscillations. The quasi-biennial oscillation is also apparent in the 20°N section, but with greatly reduced amplitude. All three sections also show evidence of a semiannual oscillation first noted by Reed [1965b, 1966]. This is most pronounced in the 20° section, but it also shows up quite strongly above 30 km in the 3°S section.

Further documentation of the semiannual oscillation at the higher levels is provided in Figure 4, which is based on monthly mean rawinsonde and rocketsonde data at the Canal Zone (9°N) and Ascension Island (8°S). Between 30 and 40 km, there is a transition from a lower stratospheric regime in which the quasi-biennial oscillation is dominant to an upper regime in which the semiannual cycle is dominant. There is some evidence of a coupling between
Fig. 4. 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 (9°N) and Ascension Island (8°S) were averaged together to remove all fluctuations with odd symmetry about the equator. Above 35 km this procedure could not be used because rocket data were only available for Ascension Island. At these levels the annual cycle was removed by harmonic analysis. Some minor smoothing was done to make the analyses compatible at 35 km.

the two oscillations in the sense that the shear zones that mark the leading (lower) edges of the westerly regimes of the quasi-biennial oscillation seem to appear first in association with westerly regimes of the semiannual oscillation. This apparent relationship may have important implications on the mechanism for the quasi-biennial cycle. This idea is discussed further in section 5.

Figure 5 shows a time-height section of zonal wind at the Canal Zone station (9°N) for a period extending over about 20 years. To continue the record beyond June 1970, it was necessary to substitute data for Kwajalein, which is located at very nearly the same latitude. The seasonal cycle and long-term means were removed by subtracting the 15-year (1956–1970) averages of the means for each month of the year. At 10 mb and above, it was necessary to use a shorter averaging period, since routine data were not available at these levels until the early 1960's.

Prior to 1963 the period of individual cycles is only slightly >2 years,
whereas through the latter part of the 1960's there are several cycles with periods closer to $2\frac{1}{2}$ years. This slight lengthening of the period after 1963 gave rise to speculation among some investigators that the quasi-biennial oscillation had disappeared. In retrospect it is clear that it had not disappeared unless one clings to the narrow definition of the term and insists that $2\frac{1}{2}$-year cycles are fundamentally different from 2-year ones. From a physical point of view it appears more plausible to view the quasi-biennial oscillation as a prominent feature of the entire record and to recognize that the period of individual cycles can vary somewhat. All cycles display the same characteristic downward propagation. The shear zone that marks the leading edge of westerly regimes is usually sharper than the corresponding easterly shear zone.

The height and latitudinal dependence of the amplitude and phase of the quasi-biennial oscillation are depicted in Figure 6, which was prepared by Reed on the basis of a harmonic analysis of monthly mean zonal wind data for 18 tropical stations. The oscillation is centered on the equator with a half width of $10^\circ$–$15^\circ$ of latitude. Amplitudes are large and nearly constant with heights $>22$ km, but they drop off rapidly as the wind regimes propagate below this...
Fig. 6. Phase and amplitude of the quasi-biennial oscillation in zonal wind (adapted from U.S. Navy Weather Research Facility [1964]). Phases are indicated by dashed lines spaced at intervals of 1 month, where time increases downward. Amplitude in meters per second is given by solid lines.

level. Phase propagates downward at an average rate of about 1 km month$^{-1}$. It is evident from the time-height sections shown in Figures 1-5 that there is a good deal of variability in the rate of downward propagation of individual wind regimes. In general, there is a tendency for westerly regimes to propagate downward somewhat more rapidly than easterly regimes.

The distribution of amplitude and phase in the semiannual cycle is somewhat less clearly defined because of the sparsity of data at rocket levels. Below 40 km, van Loon et al. [1972] have shown that amplitude tends to increase with latitude, at least in the northern hemisphere tropics. This idea is supported by Figure 3, which shows the clearest evidence of the semiannual cycle in the 20° section. However, Reed [1973] has shown that in the upper stratosphere the cycle is strongest at equatorial latitudes with amplitudes as large as 25 m sec$^{-1}$ near 50 km. Like the quasi-biennial oscillation but to a much lesser degree the semiannual cycle exhibits downward phase propagation; i.e., the phase propagates through a much smaller fraction of a cycle within the range of levels in which the oscillation is observed.

It can be demonstrated from scaling considerations that the zonal wind and temperature fields within the quasi-biennial oscillation and the semiannual
cycle must be in thermal wind equilibrium to within a few kilometers of the equator. Therefore the existence of fluctuations in zonally averaged temperature with corresponding periods is implied. Reed [1962a, b] has shown observational evidence of a quasi-biennial temperature oscillation that has the proper phase relationship to the wind oscillation to satisfy the thermal wind relationship. At the equator, warm temperature anomalies of a few degrees centigrade coincide with the layers in which westerly winds are increasing with height and vice versa. At subtropical latitudes, there is evidence of a small temperature oscillation with a reversed phase relative to that of the equatorial temperature oscillation [U.S. Navy Weather Research Facility, 1964]. Van Loon et al. [1972] have documented the geostrophic nature of the semiannual cycle in the lower stratosphere.

b. Theory. Early attempts at a theoretical explanation of the quasi-biennial oscillation were generally based on thermal forcing by some hitherto undetected periodicity in solar output. Possible candidates for a forcing mechanism included a neglected cycle in sunspot numbers [Shapiro and Ward, 1962], the fifth and sixth harmonics of the 11.2-year sunspot cycle [Probert-Jones, 1964], a 26-month oscillation in apparent solar diameter [Newell, 1964], and a beat frequency between the solar rotation and the lunar synodic periods (R. S. Lindzen, personal communication, 1965). Staley [1963] and Lindzen [1965, 1966] examined the atmospheric response to a periodic thermal drive and showed that under certain conditions downward propagating oscillations in temperature and zonal wind could be produced.

In the theories based on thermal forcing the zonal wind field is coupled to the heating field through a system of zonally symmetric circulation cells driven by differential heating. These motions produce local changes in zonal wind by advecting absolute angular momentum in the meridional plane. In a rotating frame of reference the local changes in angular momentum arise mainly as a result of Coriolis torques produced by the mean meridional circulation. This is known to be a powerful mechanism for bringing about changes in zonal wind at latitudes away from the equator. However, it is considerably less effective at very low latitudes, where the Coriolis parameter is small. Moreover, it can be shown that mean meridional circulations are incapable of producing any sizable westerly accelerations within a belt of latitude centered on the equator. For a mean equatorward motion to advect air with greater zonally averaged westerly angular momentum into an equatorial strip, there would have to be a substantial increase of westerly angular momentum with latitude. Such a zonal flow configuration would violate one of the conditions for the stability of a balanced zonally symmetric vortex. Thus it is not possible to account for the zonal momentum budget of the oscillation without relying on zonally asymmetric motions to provide a source of westerly momentum at the equator.

The need for a momentum source prompted Tucker [1964] to examine the transient eddy statistics for evidence of long-term fluctuations in the meridional flux divergence of zonal momentum in the tropical stratosphere. At the 25-km level he found evidence of fluctuations that seemed to be related to the quasi-biennial oscillation, the maximum flux divergence occurring simultaneously with
the maximum easterlies. Using a larger data sample from northern hemisphere stations, Wallace and Newell [1966] confirmed the existence of long-term variations in the poleward flux of angular momentum and showed that they occurred with simultaneous phase throughout the 20- to 30-km layer over a broad range of latitudes extending from about 25°N to 50°N. These variations resulted from year to year differences in the winter stratospheric circulation. During the period analyzed those winters with equatorial easterlies in the 25- to 30-km layer tended to have stronger eddy activity and larger poleward fluxes of westerly angular momentum than winters with westerlies in that layer. The period of record was not long enough to establish the statistical significance of this relationship, but the observations are supported by Dickinson's [1968a] theoretical analysis, which indicates that the distribution of zonal wind in the equatorial stratosphere should strongly modulate the planetary waves propagating upward from the troposphere in the subtropics during the winter season.

Given the existence of the type of forcing envisioned by Tucker and Wallace and Newell, it still remained to be demonstrated how an oscillation in momentum fluxes that was simultaneous at all levels could generate a downward propagating zonal wind oscillation. Tucker [1964] had previously suggested that the observed downward propagation was mainly a result of advection produced by a mean sinking motion throughout this region. Wallace [1967] supported the notion of downward advection by a mean sinking motion at the equator, but he recognized that uniform subsidence throughout the tropical stratosphere would have catastrophic effects on the steady state momentum budget, since continuity considerations would require the existence of prohibitively large mean meridional motions in the subtropics, which in turn would lead to very strong zonal winds in the long-term mean. It seemed more plausible that the downward propagation was due to a combination of advection near the equator and Coriolis torques away from the equator. Dickinson [1968b] has elaborated further on the possible role of Coriolis torques in this connection.

To test this hypothesis, Wallace and Holton [1968] developed a diagnostic numerical model of the quasi-biennial oscillation based on the equations of motion, the continuity equation, and the thermodynamic energy equation for a geostrophically balanced vortex. A time-varying distribution of momentum flux divergence was specified in accordance with the observations of Wallace and Newell [1966]. The fluctuations were assumed to be simultaneous at all latitudes and levels and to be confined to the region above 25 km. After performing a large number of experiments with various distributions of steady state mean meridional circulations, they concluded that it was not possible to simulate the observed downward propagation of the wind regimes. Coriolis torques proved to be much less effective at transporting momentum downward then had previously been supposed, and as a result the wind regimes attenuated much too rapidly as they propagated downward below 25 km. (Holton [1968] later demonstrated this more clearly by using a simplified model.) Through the use of the model it was also possible to demonstrate that the size of the fluctuations in heating rates that would be required to drive the quasi-biennial
oscillation would be on the order of 1.5°C day⁻¹, which is much larger than earlier estimates. This would require enormous variations in solar output, which should be easily detectable if they existed. On the basis of this evidence, Wallace and Holton concluded that the only way to account for the zonal wind changes in the tropical stratosphere is to invoke a source of zonal momentum that propagates downward. Within a year after the publication of their paper, observational studies of vertically propagating planetary wave disturbances provided evidence of just such a momentum source. I will defer the discussion of subsequent developments leading to a theory of the quasi-biennial oscillation until I have reviewed the structure and dynamics of equatorial stratospheric wave disturbances.

Before the conclusion of this section on the mean zonal flow the current status of the theory of the semiannual cycle should be mentioned. It has commonly been assumed that at the upper levels, where the cycle is largest at equatorial latitudes, the zonal wind fluctuations are a consequence of the fact that there is a semiannual cycle in incoming solar radiation in the equatorial region. This explanation seems satisfactory enough until one examines the balance requirements for angular momentum. Just as in the case of the quasi-biennial oscillation, there is the problem of explaining the rather abrupt appearance of westerly momentum at the equator without recourse to a momentum source related to zonally asymmetric motions. By the use of a diagnostic model similar to that of Wallace and Holton, Meyer [1970] has shown that the response of the equatorial wind field to the semiannual cycle in solar heating is very small. He concluded that a semiannual periodicity in momentum flux divergence is needed to account for the observed zonal wind changes. At the time of this writing the nature of this momentum source is not known. Meyer suggested that meridional fluxes produced by the diurnal tide may be responsible. Within the 40- to 60-km layer, both the diurnal and the semidiurnal tides are large enough to be of potential importance in the momentum budget, and either of them could conceivably contribute to a semiannual variation in momentum flux divergence at the equator.

In the lower stratosphere, where the semiannual cycle reaches its maximum amplitudes in extratropical latitudes, the underlying dynamics are apparently quite different. The middle-latitude amplitude maximum at these levels suggests that the half-yearly periodicity may be closely related to the occurrence of the midwinter breakdowns in the so-called 'polar night jet.' This phenomenon is discussed in detail by Matsuno [1970, 1971].

3. WAVE DISTURBANCES

a. Observational evidence. Almost all the early observational studies of winds in the tropical stratosphere were based on monthly mean data. The first systematic effort to analyze daily wind data over an extended period was reported on by Yanai and Maruyama [1966] and Maruyama and Yanai [1967]. Figure 7, reproduced from their second paper, shows a time-height section of the horizontal wind vector at stratospheric levels over Canton Island (3°S).
Layers of alternating northerly and southerly wind regimes can be observed to propagate downward. At any given level the meridional wind component exhibits systematic fluctuations with periods on the order of 4–5 days. A subsequent study by Maruyama [1967] showed that the fluctuations are produced by westward propagating wave disturbances, which are observed quite generally at stations in the central western Pacific within 10° of the equator during westerly regimes of the quasi-biennial oscillation. Maruyama's paper gave the first detailed account of the structure of these disturbances as manifested in the wind, geopotential height, and temperature fields.

The horizontal structure of the waves is shown in Figure 8b. The meridional wind component is seen to have even symmetry about the equator, whereas the zonal component and the geopotential height field have odd symmetry. A longitude-height cross section through the waves at a latitude north of the equator is shown in Figure 9. The vertical components of the vectors in the figure indicate the axes of the maximum vertical motion, the horizontal components indicate the axes of the maximum zonal motion, and the arrows directed in and out of the paper indicate the axes of the maximum meridional motions, southerly (northward) winds being directed into the paper. We note that the fluctuations in vertical velocity have odd symmetry with respect to the equator and are in phase with the geopotential height fluctuations. The maximum temperatures are observed concurrently with the maximum poleward wind component.

Wallace and Kousky [1968] reported on another type of wave disturbance that they identified while they were analyzing daily data taken during the
transition from an easterly regime to a westerly regime of the quasi-biennial oscillation in early 1966. During this period, time-height sections of zonal wind and temperature (Figures 10 and 11) showed evidence of systematic fluctuations with periods on the order of 15 days centered in the upper part of the easterly regime. There was no evidence of related fluctuations in the meridional wind.
component. The observed fluctuations were shown to be associated with the passage of waves, the structures of which are shown in Figures 8a and 12. In contrast to the waves described by Yanai and Maruyama these propagate eastward and exhibit even symmetry about the equator in the zonal wind component and in the geopotential height field. They resemble the former waves in the sense that they are downward propagating and they exhibit an in-phase relationship between vertical velocity and geopotential height.

Fig. 11. Time-height section of temperature at Balboa. Isotherms are placed at intervals of 2°C. Temperatures below $-80^\circ$C are shaded. Dashed lines represent axes of the more prominent easterly fluctuations shown in Figure 10.

Fig. 12. Idealized cross section along a latitude circle showing phases of the zonal wind, temperature, and pressure and vertical motion oscillations associated with Kelvin waves (adapted from Wallace and Kousky [1968]). Conventions are the same as those in Figure 9.
A brief summary of the properties of the two types of waves discussed in this section is given in Table 1. Some of this information has been gathered from time-height sections, such as those shown in Figures 7, 10, and 11. Cross-spectrum analysis techniques have also been used to provide objective estimates of wave amplitudes, periods, horizontal and vertical wavelengths, and the direction of propagation.

The Doppler-shifted phase speed is given by \( c - U \), where \( c \) is the wave speed relative to the ground (positive eastward) and \( U \) is the zonal wind speed at the level in question. The values for \( U \) used in Table 1 are typical of the levels at which the strongest waves are observed. The Doppler-shifted frequency is given by

\[
\omega' = k(c - U) = \frac{\omega(c - U)}{c}
\]

where \( k \) is the zonal wave number and \( \omega \) is the ground-based frequency.

In Table 1 the amplitudes of the geopotential height and vertical motion fluctuations within the waves were estimated indirectly. For the waves described by Wallace and Kousky the geopotential height amplitude is readily obtained by applying the zonal equation of motion at the equator to a sinusoidal zonally propagating wave. This operation yields

\[
z^* = \frac{\omega' Lu^*}{2\pi g}
\]

where \( z^* \) and \( u^* \) are the amplitudes of the geopotential height and zonal wind fluctuations, respectively; \( \omega' \) is the Doppler-shifted frequency of the waves; \( L \) is the zonal wavelength; and \( g \) is the acceleration of gravity. An analogous expression for the waves of Yanai and Maruyama can be derived from the meridional equation of motion. The amplitudes of vertical motion fluctuations

| TABLE 1. Description of Vertically Propagating Wave Modes in the Tropical Stratosphere |
|---------------------------------|-----------------|-----------------|
| Theoretical description        | mixed Rossby-gravity wave | Kelvin wave |
| Frequency \( \omega \) (ground based) | \( 2\pi/4-5 \) days | \( 2\pi/15 \) days |
| Horizontal wavelength \( L \)   | \( \sim 10,000 \) km | \( \sim 30,000 \) km |
| Zonal wave number \( k \)       | \( \sim 4 \) | 1-2 |
| Vertical wavelength \( D \)     | 4-8 km | 6-10 km |
| Structure                       | Figures 8b and 9 | Figures 8a and 12 |
| Average phase speed relative to ground | \( -23 \) m sec\(^{-1}\) | \( +25 \) m sec\(^{-1}\) |
| Average phase speed relative to zonal wind | \( -30 \) m sec\(^{-1}\)† | \( +50 \) m sec\(^{-1}\)‡ |
| Doppler-shifted frequency \( \omega' \) | \( 2\pi/3 \) days† | \( 2\pi/8 \) days‡ |
| Amplitudes                      |                 |                 |
| Zonal wind \( u^* \)           | \( 2-3 \) m sec\(^{-1}\) | \( \sim 8 \) m sec\(^{-1}\) |
| Meridional wind \( v^* \)      | \( 2-3 \) m sec\(^{-1}\) | 0 |
| Temperature \( T^* \)          | \( \sim 1^\circ \)C | \( 2^\circ-3^\circ \)C |
| Geopotential height \( z^* \)  | \( \sim 30 \) meters | \( \sim 4 \) meters |
| Vertical velocity \( w^* \)     | \( \sim 0.15 \) cm sec\(^{-1}\) | \( \sim 0.15 \) cm sec\(^{-1}\) |

† Near level of maximum westerly winds, where \( U \sim 7 \) m sec\(^{-1}\).
‡ Near level of maximum easterly winds, where \( U \sim -25 \) m sec\(^{-1}\).
for both waves can be estimated by applying the thermodynamic energy equation and neglecting the nonlinear terms. As a result,

\[ w^* = \omega^* \frac{T^*}{(\frac{\partial T}{\partial z} + c_p) / \Delta z} \]

where \( T_0 \) is the zonally averaged temperature, \( c_p \) is the specific heat, and \( z \) is the height. The term in the denominator is the mean static stability of the tropical stratosphere, which is on the order of 12°C km\(^{-1}\). For the waves described by Wallace and Kousky, \( w^* \) can also be estimated from the two-dimensional continuity equation in \( u \) and \( \psi \), since there is no meridional velocity. Thus

\[ w^* = u^* D/L \]

where \( D \) is the vertical wavelength.

b. Theory of equatorial wave disturbances. In the analysis and interpretation of the observations the investigators cited in the previous section relied heavily on insights drawn from the theory of equatorial wave disturbances. In this section I will review some of the theoretical results that are helpful in understanding the nature of the observed waves.

The existence of free barotropic modes of oscillation on an equatorial beta plane was first discussed by Rattray [1964] in an oceanographic context and later by Rosenthal [1965] and Matsuno [1966] with reference to the atmosphere. These modes are zonally propagating wavelike solutions of the linearized equations of motion in the horizontal plane. The Coriolis parameter is approximated by \( f = \beta y \), where \( \beta = 2\Omega/a \), \( \Omega \) is the angular velocity of the earth, and \( a \) is the radius of the earth.

The latitudinal structure of these modes can be described in terms of oscillatory functions that decay with increasing distance from the equator (Hermite functions). The order of the Hermite functions is designated by an integer \( n = 0, 1, 2, \ldots \) whose value indicates the number of nodes in the meridional velocity in the latitude domain. There is one additional mode designated by \( n = -1 \) for which the meridional velocity is identically zero.

For \( n \geq 1 \), there are three wave solutions corresponding to each value of \( n \): a westward propagating Rossby wave and a pair of inertia-gravity waves, one propagating eastward and one propagating westward. For the \( n = 0 \) mode, there are only two solutions, the eastward propagating gravity wave and a westward propagating 'mixed Rossby-gravity mode,' which behaves like a Rossby wave at a high zonal wave number and a gravity wave at a low zonal wave number. The horizontal structure of this mixed mode is depicted in Figure 8b. There is only one solution that corresponds to \( n = -1 \), the gravity wave that has come to be called a 'Kelvin wave' because of its resemblance to the oceanic Kelvin wave. (A shallow water gravity wave that propagates along a coastal boundary and has no velocity component normal to the coastline. The velocity component parallel to the coastline is in geostrophic balance with the pressure field. Both velocity and pressure decay exponentially with increasing distance from the coastline.) Here, the equator (where the Coriolis parameter changes sign) plays the same role as the coastal boundary in the oceanic case, and the
zonal velocity is in geostrophic equilibrium with the pressure field. For the amplitude of the wave to decay with increasing distance from the equator the Kelvin wave must propagate eastward relative to the mean zonal flow. (The eastward propagation is easily verified by considering the equation of motion for the zonal wind component at the equator together with the (geostrophic) requirement that maximum westerlies coincide with highest pressure.) The horizontal structure of the equatorial Kelvin wave is shown in Figure 8a.

Each of the equatorial beta plane solutions corresponds to the asymptotic form of a solution for the modes of oscillation of a barotropic ocean of uniform depth on a rotating sphere as the parameter $\epsilon = 4 \Omega^2 a^2 / gh$ becomes very large [Longuet-Higgins, 1968]. Here, $\Omega$ is the angular frequency of rotation, $a$ is the radius of the sphere, $g$ is the acceleration of gravity, and $h$ is the depth of the ocean. The Kelvin wave and mixed Rossby-gravity wave correspond, respectively, to the gravest eastward and westward propagating modes in the solutions of Longuet-Higgins.

Barotropic wave theory offers only a partial understanding of the nature of equatorial stratospheric waves because they are not barotropic but vertically propagating. For a more complete understanding it will be necessary to consider the results of a theoretical investigation by Lindzen [1967], which have a bearing on the vertical structure of equatorial waves.

Lindzen solved the linearized equations of motion, continuity, and thermodynamic energy on an equatorial beta plane for zonally propagating waves in a stably stratified medium. For a mean zonal current that is independent of latitude and height he was able to express each of the solutions as the product of a vertical structure function, which can be obtained by solving the vertical structure equation of classical tidal theory, times one of the latitudinal structure functions discussed previously in connection with the barotropic modes. The solutions contain an eigenvalue or constant of separation called the equivalent depth, $h_{n,k,\omega'}$, where $n$ is again the order of the Hermite function that describes the latitudinal structure, $k$ is the zonal wave number, and $\omega'$ is the Doppler-shifted frequency.

Lindzen showed that in the absence of a rigid upper boundary the only possible free modes of oscillation are the barotropic modes discussed previously. For these the appropriate equivalent depth is on the order of $10$ km, and the corresponding periods are on the order of $\leq 1$ day for the gravity waves (including the Kelvin wave) and $< 2$ days for the mixed Rossby-gravity wave. These periods are sufficiently short to preclude the possibility of identifying either of the observed waves in the equatorial stratosphere with the free modes of oscillation. The free Rossby waves for $n \geq 1$ have periods of a few days or longer. According to the analysis of Longuet-Higgins these should extend well into the middle latitudes. Madden and Julian [1972] have shown evidence of the existence of a global scale oscillation in sea level pressure that may correspond to the $n = 1$ free barotropic Rossby mode.

Lindzen showed that in addition to the free barotropic modes there is also the possibility of forced modes that can exhibit either a sinusoidal or an exponential height dependence depending on the frequency of excitation. The vertically
propagating modes are capable of transferring energy in the vertical in contrast to the exponential modes, which are 'trapped' near the level of excitation. Vertical propagability is associated with the range of small positive equivalent depths. For any given mode and zonal wave number, there exists a critical frequency below which disturbances can propagate vertically.

Any combination of latitudinal mode \( n \), wave number \( k \), and Doppler-shifted frequency \( \omega' \) uniquely determines the equivalent depth, which in turn determines (1) whether the waves will be trapped or vertically propagating, (2) the vertical wavelength or \( e \)-folding depth, and (3) the latitudinal half width about the equator.

A number of consistency checks can be performed by using these relationships to verify the identity of the observed waves. Maruyama [1967] and Lindzen and Matsuno [1968] have used these three relationships as consistency checks to verify the identification of the waves observed by Yanai and Maruyama with the mixed Rossby-gravity mode. The corresponding evidence for the identification of the waves observed by Wallace and Kousky with the equatorial Kelvin wave can be found in Holton and Lindzen [1968].

The \( n = 0 \) inertia-gravity mode and the modes for which \( n \geq 1 \) have not been identified observationally. For these modes the theory described above gives some indication of the range of vertical wavelengths and latitudinal half widths that would be consistent with planetary scale waves with periods of a few days or longer. For all the inertia-gravity modes the vertical wavelengths are on the order of \( \sim 1 \) km or less, and the latitudinal half widths are generally \( <10^\circ \) and tend to decrease with increasing \( n \). These modes might possibly account for some of the fine structure observed in the high vertical resolution data from the Line Island experiment, a sample of which is shown in Figure 13 (adapted after Madden and Zipser [1970]). A comprehensive program of special high vertical resolution soundings would be required to identify the individual modes that are responsible for this fine structure.

For the \( n > 1 \) Rossby modes the situation is much different. Holton [1970] has shown that these have very long vertical wavelengths (on the order of 100 km). It can also be shown that these modes have latitudinal half widths that range up to 30\(^\circ\) or more, which means that they are not adequately described by the equatorial beta plane solutions.

There is one case in which there appears to be a direct conflict between the observations and the theory of equatorial wave disturbances. This involves results of Kousky and Wallace [1971], who noted the existence of zonal wind and temperature fluctuations with \( \sim 15 \)-day periods in the westerly regime above the transition zone during July–October 1963. These fluctuations appeared to be produced by waves with properties similar to those of the Kelvin waves observed in the easterly wind regime at the lower levels. However, the vertical wavelength was observed to be considerably longer than that of the lower waves, whereas wave theory predicts that Kelvin waves should have a much shorter wavelength in the westerly wind regime, where the Doppler-shifted frequency is much lower than it is in the easterlies, given the same ground-based frequency. Kousky and Wallace further showed that these upper waves could not possibly
Fig. 13. Time-height section (adapted from Madden and Zipser (1970)) of the meridional wind component at Palmyra in meters per second. The time of individual radiosonde releases is indicated by the arrow along the top of the section. Southward winds are shaded.
satisfy the continuity equation with \( v \) identically equal to zero. The long vertical wavelength of these disturbances and the existence of a small meridional velocity component could be explained if one identified them with the \( n = 1 \) Rossby mode [Matsuno, 1966], but this interpretation would be inconsistent with the observations, which indicate that the waves propagate eastward and are narrowly confined about the equator.

With this one possible exception the observations and theory of wave disturbances in the equatorial stratosphere seem to be in mutual agreement. There is strong observational evidence that the two gravest modes predicted by the linear wave theory (i.e., the Kelvin wave and the mixed Rossby-gravity wave) are the dominant types of wave motion in this region. The structure of these disturbances is consistent with an interpretation in terms of forced modes that propagate energy vertically away from the level of excitation. Both types of waves exhibit an in-phase relation between vertical velocity and geopotential height, which is indicative of an upward flux of wave energy from a source of excitation in the troposphere.

c. Possible sources of excitation. The possibility of an in situ energy source for wave disturbances in the tropical stratosphere (i.e., barotropic or baroclinic instability) would appear to be ruled out on the basis of the observed wave structure. Neither the Kelvin wave (which has no meridional wind component) nor the mixed Rossby-gravity wave (which exhibits a quadrature relationship between the zonal and meridional wind components) is capable of transporting zonal momentum in the meridional direction, as would be expected for barotropically unstable waves. In both types of waves the fluctuations between vertical motion and temperature appear to be in quadrature in contrast to the in-phase relationship, which would be expected for waves that draw their energy from baroclinic instability.

By a similar line of reasoning it can be argued that the structure of waves is not compatible with a primary source of energy in the middle-latitude stratosphere. An equatorward flux of wave energy would require an out of phase relationship between the fluctuations in geopotential height and the meridional wind component that is not observed.

Thus it appears that the only mechanism that is capable of providing a continuous source of excitation for the equatorial waves is the upward flux of wave energy from below, which is a result of a positive correlation between vertical motion and geopotential height. This view of the energetics is in full accord with the observed structure of the waves.

The upward flux of wave energy through the tropopause per unit area is related to the covariance between vertical velocity and geopotential height. Noting that these fluctuations are in phase, we can express the flux as

\[
F = \frac{1}{2} \rho_0 g w^* z^* \tag{5}
\]

Using the values in Table 1 for \( w^* \) and \( z^* \) and a mean density \( \rho_0 = 0.18 \text{ kg m}^{-3} \) for the tropopause level, we obtain fluxes of \( \sim 0.04 \) and \( \sim 0.005 \text{ w m}^{-2} \) for the Kelvin wave and mixed Rossby-gravity wave. The latter value is in close agreement with that obtained by Yanai and Hayashi [1969] on the basis of direct
calculations. These fluxes are small in comparison with the typical rates of generation of kinetic energy in the tropical troposphere, which range from \( \sim 0.2 \text{ w m}^{-2} \) in synoptic scale disturbances [Wallace, 1971] to \( >1 \text{ w m}^{-2} \) in the monsoons. However, they are large enough to be of considerable importance in the energetics of the stratosphere. For example, for the Kelvin wave, if the stratosphere were initially at rest and forcing at the tropopause were 'turned on' at the rate of 0.04 w m\(^{-2}\), as is indicated above, within the period of one wave passage sufficient energy would be transferred upward to account for the observed wave amplitude (~8 m sec\(^{-1}\)) throughout the layer extending up to one wavelength (~8 km) above the tropopause. This finding is consistent with Lindzen's theoretical analysis [Lindzen, 1970, equations 50–52], which shows that for the Kelvin wave the upward component of the group velocity is exactly equal to the downward component of the phase velocity. (From Lindzen’s equations 49 and 51 for the mixed Rossby-gravity wave it can be shown that the upward group velocity is about \( \frac{1}{4} \) as large as the downward phase velocity. Thus it takes about three wave passages for these waves to replenish their kinetic energy through a depth of one vertical wavelength by means of the vertical flux of wave energy. This finding is also consistent with the values in Table 1.)

Further evidence of a tropospheric energy source was provided by observational studies of Yanai and Hayashi [1969], Nitta [1970], and Yanai and Murakami [1970], which showed that the mixed Rossby-gravity wave in the lower stratosphere is coherent with wind fluctuations at tropospheric levels.

There is also evidence of spectral peaks in tropospheric wind fluctuations with periods similar to those in the stratosphere. The existence of a 4- to 5-day peak in spectra of the meridional wind component is documented in numerous studies, and there is also mention of a 10- to 15-day peak in the zonal wind spectra in studies of Wallace and Chang [1969] and Yanai and Murakami [1970]. These observations together with the observed coherence between tropospheric and stratospheric wind fluctuations might be taken as evidence that the Kelvin wave and mixed Rossby-gravity wave exist as identifiable wave structures in the troposphere as well as in the stratosphere, so that the stratospheric waves are simply the upward extension of the tropospheric ones. However, there are indications that the relation between stratospheric and tropospheric disturbances may not be this simple.

Previously [Wallace, 1971, pp. 603–604], I noted that the 4- to 5-day waves appear to be associated with two types of disturbances: a westward propagating synoptic scale wave and a planetary scale wave that appears to be related to the mixed Rossby-gravity wave in the stratosphere. The cloud patterns associated with the planetary scale wave appear to be confined to the western Pacific region, where they assume the form of a standing wave oscillation in the longitude domain, with no evidence of westward propagation. The oscillation appears to have odd symmetry about the equator. Holton has subsequently suggested that this 'standing wave oscillation' might be produced by the interaction between the westward propagating synoptic scale disturbances mentioned previously and the prominent features in the steady state monsoon circulation. Wallace [1972] has presented observational evidence in support of this view.
Thus it appears that the mixed Rossby-gravity wave in the lower stratosphere may be a response to tropospheric forcing with the same frequency but with a quite different horizontal structure.

Holton [1972] has attempted to simulate this response in a linearized, spectral primitive equation model. The forcing is specified in terms of a diabatic heat source that represents the release of latent heat in the observed tropospheric waves. The spatial distribution of the heat source is modeled in accordance with observations of the standing wave oscillation in tropical cloudiness in the western Pacific as described by Wallace [1971]. The localized standing wave forcing function is expanded in terms of a Fourier series of eastward and westward propagating waves, and the response to each wave component is calculated separately. Because of the rotation of the earth the stratospheric response to an oscillating heat source that is antisymmetric about the equator is largest for the westward propagating waves that assume the form of the mixed Rossby-gravity mode with a structure very similar to the observed one. Despite the fact that the largest forcing is in wave number 1, wave numbers 2–4 show the largest response in agreement with observations. Energy is dispersed longitudinally away from the localized source region, so that at the higher levels the waves are observed at all longitudes in the model. The stratospheric response was much stronger for a mean zonal wind profile with westerlies in the lower stratosphere than for one with easterlies. This finding is consistent with the fact that these waves have been observed only when mean westerlies prevail in the lower stratosphere.

The situation with regard to the Kelvin wave is less clear because of the lack of definitive observations in the troposphere. There is little question that there are substantial low-frequency large-scale zonal wind fluctuations at equatorial stations. However, these appear to be coherent with fluctuations in the meridional wind component [Wallace and Chang, 1969], which would preclude the possibility of their being a manifestation of Kelvin waves. It seems more likely that they are associated with slow changes in the monsoon patterns than with any identifiable zonally propagating wave structure.

Holton [1972, 1973] found that the stratospheric response to a standing wave tropospheric heat source with even symmetry about the equator is mainly in the form of eastward propagating Kelvin waves. The amplitude of the response is a function of the vertical profile of mean zonal wind and the frequency and wave number of the forcing. The main results can be summarized as follows.

1. For a given vertical profile of mean zonal wind and forcing with a given wave number, there is significant Kelvin wave response in the lower stratosphere only for a limited range of frequencies of forcing. In this sense the atmosphere behaves as a band-pass filter.

2. For a given vertical wind profile the amplitude of the stratospheric response appears to be independent of the zonal wave number of the forcing, whereas the frequency of the maximum response is directly proportional to the wave number. In other words the frequency response appears to be a function of the phase speed $\omega / k$ of the forcing.
3. For a vertical profile with mean easterlies in the layer below 25 km and westerlies above, there is a strong response to wave number 1 forcing at periods near 15 days in the westerly shear zone near 25 km. For a profile with westerlies below 25 km the maximum response to wave number 1 forcing is considerably smaller than that in the former case and is shifted toward higher frequencies.

Holton interprets these results as indicating that the stratospheric response to tropospheric forcing is large only when the vertical wavelength of the forced waves is compatible with the vertical structure of the forcing. The vertical wavelength of the forced waves is determined by their equivalent depth, which for the Kelvin wave is a unique function of the Doppler-shifted phase speed $\omega' / k$. The maximum response occurs when the Doppler-shifted phase speed of the forcing is on the order of 20–40 m sec$^{-1}$. This is consistent with a vertical wavelength of 6–10 km for the forced waves, which corresponds closely to the vertical scale of the assumed distribution of condensation heating in the troposphere.

Holton's calculations also provide some insight into the question of why the predominant wavelength of the Kelvin waves in the lower stratosphere is zonal wave number 1. Only at this wave number does his model show significant stratospheric response to forcing with periods >10 days. Therefore, if large-scale heating in the troposphere exhibits a 'red noise' spectrum in the time domain, as is indicated in some of the spectral studies [e.g., Wallace and Chang, 1972], the resulting stratospheric response should exhibit a red noise spectrum in the space domain. Holton's results also offer a possible explanation of the observed spectral peak around periods of 15 days, since this is the lowest frequency at which a strong Kelvin wave response is possible for realistic vertical profiles of diabatic heating in the troposphere. Furthermore, if there is any tendency for a dominance of the lowest wave numbers in the tropospheric forcing, such as might result from pulsations in the intensity of the monsoon circulations, it should reenforce the dominance of wave number 1 and the 15-day period in the stratospheric Kelvin waves.

Using a similar model with a somewhat different distribution of tropospheric heat sources, Murakami [1972] was also able to obtain a quite realistic simulation of the Kelvin wave in the lower stratosphere. The heat source was assumed to have the form of an eastward propagating wave with maximum amplitude at 20° latitude. Thus it appears that Kelvin waves can be generated in response to tropospheric forcing with a wide variety of latitudinal distributions. The only crucial requirement is that there be a component that has even symmetry about the equator.

Thus the theory of equatorial wave disturbances provides a rational explanation for many of the observed features of wave disturbances in the tropical lower stratosphere. Whatever the actual distribution of forcing in the troposphere it can be viewed as a linear combination of the 'normal mode' solutions to the equations of motion on an equatorial beta plane. The Kelvin wave and mixed Rossby-gravity wave are the only normal mode solutions for which the forced waves have vertical wavelengths that are long enough to be comparable to the
vertical scale of the profile of diabatic heating (or vertical motion) in the troposphere. Therefore it is these modes that are excited most strongly. The dominance of wave number 1 Kelvin waves with a 15-day period appears to be a consequence of tropospheric forcing with a red noise spectrum in either the time or the space domains or possibly both. The dominance of wave numbers 3–5 and the 4- to 5-day period in the mixed Rossby-gravity wave appears to be a result of a somewhat more specific distribution of tropospheric forcing. There is observational evidence that suggests that the interaction between the transient disturbances along the intertropical convergence zone and the steady state monsoon flow yields a distribution of forcing with the appropriate space and time scales to excite mixed Rossby-gravity waves with properties similar to the observed. The range of zonal wind speeds in the lower stratosphere is such that the response to a given forcing increases with the Doppler-shifted phase speed $c - U$. This explains why Kelvin waves are observed when the mean zonal wind in the lower stratosphere is easterly, whereas mixed Rossby-gravity waves are observed during periods of westerlies.

4. INTERACTION BETWEEN THE WAVES AND THE MEAN ZONAL FLOW

The vertically propagating waves discussed in the previous section produce a substantial vertical flux of zonal momentum. Under certain conditions, which depend on the distribution of mean zonal wind, this momentum can be given up to the mean zonal flow. The absorption of momentum by the zonal flow produces a net easterly or westerly acceleration, which alters the original mean zonal wind distribution. The purpose of this section is to discuss the mechanisms whereby the waves interact with the mean zonal flow and to show how these interactions produce the observed downward propagation of shear zones.

From Figures 9 and 12 it is evident that the vertical structures of the Kelvin wave and the mixed Rossby-gravity wave are both characterized by an in-phase relation between fluctuations in zonal wind and vertical velocity. Thus both waves produce an upward flux of relative westerly momentum that is given by $\rho u^* w^*/2$.

Maruyama [1968] developed a spectral technique for evaluating this term on the basis of time series data for zonal wind and temperature. For the mixed Rossby-gravity wave he computed a value on the order of $2 \times 10^{-3}$ m$^2$ sec$^{-2}$. Using the same technique, Kousky and Wallace obtained values of up to $\sim 6 \times 10^{-3}$ m$^2$ sec$^{-2}$ during periods of strong Kelvin wave activity.

For the mixed Rossby-gravity wave, there is an additional effect associated with a mean meridional circulation driven by the wave. That such a circulation exists can be seen from a careful inspection of Figures 8b and 12. At any given level, air parcels experience their maximum equatorward motion ¾ wave-length after their maximum upward motion. Thus the air trajectories within the waves describe ellipses in the meridional plane, the rising motion taking place somewhat further from the equator than the corresponding sinking motion. Because of this difference in latitude the rising parcels carry upward with them
smaller amounts of westerly angular momentum associated with the earth's rotation than the sinking parcels carry downward. Thus the overall effect of this meridional circulation is to produce a net upward transport of easterly angular momentum. This predominates over the effect mentioned previously, so that the total effect of the mixed Rossby-gravity wave is to produce an upward flux of easterly momentum [Lindzen, 1971]. By using (75-78) in Lindzen's paper together with reasonable estimates of the zonal wavelength and the Doppler-shifted phase speed, we can compute the ratio of the net upward flux of easterly momentum to the observed upward flux of relative westerly momentum. This ratio is quite sensitive to the assumed value of \( c - U \). For \( c - U \) near \(-30 \text{ m sec}^{-1}\), as given in Table 1, the corresponding ratio is on the order of 1.5:1, and for \( c - U = -37 \text{ m sec}^{-1}\) it is about 3:1. Applying these ratios to the previously cited observational results of Maruyama [1968], we are led to conclude that the net upward flux of easterly momentum by the mixed Rossby-gravity wave is on the order of \( 3 \times 10^{-3} \text{ m}^2 \text{ sec}^{-2} \), but it could conceivably be twice this value in regions of strong westerlies.

To determine how this flux of momentum is imparted to the mean zonal flow, it will be helpful to consider the behavior of waves with a fixed ground-based frequency as they propagate through a layer in which there is vertical shear of the mean zonal wind. Under these conditions the structure of the waves changes with height as they adapt to the local Doppler-shifted frequency, which is proportional to the difference between the zonal propagation speed of the waves and the mean zonal wind speed at the level in question.

In the previous section it was shown that for observed ranges of ground-based frequencies and mean zonal winds large values of \( |c - U| \) are required in the lower stratosphere for there to be a strong response to tropospheric forcing. Thus the situation in which \( |c - U| \) is large in the lower stratosphere and decreases with height is of special interest in this regard. For the Kelvin wave, this corresponds to a mean zonal wind profile with lower stratospheric easterlies surmounted by a westerly shear zone, as shown in Figure 10. For the mixed Rossby-gravity wave, just the opposite configuration is required. Let us first consider the Kelvin wave.

a. Kelvin wave. Eliassen and Palm [1960] studied the vertical propagation of gravity waves in a mean flow with vertical wind shear. Their formulation can be applied directly to the Kelvin wave to show that in the absence of damping the latitudinal integral of the vertical flux of zonal momentum is invariant with height provided that \( U \neq c \). The corresponding vertical flux of wave energy is directly proportional to \( |c - U| \). An immediate consequence of this relation is that wave energy cannot propagate through the 'critical level' at which \( U = c \). Booker and Bretherton [1967] showed that at such a level there is an abrupt transfer of momentum from the waves to the mean zonal flow.

Observations indicate that the Kelvin wave is apparently absorbed by the mean zonal flow without encountering a critical level. For a wave number 1 Kelvin wave with a 15-day period the ground-based phase speed is \( \sim 30 \text{ m sec}^{-1} \). Mean westerly winds rarely if ever reach this strength in the lower stratosphere. Studies by Maruyama [1969] and Kousky and Wallace [1971] indicate that
the amplitude of the Kelvin waves decreases markedly with height near the base of the upper-level westerlies, where \( U \sim 0 \).

There are two possible processes that could prevent the Kelvin waves from penetrating very far into the upper-level westerlies. First, as the Kelvin waves pass upward from the easterly wind regime in the lower stratosphere into the zone of strong westerly wind shear, their structure begins to change. To maintain a constant upward flux of westerly momentum and at the same time decrease the flux of wave energy in proportion to \(|c - U|\), it is necessary for \( u^* \) to increase while \( w^* \) decreases. From (4) it can be seen that a larger ratio of \( u^* / w^* \) implies a more horizontal inclination of the wave axes. According to Lindzen [1971],

\[
\begin{align*}
  u^* &\propto |c - U|^{-3/4} \\
  D &\propto |c - U|
\end{align*}
\]  

Both these effects serve to increase the vertical wind shear in the waves, which is proportional to the ratio \( u^* / D \). For sufficiently large values of vertical wind shear the Richardson number

\[
Ri = \frac{g}{T_0} \frac{\rho \frac{\partial T}{\partial z}}{|\partial V/\partial z|} \]

will approach \( \frac{1}{4} \), which is the critical value for the onset of shear instability. (Here, \( V \) is the horizontal wind vector.)

By the combination of (6), (7), and (8), it can be shown that to a close approximation \( Ri \propto |c - U|^3 \). Thus the modest decrease in \( c - U \) that the waves encounter as they propagate from the easterly wind regime up to the level of zero zonal wind produces nearly an order of magnitude decrease in the Richardson number in the waves. The decrease in density with height should also contribute to increasing the zonal wind amplitude and lowering the Richardson number.

Even with all these effects tending to lower the Richardson number, quite low values of \(|c - U|\) (i.e., rather strong westerlies) would be required to bring it to the critical value were it not for the additional effect of the strong vertical shear of the mean zonal wind in the transition zone. Kousky [1970] has shown that the combination of the shear of the mean zonal wind and the shear in the waves is sufficient to produce Richardson numbers near the critical value at levels near the transition from low-level easterlies to upper-level westerlies. It appears likely that in the presence of this strong shear the waves break down into smaller-scale Kelvin-Helmholtz waves and turbulence within this transition zone.

Second, the upward component of the group velocity of the Kelvin wave is proportional to \((c - U)^2\) [Lindzen, 1970]. Therefore any waves that are able to penetrate into the upper-level westerlies do so at a substantially reduced group velocity that renders them much more susceptible to dissipative processes [Lindzen, 1970, 1971]. As evidence of the importance of these processes it is worth noting that above 25 km, where \( U > 0 \), the characteristic time scale for radiative relaxation is comparable to the Doppler-shifted period of the Kelvin
wave, which (as was shown previously) is the time required for wave energy to propagate upward through one vertical wavelength.

b. Mixed Rossby-gravity wave. The two mechanisms described above whereby Kelvin waves are absorbed by the mean zonal flow in the absence of a critical level are also operative for the mixed Rossby-gravity wave with minor modifications. The shear instability mechanism may be relatively less important for the Rossby-gravity wave than for the Kelvin wave because the wave amplitudes are smaller and the intensity of the easterly shear zones is usually not as large as that of the westerly ones, as is seen in Figures 3-5. On the other hand it appears that the mixed Rossby-gravity wave is particularly susceptible to radiative damping because the vertical component of the group velocity is smaller than that of the Kelvin wave and decreases in proportion to \((c - U)^4\) as the waves approach their critical level [Lindzen, 1970].

c. Effect of wave absorption on the mean zonal flow. Regardless of the absorption mechanism, any layer in which vertically propagating waves are absorbed by the mean zonal flow experiences a convergence of the vertical flux of westerly (or easterly) momentum, since a larger flux is entering the layer from below than is exiting through the top. This produces an acceleration of the mean zonal wind, which is westerly for the Kelvin wave and easterly for the mixed Rossby-gravity wave.

Given the existence of a mean zonal wind profile having a shear zone of either sign, the zonal momentum carried aloft by the waves should tend to be deposited near the transition between easterlies and westerlies. In either case the resulting acceleration is in the direction of the wind above the shear zone (i.e., westerly for the Kelvin wave and easterly for the mixed Rossby-gravity wave). Thus the absorption of the waves acts to produce a downward displacement of the shear zone. This in turn causes the waves to be absorbed at lower levels, so that the shear zone is lowered further. The process continues until the shear zone approaches the tropopause level. The rate of descent is governed by the intensity of the waves.

5. MECHANISM FOR THE QUASI-BIENNIAL OSCILLATION

By a rather straightforward extension of the arguments of the previous section it is possible to arrive at a plausible explanation for the quasi-biennial oscillation. The mechanism that I will be describing was first proposed by Lindzen and Holton [1968] and was revised by Holton and Lindzen [1972].

Let us begin by assuming a mean zonal wind profile with easterlies at all levels. This will allow Kelvin waves to propagate vertically with only weak absorption due to radiative damping, which will be most pronounced above 25 km, where the characteristic time scale for radiative damping is relatively short. Given a sufficiently long time, the absorption of Kelvin waves at the higher levels would probably give rise to a westerly wind regime there. In any event the semiannual oscillation at levels above 35 km will provide an upper-level westerly shear zone within a few months if one does not already exist. Once the shear zone is established, it will begin to propagate downward as a
result of the enhanced absorption of Kelvin waves that takes place within it. Kousky and Wallace [1971] have shown that the rate of convergence of the flux of westerly momentum into descending westerly shear zones is consistent with the accelerations of the mean zonal wind. Given the observed level of Kelvin wave activity, it should take the shear zone 8–12 months to propagate downward from the 30-km level to near the tropopause.

When the descending westerly wind regime reaches the tropopause level, the mixed Rossby-gravity waves, which were previously absorbed by the lower stratospheric easterlies, become free to propagate energy upward until they encounter an upper-level easterly shear zone.

Upper-level easterlies can arise from either of the mechanisms mentioned previously or from the advection of air with lower angular momentum by mean meridional motions. Once an upper-level easterly regime is established, the downward propagation proceeds as mentioned previously until the cycle is completed.

The period of the cycle in Holton and Lindzen's [1972] model can be altered by changing the intensity of the momentum sources, which are specified as external parameters. The range of observed values for the momentum fluxes is consistent with an oscillation having an average period of 2 years or a little longer. The wide range of observed periods for individual wind regimes indicates that the intensity of the momentum sources probably varies somewhat from cycle to cycle.

There is strong observational support for all the crucial requirements of Lindzen and Holton's mechanism: (1) upward fluxes of zonal momentum of the required magnitude, (2) absorption of the waves and convergence of zonal momentum in the shear zones, and (3) a semiannual oscillation that provides alternating easterly and westerly shear zones at the 35-km level. There is still some uncertainty regarding some of the details, e.g., the relative importance of the various absorption mechanisms and the roles of mean meridional motions and meridional fluxes of zonal momentum. There is also some question as to which dynamical processes are responsible for the attenuation of the oscillation as the wind regimes approach the tropopause level. However, despite these remaining uncertainties, there seems to be little doubt that a mechanism very similar to the one described by Lindzen and Holton must be responsible for the major features of the quasi-biennial oscillation.

6. SUMMARY AND OUTLOOK FOR FUTURE RESEARCH

The properties of the mean zonal flow and of the wave disturbances in the tropical stratosphere are now well documented observationally. With the one exception pointed out in section 3b, there is excellent agreement between observations and theory with respect to the structure of the wave disturbances. There is strong observational support for a theory that views quasi-biennial oscillation as a consequence of the interactions between the waves and the mean zonal flow.

Thus it appears that largely within the past 5 years the major problems
concerning the dynamics of the tropical lower stratosphere have been solved at least in principle. Final verification and refinement of existing theoretical ideas will probably have to await more comprehensive modeling studies of the entire tropical general circulation. Work of this type is already in progress at the Geophysical Fluid Dynamics Laboratory of the National Oceanic and Atmospheric Administration and at the National Center for Atmospheric Research. The present high resolution models with about three levels in the lower stratosphere are capable of simulating some of the observed features of the Kelvin wave and the mixed Rossby-gravity wave. Modeling of this type will be especially helpful in clarifying the nature of the relation between tropospheric and stratospheric disturbances. It should also be possible to simulate Lindzen and Holton's mechanism for the quasi-biennial oscillation in a general circulation model with higher vertical resolution in the stratosphere.

It is likely that the upper stratosphere will receive increased emphasis in future observational studies when better-quality rocketsonde data become available. At the present time, very little is known about the wave disturbances in this region because previous rocketsonde firings have been too widely spaced in time to provide more than climatological information. A well-coordinated observing program with daily soundings at a few well-located tropical stations could do much toward eliminating this deficiency.

There is also strong justification for a special program of observations to study the effect of wave disturbances on the vertical profile of ozone in the lower stratosphere. The vertical velocities in the waves are large enough to produce ozone fluctuations that are well within the measurable range. The phase relationship between ozone and temperature in the waves should provide valuable information on the radiative and photochemical time scales in this region as a function of height. What is needed to document the ozone fluctuations is an extended period of daily high vertical resolution measurements of ozone, temperature, and wind at one or more stations within about 8° of the equator.

The question of possible relations between the quasi-biennial oscillation and the circulation in other regions of the atmosphere is also in need of further clarification. Dickinson [1968a] has shown that the zonal flow configuration in the tropics influences the leakage of wave energy from the stationary planetary scale disturbances in the winter stratosphere. On this basis he hypothesized that the quasi-biennial oscillation should produce a significant modulation of the upward flux of wave energy from the troposphere over a broad range of latitudes. Conceivably, this could influence the character of the stratospheric circulation throughout the winter hemisphere. Independently, it has been suggested on the basis of observational studies [e.g., Labitzke, 1966; Barbé and Reininger, 1966] that there may indeed be some relationship between the phase of the quasi-biennial oscillation and the occurrence of sudden warmings in the winter stratosphere. There is also some evidence that ozone amounts in the middle-latitude stratosphere show a quasi-biennial periodicity [Funk and Garnham, 1962; Ramanathan, 1963]. With almost 20 years of stratospheric wind data now available it should be possible to establish fairly conclusively whether these alleged relationships are statistically significant. Should this prove to be true,
there will be a need for further theoretical investigations of the coupling between the zonal flow in the tropics and the circulation at higher latitudes.

There is also some evidence of an apparent quasi-biennial periodicity in long records of surface data [e.g., Landsberg, 1962]. The prospects of relating this to the equatorial stratospheric oscillation appear to be somewhat more remote, since the phenomenon is only weakly evident in the tropospheric time series that have been analyzed thus far. However, there is some possibility that further studies with refined analysis techniques will be able to define and isolate more clearly the tropospheric oscillation, so that it can be related to year to year changes in the stratospheric circulation.

Acknowledgments. I wish to thank Drs. J. R. Holton, R. E. Dickinson, and R. J. Reed for their helpful suggestions. The work was supported by the National Science Foundation under grant GA-32439. This paper is contribution 274 of the Department of Atmospheric Sciences, University of Washington, Seattle, Washington.

REFERENCES


(Received September 29, 1972; revised December 29, 1972.)