ADVANCED LECTURES 1987

THE MIDDLE ATMOSPHERE

1 OBSERVED STRUCTURE OF THE MIDDLE ATMOSPHERE: CLIMATOLOGY AND SEASONAL EVOLUTION

1.1 INTRODUCTION

The section of the atmosphere above the tropopause and below a height of about 85 km is usually referred to as the middle atmosphere. In the lower part, known as the stratosphere, the temperature generally increases with height while above, in the mesosphere, it decreases with height (Fig. 1.1). Although the region contains only about one sixth of the mass of the earth's atmosphere, it attracts much attention from meteorologists and atmospheric chemists. This interest is stimulated, in part, by concern about the ozone layer in the stratosphere, a possibly fragile layer of trace gas which shields the earth's surface from the harmful effects of the sun's ultraviolet radiation.

The first evidence that atmospheric temperature did not decrease monotonically with height from the ground to cold space was obtained in 1902 by the French meteorologist, Teisserenc de Bort. He used unmanned balloons, a measurement technique that is still in use today. Further progress came with two companion inventions: the rocket and telemetry. In conjunction with radiosonde ascents into the lower stratosphere, a three-dimensional picture of temperature and wind structure in the middle atmosphere began to be built up in the 1950s.

A limitation of such data is that their spatial coverage is not uniform. Measurements are sparse over the oceans, and so the southern hemisphere is especially poorly covered. In the last 15 years or so, measurements from satellites have transformed the situation. It is now possible to map the temperature distribution with good coverage in three

dimensions, and to follow its changes day by day. Wind fields can be obtained approximately from temperature fields (away from the equator), affording a whole range of studies of the dynamics of the middle atmosphere. A major advantage of using satellite measurements is that, being obtained from a single instrument, they have uniform quality. So there is no need to cross calibrate different types of *in situ* measurements.

Useful supplementary information, particularly on atmospheric motions with small spatial or temporal scales, comes from two other measuring techniques: radar and lidar. The MST (Mesosphere-Stratosphere-Troposphere) radar obtains echoes from atmospheric irregularities and fluctuations in electron density (the latter only in the mesosphere). The Doppler shift of the echoes gives the line-of-sight velocity. The structures of tides and gravity waves in the middle atmosphere can be obtained from radar data. Lidars provide complementary information to that from radars. They use Rayleigh scattering from atmospheric molecules to measure density and temperature.

Observational data for the middle atmosphere are being augmented by results of simulations with numerical models. A number of general circulation models with levels in the middle atmosphere have been integrated through a year or more with seasonal variations in the thermal forcing (see lecture 10). A likely future development is the assimilation of observations by a numerical model of the middle atmosphere (as is done at present in operational models used for weather forecasting), with the aim of improving the quality of analysed meteorological fields. The Meteorological Office has plans to do this as part of its contribution to an international project involving a new satellite - the Upper Atmosphere Research Satellite, scheduled for launch in 1990.

In the rest of this lecture, we shall consider the main features of the thermal and wind structures of the middle atmosphere, noting variations that occur on seasonal and longer time periods, and pointing out the main differences between the northern and southern hemispheres.

1.2 CLIMATOLOGICAL MEANS

1.2.1 Zonally averaged structure of temperature and wind

Fig. 1.2 shows a representative cross-section of the zonally averaged temperature field for January (panel (a)) and July (panel (b)). As expected from radiative considerations, the highest temperatures are at about 1 mb (near 50 km) over the summer pole, where the heating due to the absorption of solar radiation by ozone is greatest. At the mesopause (near 0.01 mb, about 85 km), however, the situation is reversed: temperatures increase equatorwards from a very cold summer pole. In the lower stratosphere, the main features are the cold tropical tropopause and the mid-latitude temperature maximum in the winter hemisphere.

There are pronounced differences between hemispheres during winter in the stratosphere and lower mesosphere at high latitudes (reflecting different levels of dynamical activity). In the northern hemisphere, polar temperatures are warmer through most of the stratosphere and colder near the stratospause than they are in the southern hemisphere. During summer, the stratopause is about 5 K warmer than it is in the northern hemisphere (probably because the earth is closer to the sun in January than it is in July).

The temperature structure shown in Fig. 1.2 departs significantly from that of a hypothetical atmosphere in radiative-convective equilibrium. The

winter pole in the middle atmosphere is as much as 100 K warmer than the equilibrium value. A task for dynamical theory is to explain differences like this.

The zonally averaged, zonal wind distribution is shown in Fig. 1.3. Zonal winds in winter and summer are westerly and easterly, respectively, with maxima at low latitudes in the lower mesosphere. Easterlies occupy equatorial latitudes in the stratosphere at the solstices, except in the westerly phase of the quasi-biennial oscillation (see below). In the middle and upper mesosphere both easterly and westerly winds decrease rapidly with height. This is thought to be due to the drag exerted by the dissipation of gravity waves, which propagate upwards from the troposphere (lecture 7). A prominent feature is that zonal winds are typically much stronger during winter in the southern hemisphere than they are during winter in the northern hemisphere. This is consistent (through thermal wind balance) with the colder polar temperatures in the southern hemisphere.

1.2.2 Monthly-mean horizontal fields

The monthly mean circulation in the stratosphere during mid winter is illustrated in Fig. 1.4. The main features of the circulation have a planetary scale; variations around latitude circles can mostly be represented by harmonics with zonal wavenumbers 1 and 2 (notice the absence of scales of motion typical of weather systems in the troposphere). Large-scale disturbances in the stratophere are forced in the underlying troposphere. In the northern hemisphere (panel (a)), the circulation is dominated by a cold, cyclonic, polar vortex, displaced somewhat from the pole. A persistent feature near 180°E is the so-called Aleutian High, a strong anticyclone which fluctuates in intensity throughout the winter. The

mid-winter circulation in the southern hemisphere (panel (b)), by contrast, is far less disturbed, and the flow is much more zonal.

In summer, disturbances in both hemispheres are very weak. A warm anticyclone lies almost directly over the pole, and streamlines of the time-mean flow are so closely coincident with latitude circles that it is not worth showing a figure!

1.3 THE SEASONAL CYCLE

The march of solar heating drives a seasonal cycle in the middle atmosphere. In the stratosphere, cold polar temperatures and a strong westerly vortex in winter are replaced by warm polar temperatures and a weak easterly vortex in summer. Because of dynamical disturbances, however, the seasonal cycle does not proceed uniformly.

Fig. 1.5 shows the annual march of zonal mean radiance (proportional to temperature) near the north and south poles in the upper stratosphere. Winter temperatures are much more variable in the northern hemisphere because of 'sudden warmings' (lecture 2). The largest temperature oscillations in the southern hemisphere occur during the transition in spring from winter westerlies to summer easterlies.

Fig. 1.6 shows the evolution of zonal-mean wind in the upper stratosphere. In the northern hemisphere, the zonal-mean westerlies are strongly disrupted by disturbances (e.g. during January-February, 1981). Should these disturbances occur in late winter, the westerly circulation may never recover before summer easterlies are introduced by the evolving radiation field. These 'final warmings' occur earlier in the seasonal cycle in the northern hemisphere than in the southern hemisphere where winds evolve more regularly.

1.3.1 The semi-annual oscillation

In equatorial regions the seasonal cycle of winds in the upper stratosphere is dominated by a semi-annual oscillation (SAO), which has a maximum amplitude of about 35 ms^{-1}. In Fig. 1.6, easterly winds penetrate from the summer hemisphere into the winter hemisphere twice a year; the easterlies spread further across the equator during the southern hemisphere summer. A time-height section of the SAO at the equator is shown in Fig. 1.7. Near the stratopause, the westerly phase of the oscillation appears at the equinoxes and propagates downwards. Westerly flows at the equator have angular momentum in excess of that due to the earth's rotation, and thereby pose interesting dynamical questions.

1.3.2 Inter-annual variability and the quasi-biennial oscillation

1.3.2.1 Extra-tropical latitudes

The circulation of the middle atmosphere has strong inter-annual variability in winter. Fig. 1.8 is a frequency distribution of temperature at 30 mb for both poles, based on nearly 30 years' data. During summer when prevailing winds are easterly, tropospheric disturbances do not penetrate far into the stratosphere (see lecture 6 for a theoretical explanation). So conditions are quiet and inter-annual variability is small, especially at mid latitudes. In the northern winter and in spring, the frequency distribution is broad: some winters are much more disturbed than others. It is found that zonal wave 1 is strong in disturbed winters.

Inter-annual variability is much smaller during southern winters because large-scale waves are weaker than they are in the northern hemisphere. Notice the greater variability in spring when disturbances are strong.

1.3.2.2 The quasi-biennial oscillation

In the equatorial stratosphere there is a long period variation of a very different kind to that found at mid latitudes - the quasi-biennial oscillation (QBO). Monthly-mean zonal winds in the tropical lower and middle stratosphere have a strong quasi-cyclic behaviour (Fig. 1.9): the winds alternate between easterlies and westerlies of about 20 to 30 ms^{-1} with a period that ranges between about 23 and 34 months. These jets descend slowly, successively being replaced from above by one of opposite sign. There is some evidence that the tropical QBO contributes to the inter-annual variability in the extra-tropics by influencing the planetary-scale disturbances.

FURTHER READING

Andrews, D.G., Holton, J.R. and Leovy, C.B. (1987). 'Middle atmosphere dynamics', Academic Press, New York.

Holton, J.R. (1979). 'An introduction to dynamic meteorology', Chap. 11, 2nd ed., Academic Press, New York.

W.M.O. Global Ozone Research and Monitoring Project (1985). 'Atmospheric Ozone 1985', Report No. 16, Vol. 1.

REFERENCES

Hirota, I. (1978). Equatorial waves in the upper stratosphere and mesosphere in relation to the semiannual oscillation of the zonal wind. J. Atmos. Sci., **35**, 714-722.

Hirota, I., Hirooka, T. and Shiotani, M. (1983). Upper stratospheric circulation in the two hemispheres observed by satellites. Quart. J. Roy. Met. Soc., **109**, 443-454.

Labitzke, K. and Naujokat, B. (1983). On the variability and on trends of temperature in the middle stratosphere. Beitr. Phys. Atmosph., **56**, 495-507.

Naujokat, B. (1981) Long-term variations in the stratosphere of the northern hemisphere during the last two sunspot cycles. J. Geophys. Res., **86**, 9811-9816.

Naujokat, B. (1986) An update of the observed QBO of the stratospheric winds over the tropics. J. Atmos. Sci., **43**, 1873-1877.

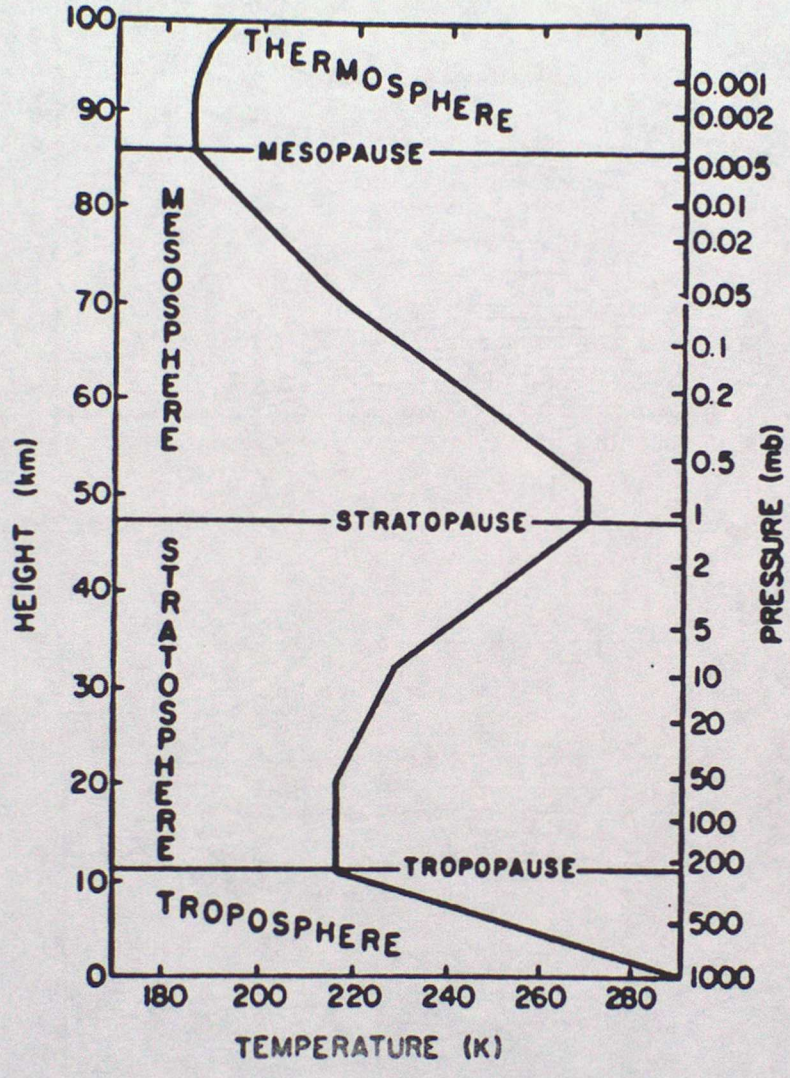
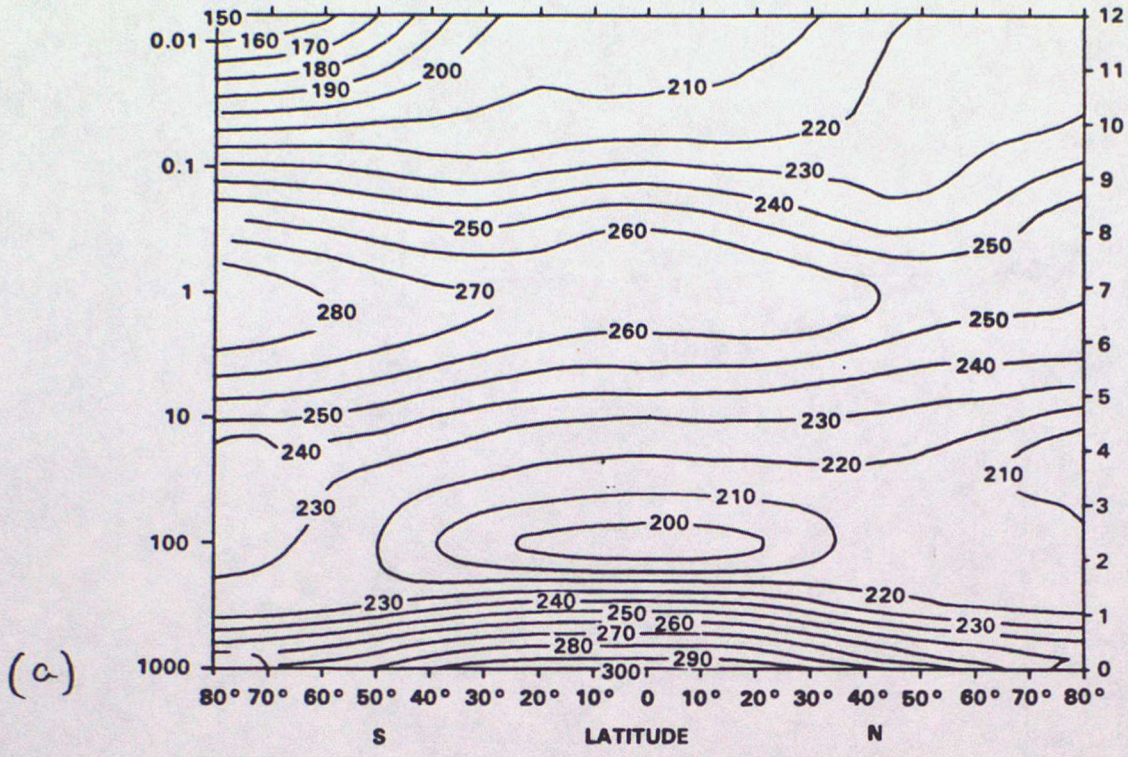


Fig. 1.1 Schematic mid-latitude temperature profile.



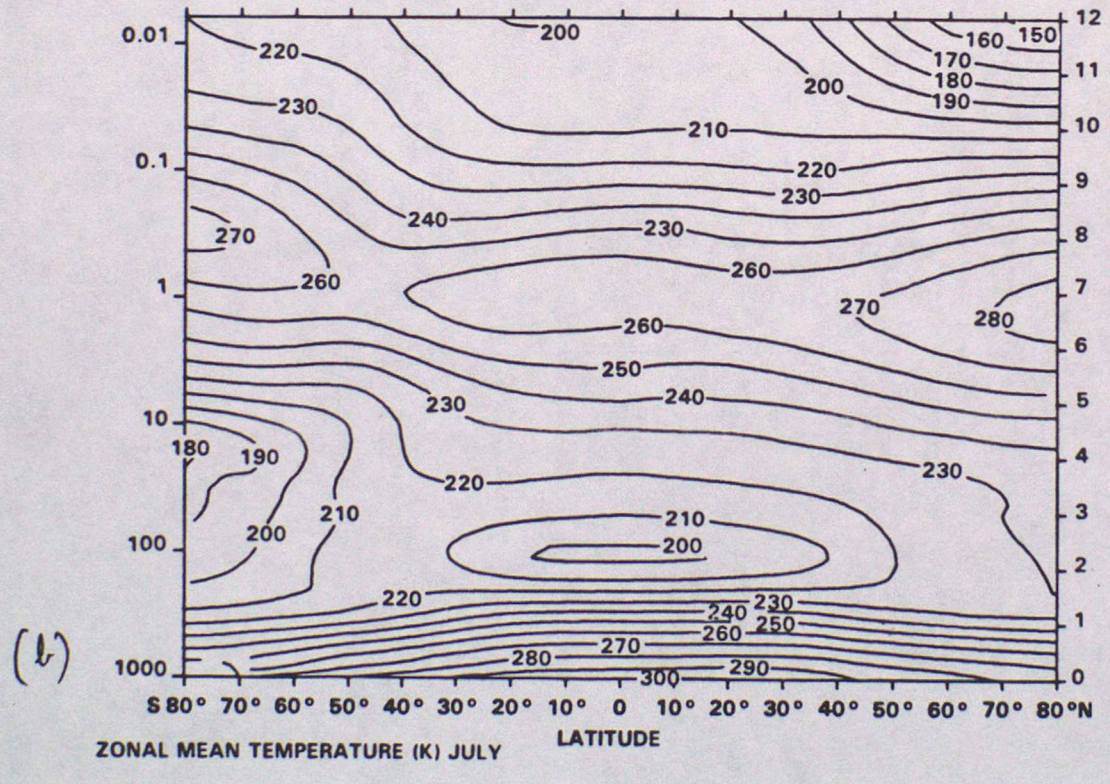


Fig. 1.2 Cross sections of zonal-mean temperature (K) for an average over 5 years of the monthly means for (a) January and (b) July. The data are for the combined SCR/PMR retrieval made at the University of Oxford for the period January 1973 to December 1974, and July 1975 to June 1978. (Supplied by J.J. Barnett and M. Corney)

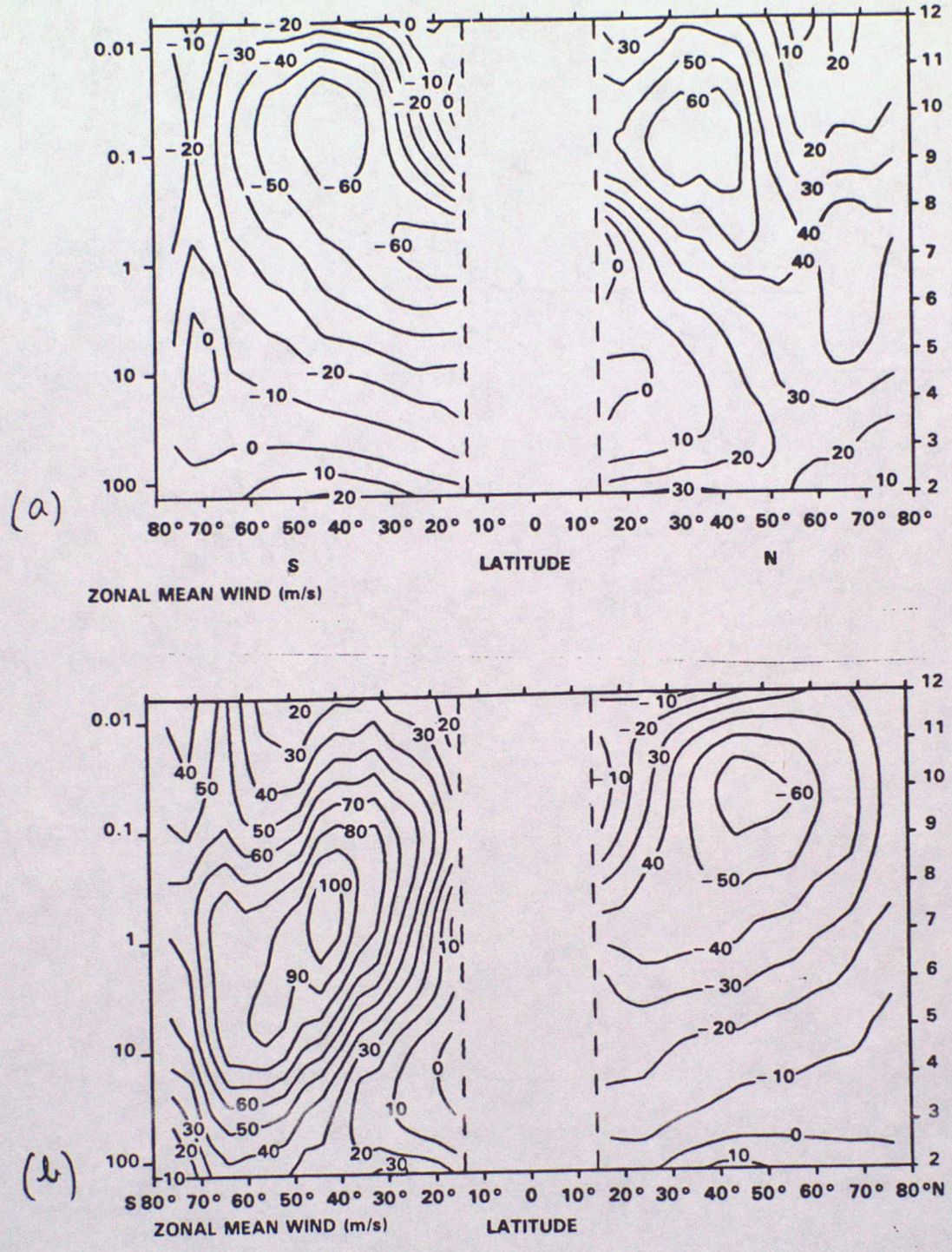
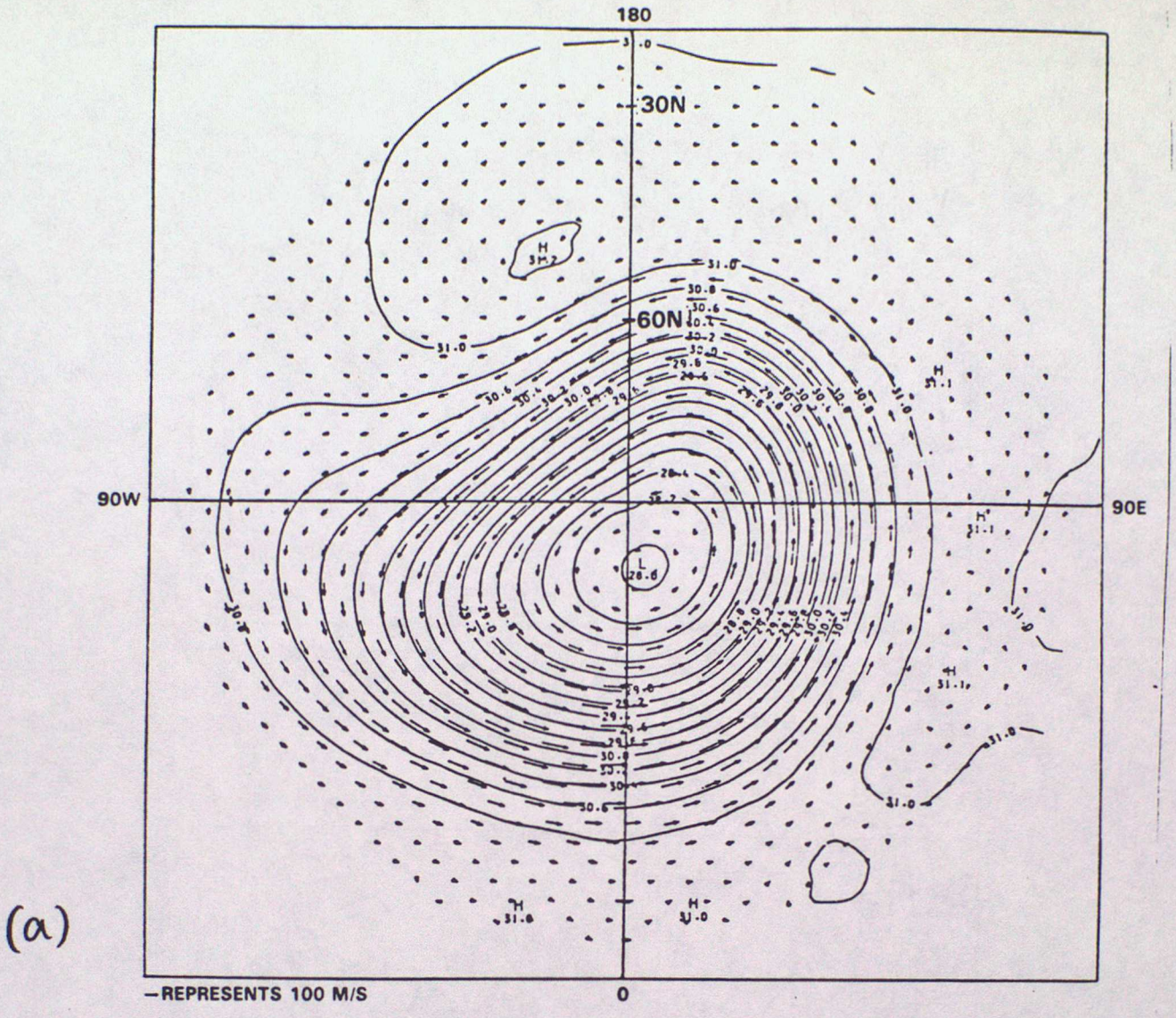
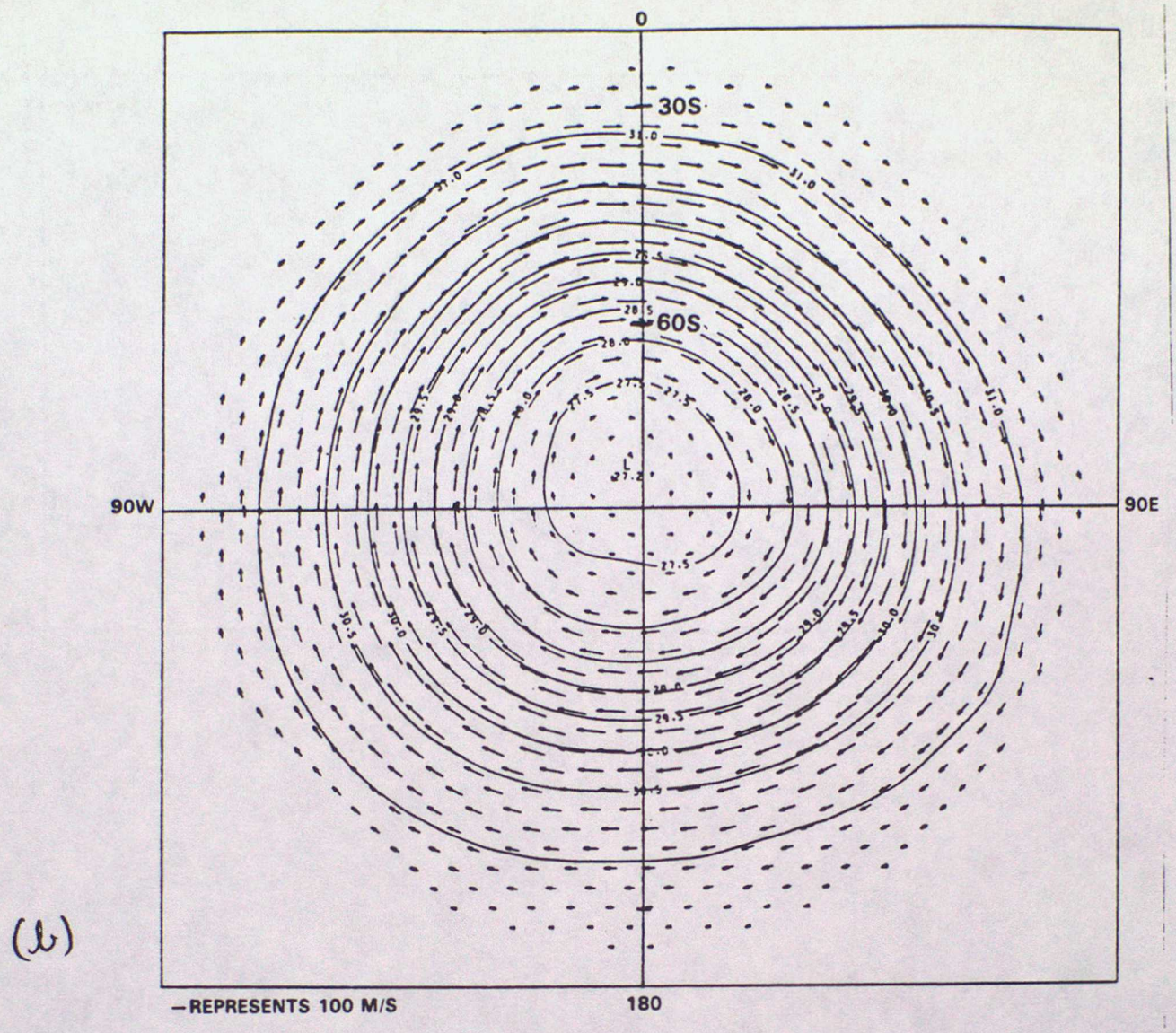


Fig. 1.3 Cross sections of zonal-mean geostrophic wind (m s^{-1}) for (a) January and (b) July. Data source as in Fig. 1.2.



(a)

Fig. 1.4(a) Polar stereographic map at 10 mb of monthly mean geopotential height (km) for the northern hemisphere for January 1981. Data obtained from a stratospheric sounding unit on the satellite NOAA-6. Analysis made by the Middle Atmosphere Group, Meteorological Office.



(b)

Fig. 1.4(b) As for Fig. 1.4(a) but for the southern hemisphere for July 1981.

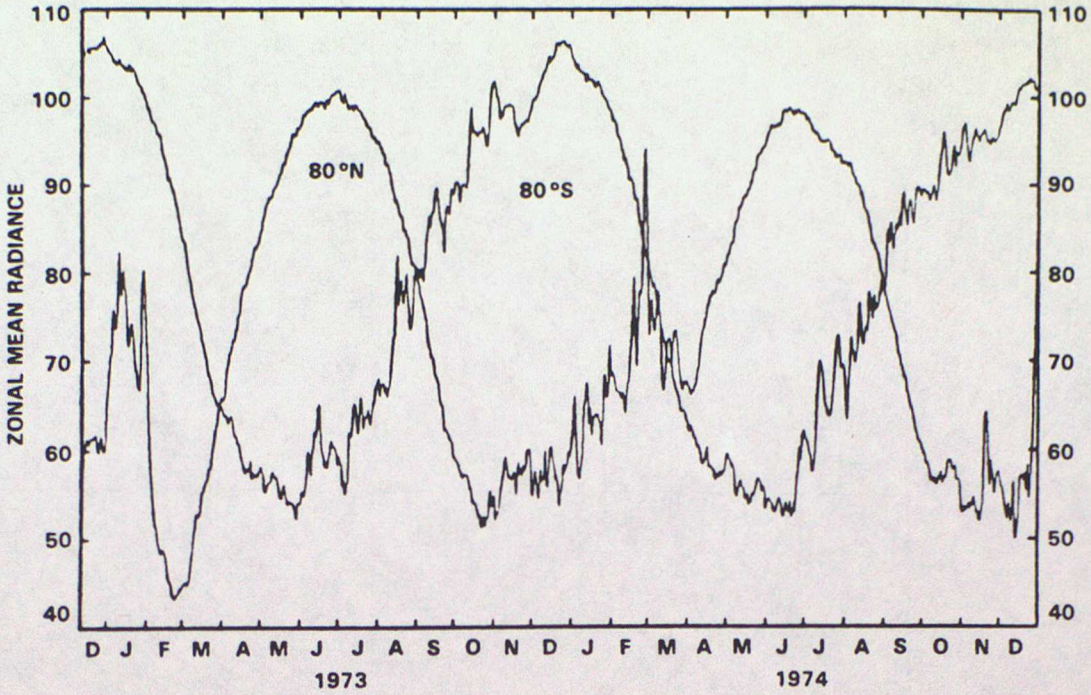


Fig. 1.5 Annual march of zonal-mean radiance observed by the Nimbus 5 SCR channel B12 for 80°N and 80°S. Units are mW m^{-2} $ster^{-1}$. Taken from Hirota et al. (1983).

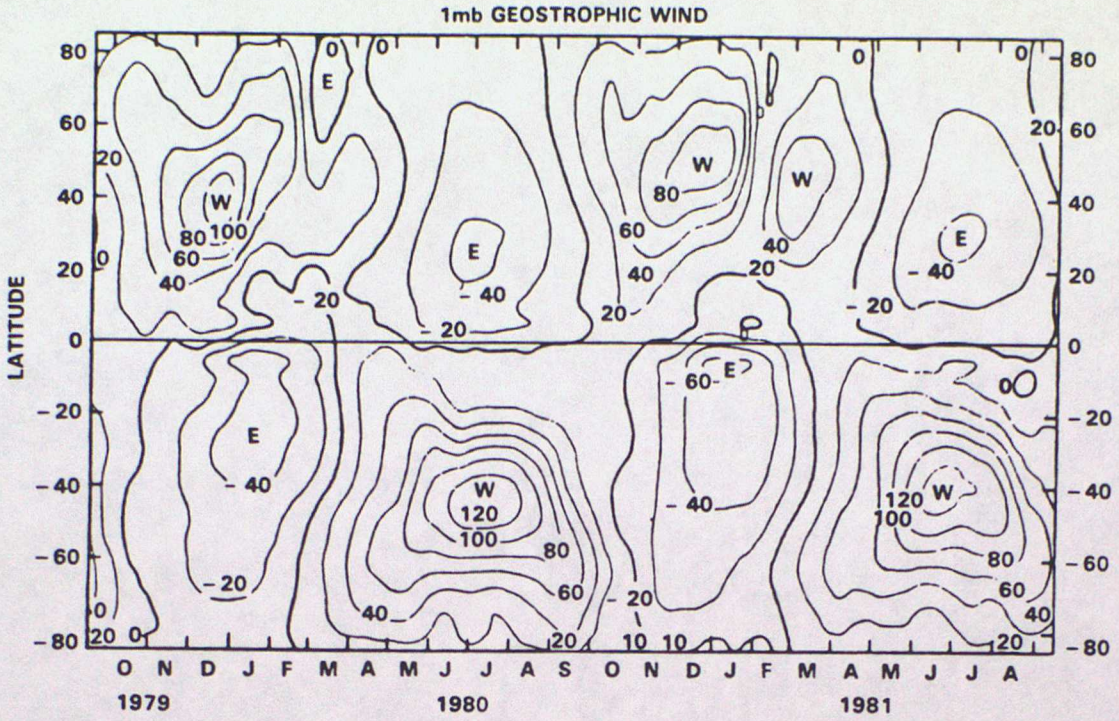


Fig. 1.6 Latitude-time section of the zonal-mean geostrophic wind at the 1 mb level estimated from the 20-day average height field observed by the Tiros-N SSU. Units are m s^{-1}. Positive values denote westerly winds. Taken from Hirota et al. (1983).

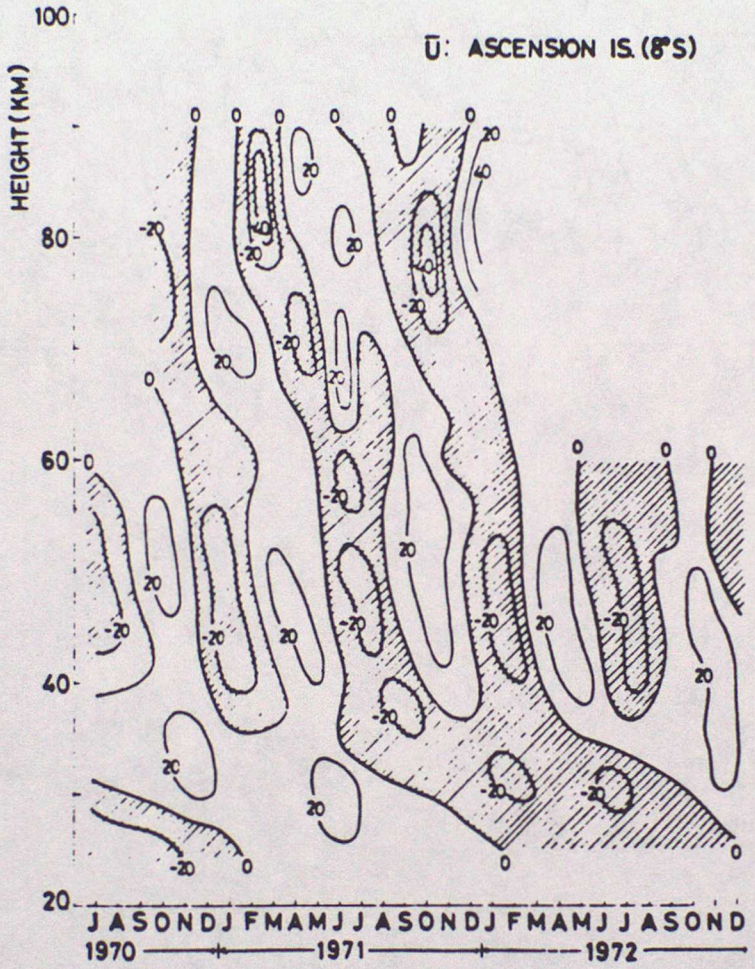


Fig. 1.7 Semiannual oscillation at Ascension Island (8°S). Taken from Hirota (1978).

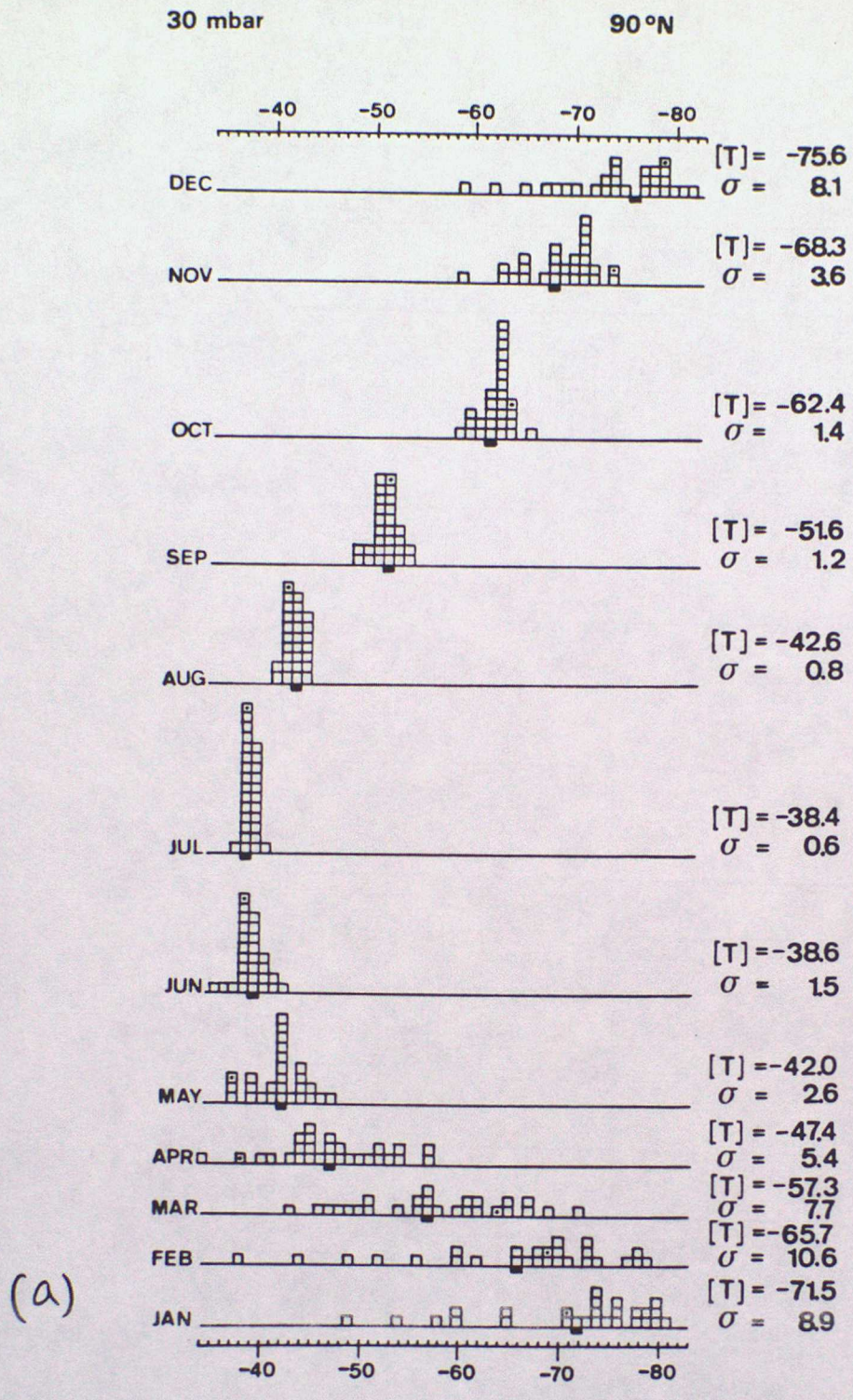


Fig. 1.8(a) Frequency distribution of the monthly mean 30 mb temperatures (°C) at the North Pole made from radiosonde data for the period July 1955 to December 1982. The interval is 1°C. The long-term average [T] is given on the right-hand side of the picture together with the standard deviation σ, and [T] is also marked as a black box in the frequency distribution. (After Naujokat, 1981, and Labitzke and Naujokat, 1983)

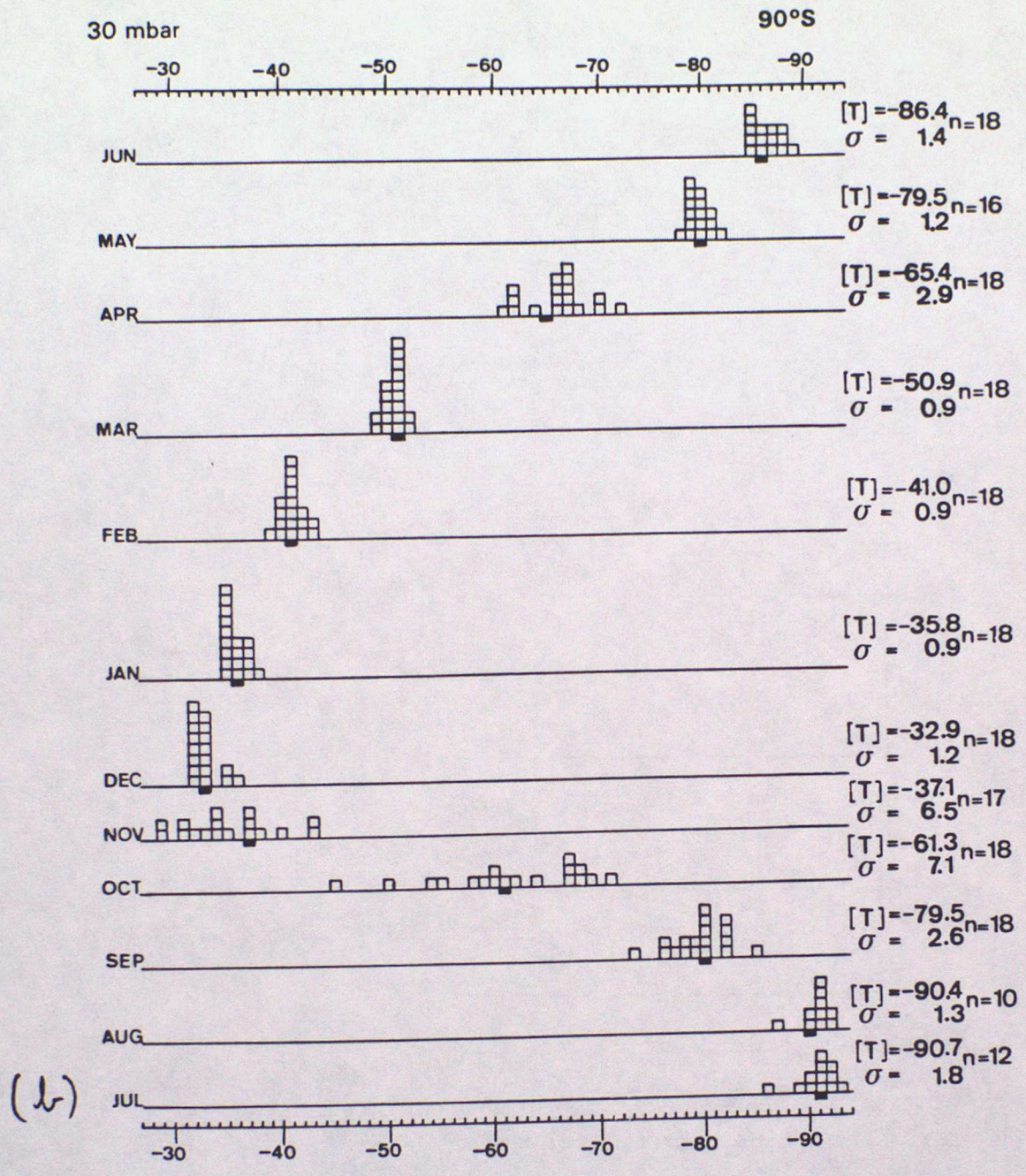


Fig. 1.8(b) As for Fig. 1.8(a) but for the South Pole for the period 1961 to 1978 using, for each month, data for the number of years specified by n.

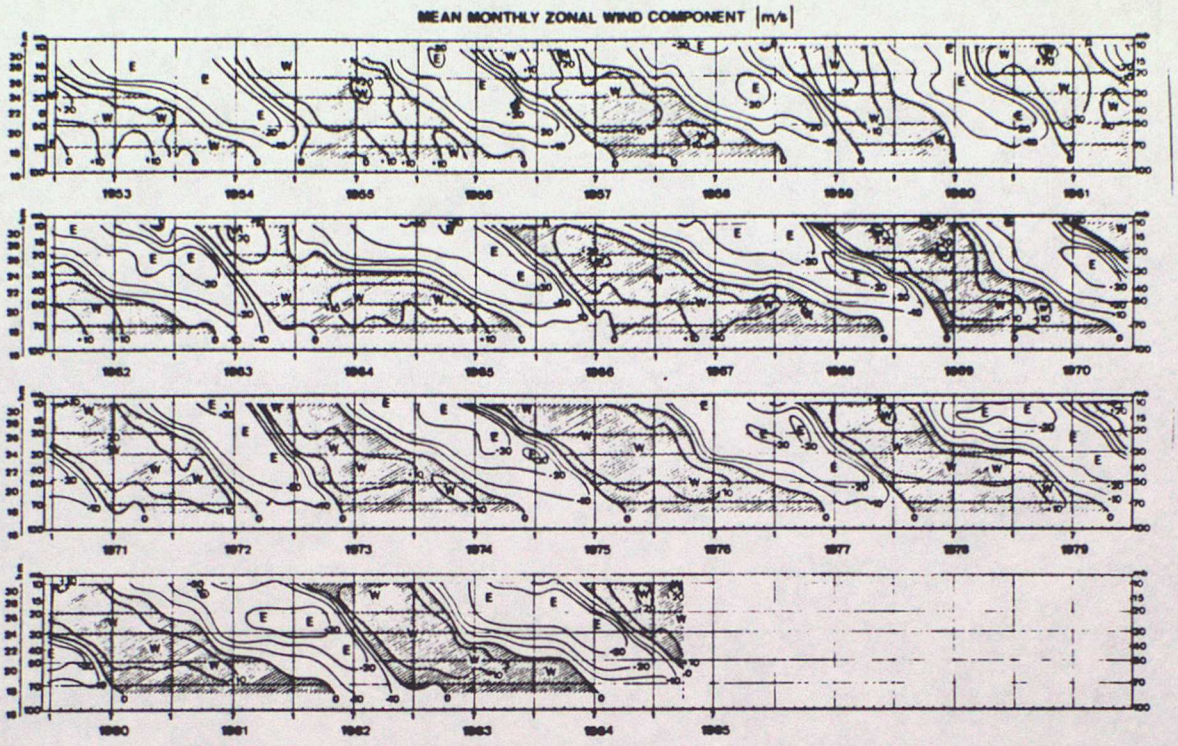


Fig. 1.9 Time-height cross-section of monthly mean zonal winds (m s^{-1}) at equatorial stations, calcualted from daily values:

Jan. 1953 - Aug. 1967, Canton Island, 3°S/172°W

Sep. 1967 - Dec. 1975, Gan/Maledive Islands, 1°S/73°E

Jan. 1976 - Apr. 1985, Singapore, 1°N/104°E

Note the downward propagating westerly (W) and easterly (E) regimes. (After Naujokat, 1986)

2 OBSERVATIONS OF TRANSIENT DISTURBANCES IN THE MIDDLE ATMOSPHERE

2.1 INTRODUCTION

In lecture 1 we considered some of the salient features of the climatology of the middle atmosphere as revealed by monthly-mean fields of winds and temperature. But a wide range of interesting dynamical bahaviour happens on time scales of less than a month. There are dramatic sudden warmings which develop over a week or so; there are large-scale travelling waves (lost in a time-mean view); ephemeral gravity waves exert a brake on the winds where they dissipate (particularly in the mesosphere); travelling Kelvin and Rossby gravity waves are thought to be important in the semi-annual and quasi-biennial oscillations; and there are atmospheric tides. The term 'transient disturbance' is sometimes used (perhaps rather loosely) to embrace all of these.

2.2 TRANSIENT DISTURBANCES

2.2.1 Stratospheric sudden warmings

The most spectacular departures from a regular seasonal cycle in the middle atmosphere occur during 'sudden warmings' in the stratosphere in winter, particularly in the northern hemisphere. The westerly circulation in the stratosphere is disrupted on a very large scale, and locally temperatures may rise by 80 K or more in a matter of a week or so (temperatures have been recorded to exceed +20°C during strong events in the northern stratosphere). Though much weaker, the atmospheric disturbance can be detected at the mesopause and above.

A warming is classified as major if it is strong enough for zonal-mean westerlies to be replaced by easterlies at and above 10 mb (30km), and poleward of 60° latitude. Otherwise it is called a minor warming. This distinction should not be applied too rigidly as the two types have much in common.

Warmings are also broadly divided into 'wave-1' and 'wave-2' types. In the former, the Aleutian High grows and the cyclone is displaced farther from the pole than it normally is (see Fig. 1.4(a)); in the latter, the cyclone may split as two anticyclones develop about 180° apart in longitude.

A striking example of a wave-2 major warming took place in the northern hemisphere at the end of December 1984. Fig. 2.1(a) is a map of geopotential-height and temperature at 10 mb on 28 December. The Aleutian high lies near the dateline as usual, but there is another strong (non-climatological) anticyclone developing near 0°E. The cyclone is highly elongated and just about to split into two separate vortices. Notice that the two pools of warm air lie to the west of the anticyclones. This is a characteristic of upward-propagating disturbances (lecture 6). By 2 January 1985, the cyclone has split (Fig. 2.1(b)). The two anticyclones have merged to form one huge anticyclone of warm air over the pole, flanked by a pair of cyclones.

During warmings there is a large exchange of air between high and low latitudes (there may be considerable poleward transport of photochemically important trace chemicals such as ozone). The movement of air can be followed using maps of Ertel's potential vorticity on isentropic surfaces. Over short periods (a week or so in the stratosphere), contours of potential vorticity on such surfaces are approximately material lines. An example for an early winter warming in the northern hemisphere is shown in

Fig. 2.2. A mass of air with low values of potential vorticity is being injected from low latitudes into high latitudes. Associated with it is a strong anticyclonic circulation with a long tongue of high potential vorticity around its southern flank. The term 'wave breaking' has been applied to the extreme and irreversible buckling of potential vorticity contours such as that depicted in the figure.

Stratospheric warmings are thought to be caused by planetary-scale disturbances in the troposphere which propagate upwards in winter, when wind conditions are right for this to happen (see lecture 6). Major warmings, involving the replacement of polar westerlies by easterlies, do not occur in mid winter in the southern hemisphere. This is probably because the southern troposphere does not contain such persistent large-scale disturbances as does its northern counterpart.

2.2.2 Travelling waves

A Fourier analysis of meteorological fields in the stratosphere will generally yield some zonal harmonics with phases (longitude of peaks or troughs) that change with time. These harmonics might be called travelling waves. Whether these 'waves' are waves in the normal physical sense, with theoretically predicted properties, is another matter altogether.

Some travelling waves in the stratosphere may be associated with theoretically predicted 'normal modes'. (Normal modes have properties independent of details of forcing, unlike the forced planetary waves involved in sudden warmings). Typically, they have much smaller amplitudes than disturbances considered above, and can be studied fairly well with linear wave theory.

The most widely documented normal mode is the 5-day wave in zonal wavenumber 1. It is global, symmetric about the equator, and propagates westward around the earth once in 5 days. Satellite data have been Fourier analysed and time filtered to produce the example shown in Fig. 2.3. The analogue of this wave in wavenumber 2 has also been found. It has a 4-day period.

Other waves in the stratosphere have been tentatively identified as normal modes. Salby (1984) gives a comprehensive review.

2.2.3 Gravity waves

Gravity waves are believed to play an important role in determining the large-scale circulation and temperature structure of the middle atmosphere (particularly the mesosphere), where the turbulence generated by breaking waves induces a drag on the mean flow. The main sources of these waves are thought to be in the lower atmosphere, and include topography, frontal disturbances, convective activity, geostrophic adjustment and shear instabilities.

The first information on gravity waves came from rockets, but now radars are providing valuable data, albeit with limited geographical coverage. Gravity waves have horizontal wavelengths of about 100-200 km, vertical wavelengths of about 5-15 km, phase speeds of up to about 80 ms^{-1} and periods of a few minutes to an hour or so. Radar measurements indicate that the drag they exert on the mean winds in the mesosphere is about 20 to 80 $m\ s^{-1}\ day^{-1}$.

2.2.4 Kelvin waves and Rossby-gravity waves

In the lower equatorial stratosphere, two basic types of planetary-scale wave have been identified from radiosonde measurements: the eastward-travelling Kelvin wave and the westward-travelling Rossby-gravity wave. They are believed to be forced by convection in the tropical troposphere (though the precise forcing mechanisms are not well understood), from where they propagate upward into the middle atmosphere.

Some observations that seem to show Kelvin waves are shown in Fig. 2.4, and further details ot the two types of travelling wave are given in the table below.

Observed equatorial waves in the lower stratosphere	Kelvin	Rossby-gravity
Period (ground based)	15 days	4 - 5 days
Zonal wavenumber	1 - 2	4
Vertical wavelength	6 - 10 km	4 - 8 km
Phase speed w.r.t. ground	+25 m s^{-1} (eastward)	-23 m s^{-1} (westward)
Observed when basic flow is	westward	eastward

2.2.5 Atmospheric tides

Tides in the upper mesosphere show up clearly in radar observations of drifting meteor trails. In the middle atmosphere, the dominant tidal motions are diurnal and semi-diurnal modes driven by solar heating (mainly of water vapour in the troposphere and ozone in the stratosphere). In the stratosphere, amplitudes of tidal wind oscillations are small (less than about 1-5 m s^{-1}). But in the mesosphere, amplitudes can be large, and it has been estimated that the dissipation of tides leads to the generation of easterly winds of up to 60 m s^{-1} at equatorial latitudes.

REFERENCES

Clough, S.A., Grahame, N.S. and O'Neill, A. (1985). Potential vorticity in the stratosphere derived using data from satellites. Quart. J. Roy. Met. Soc., **111**, 335-358.

Fairlie, T.D.A. and O'Neill, A. (1987). The stratospheric major warming of winter 1984/85: observations and dynamical inferences. Submitted to Quart. J. Roy. Met. Soc..

Hirota, I. and Hirooka, T. (1984). Normal mode Rossby waves observed in the upper stratosphere. Part I: First symmetric modes of wavenumbers 1 and 2. J. Atmos. Sci., **41**, 1253-1267.

Salby, M.L. (1984). Survey of planetary-scale traveling waves: the state of theory and observations. Rev. Geophys. Space Phys., 22, 209-236.

Wallace, J.M. and Kousky, V. (1968). Observational evidence for Kelvin waves in the tropical stratosphere. J. Atmos. Sci., **25**, 900-907.

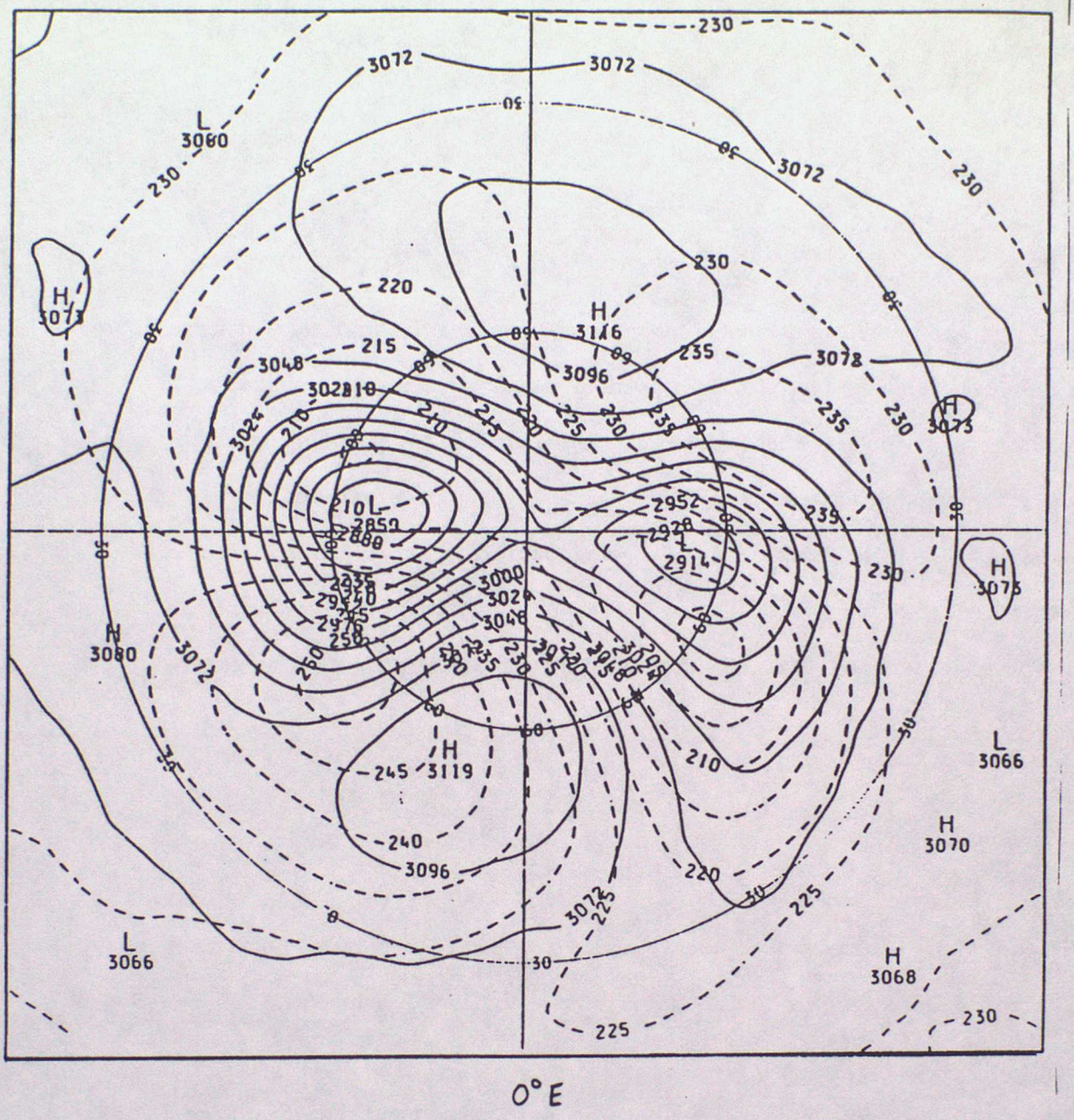


Fig. 2.1(a) Synoptic map at 10 mb of geopotential height (solid lines) and temperature (dashed lines) for 28 December 1984. Units: dam for geopotential height; K for temperature. (Taken from Fairlie and O'Neill, 1987)

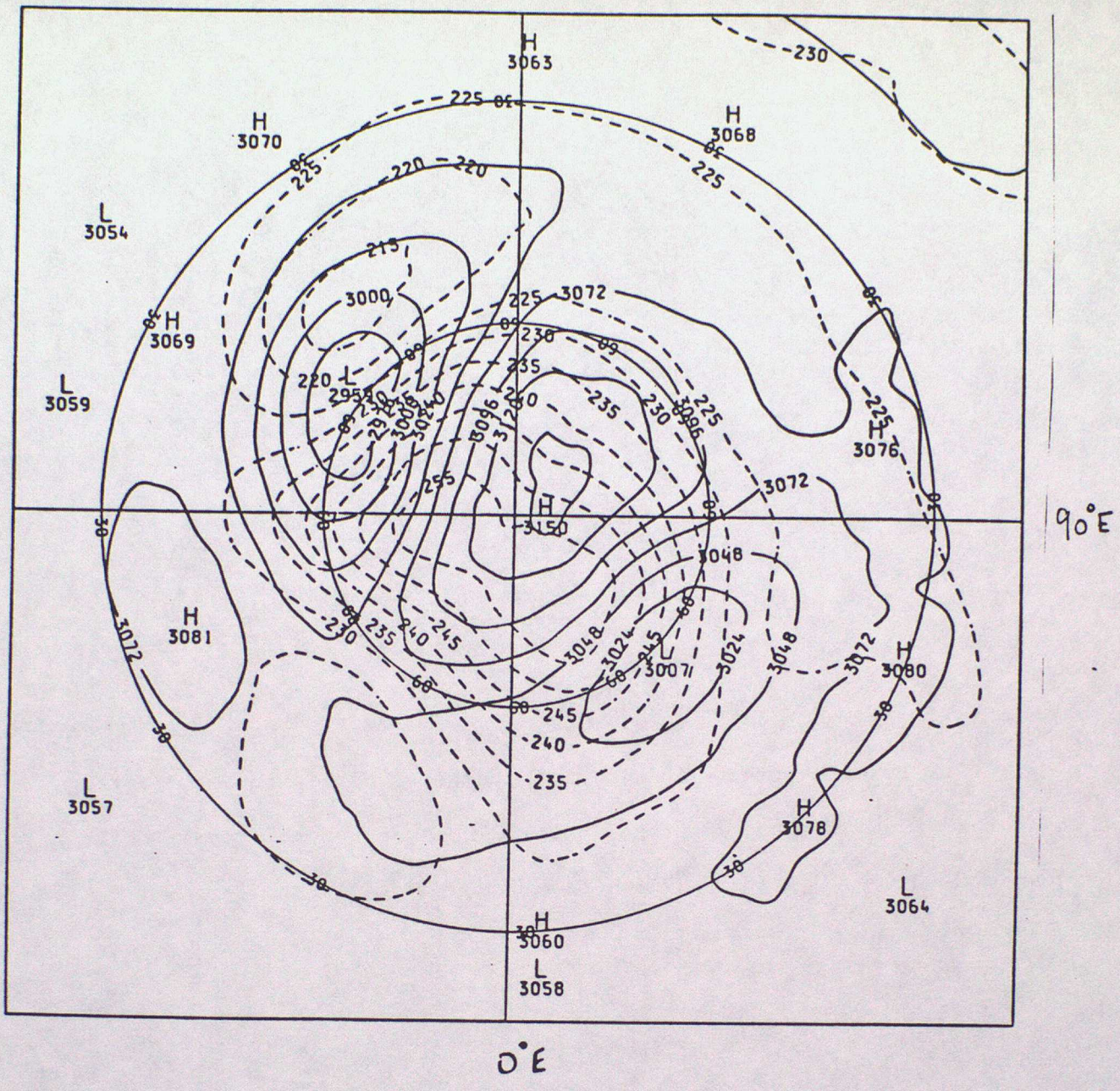


Fig. 2.1(b) As for Fig. 2.1(a) but for 2 January 1985.

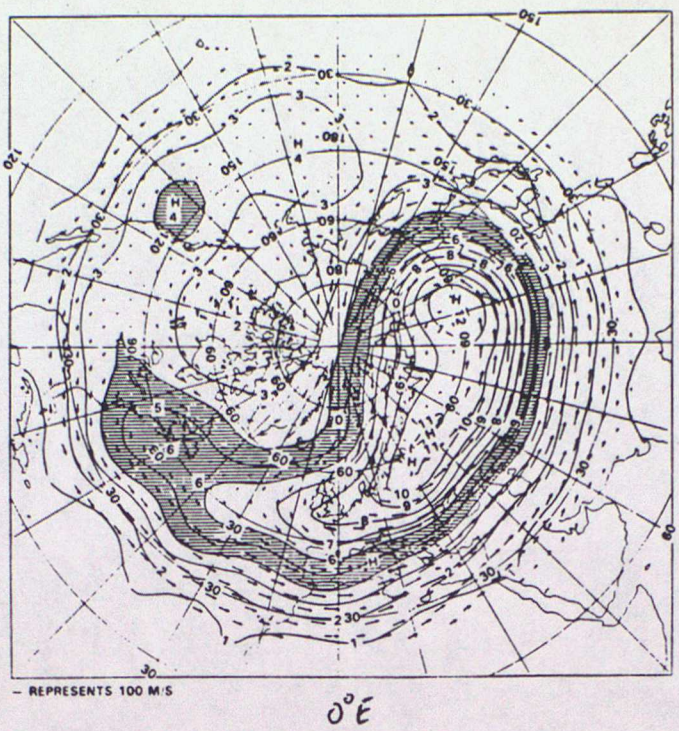


Fig. 2.2 Ertel's potential vorticity and geostrophic winds evaluated on the 850 K isentropic surface near 10 mb on 4 December 1981. Units are K m^2 kg^{-1} s^{-1} x 10^{-4}. (Taken from Clough et al., 1985)

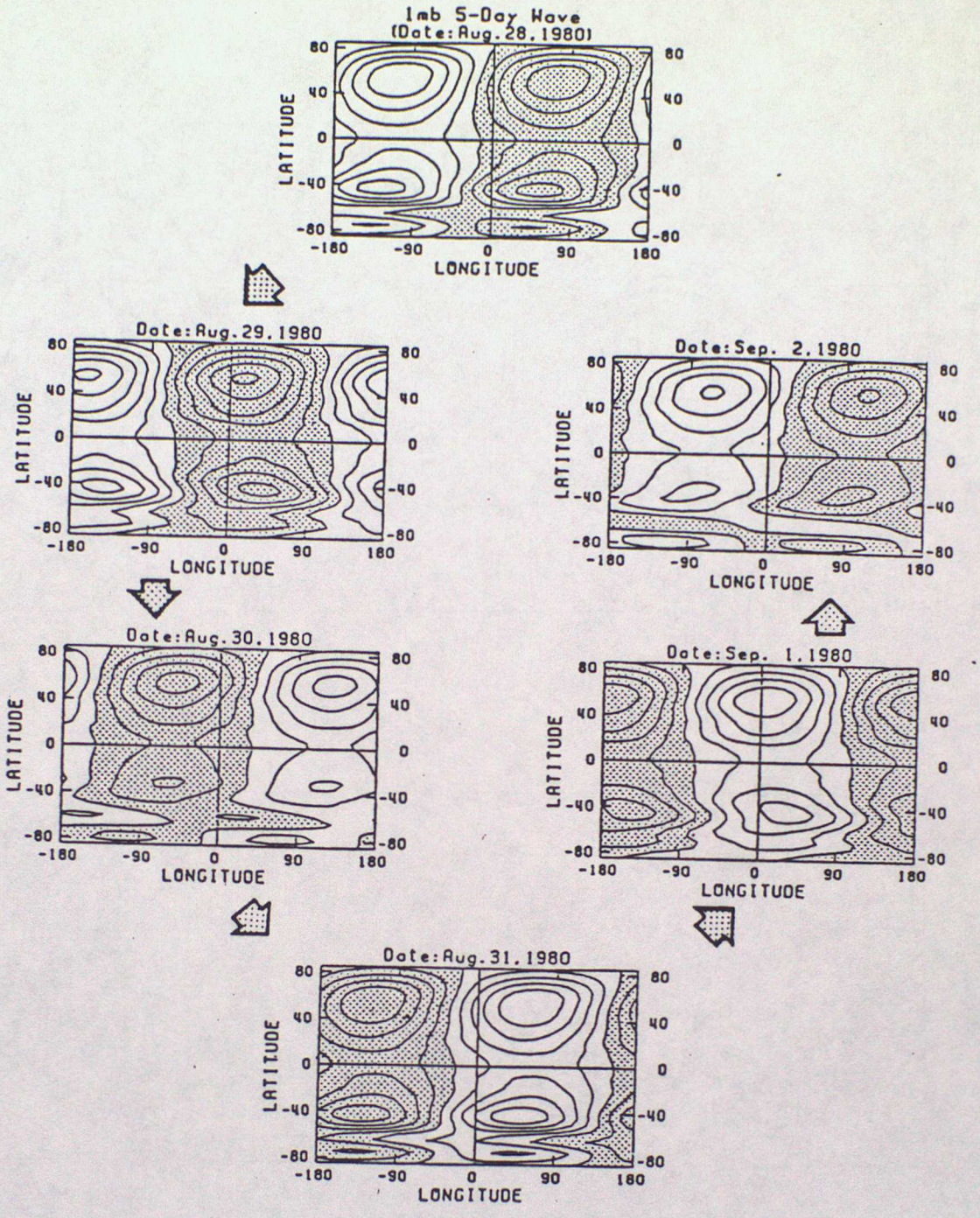
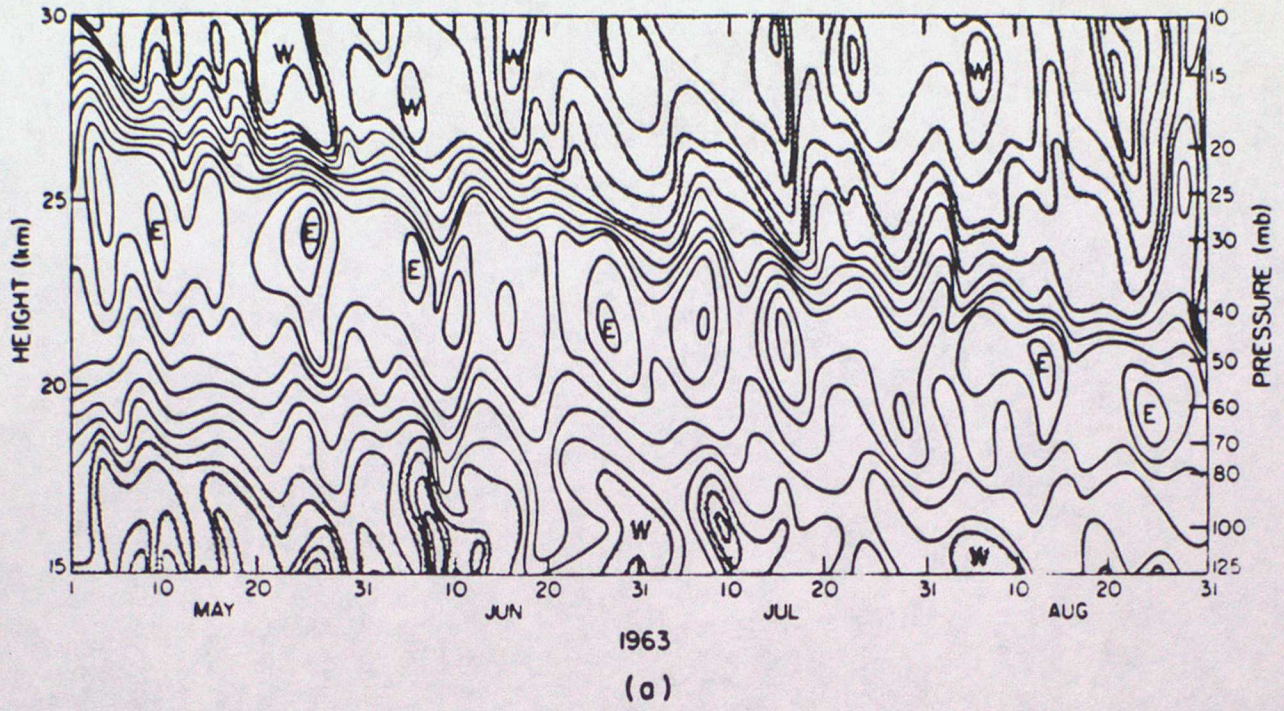


Fig. 2.3 The travelling pattern of the 1 mb 5-day wave for the six successive days of 28 August to 2 September 1980. Shaded areas denote negative height values. The contour interval is 20 m. Taken from Hirota and Hirooka (1984).



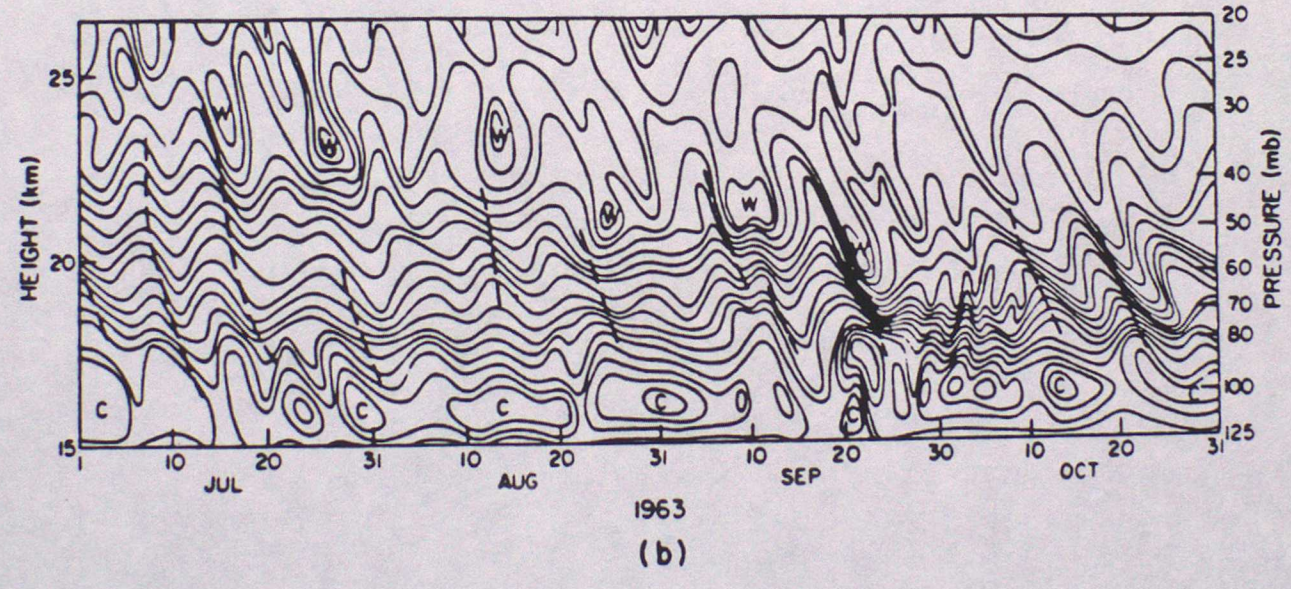


Fig.2.4 Time-height sections for the equatorial lower stratosphere, showing evidence of Kelvin-wave activity. (a) Zonal wind and (b) temperature at Canton Island (3°S0. Note the westerly phase of the QBO encroaching from upper levels in (a). Taken from Wallace and Kousky (1968).

3. THE PRIMITIVE AND QUASIGEOSTROPHIC EQUATIONS IN LOG-PRESSURE COORDINATES

3.1 THE PRIMITIVE EQUATIONS

The atmosphere is a shallow envelope of compressible gas surrounding a rotating planet. The equations of motion for such a gas are well known, but are more complicated than necessary for understanding the large and medium scale meteorological phenomena that will be considered in these lectures. Scaling arguments (e.g. Holton, 1979, § 2.4) show that, for such phenomena, the vertical momentum equation reduces to hydrostatic balance, and the Coriolis force associated with the horizontal components of the earth's rotation vector can be neglected. The resulting set of equations are called the *primitive equations*. (See Phillips, 1973 and Gill, 1982 for derivations.) For simplicity, we shall mostly use the *beta-plane approximation*, which means that, rather than employing the full spherical geometry, we work on a tangent plane (with x = eastward distance, y = northward distance and z* = geometric height) and approximate the vertical component of the earth's rotation vector $f = 2\Omega\sin\phi$ by the linear function $f = f_0 + \beta y$; here f_0 and β are constants and $\Omega = 2\pi\,(\text{day})^{-1}$ and ϕ is latitude.

Even after these simplifications, the primitive equations are quite cumbersome. They can however be made slightly neater by changing the vertical coordinate. We note first that hydrostatic balance in the vertical is expressed by

$$\rho g = -\partial p/\partial z^* \,, \tag{3.1a}$$

and the ideal gas equation of state by

$$p = RT\rho \,, \tag{3.1b}$$

where ρ is density, g the gravitational acceleration, p pressure, T temperature and R the gas constant for dry air (≈ 287 J kg^{-1} K^{-1}).

Consider now an isothermal atmosphere at constant temperature T_s and with constant pressure p_s at the ground, which is assumed to be flat and located at

$z^* = 0$. Eqs. (3.1a,b) show that

$$p = p_s e^{-z^*/H} \ , \quad \rho = \rho_s e^{-z^*/H} \ ,$$

where $H = RT_s/g$ is the *scale height* and $\rho_s = p_s/RT_s$. Equivalently, $z^* = -H \ln(p/p_s)$ gives the relationship between geometric height and pressure for this isothermal atmosphere.

Next consider a non-isothermal atmosphere: from (3.1a) it can easily be seen that $-\ln(p)$ increases monotonically with z^*. Choose T_s to be some representative middle atmosphere temperature (say 240 K) and define $H = RT_s/g$ ($\approx$ 7 km if T_s = 240 K). Then *define* a "log-pressure" coordinate, with dimensions of height,

$$z \equiv -H \ln(p/p_s) \ ,$$

which increases monotonically with z^* , and equals z^* if $T = T_s$ everywhere. (Typical stratospheric and mesospheric variations of T are usually small enough that $z \approx z^*$ throughout the middle atmosphere if T_s is chosen judiciously.) Note that $p = p_s \exp(-z/H)$. It turns out that if z is used in place of z^* as a vertical coordinate, with x,y,t as the other independent variables, if $w \equiv Dz/Dt$ is used in place of the vertical velocity, where

$$D/Dt \equiv \partial/\partial t + u\partial/\partial x + v\partial/\partial y + w\partial/\partial z$$

is the material derivative, and if $\Phi \equiv gz^*$ is the geopotential, the primitive equations reduce to

Horizontal momentum:

$$Du/Dt - fv + \Phi_x = 0 \ , \qquad (3.2a)$$

$$Dv/Dt + fu + \Phi_y = 0 \ , \qquad (3.2b)$$

Hydrostatic balance:

$$\Phi_z = RT/H \ , \qquad (3.2c)$$

Mass continuity:

$$u_x + v_y + \rho_0^{-1}(\rho_0 w)_z = 0 \ , \qquad (3.2d)$$

Thermodynamics:

$$D\theta/Dt = (J/c_p)e^{\kappa z/H} \equiv Q \ , \qquad (3.2e)$$

where friction has been ignored, but diabatic heating included. (See, e.g. Holton, 1975.) Here suffixes x,y,z denote partial derivatives, the potential temperature θ is defined by

$$\theta = T(p/p_s)^{-\kappa} = Te^{\kappa z/H} \quad , \text{ where } \kappa = R/c_p \approx 2/7, \tag{3.2f}$$

and J is the diabatic heating rate per unit mass, which in the middle atmosphere equals the net *radiative* heating or cooling rate. The basic density is defined by $\rho_0(z) \equiv \rho_s \exp(-z/H)$ and also equals p/RT_s : some authors use p instead of ρ_s here. Note that in terms of temperature, (3.2e) becomes

$$DT/Dt + \kappa Tw/H = J/c_p \ . \tag{3.2e'}$$

3.2 BOUNDARY CONDITIONS

To solve (3.2), or any approximate set of equations derived from them, it is necessary to apply suitable boundary conditions. These will depend on the physical problem being considered, and some common examples are given below.

3.2.1 Conditions at the lower boundary

(a) If the lower boundary is the *ground*, the shape of the topography should be specified in terms of the geometric height z*, rather than z (this is a slight inconvenience of log-pressure coordinates): for example

$$z^* = h(x,y,t) \quad \text{at the ground.}$$

(The t-dependence is a mathematical device that is sometimes useful in idealised initial-value problems.) Since the ground is a material surface, the kinematic boundary condition is

$$D(z^* - h)/Dt = 0 \text{ at } z^* = h.$$

In terms of $\Phi = gz^*$ we have

$$D\Phi/Dt = gDh/Dt \text{ at } \Phi = gh \ . \tag{3.3a}$$

Flat ground can be taken as $h = 0$.

(b) Some models specify the geopotential or geometric height of a given log-pressure level $p = p_0$, e.g.

$$\Phi(x,y,z_0,t) = F(x,y,t) \ , \qquad (3.3b)$$

where $z_0 = -H \ln(p_0/p_s) = \text{constant}$.

(c) In simple mathematical models one sometimes specifies the log-pressure at a lower material boundary. Thus $z = \zeta(x,y,t)$, say, at the boundary, and the kinematic boundary condition is

$$w \equiv Dz/Dt = D\zeta/Dt \text{ at } z = \zeta(x,y,t). \qquad (3.3c)$$

3.2.2 Conditions at the upper boundary

Some idealised models of the troposphere impose a rigid lid at the tropopause, which is supposed to partially represent the effect of the stable stratosphere above. However, when considering the stratosphere and mesosphere we must generally regard them to be unbounded as z^* and z tend to infinity. The boundary condition at great heights then falls into one of two categories:

(a) The disturbance tends to zero as $z \to \infty$ (where a suitable measure of the disturbance might be its energy per unit volume).

(b) A *radiation condition* applies: that is, "information" is transferred *upwards*, not *downwards*, at large z.

These two possibilities will be discussed in Lecture 6, when planetary waves are investigated.

3.2.3 Conditions at side boundaries

Conditions here depend on the geometry of the problem. On the sphere, it is only necessary that all physical variables are bounded at the poles. In idealised "channel" models (with vertical walls at $y = y_1$, y_2, say) $v = 0$ is taken on the walls.

3.3 STATIC STABILITY

It is useful to have a measure of the gravitational stability of a stratified atmosphere to small disturbances. To obtain this, we introduce a reference temperature profile $T_0(z)$ (which could, for example, be the global horizontal average of the observed temperature at each z or the schematic profile of Fig.1.1). We then define

$$N^2(z) = (R/H)[T_{0z} + \kappa T_0/H]. \tag{3.4}$$

We can also introduce the reference potential temperature $\theta_0 = T_0 e^{\kappa z/H}$; then

$$\theta_{0z}(z) = HN^2 R^{-1} e^{\kappa z/H} . \tag{3.5}$$

N^2 and θ_{0z} are measures of static stability; it can be shown that if they are negative, the atmosphere is gravitationally unstable, with "heavy" fluid over "light". (For a compressible fluid it is the vertical stratification of potential temperature, or entropy, rather than density, that determines whether the fluid is statically stable or unstable.) Viewed on a large scale, N^2 and θ_{0z} are basically positive in the stratosphere, which are therefore *stably* stratified in general; a small-scale process that violates this is mentioned in Lecture 7. [Note that N differs slightly from the usual definition of the buoyancy, or Brunt-Väisälä, frequency, $(g\theta_{0z}/\theta_0)^{\frac{1}{2}}$; see Gill (1982), p. 184.]

3.4 QUASIGEOSTROPHIC THEORY

The primitive equations (3.2) are still complicated, despite the simplifications that have gone into them. Moreover, they are capable of describing a very wide range of atmospheric flows, from slow motions of global scale to quite rapid, medium-scale disturbances. To focus on the larger-scale, slower motions, at least in extratropical regions, we can introduce further approximations to obtain the *quasigeostrophic equations*. We shall proceed in a non-rigorous manner: for a more detailed approach see, e.g., Pedlosky (1979) or Gill (1982).

Our starting-point is the fact that for large-scale, low-frequency, extra-tropical flows, approximate *geostrophic balance* holds: that is, the Coriolis terms (-fv, fu) in (3.2a,b) are roughly balanced by the horizontal gradients of geopotential. Hence the horizontal wind (u,v,0) satisfies

$$u \approx u_g \ , \quad v \approx v_g \ ,$$

where the *geostrophic wind* (u_g, v_g) is defined in terms of the geopotential by

$$(u_g, v_g) = (-\psi_y \ , \ \psi_x) \ , \tag{3.6a}$$

where

$$\psi = f_0^{-1}(\Phi - \Phi_0) \tag{3.6b}$$

is the geostrophic streamfunction and $\Phi_0(z)$ is a suitable reference geopotential profile, such as

$$\Phi_0(z) = \int_0^z T_0(z')dz' \ .$$

From hydrostatic balance (3.2c) we have

$$T_e \equiv T - T_0 = HR^{-1}f_0\psi_z \ , \tag{3.6c}$$

and

$$\theta_e \equiv \theta - \theta_0 = HR^{-1}f_0e^{\kappa z/H}\psi_z \ . \tag{3.6d}$$

Combining (3.6a,c) to eliminate ψ, we obtain the *thermal wind* equations

$$u_{gz} = -(R/Hf_0)e^{-\kappa z/H}\,\theta_y \ , \quad v_{gz} = (R/Hf_0)e^{-\kappa z/H}\,\theta_x \ , \tag{3.6e,f}$$

which relate vertical wind shears to horizontal (potential) temperature gradients. From (3.6a), $u_{gx} + v_{gy} = 0$, so by the continuity equation (3.2d) the geostrophic wind is associated with a vertical "velocity" w_g that satisfies $(\rho_0 w_g)_z = 0$. To ensure that w_g is bounded as $z \to \infty$, we must therefore take $w_g \equiv 0$.

It will be noticed that (3.6a,b) are approximate versions of the horizontal momentum equations (3.2a,b) in which Du/Dt and Dv/Dt are ignored and $f = f_0 + \beta y$ is replaced by f_0. To investigate the time development of the geostrophic flow, we must go to a better approximation. This can be done by defining *ageostrophic* velocities (suffix a) by

$$u = u_g + u_a \ , \quad v = v_g + v_a \ , \quad w = w_a \ ,$$

where $|u_a| \ll |u_g| - U$, $|v_a| \ll |v_g| - U$; i.e., the ageostrophic terms (u_a, v_a) are much smaller than the geostrophic wind, which is taken to be of order U. Now let L be a typical horizontal length scale, so that $\partial/\partial x$, $\partial/\partial y$ are $O(L^{-1})$, and suppose that

$$\text{(i) } Ro \equiv U/f_0 L \ll 1 \ , \quad \text{(ii) } \partial/\partial t \ll f_0 \ , \quad \text{(iii) } \beta y \ll f_0 \ ,$$

all hold. (Ro is called the *Rossby Number*.) It can be shown (e.g. Gill, 1982, p. 498) that (3.6a,b) are then a valid first approximation and that (3.2a,b,d,e) then give the *quasi-geostrophic* set

$$D_g u_g - f_0 v_a - \beta y v_g = 0 \ , \tag{3.7a}$$

$$D_g v_g + f_0 u_a + \beta y u_g = 0 \ , \tag{3.7b}$$

$$u_{ax} + v_{ay} + \rho_0^{-1} (\rho_0 w_a)_z = 0 \ , \tag{3.7c}$$

$$D_g \theta + w_a \theta_{0z} = Q \ , \tag{3.7d}$$

approximately, where

$$D_g \equiv \partial/\partial t + u_g \partial/\partial x + v_g \partial/\partial y$$

is the time derivative following the geostrophic wind, and (3.6a,b) have been used to cancel out the leading-order geostrophic-balance terms in (3.2).It is assumed that the departure θ_e from the reference θ_0 is always small, in the sense that $|\theta_{ez}| \ll \theta_{0z}$, so that $w_a \theta_z$ can be replaced by $w_a \theta_{0z}$, as in (3.7d). This is a fair approximation in the middle atmosphere.

The quasigeostrophic equations (3.7) are still quite complicated; however, they can be combined to yield a single useful and illuminating equation, (3.11). First, we obtain the *vorticity equation*

$$D_g\zeta_g = f_0\rho_0^{-1}(\rho_0 w_a)_z \tag{3.8}$$

by taking ∂(3.7b)/∂x - ∂(3.7a)/∂y and using the identities $(\underline{u}_g)_x.\nabla v_g \equiv 0$, $(\underline{u}_g)_y.\nabla u_g \equiv 0$, which follow from (3.6), and $D_g(f_0 + \beta y) \equiv \beta v_g$, together with (3.7c). Here

$$\zeta_g \equiv f_0 + \beta y + v_{gx} - u_{gy} \equiv f_0 + \beta y + \psi_{xx} + \psi_{yy}$$

is the geostrophic approximation to the *absolute vorticity*, $f + v_x - u_y$. The term on the right of (3.8) is called a *stretching* term, since it can generate vorticity by differential vertical motion. Next, multiplication of (3.7d) by ρ_0/θ_{0z} (a function of z, which can be taken through D_g) gives

$$D_g(\rho_0\theta/\theta_{0z}) + \rho_0 w_a = \rho_0 Q/\theta_{0z} \,. \tag{3.9}$$

The z-derivative of this can be combined with (3.8) and the relation

$$\theta/\theta_{0z} = f_0\psi_z/N^2 \,, \tag{3.10}$$

which follows from (3.5) and (3.6d), to give the important *potential vorticity equation*

$$D_g q_g = f_0\rho_0^{-1}(\rho_0 Q/\theta_{0z})_z \quad (\equiv r, \text{ say}), \tag{3.11}$$

where

$$q_g \equiv \zeta_g + f_0^2\rho_0^{-1}(\rho_0\psi_z/N^2)_z$$

$$\equiv f_0 + \beta y + \psi_{xx} + \psi_{yy} + \rho_0^{-1}(\rho_0\varepsilon\psi_z)_z \tag{3.12}$$

is the *quasigeostrophic potential vorticity* and $\varepsilon(z) \equiv f_0^2/N^2(z)$. Eq.(3.11) gives the time-development of q_g. In particular, if Q = 0 (so that the flow is adiabatic as well as frictionless) $D_g q_g = 0$, and q_g is conserved following the geostrophic wind. Given q_g we can in principle invert the operator on the

right of (3.12), using appropriate boundary conditions, to get ψ and hence u_g , v_g , θ etc., from (3.6).

Another useful equation is the *omega equation*, obtained by eliminating the $\partial/\partial t$ terms in (3.7) to get a diagnostic equation (involving no time derivatives) for w_a in terms of ψ: see, e.g., Holton (1979).

REFERENCES

Gill, A.E., 1982. *Atmosphere-ocean dynamics* (Academic Press).

Holton, J.R., 1975. *The dynamic meteorology of the stratosphere* (Met. Monograph *37*, American Meteorological Society).

Holton, J.R., 1979. *An introduction to dynamic meteorology* (Academic Press).

Pedlosky, J., 1979. *Geophysical fluid dynamics* (Springer).

Phillips, N.A., 1973. Principles of large scale numerical weather prediction. In *Dynamic meteorology*, ed. P. Morel (Reidel), pp. 1-96.

BIBLIOGRAPHY

An expanded version of the material of this lecture appears in

Andrews, D.G., Holton, J.R. and Leovy, C.B., 1987, *Middle Atmosphere Dynamics*, (Academic Press) [hereafter called AHL], §§ 3.1, 3.2.

4. ZONAL MEANS AND EDDIES

4.1 BASIC IDEAS

Many of the middle atmosphere phenomena to be considered in these lectures can be regarded as involving the two-way interaction between a mean flow and disturbances ("waves" or "eddies") that are superimposed upon it. We shall mostly be concerned with cases where the mean is a *zonal* mean, to be denoted by an overbar. On the sphere, the zonal mean is unambiguously defined as the average around a latitude circle. On the beta-plane we can attempt to mimic the sphere by supposing that everything is periodic in x, with period X, say: thus $u(x) = u(x+X)$, for example. In this case

$$\bar{u}(y,z,t) \equiv X^{-1}\int_0^X u(x,y,z,t)\, dx$$

is the zonal average. We denote the eddy or wave departure from the mean by a prime:

$$u'(x,y,z,t) \equiv u - \bar{u} .$$

It should be emphasised that the zonal mean is a purely mathematical definition. It has the advantage of being a convenient way of organising, compressing and summarising large quantities of atmospheric data. On the other hand it suffers from the disadvantage of not necessarily representing the best physical separation of eddy and mean quantities in all cases
— for example when the flow is strongly zonally-asymmetric. This is particularly true near the poles.

We now proceed to examine the equations governing zonal-mean quantities like $\bar{u}$, $\bar{\theta}$, etc. For simplicity we shall focus on quasigeostrophic flow, as described by the set (3.7). [For brevity, the suffix g on geostrophic quantities will mostly be dropped, although the suffix a will be retained.] First, (3.6a) can be used to rewrite the nonlinear terms in $D_g u_g$ and the term $\beta y v_g$ in (3.7a), giving

$$u_t + (u^2)_x + (uv)_y - f_0 v_a - \beta y \psi_x = 0.$$

Noting that $\overline{(\ldots)}_x \equiv 0$, and in particular the important fact that the zonal mean northward geostrophic wind is identically zero,

$$\bar{v}_g = \bar{v} = \bar{\psi}_x \equiv 0,$$

and using the relation $\overline{uv} = \overline{(\bar{u} + u')(\bar{v} + v')} = \bar{u}\,\bar{v} + \overline{u'v'} = \overline{u'v'}$ (since $\overline{u'} = 0$, $\bar{v} = 0$), we obtain the zonal-mean zonal momentum equation

$$\bar{u}_t - f_0\bar{v}_a = -\,(\overline{v'u'})_y \,. \tag{4.1a}$$

Likewise, from (3.7d,c) and (3.6e) we get the zonal-mean thermodynamic, continuity and thermal wind equations

$$\bar{\theta}_t + \theta_{0z}\bar{w}_a = -(\overline{v'\theta'})_y + \bar{Q} \,, \tag{4.1b}$$

$$\bar{v}_{ay} + \rho_0^{-1}(\rho_0\bar{w}_a)_z = 0 \,, \tag{4.1c}$$

$$f_0\bar{u}_z = -\,RH^{-1}e^{-\kappa z/H}\,\bar{\theta}_y \,. \tag{4.1d}$$

The terms $\overline{v'u'}$ and $\overline{v'\theta'}$ are called the northward *eddy fluxes of momentum and heat*, respectively (more correctly of zonal velocity and potential temperature). Supposing these fluxes and $\bar{Q}$ to be given, and suitable boundary conditions to be imposed, we have a closed set of four equations for $\{\bar{u}, \bar{\theta}, \bar{v}_a, \bar{w}_a\}$. (The zonal mean of the y-momentum equation (3.7b) gives $\bar{u}_a$, if further eddy fluxes are specified; however, this quantity will not be needed here.) From (4.1) [using (4.7), below] or from the zonal mean of (3.11), it is easy to derive the mean potential vorticity equation

$$\bar{q}_t + (\overline{v'q'})_y = f_0\rho_0^{-1}(\rho_0\bar{Q}/\theta_{0z})_z \,. \tag{4.2}$$

4.2 TRANSFORMED EULERIAN-MEAN EQUATIONS

The zonal average defined above is an example of an *Eulerian* mean, since it averages over a set of points fixed in (x,y,z)-space. Likewise, eqs.(4.1) are sometimes called the Eulerian-mean equations. There are some advantages, however, in transforming the latter, as follows.

First, we define a *residual mean meridional circulation* ($\bar{v}^*$, $\bar{w}^*$) by

$$\bar{v}^* \equiv \bar{v}_a - \rho_0^{-1}(\rho_0\overline{v'\theta'}/\theta_{0z})_z \ , \ \bar{w}^* \equiv \bar{w}_a + (\overline{v'\theta'}/\theta_{0z})_y \ .$$

These can be used to substitute for $(\bar{v}_a, \bar{w}_a)$ in (4.1) to obtain the *transformed Eulerian-mean* (TEM) equations

$$\bar{u}_t - f_0\bar{v}^* = \rho_0^{-1}\nabla.\underset{\sim}{F} \ , \tag{4.3a}$$

$$\bar{\theta}_t + \theta_{0z}\bar{w}^* = \bar{Q} \ , \tag{4.3b}$$

$$\bar{v}^*_y + \rho_0^{-1}(\rho_0\bar{w}^*)_z = 0 \ , \tag{4.3c}$$

and, as before,

$$f_0\bar{u}_z = -RH^{-1}e^{-\kappa z/H}\bar{\theta}_y \ . \tag{4.3d}$$

Here

$$\nabla.\underset{\sim}{F} \equiv -(\rho_0\overline{v'u'})_y + (\rho_0 f_0\overline{v'\theta'}/\theta_{0z})_z \tag{4.4a}$$

is the divergence of the quasigeostrophic form of the *Eliassen-Palm flux* (EP flux) [see Eliassen and Palm, 1961],

$$\underset{\sim}{F} \equiv (0, -\rho_0\overline{v'u'}, \rho_0 f_0\overline{v'\theta'}/\theta_{0z}) \ . \tag{4.4b}$$

Note that in (4.3) the only explicit appearance of eddy fluxes is in $\rho_0^{-1}\nabla.\underset{\sim}{F}$ in the zonal momentum equation (4.3a): no eddy-forcing terms occur on the right of the thermodynamic equation (4.3b). Thus, as far as their effects on the mean-flow "tendencies" $(\bar{u}_t, \bar{\theta}_t)$ and residual circulation $(\bar{v}^*, \bar{w}^*)$ are concerned, the eddy momentum and heat flux do not act separately [as might be thought from (4.1)], but in the combination $\nabla.\underset{\sim}{F}$ given by (4.4a). This is emphasised by solving (4.3) [or (4.1)] to find the mean tendencies and the residual circulation. For example it is easy to show that

$$\rho_0[\partial^2/\partial y^2 + \rho_0^{-1}\partial/\partial z(\rho_0\varepsilon\partial/\partial z)]\bar{u}_t = \{\nabla.\underset{\sim}{F}\}_{yy} - (\rho_0 f_0\bar{Q}/\theta_{0z})_{yz} \ , \tag{4.5}$$

$$\rho_0[\partial^2/\partial y^2 + \rho_0^{-1}\partial/\partial z(\rho_0\varepsilon\partial/\partial z)]f_0\bar{v}^* = -\{\rho_0\varepsilon(\nabla.\underset{\sim}{F}/\rho_0)_z\}_z - (\rho_0 f_0\bar{Q}/\theta_{0z})_{yz}. \tag{4.6}$$

Here the "rectified" effects of the eddies are expressed by the terms involving $\nabla.\underline{F}$ on the right, and the effects of radiative heating by the terms in $\bar{Q}$. Note that these forcing terms produce *nonlocal* responses in $\bar{u}_t$ and $f_0\bar{v}^*$, since the operator in square brackets on the left is elliptic. To solve for $\bar{u}_t$ and $f_0\bar{v}^*$, given the right-hand sides, boundary conditions must be imposed.

An important alternative form for $\rho_0^{-1}\nabla.\underline{F}$ can be derived by simple manipulations using the following identities (which stem from (3.6a), (3.10) and (3.12)):

$$u' = -\psi'_y \ , \quad v' = \psi'_x \ , \quad \theta'/\theta_{0z} = f_0\psi'_z/N^2,$$

$$q' = \psi'_{xx} + \psi'_{yy} + \rho_0^{-1}(\rho_0\varepsilon\psi'_z)_z \ .$$

Integrations by parts and the fact that $\overline{(\ldots)}_x = 0$ give

$$\overline{v'q'} = \rho_0^{-1}\nabla.\underline{F} \ ; \qquad (4.7)$$

the term on the left is the northward eddy flux of potential vorticity. [Note that (4.7) implies that (4.5) is the y-derivative of (4.2).]

4.3 THE GENERALISED ELIASSEN-PALM RELATION

We have seen that $\nabla.\underline{F}$ plays an important role in the forcing of the mean flow by the waves. It will now be shown that this quantity also figures in an equation describing the propagation of the waves themselves. This will reveal the kinds of physical process which contribute to $\nabla.\underline{F}$.

We start with the potential vorticity equation (3.11). We drop the suffix g and split each quantity into a mean and an eddy part, e.g. :

$$u = \bar{u} + u' \ , \quad v = \bar{v} + v' = v' \ , \quad r = \bar{r} + r' = r' \ ,$$

where $\bar{v} \equiv 0$ identically, as before, and $\bar{r}$ will be *assumed* zero. We also suppose that the primed quantities u' , v' , etc., are small compared with $\bar{u}$ in the sense that $|u'| \,/\, |\bar{u}| = O(a)$, $a \ll 1$, say. Then, to leading order in a we obtain the linearised equation

$$(\partial/\partial t + \bar{u}\partial/\partial x)q' + v'\bar{q}_y = r' + O(a^2) \qquad (4.8)$$

[$\bar{q}_x \equiv 0$, of course, and $\bar{q}_t = O(a^2)$ by (4.2) with $\bar{r} = 0$]. By taking $q'/\bar{q}_y$ times (4.8) and averaging and manipulating, we obtain

$$\partial(\tfrac{1}{2}\rho_0\overline{q'^2}/\bar{q}_y)/\partial t + \nabla.\underline{F} = \rho_0\,\overline{r'q'}/\bar{q}_y + O(a^3)\ , \qquad (4.9)$$

using (4.7). This is the quasigeostrophic version of the *generalised Eliassen-Palm theorem* (GEP theorem) [Andrews and McIntyre, 1976; McIntyre 1980; Andrews 1987, etc.]. When dissipation is absent ($r' = 0$) and $O(a^3)$ terms are neglected, it takes the form of a *conservation law*

$$\partial A/\partial t + \nabla.\underline{F} = 0\ ; \qquad (4.10)$$

such conservation laws are of fundamental importance in many branches of physics. The "density" $A = \tfrac{1}{2}\rho_0\overline{q'^2}/\bar{q}_y$ is sometimes called the (quasigeostrophic) *wave-activity density*: when $\bar{q}_y > 0$ it is a useful measure of wave amplitude. Eq. (4.9) shows how $\nabla.\underline{F}$ [and hence the eddy-forcing of mean-flow changes as given, for example, by (4.5)] depends on *wave transience* (the $\partial/\partial t$ term), *wave dissipation or forcing* (the r' term: collectively called *non-conservative* wave effects) and *wave nonlinearity* ($O(a^3)$ terms). Eliassen and Palm (1961) considered the case when all three of these are absent, and showed that $\nabla.\underline{F} = 0$ then.

A useful alternative form of A is in terms of the *northward parcel displacement* η' , defined by

$$(\partial/\partial t + \bar{u}\partial/\partial x)\eta' = v' + O(a^2)\ . \qquad (4.11)$$

From (4.8) and (4.11) it follows that $q' = -\eta'\bar{q}_y + O(a^2)$ if $r' = 0$, and hence $A = \tfrac{1}{2}\rho_0\overline{\eta'^2}\bar{q}_y$. This can be used in (4.9) when $r' = 0$ provided that a suitable change is made to the non-conservative term on the right.

It will be noticed that for *steady, conservative, linear* waves, when Eliassen and Palm's result $\nabla.\underline{F} = 0$ holds, the wave-forcing on the right of the TEM equations (4.3) vanishes. Provided that boundary conditions are similarly unforced, and if $\bar{Q} = 0$, it can easily be seen that a possible state is the

steady one $\bar{u}_t = \bar{\theta}_t = \bar{v}^* = \bar{w}^* = 0$. This is an example of the *nonacceleration theorem* or *Charney-Drazin theorem*. Note that in this case we may still have nonzero eddy-flux terms on the right of the untransformed eqs.(4.1), and $\bar{v}_a$, $\bar{w}_a \neq 0$.

REFERENCES

Andrews, D.G., 1987. On the interpretation of the Eliassen-Palm flux divergence. *Quart. J. R. Met. Soc.* **113** 323-338.

Andrews, D.G. and McIntyre, M.E., 1976. Planetary waves in horizontal and vertical shear: the generalized Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.* **33**, 2031-2048.

Eliassen, A. and Palm, E., 1961. On the transfer of energy in stationary mountain waves. *Geofys. Publ.* **22**, No. 3, 1-23.

McIntyre, M.E., 1980. An introduction to the generalized Lagrangian-mean description of wave, mean-flow interaction. *Pure Appl. Geophys.* **118**, 152-176.

BIBLIOGRAPHY

For further details, see AHL, §§ 3.3, 3.5, 3.6.

5. SOME SIMPLE ZONALLY-AVERAGED MODELS

In this lecture we shall consider some very simple "back of the envelope" models of the stratosphere and mesosphere, in an attempt to isolate the basic physical and dynamical processes that determine the zonally-averaged state of this part of the atmosphere. These models will turn out to illuminate some fundamental principles which can be obscured by more complex models. They also point the way to methods of interpreting the more complicated models and the atmosphere itself.

Three different models will be considered; each will be based on a slightly modified version of the TEM equations (4.3), namely

$$\bar{u}_t - f_0\bar{v}^* = \rho_0^{-1}\nabla.\underline{F} + \bar{X} \equiv D \ , \tag{5.1a}$$

$$\bar{\theta}_t + \theta_{0z}\bar{w}^* = \bar{Q} \ , \tag{5.1b}$$

$$\bar{v}^*_y + \rho_0^{-1}(\rho_0\bar{w}^*)_z = 0 \ , \tag{5.1c}$$

$$f_0\bar{u}_z = -RH^{-1}e^{-\kappa z/H}\,\bar{\theta}_y \ . \tag{5.1d}$$

Here an extra "friction" term $\bar{X}$ is included, which will represent an extra forcing effect. It will later (in Lecture 7) be suggested that a possible physical process producing a significant $\bar{X}$ in the mesosphere is the "breaking" of small-scale gravity waves there. (Such waves are not described by quasi-geostrophic dynamics, and their effects are not formally included in $\nabla.\underline{F}$ here.)

5.1 STEADY-STATE MODEL

If we set $\partial/\partial t = 0$ in (5.1a,b) we obtain

$$-f_0\bar{v}^* = D \ , \quad \theta_{0z}\bar{w}^* = \bar{Q} \tag{5.2a,b}$$

which, when substituted into (5.1c) give

$$-f_0^{-1}D_y + \rho_0^{-1}(\rho_0\bar{Q}/\theta_{0z})_z = 0 \ , \tag{5.2c}$$

showing how the net radiative heating must be related to the wave and frictional effects $\rho_0^{-1}\nabla.\underline{F} + \overline{X} = D$.

If waves and friction are absent, so that D = 0, then $\overline{v}^* = 0$ by (5.2a); from (5.1c) $\overline{w}^*$ must therefore also vanish, if it is not to grow exponentially with z. Then (5.2b) implies that the atmosphere will be in *radiative equilibrium*, with $\overline{Q} = 0$ and $\overline{T} = T_r(y,z)$, say. Associated with this radiative equilibrium temperature we must have a zonal wind $u_r(y,z)$ satisfying thermal wind balance,

$$f_0 \partial u_r/\partial z = - RH^{-1} \partial T_r/\partial y \ .$$

(We can define u_r completely by requiring it to vanish, or equal a zonal-mean climatological value, at z = 0.)

The situation changes if D is nonzero and y-dependent; by (5.2c) $\overline{Q}$ cannot generally be zero now, so $\overline{T} \neq T_r$ and $\overline{u} \neq u_r$. Moreover $\overline{v}^*$ and $\overline{w}^*$ are now nonzero, by (5.2a,b). Because of the balance between $\theta_{0z}\overline{w}^*$ and $\overline{Q}$ in (5.2b), $(\overline{v}^*, \overline{w}^*)$ is sometimes called a *diabatic circulation* in the present context, and in the past has often been thought of as "driven" by the net radiative heating. However, it is in fact more appropriate to regard this circulation and the net heating as *eddy-driven*, since they both vanish when D = 0, as shown above.

5.2 TIME-DEPENDENT, ZONALLY-AVERAGED MODEL WITH NO WAVES OR FRICTION

Suppose we now set D = 0, but allow time-dependence by taking $\overline{Q} = \overline{Q}(y,z,t)$. This could be due to the seasonal variations in solar position, for example. More precisely, we can regard $T_r = T_r(y,z,t)$ as the "radiatively-determined" temperature calculated from a time-dependent radiative (and perhaps photochemical) model: examples are shown in Figs. 5.1a and 5.2a. Putting D = 0 in (5.1) we obtain a version of (4.5):

$$[\partial^2/\partial y^2 + \rho_0^{-1}\partial/\partial z(\rho_0 \varepsilon \partial/\partial z)]\overline{u}_t = -\rho_0^{-1}(\rho_0 f_0 \overline{Q}/\theta_{0z})_{yz} \ . \qquad (5.3)$$

Further progress is difficult without a simple parameterization of $\overline{Q}$. We

shall adopt the *Newtonian cooling* form

$$\bar{Q} \equiv (J/c_p)e^{\kappa z/H} = -(\bar{\theta} - \theta_r)/\tau_r$$

where $\theta_r(y,z,t) = e^{\kappa z/H}T_r$ and $\tau_r(z)$ is a radiative relaxation time. This parameterization mimics, albeit very crudely, the "radiative spring" (Fels, 1985) whereby temperatures are relaxed towards their radiatively-determined values. (It also clearly illustrates how net heating arises when the actual state is out of radiative balance.) Using the thermal wind equation and a little manipulation, we then obtain

$$[\partial^2/\partial y^2 + \rho_0^{-1}\partial/\partial z(\rho_0\varepsilon\partial/\partial z)]\bar{u}_t + \rho_0^{-1}\{\rho_0\varepsilon(\bar{u} - u_r)_z/\tau_r\}_z = 0 \ . \qquad (5.4)$$

In this equation we can regard that part of the radiative term involving u_r as providing the *forcing*, while $\bar{u}$ represents the *response*. Since $\partial T_r/\partial y \propto \partial u_r/\partial z$ this forcing is zero unless latitudinal gradients of T_r are present. In general (5.4) suggests that $\bar{u}$ will follow u_r, but will be somewhat lagged in time and somewhat differently distributed in space. The dynamics provides a kind of "inertia", and since $\bar{u} \neq u_r$ in general, $\bar{T} \neq T_r$, so $\bar{Q} \neq 0$ and hence $(\bar{v}^*, \bar{w}^*) \neq 0$ by (4.6) and (5.1c). The non-vanishing of $\bar{Q}$ and the residual circulation again depend on *dynamical*, rather than purely *radiative*, processes.

It is interesting to investigate how good a model this is of the zonally-averaged middle atmosphere. This can be done by simple scaling arguments. We assume that L is a typical y-scale and that h = O(H) is a typical height scale. Very roughly we have $f_0^2L^2 \sim N^2H^2$ in much of the middle atmosphere, so that the operator in square brackets in (5.4) is $O(L^{-2})$. Likewise $\rho_0^{-1}\partial_z(\rho_0\varepsilon\partial_z)$ is $O(f_0^2N^{-2}H^{-2}) \sim O(L^{-2})$ so that, from (5.4),

$$\bar{u}_t \sim |u_r - \bar{u}|/\tau_r$$

in crude order-of-magnitude terms. However, if $\Delta\bar{u}$ is the maximum annual variation of $\bar{u}$ and τ is a seasonal timescale (say 3 months) we have $\bar{u}_t \sim \Delta\bar{u}/\tau$. On the other hand typical relaxation times τ_r are a few days ($\ll \tau$), except perhaps in the polar night, so that

$$|u_r - \bar{u}| \sim (\tau_r/\tau)\Delta\bar{u} \ll \Delta\bar{u} \ ,$$

and departures of $\bar{u}$ from the radiatively-determined value u_r are much less than the maximum annual swing $\Delta\bar{u}$ in this model. In this sense we can say that $\bar{u} \approx u_r$; a similar argument (see AHL § 7.2.2) shows that $\bar{T} \approx T_r$. (This would not be true if $\tau \sim \tau_r$.) However, by comparing Fig. 5.1a with 5.1b and Fig. 5.2a with 5.2b, we can see that the middle atmosphere is *not* everywhere close to radiative balance. For example $\bar{T}$ is far from T_r in the polar night stratosphere and in the upper mesosphere. These discrepancies mean that the present model fails to explain some of the most prominent features of the observed middle atmosphere. We therefore turn to a more complex model.

5.3 TIME-DEPENDENT, ZONALLY-AVERAGED MODEL, INCLUDING WAVES AND FRICTION

If we include waves and time-dependence, and retain the Newtonian cooling expression for $\bar{Q}$, (5.4) is replaced by

$$\underset{[1]}{[\partial^2/\partial y^2 + \rho_0^{-1}\partial/\partial z(\rho_0\varepsilon\partial/\partial z)]\bar{u}_t} + \underset{[2]}{\rho_0^{-1}\{\rho_0\varepsilon(\bar{u} - u_r)_z/\tau_r\}_z} = \underset{[3]}{D_{yy}} \qquad (5.5)$$

[cf.(4.5)]. We introduce a third timescale τ_w , representing the time over which the wave or friction term D varies. Except for rapid wave events like sudden warmings we have $\tau_w \sim \tau \gg \tau_r$; alternatively we can average over a time $\tau_w \sim \tau$ to smooth out such events. A scaling argument like that in §5.2 gives the ratio of terms in (5.5) as

$$[1] : [2] : [3] = \bar{u}/\tau : |u_r - \bar{u}|/\tau_r : \Delta D$$

if $f_0^2L^2 \sim N^2H^2$ and ΔD is the variation in D over time τ_w . In §5.2 we saw that $|u_r - \bar{u}|/\tau_r \gg \bar{u}/\tau$ in the polar night stratosphere or the upper mesosphere, so here the balance must be between terms [2] and [3]. Thus we expect

$$\rho_0^{-1}\{\rho_0\varepsilon(\bar{u} - u_r)_z/\tau_r\}_z = D_{yy} \qquad (5.6)$$

to hold approximately in such regions; in other words the term in $\bar{u}_t$ in (5.5) is small, so the balance examined in §5.1 and expressed by (5.2) holds approximately there. Indeed (5.6) can be obtained from the y-derivative of (5.2c) if $\bar{Q}$ is given the Newtonian cooling form and thermal wind balance is

used.

These simple models suggest that wave motions and friction, as represented by the term $D \equiv \rho_0^{-1}\nabla.\underline{F} + \overline{X}$, are likely to be important in various parts of the middle atmosphere for accounting for departures from the radiatively-determined state. There is clearly much scope for more sophisticated models, including for example better radiative parameterizations, coupled photochemistry, more detailed dynamics (using the primitive equations in spherical geometry[1]) and so on, up to the full two- and three-dimensional general circulation models (GCMs) of the middle atmosphere that are currently in operation (see Lecture 10). Nevertheless it is important to bridge the gap between the simple models described in this lecture and the GCMs with a hierarchy of intermediate models.

The question remains as to what wave motions are responsible for producing the requisite D. Candidates include:

(a) Planetary (Rossby) waves in the winter stratosphere,

(b) Gravity waves and tides in the upper mesosphere.

These will be discussed in later lectures. We note here that the GEP theorem (4.9) has already pinpointed the important physical processes that give a nonzero $\nabla.\underline{F}$, namely departures of the wave motion from steady, conservative, linear conditions. Transient effects are to some extent averaged out in the approach outlined above, and nonlinearity and dissipative processes associated with the waves are likely to be the most important factors determining $\nabla.\underline{F}$; a similar statement is true of $\overline{X}$, also.

REFERENCES

Barnett, J.J. and Corney, M., 1985. Middle atmosphere reference model derived from satellite data. *Handbook for MAP*, Middle Atmosphere Program, Vol. 16, pp. 47-85.

Fels, S.B., 1985. Radiative-dynamical interactions in the middle atmosphere. *Adv. Geophys.* **28A**, 277-300.

Held, I.M. and Hou, A.Y., 1980. Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.* **37**, 515-533.

---

[1]For example the set (5.1) does not apply near the equator, and other dynamical regimes are possible, such as the nonlinear Hadley circulations studied by Held and Hou (1980), which are thermally-driven and almost inviscid.

Shine, K.P., 1987. The middle atmosphere in the absence of dynamical heat fluxes. *Quart. J. Roy. Met. Soc.* 113, 603-633.

BIBLIOGRAPHY

For further detail, see

Andrews, D.G., 1987. The influence of atmospheric waves on the general circulation of the middle atmosphere. To appear in *Phil. Trans. Roy. Soc. A.* (Also Met. O. 20, DCTN 50.)

AHL, Chapter 7.

World Meteorological Organization: Global Ozone Research and Monitoring Project. Report No. 16: Atmospheric Ozone 1985. (Hereafter called WMO_3.)

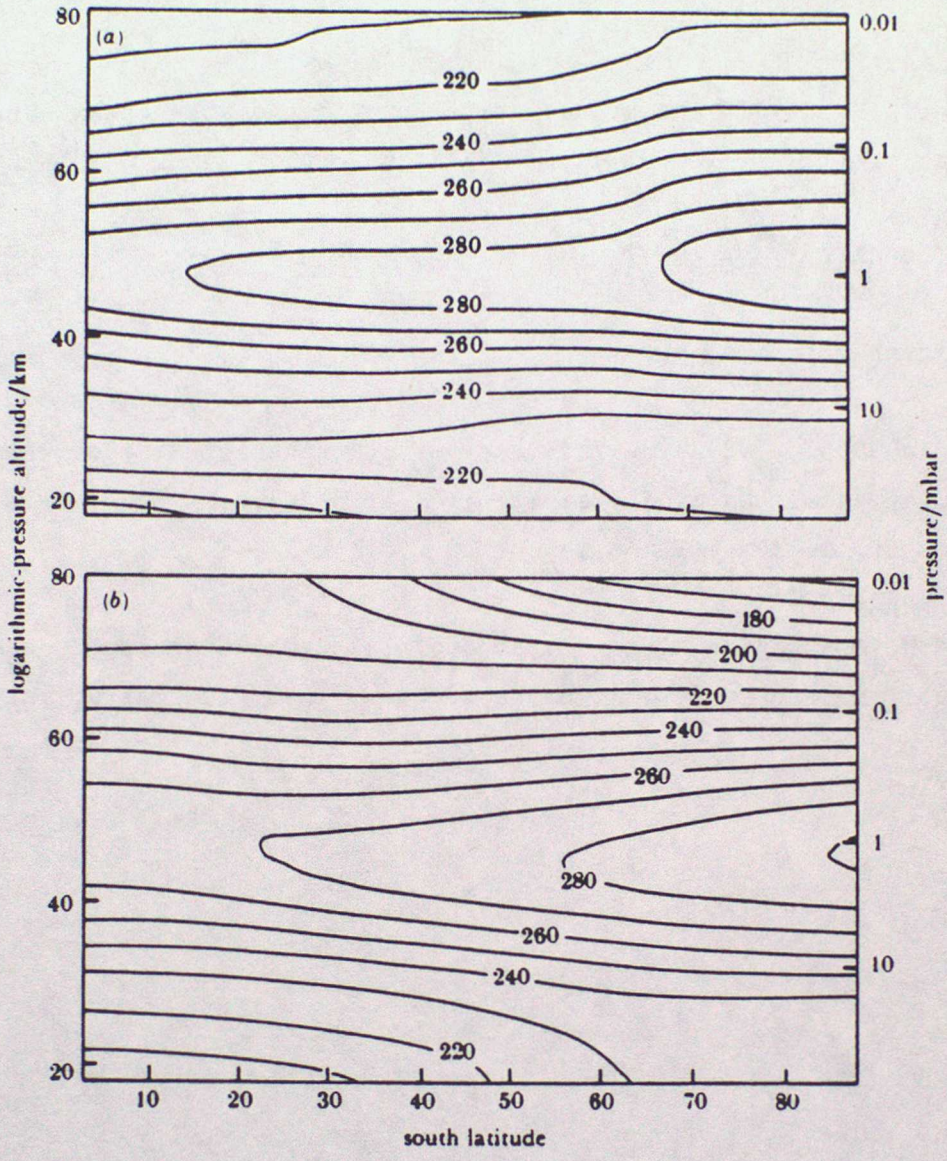


Fig. 5.1. (a) Radiatively-determined temperatures (K) for the Southern Hemisphere middle atmosphere on 21 December (near summer solstice). (b) Observed monthly-mean, zonal-mean temperatures for the Southern Hemisphere middle atmosphere in December, interpolated from the climatology of Barnett and Corney (1985). (Redrawn, with some smoothing of contours in (a), from Shine 1987.)

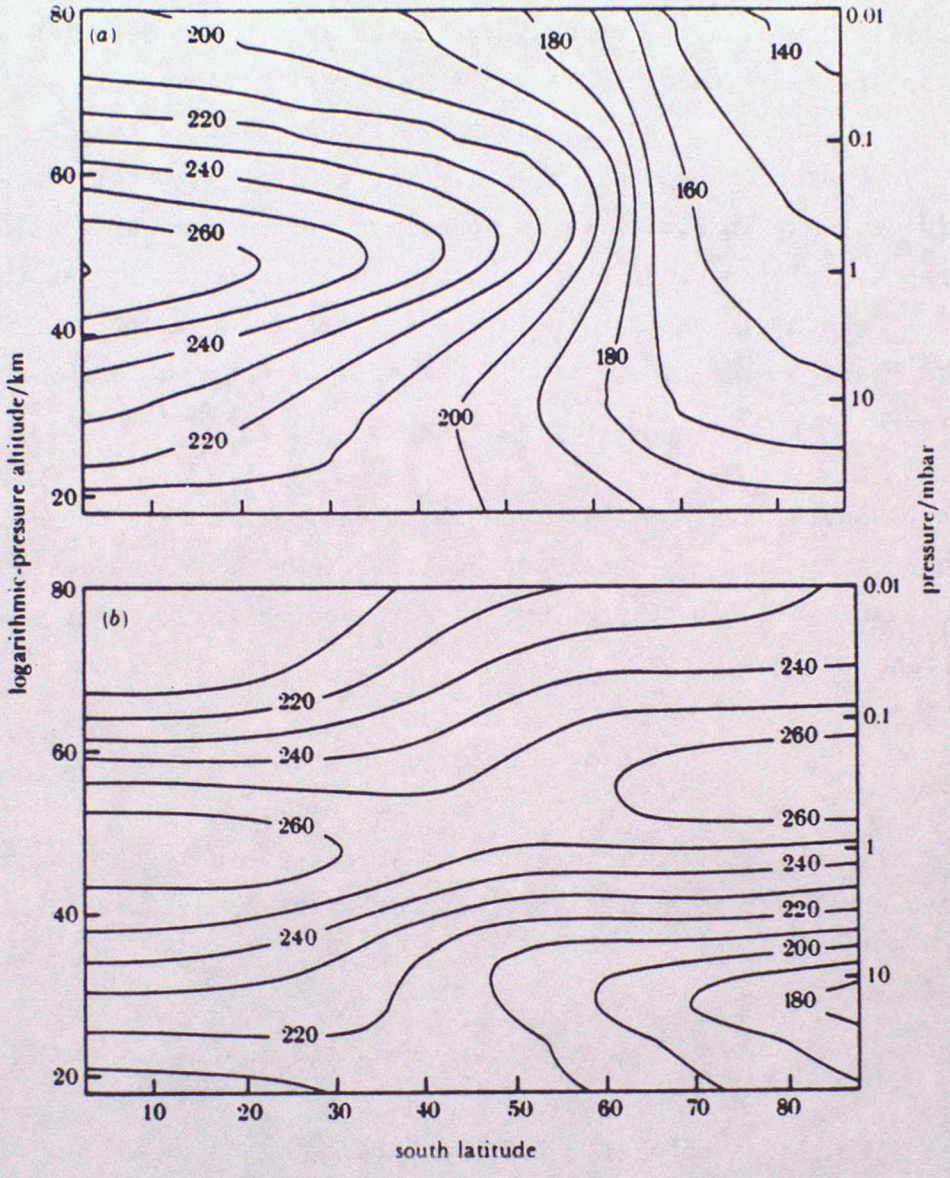


Fig. 5.2. (a) As Fig. 5.1a but on 19 June (near winter solstice). (b) As Fig. 5.1b but for June.

6. EXTRATROPICAL PLANETARY WAVES

Planetary, or Rossby, wave disturbances are common features of the winter stratosphere as revealed for example by satellite data. They can be separated into *stationary* waves, whose phases are fixed relative to the ground, and *travelling* waves, whose phases move. The kind of pattern that appears in the monthly-mean map in Fig. 1.4a is sometimes regarded as a "stationary-wave" pattern; travelling waves have been mentioned in § 2.2.2. Another classification is in terms of whether the waves are *forced* or *unforced*. We shall consider forced planetary waves (especially of the stationary variety) in this lecture.

Global-scale waves of this type obey quasigeostrophic dynamics to quite a good approximation. We shall suppose that they represent small departures from a mean zonal flow $\bar{u}(y,z)$, so that the potential vorticity equation (3.11) linearises to

$$(\partial/\partial t + \bar{u}\partial/\partial x)q' + v'\bar{q}_y = 0 \ , \tag{6.1}$$

[cf. (4.8)], where diabatic and nonlinear effects have been neglected. Here the mean northward potential vorticity gradient (sometimes called "effective beta") is

$$\bar{q}_y = \beta - \bar{u}_{yy} - \rho_0^{-1}(\rho_0 \varepsilon \bar{u}_z)_z \tag{6.2}$$

and the disturbance potential vorticity is

$$q' = \psi'_{xx} + \psi'_{yy} + \rho_0^{-1}(\rho_0 \varepsilon \psi'_z)_z \ ; \tag{6.3}$$

these derive from (3.12) and (3.6a).

For simplicity, we suppose that the waves are forced from below by fluctuations in the height of some isobaric surface $p = p_0$, perhaps in the lower stratosphere near 100 mb. Thus we specify $\Phi(x,y,z_0,t)$, where $z_0 = -H \ln(p_0/p_s)$; if we choose the reference pressure p_s to equal p_0, then $z_0 = 0$. In terms of ψ [cf. (3.6b)] the linearised lower boundary condition can then be written

$$\psi' = \psi_0'(x,y,t) \text{ at } z = 0 . \tag{6.4}$$

We shall impose vertical sidewalls at latitudes y = 0, L, so that

$$\psi_x' = v' = 0 \text{ at } y = 0,L \tag{6.5}$$

[see §3.2.3]; the upper boundary condition [§3.2.2] remains to be investigated.

6.1 SIMPLE EXAMPLES

To make things as easy as possible, we first restrict attention to a uniform mean zonal flow; thus $\bar{u}$ = constant and $\bar{q}_y = \beta$. We also take N^2 = constant, so that ε = constant. This was one of the cases investigated in the celebrated paper by Charney and Drazin (1961). We take the boundary-forcing term to be

$$\psi_0' = \text{Re } \hat{\psi}_0 \, e^{ik(x-ct)} \sin ly , \tag{6.6}$$

which represents a wavy pattern of zonal wavelength $2\pi k^{-1}$ and meridional wavelength $2\pi l^{-1}$, moving with phase speed c (eastwards if $c > 0$, westwards if $c < 0$). For consistency with this we look for solutions of (6.1) in the form

$$\psi' = \text{Re } \hat{\psi}(z) \, e^{ik(x-ct)} \sin ly , \tag{6.7}$$

which satisfies (6.4) if $\hat{\psi}(0) = \hat{\psi}_0$ and (6.5) if $l = n\pi L^{-1}$, where n is a nonzero integer [or if sin ly is replaced by unity in (6.6) and (6.7)]. Substitution of (6.7) into (6.1) and use of (6.3) leads to the following second-order ordinary differential equation for $\hat{\psi}(z)$:

$$\hat{\psi}_{zz} - H^{-1}\hat{\psi}_z + B\hat{\psi} = 0 , \tag{6.8}$$

where

$$B \equiv \varepsilon^{-1}[\beta(\bar{u} - c)^{-1} - (k^2 + l^2)] ,$$

and where we have used the definition $\rho_0 = \rho_s e^{-z/H}$. The lower boundary

condition is

$$\hat{\psi} = \hat{\psi}_0 \text{ at } z = 0 \ , \tag{6.9}$$

while the upper boundary condition is to be determined.

Looking for solutions $\hat{\psi} \propto e^{\lambda z}$ we find

$$\lambda = (2H)^{-1} \pm [(2H)^{-2} - B]^{\frac{1}{2}} \ .$$

Two possibilities arise, according as the term inside the radical is positive or negative:

(a) $(2H)^{-2} - B \equiv \mu^2 > 0$: In this case $\lambda = (2H)^{-1} \pm \mu$, where $\mu = + [(2H)^{-2} - B]^{\frac{1}{2}}$, and $\hat{\psi} = \hat{\psi}_0 \exp([z/2H] \pm \mu z)$. The wave kinetic energy per unit volume equals $\frac{1}{2}\rho_0(\psi_x'^2 + \psi_y'^2)$, and will be proportional to $\exp(\pm 2\mu z)$ here. (The vertical component of the EP flux has the same z-dependence.) Clearly the lower sign must be chosen, in accordance with condition (a) of §3.2.2 . Then

$$\psi' = \exp([z/2H] - \mu z)\, \psi_0'(x,y,t) \ ; \tag{6.10a}$$

this represents a *vertically-trapped* mode, and has no phase tilt with height. Note that ψ' , u' , v' etc. actually grow with height if $\mu < (2H)^{-1}$, i.e. if $B > 0$; however the decreasing density ρ_0 more than compensates for this in the kinetic energy density.

(b) $(2H)^{-2} - B \equiv - m^2 < 0$: Now $\lambda = (2H)^{-1} + im$, where $m = \pm [B - (2H)^{-2}]^{\frac{1}{2}}$ and

$$\psi' = \text{Re}\, \hat{\psi}_0 \exp\{[z/2H] + i(kx - kct + mz)\} \sin ly \ ; \tag{6.10b}$$

the presence of im in the exponent indicates vertical *propagation*; phase lines are no longer vertical, but tilt with height (see below). We can use the definitions of B and m to obtain the equation

$$\beta/(\bar{u} - c) = k^2 + l^2 + \varepsilon[m^2 + (2H)^{-2}] > 0 \ ; \tag{6.11}$$

the right-hand side is clearly positive in this propagating case ($m^2 > 0$) and hence we must have $\bar{u} > c$. In other words the phase speed must be *westward* with respect to the mean zonal wind for vertically-propagating Rossby waves to occur. Moreover, for fixed k, l, ε, H, and with m^2 positive,

$$\beta/(\bar{u} - c) > k^2 + l^2 + (\varepsilon/4H^2) ,$$

so that

$$0 < \bar{u} - c < \beta[k^2 + l^2 + (\varepsilon/4H^2)]^{-1} \equiv \bar{u}_c , \text{ say} . \qquad (6.12)$$

Thus vertical propagation only occurs if $\bar{u} - c$ is not too large. For stationary waves, which have c = 0,

$$0 < \bar{u} < \bar{u}_c \qquad (6.13)$$

for vertical propagation: that is, if c = 0, the winds must be westerly (i.e. eastward) and not too strong. Condition (6.12) is sometimes called the Charney-Drazin criterion (not to be confused with the Charney-Drazin *theorem* mentioned in Lecture 4 !).

To estimate the typical size of $\bar{u}_c$ we can take the following values, representative of midlatitudes: $\phi = 60°$, $\Omega \approx 7.3\times10^{-5}$ s^{-1}, a_0 (earth radius) = 6.4×10^6 m, $\beta = 2\Omega\cos\phi/a_0 \approx 1.1\times10^{-11}$ m^{-1} s^{-1}, $f = 2\Omega\sin\phi \approx 1.3\times10^{-4}$ s^{-1}, $H \approx 7\times10^3$ m, $N \approx 2\times10^{-2}$ s^{-1}, $\varepsilon = f^2N^{-2} \approx 4\times10^{-5}$, $l = \pi/L$, $L \approx 10^7$ m, $k = s/(a_0\cos\phi)$, where s = zonal wavenumber. We find $\bar{u}_c \approx 110/(s^2 + 3)$ m s^{-1}. The present model is a highly simplified one, and suggests rather low values of $\bar{u}_c$ (corresponding to a zonal jet that has been "smeared out" in the meridional plane). However, it does illustrate the fact that the "window" for propagation, given by (6.13), becomes smaller as the zonal wavenumber increases. A more complex theory suggests that only the "ultra-long waves" s = 1, 2, 3 have any chance of propagation up into the stratosphere in the winter westerlies; this is in qualitative agreement with the fact that observed stationary disturbances tend to be of rather large scale in the winter stratosphere. In summer, when $\bar{u} < 0$ in the stratosphere, no stationary forced disturbances are observed, and this is again in accord with the Charney-Drazin criterion.

It remains to determine the sign of m; the radiation condition of §3.2.2 is used for this purpose, since the wave kinetic energy density is here found to be *constant* with height, irrespective of the sign of m. We calculate the vertical component of the *group velocity* $c_g^{(z)}$, defined by

$$c_g^{(z)} \equiv \partial\omega/\partial m \ , \text{ where } \omega \equiv ck \ ,$$

[see texts on wave motion, e.g. Lighthill (1978), §3.6; also Gill (1982), §5.4]. This group velocity component must be chosen *positive*, so that information is carried upwards, rather than downwards. We can arrange (6.11) to obtain ω as a function of (k,l,m) (the Rossby-wave *dispersion relation*):

$$\omega = k\bar{u} - \beta k \{ k^2 + l^2 + \varepsilon[m^2 + (2H)^{-2}] \}^{-1} \ , \qquad (6.14)$$

so that

$$c_g^{(z)} = \partial\omega/\partial m = 2\varepsilon\beta k m \{ k^2 + l^2 + \varepsilon[m^2 + (2H)^{-2}] \}^{-2} \ , \qquad (6.15)$$

which is positive if $m > 0$ (we take $k > 0$ by convention). Thus $m = +[B - (2H)^{-2}]^{\frac{1}{2}}$ has the appropriate sign. The phase lines, $kx - kct + mz =$ constant, thus tilt *westwards* with height, as observed. [Note that for stationary waves we put $c = 0$ *after* differentiating ω to get $c_g^{(z)}$.]

Another way of arriving at the sign of m is to include small dissipation. A particularly simple case is that in which small Rayleigh friction and Newtonian cooling, with equal relaxation coefficients, are included: see, e.g., AHL, pp. 179-80.

6.2 MORE GENERAL $\bar{u}(y,z)$ AND $\bar{q}_y(y,z)$

Under certain assumptions, we can obtain some general information about propagation of Rossby waves in more complex shear flows. We return to (6.1) and substitute

$$\psi' = \mathrm{Re}\ \rho_0^{-\frac{1}{2}}\, \Psi(y,z)\ e^{ik(x - ct)} \ . \qquad (6.16)$$

We take $\varepsilon(z)$ = constant, for simplicity, and define a "stretched" vertical

coordinate $Z = Nz/f_0$. Then Ψ is found to satisfy

$$\Psi_{yy} + \Psi_{ZZ} + n_k^2 \Psi = 0 \ , \tag{6.17}$$

where

$$n_k^2(y,z) = (\bar{u} - c)^{-1}\bar{q}_y - k^2 - (\varepsilon/4H^2) \ . \tag{6.18}$$

The quantity n_k^2 is known as the *refractive index squared*, for zonal wavenumber k and phase speed c. Eq.(6.17) is identical to the equation for two-dimensional sound (or light) waves in a medium of varying refractive index, and we can make use of insights from the theory of acoustics or optics. In particular, we expect waves to propagate in regions where $n_k^2 > 0$ and avoid regions where $n_k^2 < 0$. Note that n_k^2 depends on the two-dimensional structure of $(\bar{u} - c)^{-1}\bar{q}_y$ as well as on k. Thus propagation generally depends on more complex criteria than the simple condition (6.12) of Charney and Drazin, which applies when $\bar{u}$ is constant.

The kind of theory that can be used to tackle (6.17) is analogous to "geometric optics", and supposes that the wavelength of Ψ in the meridional plane is much less than the meridional scale of the mean flow, and thus of n_k^2. This is the "WKBJ" or "Liouville-Green" approach. One looks for *locally-sinusoidal* solutions for Ψ , calculates a *local* group velocity $\underline{c}_g = (c_g^{(y)}, c_g^{(z)})$, and computes "rays", which are paths — generally curved — parallel to $\underline{c}_g$ at each point, along which information propagates. A set of *ray-tracing equations* allows a computation of how the local wavelengths, etc., vary along rays. Despite the approximations, this WKBJ theory is surprisingly accurate, and has provided helpful insights into the behaviour of observed and numerically-modelled planetary waves in the middle atmosphere.

6.3 BREAKING PLANETARY WAVES

The theory described above is linear, and can only be expected to apply to small-amplitude disturbances. For upward-propagating waves, the amplitude tends to grow exponentially with z [see (6.10b)] and nonlinear effects are likely to become important at a great enough height. An example of a non-linear planetary-wave-like phenomenon, observed in the middle stratosphere in

January 1979, was discussed by McIntyre and Palmer (1983, 1984, 1985). The presence of the wave led to a strong distortion of the basic cyclonic vortex (see Fig. 6.1; cf. Fig. 2.2). Maps of a quasi-conservative tracer (Ertel's potential vorticity: see Lecture 9) on the 850K isentropic surface (near 10 mb or 30 km) show a "tongue" of air over North America that has apparently been dragged out of the main vortex by the flow associated with the accompanying Aleutian High.

The rapid, irreversible deformation of air masses in this way is certainly a nonlinear process, and McIntyre and Palmer refer to the process as "planetary-wave breaking", by analogy with the breaking of surface waves on a beach (see also Lecture 7). Such events are likely to lead to strong mixing, thus having important implications for the transport of tracers (see Lecture 9) and also for the production of significant values of the eddy-induced zonal force $\rho_0^{-1}\nabla.\underline{F}$ in the winter stratosphere (hence leading to departures of the mean state from radiative balance there: see § 5.3).

6.4 STRATOSPHERIC SUDDEN WARMINGS

An important stratospheric phenomenon involving large-scale nonlinear disturbances is the sudden warming that was described in §2.2. A large body of theory has been developed to model the process in terms of a two-way interaction between planetary waves and the zonal-mean flow, using the kind of ideas introduced in Lecture 4 and §6.2. The planetary waves are assumed to be generated in the troposphere and to propagate strongly into the high-latitude stratosphere when zonal-mean conditions are right; they then bring about mean-flow changes there. This theory has successfully accounted for several aspects of the dynamics of sudden warmings, but more research will need to be done to elucidate their full nonlinear behaviour. A comprehensive discussion of this topic is beyond the scope of these lectures: for further details, see AHL, Chapter 6.

REFERENCES

Charney, J.G. and Drazin, P.G., 1961. Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.* 66, 83-109.

Gill, A.E., 1982. *Atmosphere-ocean dynamics* (Academic Press).

Lighthill, J., 1978. *Waves in fluids* (Cambridge University Press).

McIntyre, M.E. and Palmer, T.N., 1983. Breaking planetary waves in the stratosphere. *Nature* **305**, 593-600.

McIntyre, M.E. and Palmer, T.N., 1984. The "surf zone" in the stratosphere. *J. Atmos. Terrest. Phys.* **46**, 825-849.

McIntyre, M.E. and Palmer, T.N., 1985. A note on the general concept of wave breaking for Rossby and gravity waves. *Pure Appl. Geophys.* **123**, 964-975.

BIBLIOGRAPHY

For further information, see AHL, § 4.5 and Chapters 5 and 6.

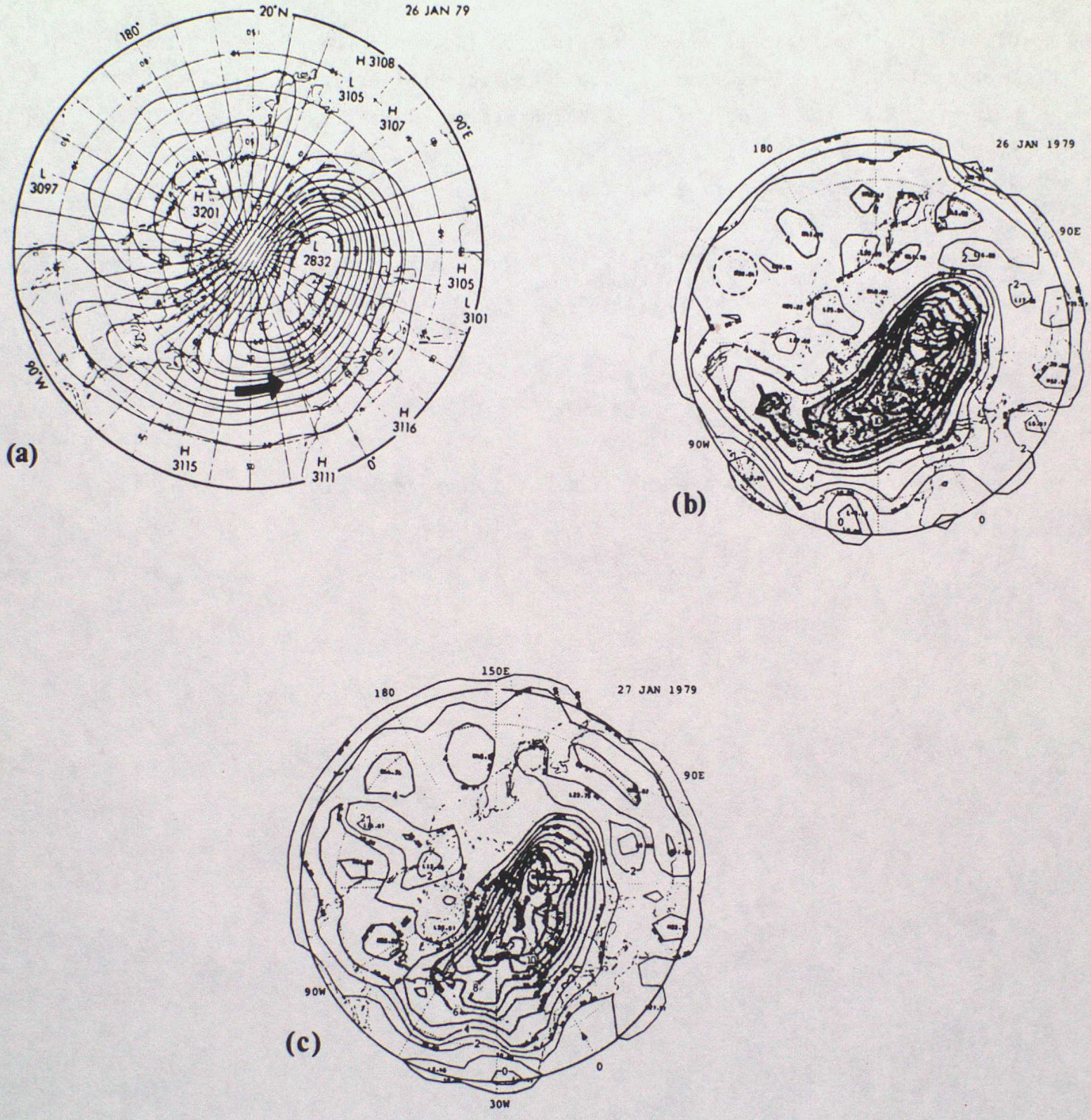


Fig. 6.1. Polar stereographic maps (outer circle: 20°N) of (a) 10 mb height (decametres) on 26 January 1979, (b) Ertel's potential vorticity on the 850K isentropic surface on 26 January 1979. (c) as for (b) but for 27 January. (From McIntyre and Palmer 1984, *q.v.* for definition of potential vorticity units.)

7. GRAVITY WAVES AND TIDES

In previous lectures it has been suggested that departures of the mean state of the polar stratosphere from radiative balance are likely to be mainly due to forcing by planetary waves. However, such an explanation cannot apply to the summer upper mesosphere, since planetary waves are of rather small amplitude there. It is believed that the deviations of the upper mesosphere from the radiatively-determined state are due principally to breaking *internal gravity waves*, and perhaps also *atmospheric tides*.

Gravity waves are a very common feature of the upper mesosphere, and some radar observations are given in Fig.7.1. They have horizontal wavelengths of up to about 100-200 km, vertical wavelengths of about 5-15 km, phase speeds of up to about 80 ms^{-1} and periods of a few minutes to an hour or so. Theory predicts that velocity and temperature amplitudes in such waves increase as $\rho_0^{-\frac{1}{2}} \propto e^{z/2H}$ as z increases, and thus they eventually "break". The process is illustrated in Fig. 7.2, which depicts some isentropes (surfaces of constant θ) which are undulating as a gravity wave propagates upwards. In the lower mesosphere, (a), the isentropes have a gentle sinusoidal variation. In the middle mesosphere, (b), isentropes are still fairly sinusoidal, but of larger amplitude. In the upper mesosphere, (c), isentropes are strongly distorted by nonlinear effects, becoming vertical at some points and leading to turbulence, small-scale mixing and dissipation. Such a process is called *gravity-wave breaking*, since it is analogous to the overturning and breaking of surface gravity waves on a shelving beach. It will tend to limit the $e^{z/2H}$ growth of the waves above the breaking level, and this has important consequences for the mean flow, as will be seen later.

The quasigeostrophic equations do not describe rapid, small-scale motions like gravity waves, so we must return to the primitive equations (3.2). Adding a forcing term F to the right of (3.2a), averaging, and using (3.2d) we obtain

$$\bar{u}_t + (\bar{u}_y - f)\bar{v} + \bar{u}_z\bar{w} = -(\overline{u'v'})_y - \rho_0^{-1}(\rho_0\overline{u'w'})_z + \bar{F}$$

in place of (4.1a). The gravity waves will turn out to make important contributions to the terms $-\rho_0^{-1}(\rho_0\overline{u'w'})_z + \bar{F} \equiv \bar{X}$, say. In considering their effects on the large-scale flow we can retain $\bar{X}$ and otherwise apply

quasigeostrophic scaling, to obtain

$$\bar{u}_t - f_0\bar{v}_a = -(\overline{u'v'})_y + \bar{X}$$

[cf.(4.1a)] or, in TEM form,

$$\bar{u}_t - f_0\bar{v}^* = \rho_0^{-1}\nabla.\underline{F} + \bar{X} ,$$

which is precisely (5.1a). Our task is now to estimate $\bar{X}$ for breaking gravity waves.

We use a simplified form of the non-rigorous but plausible approach due to Lindzen (1981). We assume the waves to be two-dimensional (so that $\partial/\partial y = 0$, $v' = 0$) and for simplicity linearise about a state of rest. Since the waves are too small in scale to feel the earth's rotation we can take $f = 0$. Then (3.2) reduce to

$$u'_t + \Phi'_x = 0 , \quad u'_x + \rho_0^{-1}(\rho_0 w')_z = 0 , \quad \Phi'_{zt} + N^2 w' = 0 , \quad (7.1a,b,c)$$

where θ', T' have been eliminated. If a lower boundary condition is

$$w' = w_0 \cos k(x - ct) \quad \text{at } z = 0$$

(where p_s can be chosen so that $z = 0$ is near the stratopause), it is easy to show that a wave solution is

$$w' = w_0\, e^{z/2H} \cos[k(x - ct) + mz] , \qquad (7.2)$$

where $m^2 = (N^2/c^2) - (2H)^{-2}$. Now it was mentioned above that $|c| \lesssim 80$ m s^{-1}, so we have

$$|c| \ll 2NH \sim 280 \text{ m s}^{-1} \qquad (7.3a)$$

(using $N \sim 2\times10^{-2}$ s^{-1} , $H \sim 7\times10^3$ m) and hence $(2H)^{-2} \ll N^2/c^2 \approx m^2$. Thus

$$|m| \gg (2H)^{-1} , \qquad (7.3b)$$

or, equivalently, the vertical wavelength (observed to be 5-15 km) satisfies

$2\pi/|m| \ll 4\pi H$ - 80 km. Thus to a good approximation we can take

$$m = -N/c \ , \tag{7.4}$$

where it can be checked that the minus sign implies an *upward* group velocity. The inequality (7.3b) also implies that the continuity equation (7.1b) reduces to $u'_x + w'_z = 0$, approximately; thus u' has a similar structure to (7.2) where $u' \approx -mw'/k$, and so by (7.2), $\rho_0 \overline{u'w'} = -\rho_s m w_0^2/2k$ = constant. Therefore

$$\rho_0^{-1}(\rho_0 \overline{u'w'})_z = 0$$

for the waves considered here. This is the primitive-equation version of the EP theorem in the present case.

We now attempt a mathematical description of the wave-breaking process. First, it can be shown from (7.1) and (7.2) that

$$\Phi' \approx (N/k) w_0 e^{z/2H} \cos[k(x - ct) + mz] \ . \tag{7.5}$$

Moreover, by (3.2c), (3.2f) and (3.5),

$$\theta_z = \theta_{0z} + \theta'_z = HR^{-1}e^{\kappa z/H}[N^2 + \Phi'_{zz} + \kappa H^{-1}\Phi'_z] \ .$$

Noting from (7.5) that Φ' grows exponentially with z, we can expect that the terms in Φ' will at a sufficient altitude become large enough to cancel N^2 at some values of x, t, at the "crests" of the waves, giving $\theta_z = 0$ there, and thus local breaking, in Lindzen's sense. This approach is not strictly self-consistent, since the linear solutions will break down before such a height is reached, and (7.5) will cease to be valid. Nevertheless we shall assume that the linear approximation is not too bad, and define a *breaking level* z_b by

$$\{\max |\Phi'_{zz} + \kappa H^{-1}\Phi'_z| \}_{z = z_b} = N^2 \ .$$

We find

$$z_b \approx 2H \ln |Nk/m^2 w_0| = 2H \ln |c^2 k/N w_0| \ , \tag{7.6}$$

using (7.4). Note that z_b depends on c and k; it increases as w_0 decreases because smaller-amplitude waves need to penetrate to greater heights before they have grown enough to break.

We next model what happens above the breaking level. The turbulence that presumably sets in leads to diffusion of heat and momentum, and a crude way of parameterizing this diffusion is to modify (7.1a,c) thus:

$$u'_t + \Phi'_x = Ku'_{zz} \ , \quad \Phi'_{zt} + N^2 w' = K\Phi'_{zzz} \ , \ z > z_b \ , \qquad (7.7a,b)$$

where K is a constant diffusion coefficient. If K is chosen so small that

$$m^2 K \ll |\omega| \ , \qquad (7.8)$$

where $\omega \equiv ck$, it is straightforward to show that above z_b

$$w' = w_0 \exp\{[z/2H] - (KN^3/c^4k)(z - z_b)\} \cos [k(x - ct) + mz] \qquad (7.9)$$

in place of (7.2): the extra exponential factor involving K results from the diffusive effects due to breaking. Lindzen hypothesised that above z_b this diffusion is just sufficient to keep the waves *saturated*, i.e. on the verge of breaking; thus $\max \left|\Phi'_{zz} + \kappa H^{-1}\Phi'_z\right| = N^2$ for all $z \geq z_b$. This saturation condition means that diffusive decay is just such as to balance the $\rho_0^{-\frac{1}{2}} \propto e^{z/2H}$ growth; thus K must be chosen such that the coefficient of z in the exponential term in (7.9) vanishes, that is

$$K = c^4 k/2HN^3 \ . \qquad (7.10)$$

[This can be shown to satisfy (7.8).] Substitution of this value of K in (7.9) gives

$$w' = w_0 \exp(z_b/2H) \cos [k(x - ct) + mz] \quad \text{for } z \geq z_b \ ; \qquad (7.11)$$

thus Lindzen's hypothesis implies no further growth in amplitude above z_b ; this is roughly consistent with observed gravity-wave amplitudes in the mesosphere. By continuity we still have $u' \approx -mw'/k$ above z_b and it can be shown, using (7.11), (7.6) and (7.4) that

$$B \equiv -\rho_0^{-1}(\rho_0\overline{u'w'})_z = c^3k/2NH = N^2K/c = \text{constant} \neq 0, \text{ above } z_b \ . \quad (7.12)$$

Thus, if $c > 0$, there is a vertical wave momentum flux convergence or EP flux divergence above the breaking level, and vice versa if $c < 0$.

Lindzen also postulates that a similar diffusion acts on the mean flow. Thus $\bar{F}$, mentioned earlier, is chosen as $K\bar{u}_{zz}$, and a similar diffusive term $K\bar{\theta}_{zz}$ is supposed to appear in the mean thermodynamic equation. (However, it is not clear that the Ks should really be the same in each term.) The full parameterization for $\bar{X}$ is

$$\bar{X} \equiv -\rho_0^{-1}(\rho_0\overline{u'w'})_z + \bar{F} = (c^3k/2NH) + (c^4k/2N^3H)\bar{u}_{zz} \ ,$$

the sum of a constant force per unit mass and a diffusive term.

A crude estimate of B (using $c = 10$ m s^{-1}, $k = 2\pi/(50$ km), $H = 7$ km, $N = 2\times10^{-2}$ s^{-1}) is about 40 m s^{-1} day^{-1}, a considerable acceleration. Likewise $K \sim 10$ m^2 s^{-1}. The breaking level tends to be in the lower or middle mesosphere in winter and in the upper mesosphere in summer. Note that breaking is likely to occur only intermittently, being concentrated near the "crests" of the waves, and the estimates of B and K may need scaling down significantly to account for this.

Lindzen's (1981) paper presents a more general treatment, using WKBJ methods for waves in a basic shear flow $\bar{u}(z)$. He also considers the effects of atmospheric tides, which are essentially global-scale gravity waves forced by the daily variation of the solar heating. These, too, are expected to break in the upper mesosphere, and may lead to values of B on the order of 10-20 m s^{-1} day^{-1} above this level in equatorial regions.

REFERENCES

Lindzen, R.S., 1981. Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.* 86, 9707-9714.

Vincent, R.A. and Reid, I.M., 1983. HF Doppler measurements of mesospheric

gravity wave momentum fluxes. *J. Atmos. Sci.* **40**, 1321-1333.

BIBLIOGRAPHY

For more on gravity waves, see AHL, § 4.6, and the excellent review:

Fritts, D.C., 1984. Gravity wave saturation in the middle atmosphere: a review of theory and observations. *Revs. Geophys. Space Phys.* **22**, 275-308.

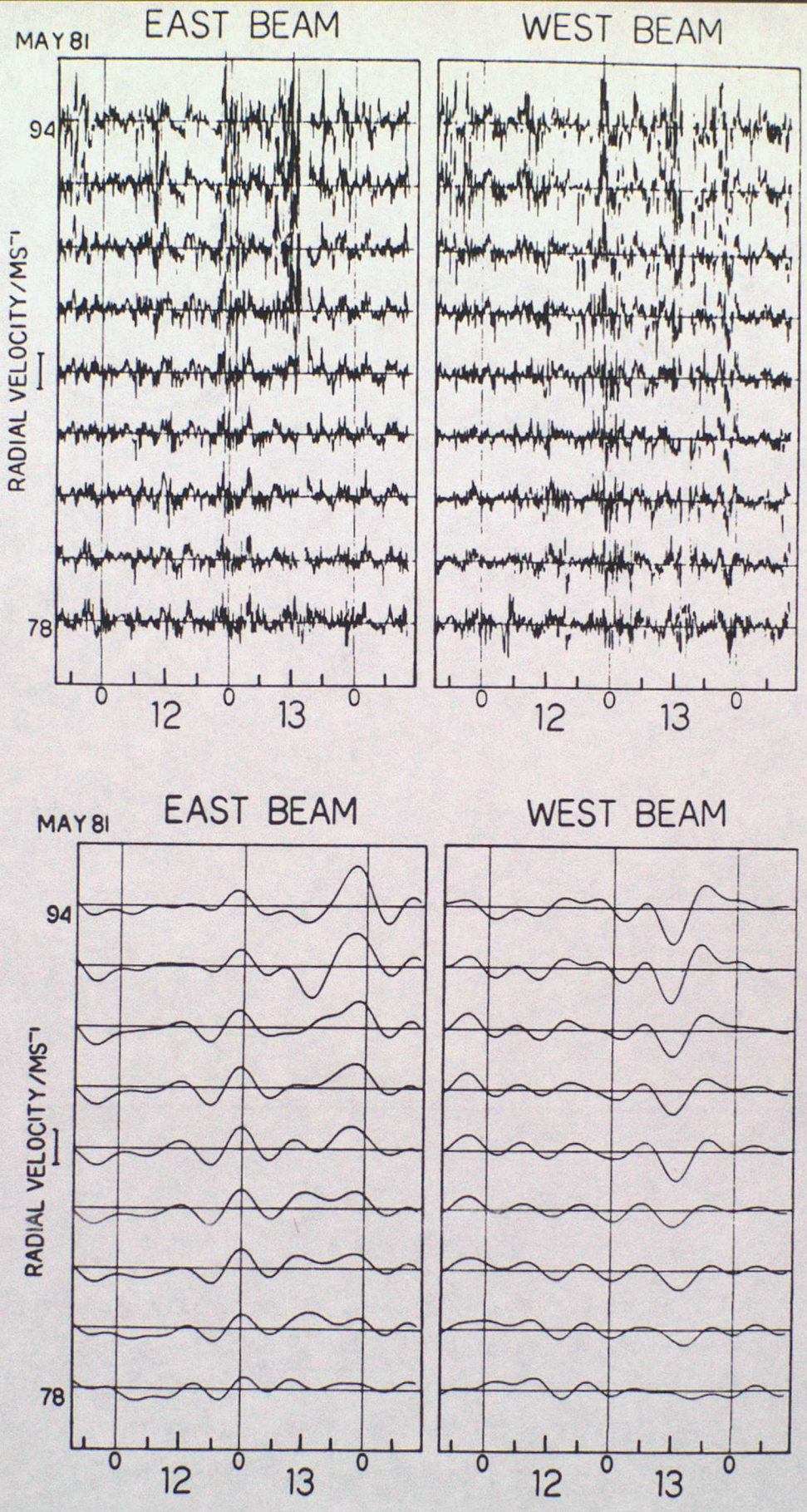


Fig. 7.1. High-frequency radar measurements of line-of-sight velocities between 78 and 94 km altitude, measured in two directions, equally inclined at small angles to the vertical. Top: data filtered to include only periods from 8 min. to 8 hr. Bottom: data filtered to include only periods longer than 8 hr. Data were collected on 11-14 May 1981. (From Vincent and Reid 1983.)

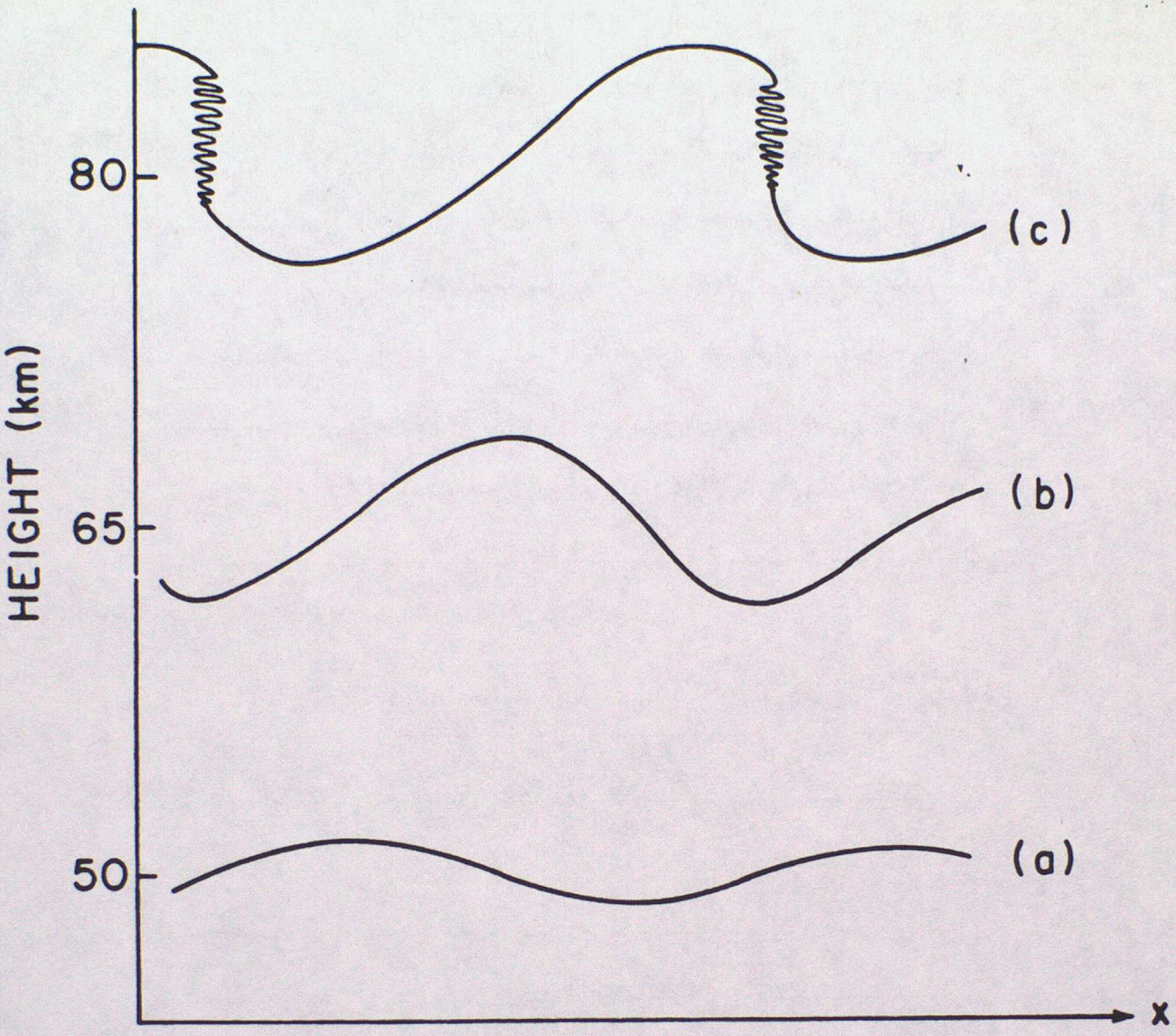


Fig. 7.2. Schematic of the breaking of vertically-propagating internal gravity waves in the mesosphere. Curves (a), (b) and (c) denote isentropic surfaces. For (a) and (b), linear nondissipative theory holds approximately. For (c), nonlinear effects are important, with turbulence occurring near the wave crests, presumably followed by small-scale mixing and dissipation. (From AHL.)

8. EQUATORIAL DYNAMICS

In this lecture we look at the low-latitude middle atmosphere, between about 20° N and S. In this region, quasigeostrophic scaling does not usually apply, and dynamical phenomena occur that are not found at higher latitudes. Examples include equatorially-trapped, planetary-scale wave motions, and long-period oscillations of the mean flow.

8.1 OBSERVED WAVE MOTIONS IN THE EQUATORIAL REGION

In the lower equatorial stratosphere, two basic types of planetary-scale waves have been observed in routine radiosonde measurements. These are the Kelvin and Rossby-gravity waves, which are trapped near the equator but propagate vertically and zonally. Some observations of these waves were mentioned in §2.2.4. Two additional classes of Kelvin wave, with larger vertical wavelengths and horizontal phase speeds, have been observed in satellite data for the upper stratosphere and mesosphere. All of these waves are believed to be forced by convective activity in the tropical troposphere (although the precise forcing mechanisms are still not well understood), whence they propagate up into the middle atmosphere.

8.2 OBSERVED OSCILLATIONS OF THE EQUATORIAL ZONAL-MEAN FLOW

Time-variations of the zonal-mean flow in the equatorial lower stratosphere are dominated by the so-called *quasi-biennial oscillation* (QBO), in which the winds fluctuate from easterly to westerly and back again with a period of about 27 months: see §1.3.2.2. The time-height section of Fig.1.9 shows that the alternate phases succeed each other in descending shear zones that move downwards at about 1 km per month. Above about 35 km, the QBO is replaced by a *semiannual oscillation* (SAO), which reaches its peak amplitude near the stratopause and again (with a phase lag) near the mesopause: see Fig. 1.7.

Two major dynamical problems in the equatorial middle atmosphere are, first, to understand the generation and propagation of the wave motions and, second, to explain the QBO and SAO. It will be found that these phenomena are

in fact intimately related.

8.3 THEORY OF EQUATORIAL WAVES

Investigation of the basic theory of equatorial waves is greatly expedited by use of the *equatorial beta-plane approximation*, in which $f \equiv 2\Omega\sin\phi$ is replaced by βy, for small latitudes ϕ, where the equatorial value of β is $2\Omega/a_0$, a_0 being the earth's radius. Because f becomes small near the equator, the Rossby number $Ro \equiv U/fL$ (see §3.4) cannot generally be expected to be small, and so the primitive equations (3.2) must be used. When linearised about a state of rest, these become

$$u'_t - \beta y v' + \Phi'_x = 0 , \tag{8.1a}$$

$$v'_t + \beta y u' + \Phi'_y = 0 , \tag{8.1b}$$

$$\Phi'_{zt} + N^2 w' = 0 , \tag{8.1c}$$

$$u'_x + v'_y + \rho_0^{-1}(\rho_0 w')_z = 0 , \tag{8.1d}$$

on the equatorial beta-plane. Here we have eliminated T and θ, and introduced the static stability defined by (3.4). Boundary conditions will be ignored for the moment.

We now take N = constant for simplicity and look for separable solutions of (8.1) in the form of zonally and vertically propagating waves:

$$\{u',v',w',\Phi'\} = e^{z/2H}\,\mathrm{Re}\,[\{\hat{u}(y),\hat{v}(y),\hat{w}(y),\hat{\Phi}(y)\}e^{i(kx + mz - \omega t)}] .$$

Then we obtain

$$-i\omega\hat{u} - \beta y\hat{v} + ik\hat{\Phi} = 0 , \tag{8.2a}$$

$$-i\omega\hat{v} + \beta y\hat{u} + \hat{\Phi}_y = 0 , \tag{8.2b}$$

$$ik\hat{u} + \hat{v}_y - i\omega m^2 N^{-2}\hat{\Phi} = 0 , \tag{8.2c}$$

after elimination of $\hat{w}$, and neglect of $(2H)^{-2}$ against m^2 in (8.2c). (The latter is justified for waves of vertical wavelength of order 10 km, as in Lecture 7.) Note that two y-derivatives appear in (8.2), indicating that a second-order ordinary differential equation will normally have to be solved for the latitudinal structure of the waves.

However, the simplest solution of (8.2) only requires the solution of a first-order ODE. This corresponds to the observed Kelvin wave, in which meridional wind fluctuations v' are found to be very small. Putting $\hat{v} = 0$ in (8.2), we get

$$-\omega\hat{u} + k\hat{\Phi} = 0 \ , \quad \beta y\hat{u} + \hat{\Phi}_y = 0 \ , \quad k\hat{u} - \omega m^2 N^{-2}\hat{\Phi} = 0 \ . \qquad (8.3a,b,c)$$

From (8.3a,c) we get $\omega = \pm Nk/m$, so that the vertical group velocity $c_g^{(z)} \equiv \partial\omega/\partial m = \mp Nk/m^2$. Observations suggest that the waves are propagating upwards from the troposphere, so that $c_g^{(z)} > 0$, and the lower sign must be chosen (if the convention $k > 0$ is retained). Thus

$$\omega = -Nk/m \ , \qquad (8.4a)$$

and

$$c_g^{(z)} = Nk/m^2 = \omega^2/Nk \ . \qquad (8.4b)$$

From (8.3a,b) we have the first-order ODE $\hat{\Phi}_y + (k\beta y/\omega)\hat{\Phi} = 0$, with solution

$$\hat{\Phi}(y) = \hat{\Phi}(0)\ \exp(-\beta k y^2/2\omega) . \qquad (8.5)$$

We must have $\hat{\Phi}$ bounded far from the equator ($|y| \to \infty$), so it is necessary that $\omega/k > 0$, implying an *eastward* phase speed; by (8.4a) we must therefore have $m < 0$. Thus the phase surfaces $kx + mz - \omega t$ = constant tilt eastward with height and move downward with time. Some idealised representations of the Kelvin-wave structure are shown in Fig. 8.1. If allowance is made for a uniform mean wind $\bar{u}$ by replacing ω by the intrinsic, or Doppler-shifted value $\omega - k\bar{u}$, one can get quite good agreement with observations.

If $\hat{v} \neq 0$, elimination of $\hat{u}$ and $\hat{\Phi}$ from (8.2) gives

$$\hat{v}_{yy} + \{ (m^2\omega^2N^{-2} - k^2 - \beta k\omega^{-1}) - \beta^2m^2N^{-2}y^2 \}\hat{v} = 0 , \qquad (8.6)$$

if $(m\omega/N)^2 \neq k^2$. Given that $\hat{v}$ is bounded at large $|y|$, this is an eigenvalue equation that also arises in the theory of the quantum harmonic oscillator, and has solutions in the form of Gaussians times Hermite polynomials. The gravest solution is

$$\hat{v} = \hat{v}(0) \exp (-\beta |m| y^2/2N) , \qquad (8.7)$$

where

$$m = - \mathrm{sgn}(\omega) N(\beta + \omega k)/\omega^2 , \qquad (8.8)$$

for upward propagation and $\omega > -\beta/k$. This is the Rossby-gravity wave; it has a westward phase speed, and details of its structure are shown schematically in Fig. 8.2. Again, a simple Doppler shift allows quite good agreement between theory and observations. For further details of this and the higher equatorial modes (which include equatorially-trapped inertio-gravity and Rossby waves), see AHL, §4.7.2 .

For zonal-mean winds $\bar{u}(y,z)$ that vary with latitude and height, this kind of simple separation of variables is not applicable. However, asymptotic theories (assuming that the shear is weak) and numerical approaches have been used to good effect in such cases.

It should be noted that the solutions given here represent *forced* waves. They can be matched to an artificial lower boundary condition like (6.6), but a more realistic model would include a suitable heating term proportional to J' on the right of (8.1c), to represent the large-scale cumulus convective heating in the upper tropical troposphere. Models that allow such a heating in the form of a standing oscillation (which can be split into two travelling forcing effects, moving eastward and westward, respectively) generate a whole set of equatorial modes, and have been moderately successful in accounting for the observed Kelvin and Rossby-gravity waves [see AHL, §4.7.3, and Salby and Garcia (1987) and Garcia and Salby (1987)].

8.4 THEORY OF THE QBO

The most obvious peculiarity of the QBO of the lower equatorial stratosphere is its 27-month period. Early theories, based for example on postulated biennial cycles in diabatic forcing or in the tropospheric circulation, were unable to reproduce the observed structure of the oscillation, and also failed to account for the non-biennial period. The currently-accepted theory, due to Lindzen and Holton (1968) and Holton and Lindzen (1972), maintains that the eastward and westward zonal wind accelerations that make up the QBO are forced primarily by nonlinear rectified effects associated with upward-propagating, dissipating, Kelvin and Rossby-gravity waves, respectively. The period (which in fact fluctuates between about 22 and 34 months) depends in a complicated way on the wave amplitudes (and hence the processes that generate the waves) and the wave-dissipation mechanisms. The full theory is fairly involved, and we shall only give a brief qualitative outline here: further details can be found in Plumb (1984) and AHL, §8.3 .

Consider first the easterly phase of the QBO, when Kelvin waves are observed in the lower equatorial stratosphere (see Fig. 8.3a). On the basis of a WKBJ theory, it can be shown that these waves have a phase speed c_K that is eastward with respect to the zonal flow $\bar{u}$ (assumed to vary slowly with z), and that their vertical group velocity $c_g^{(z)} = (\bar{u} - c_K)^2 k/N$. The latter decreases as the waves propagate up into the shear region between the lower-level easterlies and the upper-level westerlies, and as this happens the waves become increasingly susceptible to dissipation. On the other hand, during the westerly phase of the QBO, Rossby-gravity waves are found in the lower stratosphere (see Fig. 8.3b). These have a phase speed c_{RG} that is westward with respect to $\bar{u}$, and their vertical group velocity decreases as they propagate from the lower westerlies into the upper easterlies. Thus these waves, too, become increasingly dissipated in the shear zone.

The next aspect of the Holton-Lindzen mechanism concerns the mean-flow changes induced by these dissipating waves. Here the TEM formalism, as applied to the primitive equations on an equatorial β-plane, is useful. Under various scaling assumptions, including that the Richardson number $Ri \equiv (N/\bar{u}_z)^2$ is large, and that the mean-flow height scale is much greater than a vertical wavelength, the advection by the residual circulation in the zonal momentum equation becomes negligible, and we get

$$\bar{u}_t = \rho_0^{-1} \nabla . \underline{F} + \text{drag term} . \qquad (8.9)$$

We suppose that the waves and the mean flow both have latitudinal scale L, and define a meridional average

$$\langle \ldots \rangle \equiv L^{-1} \int_{-\infty}^{\infty} (\ldots)\, dy .$$

Now using WKBJ methods for the waves and a theory analogous to that of §4.3, approximate expressions for $\langle F^{(z)} \rangle$ (the meridional mean of the vertical component of the EP flux) for the Kelvin and Rossby-gravity waves can be found in terms of $\bar{u}$, wave dissipation, wave amplitudes at some lower forcing level, and phase speeds and zonal wavenumbers. By taking the meridional mean of (8.9), Holton and Lindzen arrived at a model equation of the form

$$\langle \bar{u} \rangle_t = \rho_0^{-1} [\, \langle F^{(z)} \rangle_K + \langle F^{(z)} \rangle_{RG} \,] + K \langle \bar{u} \rangle_{zz} , \qquad (8.10)$$

where the drag has been represented in terms of vertical diffusion with a small viscosity coefficient K. They integrated this equation numerically, and found that $\langle \bar{u} \rangle (z,t)$ could resemble the time-height section of Fig. 1.9 quite closely when suitable wave amplitudes and dissipation were chosen (see Fig. 8.4).

The theory of Holton and Lindzen has been extended in several ways. The most sophisticated model of the QBO to date is that of Plumb and Bell (1982), which uses the full TEM equations, rather than (8.9), and numerically calculates the time-development of $\bar{u}(y,z,t)$ and the full wave structure in this time-varying flow. So far, general circulation models have been unsuccessful in simulating the QBO, probably because of insufficient vertical resolution and over-intense wave dissipation. An ingenious laboratory experiment (using internal gravity, rather than equatorial, waves) was carried out by Plumb and McEwan (1978), and demonstrated the viability of the Holton-Lindzen wave, mean-flow mechanism in real fluids.

We finally mention the semiannual oscillations at the equatorial stratopause and mesopause. Current thinking generally ascribes the eastward accelerations at the stratopause to the "fast" Kelvin waves that have been observed by satellite instruments (see §8.1). The westward accelerations

there may be forced by cross-equatorial advection of zonal-mean easterlies by the mean meridional circulation, or driven by meridionally-propagating planetary waves from the winter hemisphere. It is possible that the mesopause oscillation, which is in antiphase with the stratopause oscillation, is mainly driven by breaking gravity waves that have been selectively filtered by passing through the underlying mean winds (see e.g. AHL, §§8.5.1, 8.5.2). Gravity waves may also play a role in forcing the stratopause oscillation.

REFERENCES

Garcia, R.R. and Salby, M.L., 1987. Transient response to localized episodic heating in the tropics, Part II. Far-field behaviour. *J. Atmos. Sci.* **44**, 499-530.

Holton, J.R., 1975. *The dynamic meteorology of the stratosphere and mesosphere.* (Meteorological Monographs **37**, American Meteorological Society.)

Holton, J.R. and Lindzen, R.S., 1972. An updated theory for the quasi-biennial cycle of the tropical stratosphere. *J. Atmos. Sci.* **29**, 1076-1080.

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

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

Plumb, R.A., 1984. The quasi-biennial oscillation. In *Dynamics of the middle atmosphere*, ed. J.R. Holton and T. Matsuno. (Terrapub), pp. 217-251.

Plumb, R.A. and Bell, R.C., 1982. A model of the quasi-biennial oscillation on an equatorial beta-plane. *Quart. J. Roy. Met. Soc.* **108**, 335-352.

Plumb, R.A. and McEwan, A.D., 1978. The instability of a forced standing wave in a viscous stratified fluid: a laboratory analogue of the quasi-biennial oscillation. *J. Atmos. Sci.* **35**, 1827-1839.

Salby, M.L. and Garcia, R.R., 1987. Transient response to localized episodic heating in the tropics, Part I. Excitation and short-time near-field behavior. *J. Atmos. Sci.* **44**, 458-498.

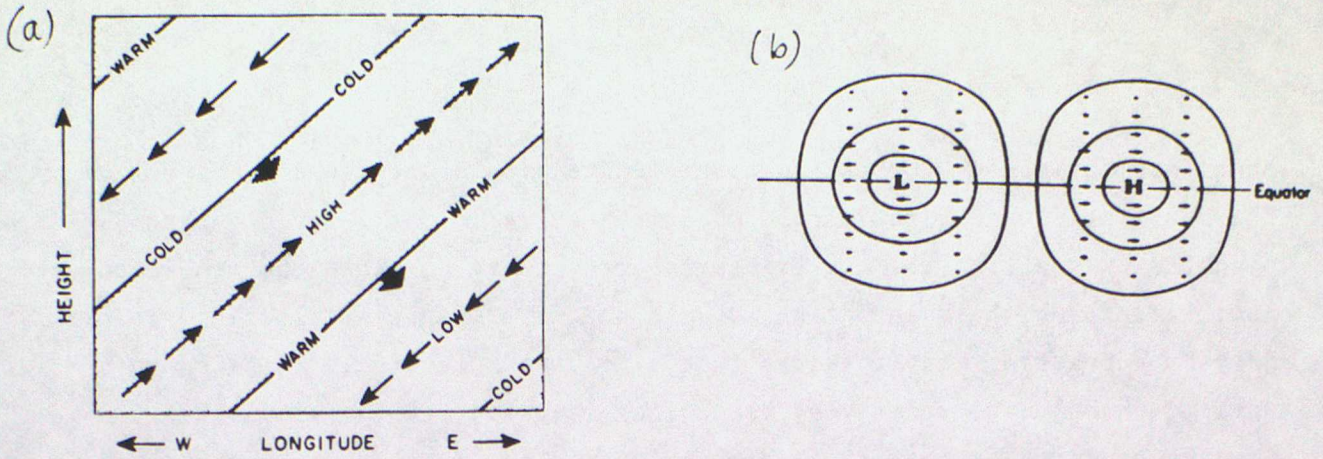


Fig. 8.1. (a) Schematic height-longitude section showing phase relations between velocity, geopotential and temperature fluctuations at the equator in an upward-propagating Kelvin wave. Thin sloping lines denote surfaces of constant phase and thick arrows show direction of phase propagation. (From Holton 1975.) (b) Schematic longitude-latitude section of geopotential and horizontal wind fluctuations for the Kelvin wave. (From Matsuno 1966.)

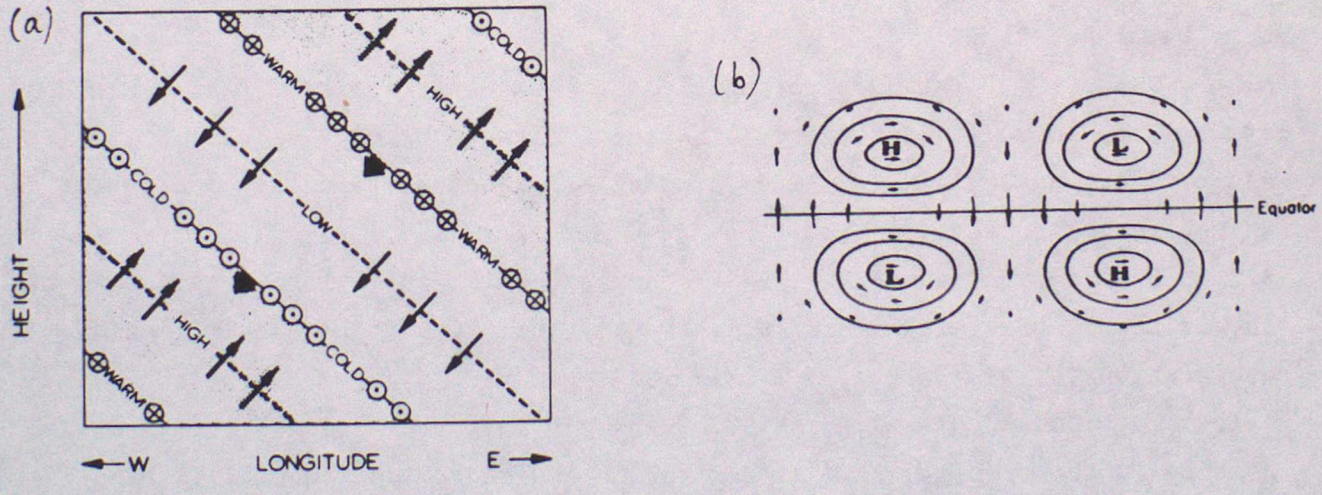


Fig. 8.2. (a) As for Fig. 8.1a but for a Rossby-gravity wave of westward phase speed ($-\beta/k < \omega < 0$) north of the equator. (From Holton 1975.) (b) As for Fig. 8.1b but for the Rossby-gravity wave. (From Matsuno 1966.)

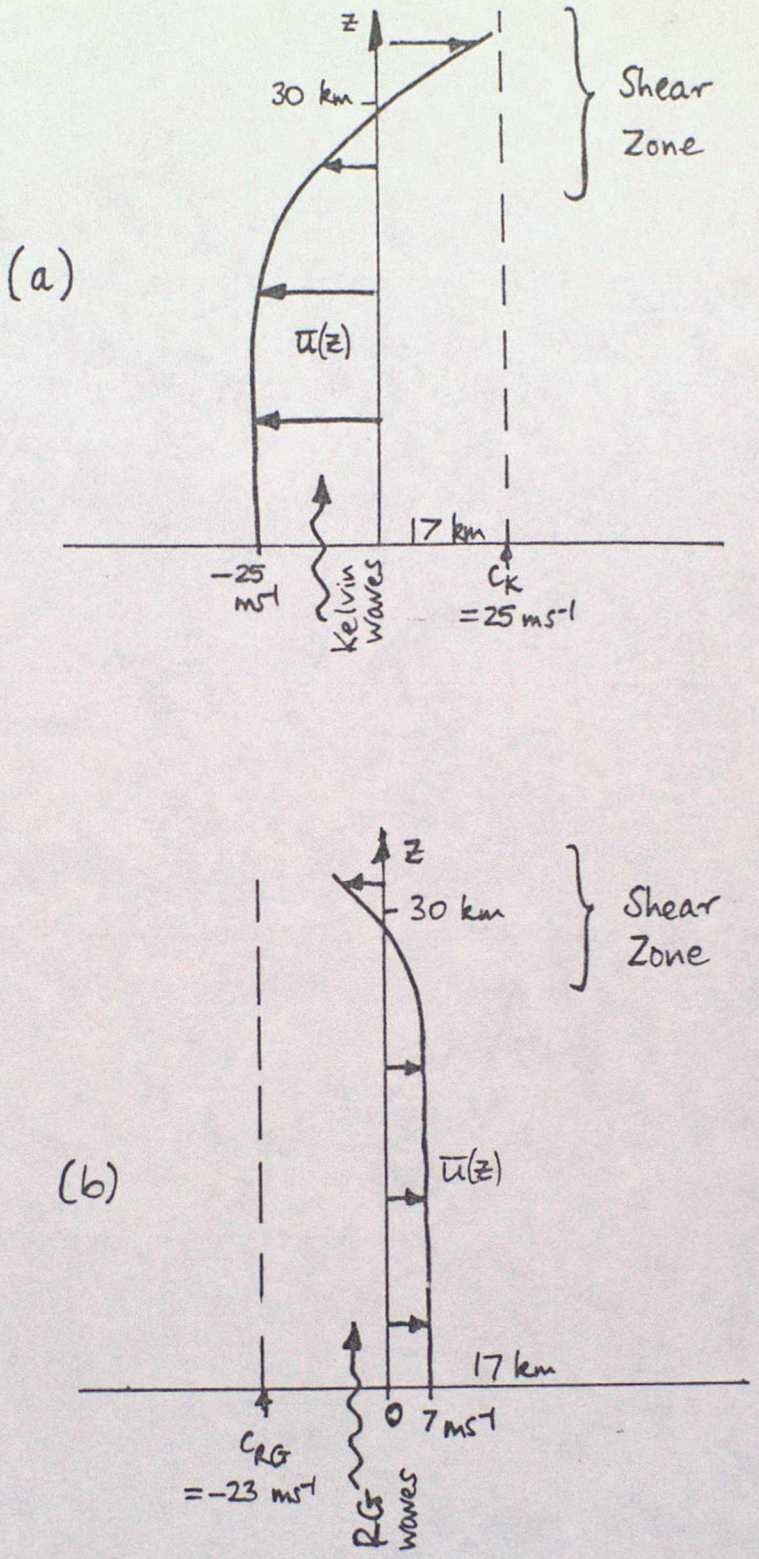


Fig. 8.3. Schematic diagrams of (a) Kelvin waves propagating up into lower-stratospheric zonal-mean easterlies and (b) Rossby-gravity waves propagating up into westerlies. Each type of wave becomes increasingly dissipated in the shear zone, where the zonal-mean winds reverse in sign. See text for further details.

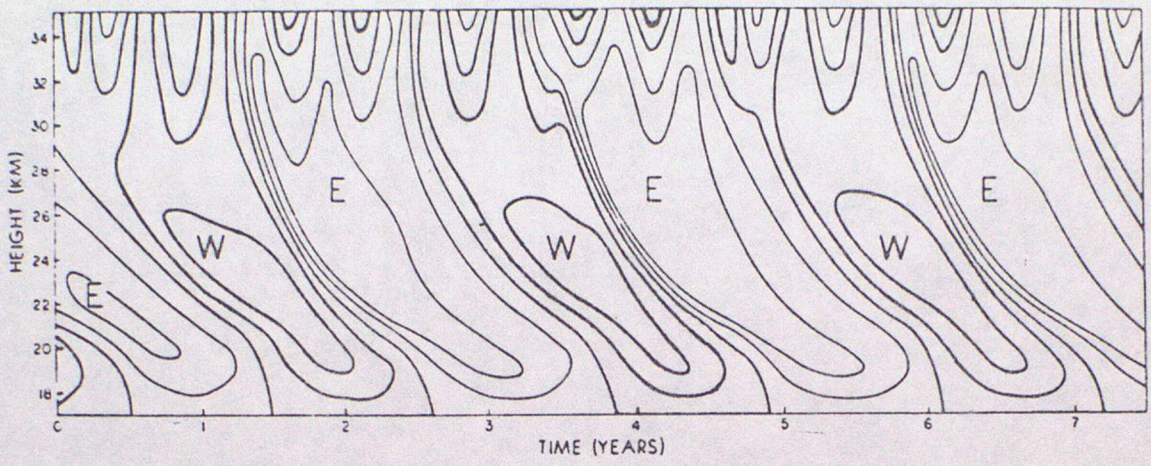


Fig. 8.4. Time-height section of mean zonal wind at the equator in the Holton-Lindzen QBO model. Contours are at 10 m s^{-1} intervals and a semiannual oscillation is imposed at 35 km.

9. TRACER TRANSPORT

9.1 SOME BASIC DEFINITIONS

The most basic concept in the study of tracer transport is that of a *materially-conserved tracer* (often called a *conservative tracer*), the amount of which remains constant in time following each air parcel. If μ is the "amount" of conservative tracer per unit mass of air, then such a tracer satisfies

$$D\mu/Dt \equiv \partial\mu/\partial t + u\partial\mu/\partial x + v\partial\mu/\partial y + w\partial\mu/\partial z = 0 \ , \tag{9.1}$$

in the log-pressure coordinates of §3.1 . In practice no physical tracer is precisely conservative, and we have

$$D\mu/Dt = S \ , \tag{9.2}$$

where S represents sources and sinks of tracer. Using the mass-continuity equation (3.2d) this can also be written in "flux" form,

$$(\rho_0\mu)_t + (\rho_0\mu u)_x + (\rho_0\mu v)_y + (\rho_0\mu w)_z = \rho_0 S \ . \tag{9.3}$$

For many atmospheric tracers S is weak in the sense that relative changes in μ following an air parcel are small over relevant advection timescales: μ is then called a *quasi-conservative* or *long-lived* tracer.

Another classification is into *dynamically-active* tracers, whose dynamical, thermodynamic or chemical properties allow them to influence the flow field that transports them and *dynamically-passive* tracers, which have no such effect. The latter could, however, be *chemically* active.

9.2 SOME EXAMPLES OF ATMOSPHERIC TRACERS

The most well-known atmospheric tracer is the potential temperature θ, defined by (3.2f). It satisfies (3.2e), which is of the form (9.2) with $S = Q$. θ is quasi-conservative for sufficiently weak heating rates; since it is a basic thermodynamic variable, it is clearly a dynamically-active tracer.

Somewhat less familiar is *Ertel's potential vorticity* P, defined by

$$P \equiv \rho_0^{-1}[\ \theta_z(f - u_y + v_x) - \theta_x v_z + \theta_y u_z\ ] \qquad (9.4)$$

in log-pressure coordinates. This is materially conserved (DP/Dt = 0) if the flow is adiabatic and frictionless; in many atmospheric cases it can be quasi-conservative on timescales of a week or so. It is also a dynamically-active tracer in general: this follows from the *invertibility principle*, which states that the field of P at any instant determines the velocity and temperature fields at that instant, given suitable boundary conditions and the assumption that the flow is in a "balanced" state. For discussion of this principle and many other aspects of the use of P see Hoskins, McIntyre and Robertson (1985).

Other atmospheric tracers include the minor chemical species, with μ given by the mass or volume mixing ratio. Many such species (e.g. ozone in the lower stratosphere, methane and nitrous oxide throughout most of the stratosphere) are long-lived. Much of the current interest in tracer transport has been stimulated by the need for a better understanding of the ways in which these chemicals are carried from one part of the atmosphere to another, and especially how they move in the latitudinal and vertical directions from source regions to sinks and reservoirs. There is also the possibility that the observed movement of chemical and dynamical tracers may give useful information on dynamical processes, such as the planetary wave-breaking phenomenon mentioned in §§ 2.2, 6.3.

Of particular interest to middle-atmospheric scientists is the problem of understanding how the distribution of ozone is determined, and the factors that might affect it. Ozone absorbs solar ultra-violet radiation, thereby heating the stratosphere and also protecting the biosphere. It is only present in small amounts, less than about 10 parts per million by volume (ppmv); indeed if all the ozone in the atmosphere were brought to standard temperature and pressure, the resulting "ozone layer" would be only 3 mm thick ! Our present concern centres around the question of whether man-made pollutants, such as chlorofluoromethanes, are depleting the ozone. Whereas ozone is mainly formed by photochemistry in the sunlit part of the upper stratosphere (above about 25 km), much ozone is found at lower levels and in the darkness of the polar night. Meridional transport by atmospheric motions

$$\text{VMR} \equiv n_T/n_A = \text{MMR} \times (M_A/M_T) = p_T/p \ .$$

These distinctions should be kept clearly in mind when studying observational or model data, and when passing between the dynamical and chemical literature. For example, chemists often write equations of the form

$$\partial n_T/\partial t = \text{"flux term"} + \text{"production term"}$$

in place of equation (9.2) for the mixing ratio.

9.4 MODELS OF TRACER TRANSPORT IN THE MIDDLE ATMOSPHERE

9.4.1 One-dimensional models

One-dimensional models have been much used by atmospheric chemists to predict the vertical structure of constituents. We define a horizontal global integral by

$$\langle \ldots \rangle \equiv \iint_{\text{globe}} (\ldots)\, dx\, dy \ ; \qquad (9.6)$$

this is a function of z and t. On applying (9.6) to (9.3) and noting that the global integral of the horizontal divergence vanishes, we obtain

$$\langle \mu \rangle_t = -\rho_0^{-1} (\rho_0 \langle \mu w \rangle)_z + \langle S \rangle \ . \qquad (9.7)$$

The term $\langle \mu w \rangle$ represents the globally-integrated vertical flux of tracer. To obtain a prognostic equation for $\langle \mu \rangle$ it is usually assumed that this flux can be represented by diffusion down the (vertical) gradient of $\langle \mu \rangle$; thus

$$\langle \mu w \rangle = - K(z) \partial \mu/\partial z \ , \qquad (9.8)$$

where K(z) is an empirical diffusion coefficient. On substituting (9.8) into (9.7) we obtain a diffusion equation, with sources and sinks, for $\langle \mu \rangle$:

$$\partial \langle \mu \rangle/\partial t = \rho_0^{-1}\, \partial/\partial z \{ \rho_0 K \partial \langle \mu \rangle/\partial z \} + \langle S \rangle \ . \qquad (9.9)$$

This approach can be generalised to allow for many constituents μ_n (n =

must therefore play an important part in determining the ozone distribution. Moreover, transport also plays an important role in distributing the species that may react with ozone.

The full problem of understanding and simulating ozone behaviour is very complex, and involves a treatment of photochemistry (perhaps including hundreds of reactions), radiative transfer, dynamics and transport. These are usually coupled in an intricate fashion, although some simplifications may result from the fact that the ozone distribution in the upper stratosphere is dominated by photochemistry, while in the lower stratosphere (where ozone is long-lived) it is dominated by transport. Various ways of modelling chemical transport are described below.

9.3 MEASURES OF TRACE CONSTITUENT CONCENTRATION

While the mass or volume mixing ratio is usually the best measure of chemical concentration for transport studies, other measures are often used by chemists. These include the number n of molecules per unit volume, the mass ρ per unit volume, and the partial pressure p. Note that

$$n_T M_T = \rho_T N_a \ , \quad n_A M_A = \rho_A N_a \ ,$$

where N_a is Avogadro's number ($\approx 6\times10^{26}$ molecules $kmol^{-1}$), M is the molecular weight and subscripts T and A denote the trace constituent and "air" excluding the trace constituent, respectively.

Since trace constituents, by definition, are present only in small amounts, $\rho_T << \rho_A \approx \rho$ and $p_T << p_A \approx p$, where ρ and p are the total air density and pressure. Then the ideal gas law gives

$$p_T = \rho_T R^* T/M_T \ , \quad p = \rho R^* T/M_A \ ,$$

where R^* is the universal gas constant. Thus the mass and volume mixing ratios satisfy

$$MMR \equiv \rho_T/\rho = (p_T/p)(M_T/M_A) \ , \qquad (9.5a)$$

Some early ideas about meridional tracer transport in the lower stratosphere were introduced by Brewer (1949) and Dobson (1956), who suggested that observed distributions of water vapour and ozone might qualitatively be accounted for by advection by a hypothetical zonally-symmetric circulation. This flow consists of a rising branch passing through the tropical tropopause, and poleward and downward flow in the extratropics; it has come to be known as a *Brewer-Dobson circulation*.

In 1961, Murgatroyd and Singleton published a diagnostic calculation of a "possible meridional circulation" in the middle atmosphere, $(\tilde{v}, \tilde{w})$ say, based on the zonal-mean thermodynamic and continuity equations

$$\bar{\theta}_t + \tilde{v}\bar{\theta}_y + \tilde{w}\bar{\theta}_z = \bar{Q} \quad , \tag{9.11a}$$

$$\tilde{v}_y + \rho_0^{-1}(\rho_0\tilde{w})_z = 0 \quad , \tag{9.11b}$$

together with boundary conditions $\tilde{v} = 0$ at the poles. The requisite $\bar{\theta}$ was obtained from observations of $\bar{T}$ and the heating $\bar{Q}$ from a radiative calculation using $\bar{T}$ and observed ozone. A sketch of the resulting circulation at the solstices is given in Figure 9.1: in the lower stratosphere it resembles the Brewer-Dobson circulation, while higher up it consists of rising motion in the summer hemisphere, cross-equatorial flow in the mesosphere and descent in the winter hemisphere. A circulation diagnosed in this way is often called a "diabatic circulation". The $\bar{\theta}_t$ term is usually small, and is sometimes omitted.

Murgatroyd and Singleton noted that their results, being based on zonal-mean observations, necessarily omitted the "eddy heat flux convergence"

$$C \equiv -(\overline{v'\theta'})_y - \rho_0^{-1}(\rho_0\overline{w'\theta'})_z$$

that would appear on the right of (9.11a) if $(\tilde{v}, \tilde{w})$ were to be replaced by $(\bar{v}, \bar{w})$. Thus their diabatic circulation differs from the Eulerian-mean circulation in regions where C is significant, such as the northern winter stratosphere: in this region Eulerian-mean ascent, rather than descent, is found. However, the term $\tilde{v}\bar{\theta}_y$ is usually small in (9.11a), which then becomes formally identical to the TEM equation (4.3b); moreover, (9.11b) is formally identical to (4.3c). Thus Murgatroyd and Singleton's circulation is in fact a

1,2,...N) where the relevant sources $\langle S_n \rangle$ may depend on all the $\langle \mu_m \rangle$; in general a different K_n should be used for each constituent.

The advantage of the 1-D models is that the transport representation is so simple that the available computer power can be devoted to consideration of many chemical species and reactions. These models are therefore valuable for testing and comparing photochemical theories. However, since they deal with *global* (or sometimes *hemispheric*) integrals, it is difficult to justify their comparison with *local* measurements. It is also difficult to give a convincing justification of the "diffusive flux" parameterization (9.8), although some rationale for this has been provided by Holton (1986) and Mahlman, Levy and Moxim (1986).

9.4.2 Two-dimensional models

A less restrictive approach is to employ the zonal (Eulerian) average, as defined in §4.1. Although this is again a non-local quantity, tracers are often fairly zonally-symmetric in the absence of strong planetary waves. Moreover, satellite measurements of composition are often averaged zonally to improve signal-to-noise levels. A study of the transport of $\bar{\mu}$ in the meridional plane is therefore of some interest.

The Eulerian zonal mean of (9.2) can be written

$$\bar{\mu}_t + \bar{v}\bar{\mu}_y + \bar{w}\bar{\mu}_z = - \left[(\overline{v'\mu'})_y + \rho_0^{-1}(\rho_0 \overline{w'\mu'})_z \right] + \bar{S} , \qquad (9.10)$$

using (3.2d). Thus the rate of change of $\bar{\mu}$ is determined not only by advection by $(\bar{v}, \bar{w})$ and the sources and sinks $\bar{S}$ but also by the "eddy flux divergence" (in square brackets). In practice, it is often found that the mean advection and eddy flux terms almost cancel: the reason is that the Eulerian mean circulation $(\bar{v}, \bar{w})$ is not independent of the eddies — indeed, it is usually largely determined by them. One should therefore not attach too much physical significance to the labels "mean transport" and "eddy transport" that are often given to these expressions. In fact, we shall see later that there are other, quite different, ways of arranging (9.10) into "mean advective transport" and "eddy flux divergence" terms. First, however, is is worth recalling some history.

comparatively localised events and handle the transport of local injections of tracer; this may make for more meaningful validation against observations. Unlike the 1D and 2D models, no artificial separation is made into zonal or global means and departures therefrom.

Owing to the large computer resources required for treating dynamics and radiation, few 3D models have yet included any but the simplest chemical schemes. A popular approach is to perform the tracer studies "off-line": the model's velocity fields are stored, and then used in a separate calculation to advect the tracer around. This allows dynamically-passive tracers to be treated, with or without active chemistry. Examples include those of Mahlman (1985) and Mahlman, Levy and Moxim (1986). Dynamically-active tracers, which influence the motion or temperature fields, cannot be studied in this way, but must be incorporated in a fully "on-line" approach; an example is that of Cariolle and Déqué (1986).

Finally, it should not be forgotten that 3D models generate vast quantities of data, which must be suitably organised and compressed if meaningful comparisons with observations are to be made, and physical interpretation carried out. There is still a need for improved diagnostics to aid in this process.

REFERENCES

Brewer, A.W., 1949. Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Met. Soc.* 75, 351-363.

Cariolle, D. and Déqué, M., 1986. Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. *J. Geophys. Res.* 91, 10825-10846.

Dobson, G.M.B., 1956. Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. Roy. Soc. Lond.* **A236**, 187-193.

Dunkerton, T.J., 1978. On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.* 35, 2325-2333.

Holton, J.R., 1986. A dynamically based transport parameterization for one-dimensional photochemical models of the stratosphere. *J. Geophys. Res.* 91, 2681-2686.

good approximation to the *residual* circulation in many cases. [It is also quite closely related to the *generalised Lagrangian-mean* circulation; see §6.5.2 of WMO_3.]

We now return to the specification of 2D models for transport and chemical studies. These tend to use either (a) the Eulerian-mean formulation or (b) the residual, diabatic or similar formulation. In each case some form of eddy flux convergence generally appears in the tracer continuity equation. The eddy flux must be parameterized in terms of the zonal-mean state with which the model is working; this is usually done by means of a two-dimensional "diffusion tensor" K, relating the vector eddy flux to the mean gradient $(\bar{\mu}_y, \bar{\mu}_z)$. Such a parameterization is difficult to justify rigorously except in one formulation of type (b), introduced by Plumb (1979) and called the *effective transport formulation*. The effective transport circulation differs slightly from the residual circulation, and considerably from the Eulerian-mean circulation; the corresponding diffusion tensor is anisotropic. The definitions of this circulation and diffusion tensor involve Lagrangian parcel displacements, and they are generally difficult to estimate from atmospheric observations. However they have been calculated from a 3D general circulation model by Plumb and Mahlman (1987), and used by them in a 2D model with some success.

Advantages of 2D models over 1D models include the facts that they can in principle represent meridional transport in a fairly self-consistent manner, and can be validated against zonal-mean observations at different latitudes. Furthermore their dynamics is still simple enough to allow detailed chemistry to be included. On the other hand, the eddy parameterizations mentioned above are essentially "non-interactive", in the sense that they cannot respond to changes in the mean state; this difficulty can only be circumvented by use of 3D models.

9.4.3 Three-dimensional models

In principle, 3D models may include the coupling between fully three-dimensional dynamics, radiation and photochemistry. A major advantage is that they explicitly describe the transport associated with the large-scale planetary waves. Of course, sub-gridscale effects, including those due to breaking gravity waves, must still be parameterized. 3D models can represent

Hoskins, B.J., McIntyre, M.E. and Robertson, A.W., 1985. On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Met. Soc.* 111, 877-946.

Mahlman, J.D., 1985. Mechanistic interpretation of stratospheric tracer transport. *Adv. Geophys.* 28A, 301-323.

Mahlman, J.D., Levy, H. and Moxim, W.J., 1986. Three-dimensional simulations of stratospheric N_2O: predictions for other trace constituents. *J. Geophys. Res.* 91, 2687-2707. (Also corrigenda, *J. Geophys. Res.* 91, 9921.)

Murgatroyd, R.J. and Singleton, F., 1961. Possible meridional circulations in the stratosphere and mesosphere. *Quart. J. Roy. Met. Soc.* 87, 125-135.

Plumb, R.A., 1979. Eddy fluxes of conserved quantities by small-amplitude waves. *J. Atmos. Sci.* 36, 1699-1704.

Plumb, R.A. and Mahlman, J.D., 1987. The zonally-averaged transport characteristics of the GFDL general circulation/transport model. *J. Atmos. Sci.* 44, 298-327.

BIBLIOGRAPHY

This lecture is based on

Andrews, D.G., 1987. Transport mechanisms in the middle atmosphere: an introductory survey. In *Transport mechanisms in the middle atmosphere*, ed. G. Visconti (Reidel), to appear. (Also Met.O. 20 DCTN 51.)

See also AHL Chapter 9 and the review :

Mahlman, J.D., Andrews, D.G., Hartmann, D.L., Matsuno, T, and Murgatroyd, R.J., 1984. Transport of trace constituents in the stratosphere. In *Dynamics of the middle atmosphere*, ed. J.R. Holton and T. Matsuno (Terrapub), pp. 387-416.

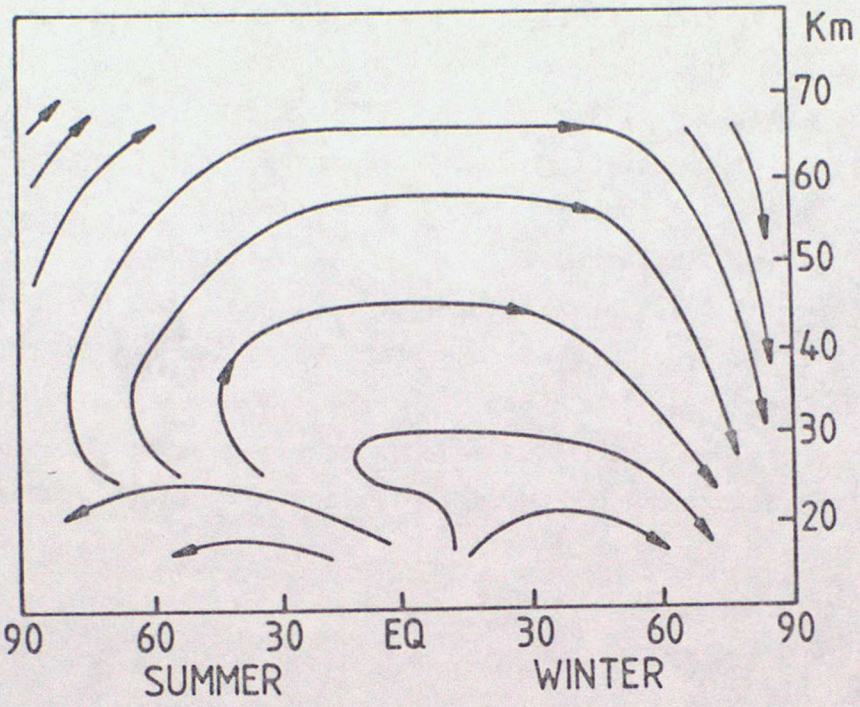


Fig. 9.1. Schematic representation of Murgatroyd and Singleton's diabatic circulation at the solstices. (After Dunkerton 1978.)

10. THE GENERAL CIRCULATION OF THE MIDDLE ATMOSPHERE

10.1 INTRODUCTION

There seems to be no very precise definition of the term "general circulation". In this lecture we shall take it to mean the large-scale flow of the atmosphere, in some loose sense. Since the troposphere certainly affects the stratosphere, and the stratosphere has at least a minor influence on the troposphere, the general circulation of the middle atmosphere cannot be considered in isolation from that of the lower atmosphere.

In modelling the middle atmosphere, it is helpful to employ a whole hierarchy of models, with the simpler ones aiding in the interpretation of the more complex ones. In Lecture 5 we considered some of the simplest such models. At the other end of the spectrum are the highly detailed *general circulation models* (GCMs), which involve numerical solution of the primitive equations with a minimum of assumptions and parameterizations. Ideally, one would wish to specify only the initial state and the boundary conditions (the solar radiation and the surface conditions); the model would then predict the global time-development of temperatures, winds, chemical concentrations, etc. In practice, limitations in our knowledge of the basic physics and in available computer resources force us to make certain compromises.

A major role of GCMs is to enable controlled experiments to be performed, to test our understanding of how the atmosphere works. With these models, hypotheses concerning cause-and-effect relationships can be examined in a way that is impossible using atmospheric observations alone. GCMs tend to be used for two broad classes of experiment: short-term runs for the detailed simulation of specific observed events (e.g. sudden warmings) and longer-terms runs from which "climatologies" can be constructed for comparison with current observational climatologies. As models become more successful in the latter mode, we shall be able to place more confidence in lengthy runs intended to predict possible changes in climate due, for example, to increasing levels of carbon dioxide or pollutant chemicals such as chlorofluoromethanes.

Most GCMs have been designed primarily for study of the troposphere, and include only a few stratospheric levels; even so, several of these have made important contributions to our knowledge of the lower stratosphere. However,

an increasing number of GCMs are now being built specifically with middle atmospheric applications in mind; some of these are described in the next section.

10.2 SOME MODELS OF THE GENERAL CIRCULATION OF THE MIDDLE ATMOSPHERE

Current general circulation models of the middle atmosphere include those being used at GFDL (see Mahlman and Umscheid, 1984), NCAR (Pitcher et al., 1983), NASA Langley Research Center (Blackshear et al., 1987), Goddard Institute for Space Studies (Rind et al., 1987), CSIRO (Hunt, 1986) and the Meteorological Office. These differ considerably in their formulation and degree of completeness, and we shall mention only two of them in detail: the GFDL model and the Met. Office model. We shall also mention a highly idealised model that may help elucidate aspects of the nonlinear dynamics of the middle atmosphere.

10.2.1 The GFDL "SKYHI" model

This 40-level model extends from the ground up to about 80 km. It uses terrain-following "sigma" coordinates below about 350 mb and pressure coordinates above. The vertical resolution is variable, but averages about 2½ km in the middle atmosphere (see Figure 10.1). The model uses finite differences in the horizontal, with a grid resolution that has varied from 9° latitude by 10° longitude in early runs, through 5°x6° and 3°x3.6° to 1°x1.2° in the latest experiments. The model includes a fairly complete set of physical parameterizations in the troposphere (except that cloudiness, pack ice and sea surface temperatures are specified from climatology). It has a state-of-the-art treatment of radiation, with radiative heating calculated every 12 hours using diurnally-averaged solar insolation. Below 35 km a zonally-symmetric ozone distribution is specified; above 35 km a simple temperature-dependent ozone photochemistry scheme is used. A nonlinear horizontal diffusion and a Richardson-number dependent vertical diffusion are included in the momentum equations. No parameterization of the drag and diffusion due to breaking gravity waves is used, although the finest-scale version of the model does explicitly represent some of these wave motions.

Most of the results published so far for the SKYHI model relate to the

5°x6° version. January-mean cross-sections of $\bar{u}$ and $\bar{T}$ are depicted in Figures 10.1 and 10.2. Note that this version produces zonal jets that are much too strong in the mesosphere compared with the observations in Figure 1.3. These are consistent with the fact that the winter stratosphere is too cold (a common problem with middle atmosphere GCMs) and the latitudinal temperature gradient in the mesosphere is not reversed, as in the observations. These deficiencies generally seem to be less pronounced in the higher-resolution versions of the model (Fels, 1985). It appears that in the lower resolution model the planetary waves are too weak in the troposphere, and are rapidly damped in the lower stratosphere, leaving no wave activity to drive the atmosphere far from the radiative state (see Lecture 5). At higher resolution the planetary waves penetrate to a greater height.

The model simulates quite well the three observed types of Kelvin wave in the equatorial middle atmosphere. It also produces equatorially-trapped inertio-gravity waves (see Lecture 8). It simulates a realistic-looking semiannual oscillation in equatorial latitudes (Figure 10.3) but not a quasi-biennial oscillation; the latter deficiency may be due to insufficient vertical resolution. Phenomena resembling minor and major stratospheric sudden warmings have arisen spontaneously during integrations of the model.

10.2.2 The Meteorological Office Middle Atmosphere model

This model does not include a troposphere, and should perhaps not be regarded as a general circulation model in the strict sense. It is forced from below by prescribing the height of the 100 mb (or sometimes 300 mb) pressure surface, using either observed or idealised data. Details of the formulation are given in Fisher (1987); in summary, the model extends to about 80 km altitude and includes 33 levels, spaced at about 2½ km in the log-pressure coordinate z. Horizontal resolution is typically 5°x5°, with fourth-order finite differencing being used. The model includes a state-of-the-art radiation code, supplied by Dr K.P. Shine of Oxford University as part of a joint Met. Office - Oxford modelling project (see Shine, 1987). The solar heating uses zonal-mean ozone prescribed from climatology. At present the model includes a Rayleigh friction in the mesosphere (with a relaxation time decreasing from about 100 days at the stratopause to a day or so at 80 km) as a crude representation of gravity-wave drag there. It is hoped to include a

more sophisticated parameterization soon.

The Met. Office model has been used for a variety of middle atmosphere studies. Several of these have been concerned with the stratospheric sudden warming phenomenon (e.g Butchart et al., 1982; Fairlie and O'Neill, 1987). Recently some annual integrations have been performed, and these are currently being diagnosed, and studied in the light of insights from simpler models (see Lecture 5). A number of more idealised experiments have been performed, in which the lower-boundary forcing has been prescribed in rather simple ways, and attempts made to interpret the resulting disturbances of the middle atmosphere in terms of wave propagation or vortex-interaction dynamics (e.g. O'Neill and Pope, 1987).

10.2.3 A barotropic model with extremely high horizontal resolution

Contrary to the general belief a few years ago, many middle atmosphere modellers now feel that a good simulation of the dynamics and transport of the stratosphere and mesosphere will require GCMs with rather high horizontal resolution, so that the details of the potential-vorticity dynamics can be represented reasonably faithfully. A model designed to investigate this idea, though certainly *not* a GCM, has been reported by Juckes and McIntyre (1987). This is a spectral, hemispheric, barotropic model, with very high horizontal resolution (equivalent to less than 1° of latitude) but representing a single layer in the vertical. The model is forced by a vorticity source in zonal wavenumber 1, with broad latitudinal scale: as Figure 10.4 shows, after 17 days, the nonlinear terms in the barotropic vorticity equation have led to the production of very fine-scale structures in the vorticity field. Analogous structures can presumably occur in the real middle atmosphere (in the isentropic potential vorticity field) and, although almost invisible to current observational techniques, are likely to have profound effects on the general circulation.

10.3 AN OVERALL DESCRIPTION OF THE GENERAL CIRCULATION OF THE MIDDLE ATMOSPHERE

In this section we attempt to give a qualitative overall picture of the large-scale circulation of the middle atmosphere, bringing together some of the topics that have been described in previous lectures.

At the most basic level, we can say that the *zonal-mean temperature structure* is quite close to being under radiative control, except in the winter stratosphere and the upper mesosphere. Thus the winter hemisphere is generally cooler than the summer, and a clear stratopause is present. The associated zonal-mean zonal winds, being in thermal-wind balance with this temperature field, and tied to comparatively low values near the tropopause, are generally westerly in winter and easterly in summer.

The presence of zonally-asymmetric wave or eddy motions modifies this basic picture somewhat; these eddies exert an effective zonal force on the mean flow, bringing about departures from the purely radiatively-determined state, and associated mean meridional circulations. In particular the winter stratosphere is warmer than expected on radiative grounds alone, and the summer mesopause is cold and the winter mesopause warm. For the same reasons, the basic winter westerlies and summer easterlies peak at about 60 km altitude and decrease to low values near the mesopause, rather than increasing with altitude throughout the middle atmosphere.

The waves that bring about this modification to the radiative state are principally large-scale planetary waves in the winter stratosphere and small-scale gravity waves in the mesosphere. Both types of wave are thought to be forced mainly in the troposphere; their propagation up through the middle atmosphere can be modified considerably by the mean state there. In general, wave amplitudes (as measured, say, by temperature or wind fluctuations) tend to grow with altitude, leading eventually to wave-breaking and nonlinear effects that inhibit further growth. The forcing of the zonal-mean state is intimately related to this "saturation" effect; the waves can usually be thought of as transferring zonal (angular) momentum from their source regions to the places where they become saturated or are otherwise dissipated (McIntyre, 1987).

The same irreversible wave processes also lead to latitudinal and vertical transport of tracers within the wave generation or dissipation regions. That breaking planetary waves can move tracers latitudinally is evident from isentropic maps like that in Figure 6.1. Such transport is temporary if the waves are reversible, but becomes quasi-permanent if the wave breakdown is irreversible. Similar sorts of ideas also apply to vertical transport of

tracers by gravity waves, although the details of the dynamics are different in important respects. A schematic illustration of the overall zonally-averaged transport in the troposphere, stratosphere and mesosphere is given in Fig. 10.5.

10.4 FUTURE PROSPECTS

We conclude these lectures with a few thoughts on the ways in which middle atmosphere research may be moving in the next few years.

10.4.1 Observational studies

The next major initiative in the area of satellite observation of the stratosphere and mesosphere is the Upper Atmosphere Research Satellite (UARS), which is due to be launched by the Space Shuttle in about 1990. This will carry several sophisticated instruments for sounding temperatures, minor chemical species, and even winds, in the middle atmosphere. The Met. Office is committed to the assimilation of data from these instruments, and the preparation of global analyses which, it is hoped, will provide a much more accurate picture of the middle atmosphere than is currently afforded by the construction of geostrophic (or, at best, gradient) winds from height fields obtained from retrieved temperatures. The Met. Office model will also be an important tool in the study and interpretation of these data.

Another satellite experiment to be performed when Shuttle flights resume is ATMOS (Atmospheric Trace Molecular Spectroscopy), in which high-resolution spectroscopic measurements are made of the atmospheric spectrum between 2 and 16 μm. The first flight, in May 1985, provided tantalising new glimpses of the chemistry and dynamics of the mesosphere.

Ground-based measurements are becoming increasingly important. For example, radars and lidars, although only providing localised information in the horizontal, have excellent vertical resolution and are now giving exciting insights into the behaviour of gravity waves in the stratosphere and mesosphere. It is to be hoped that a global-scale network of such instruments can be set up.

The prospect of new measurements must not, however, deflect us from the continued analysis and interpretation of the very large satellite data-set that already exists, and the continued expansion of climatologies from current instruments such as the SSU. Much of this information has not yet been fully exploited, or subjected to the new diagnostic approaches that have recently been suggested.

10.4.2 Modelling

Each new generation of computers offers the prospect of longer runs with GCMs of higher and higher resolution. However, even with the two CYBER 205s at GFDL, the 1°x1.2° version of the SKYHI model described in §10.2.1 only runs in approximately real time ! Another problem with such large models is the analysis of the huge amount of data that they generate. Such problems can, and must, be tackled; indeed several U.K. universities are planning to exploit the new Cray X-MP/48 at the Rutherford Appleton Laboratory for a global atmospheric modelling project (including the middle atmosphere) using an expanded version of the ECMWF forecast model. Nevertheless there is still a great need for careful, well-planned, hypothesis-testing experiments with smaller models such as the Met. Office middle atmosphere model and simpler "mechanistic" or even analytical models.

10.4.3 Theory

The simplest models just mentioned are aspects of the "theoretical framework" that is necessary for any meaningful assessment of observational or model data. For example, the simple zonally-averaged models of Lecture 5 provide a context for the discussion of zonally-averaged data, and point to the kinds of mean quadratic eddy quantities that might be of interest. Such models are of course over-simplified in many respects, and caution may have to be exercised over conclusions drawn from them, especially those conclusions involving causal relationships. While the models give a clue to the ways in which eddies may influence the mean flow, they do not usually include the back-effect of the mean flow on the waves[1]. An extension of these models to incorporate such effects would be most valuable.

---

[1] An exception is the Holton-Lindzen model of the QBO: see Lecture 8.

Another area in which simple theoretical models might be helpful is that of the interaction of dynamics and photochemistry (and perhaps radiation as well).

There is still some interest in the study of conservation laws for wave properties in the form of (4.10), and nonlinear versions for waves obeying the primitive equations have recently been found by Andrews (1987) and (in greater generality) by Haynes (1987). Laws of this kind may possibly be of value in diagnosing the behaviour of nonlinear planetary waves in models and in the atmosphere. A related topic is the investigation of conservation laws to provide diagnostics for waves on zonally-asymmetric basic flows; this is more difficult than the case where the basic flow is zonally symmetric, but some progress has been made. Also important is the study of the ways in which such waves affect the basic state.

A further area in which theoretical studies are needed is that of the dynamics of breaking waves, both gravity waves and planetary waves. This will call for analytical and numerical (and perhaps also laboratory) work. A clearer understanding of the dynamics of interacting vortices may help disentangle some of the complications of the behaviour of large-amplitude planetary waves, in which the separation into "waves" and "basic state" may be blurred.

REFERENCES

Andrews, D.G., 1987. A finite amplitude Eliassen-Palm theorem in isentropic coordinates: the time-dependent, non-conservative case. *In preparation*.

Blackshear, W.T., Grose, W.L. and Turner, R.E., 1987. Simulated sudden stratospheric warmings: synoptic evolution. *Quart. J. Roy. Met. Soc.*, to appear.

Butchart, N., Clough, S.A., Palmer, T.N. and Trevelyan, P.J., 1982. Simulations of an observed stratospheric warming with quasigeostrophic refractive index as a model diagnostic. *Quart. J. Roy. Met. Soc.* 108, 475-502.

Fairlie, T.D.A. and O'Neill, A., 1987. Aspects of dynamics of the middle atmosphere inferred using data from a satellite and a numerical model. *Phil. Trans. Roy. Soc. Lond.*, to appear. (Also Met. O. 20, DCTN 54.)

Fels S.B., 1985. Radiative-dynamical interactions in the middle atmosphere. *Adv. Geophys.* 28A , 277-300.

Fisher, M., 1987. The Met.O. 20 stratosphere-mesosphere model. Met.O. 20 DCTN 52.

Haynes, P.H., 1987. Forced, dissipative generalizations of finite-amplitude wave-activity conservation relations for zonal and non-zonal basic flows. In preparation.

Hunt, B.G., 1986. The impact of gravity wave drag and diurnal variability on the general circulation of the middle atmosphere. *J. Met. Soc. Japan* **64**, 1-16.

Juckes, M.N. and McIntyre, M.E., 1987. A high resolution one-layer model of breaking planetary waves in the stratosphere. *Nature*, to appear.

McIntyre, M.E., 1987. Dyanmics and tracer transport in the middle atmosphere: an overview of some recent developments. In *Transport processes in the middle atmosphere*, ed. G. Visconti (Reidel), to appear.

Mahlman, J.D. and Umscheid, L.J., 1984. Dynamics of the middle atmosphere: successes and problems of the GFDL "SKYHI" general circulation model. In *Dynamics of the middle atmosphere*, ed. J.R. Holton and T. Matsuno (Terrapub), pp. 501-525.

O'Neill, A. and Pope, V.D., 1987. Simulations of linear and nonlinear disturbances in the stratosphere. *Quart. J. Roy. Met. Soc.*, submitted. (Also Met.O. 20 DCTN 53.)

Pitcher, E.J., Malone, R.C., Ramanathan, V., Blackmon, M.L., Puri, K. and Bourke, W., 1983. January and July simulations with a spectral general circulation model. *J. Atmos. Sci.* **40**, 580-604.

Rind, D., Suozzo, R., Balachandran, N.K., Lacis, A. and Russell, G., 1987. The GISS global climate/middle atmosphere model with parameterized gravity wave drag. *J. Atmos. Sci.*, submitted.

Shine, K.P., 1987. The middle atmosphere in the absence of dynamical heat fluxes. *Quart. J. Roy. Met. Soc.* **113**, 603-633.

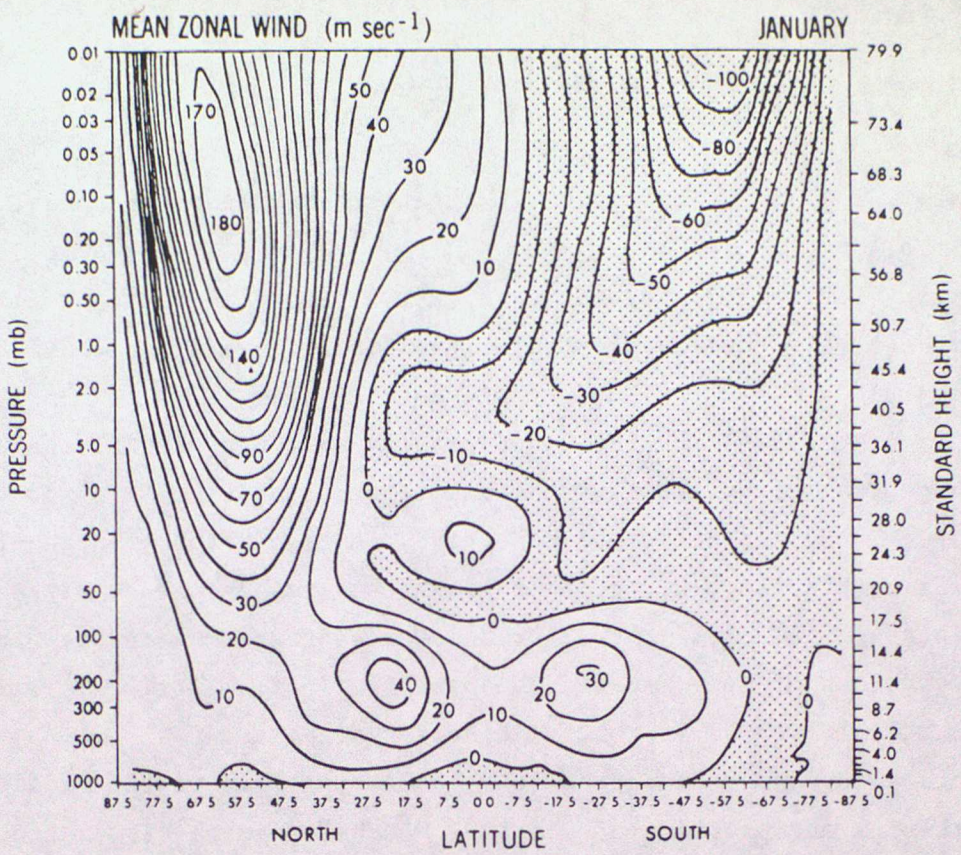


Fig. 10.1. $\bar{u}$ in m s^{-1} for January "1982" from the SKYHI model. Tick marks on the right ordinate indicate locations of model levels. (From Mahlman and Umscheid 1984.)

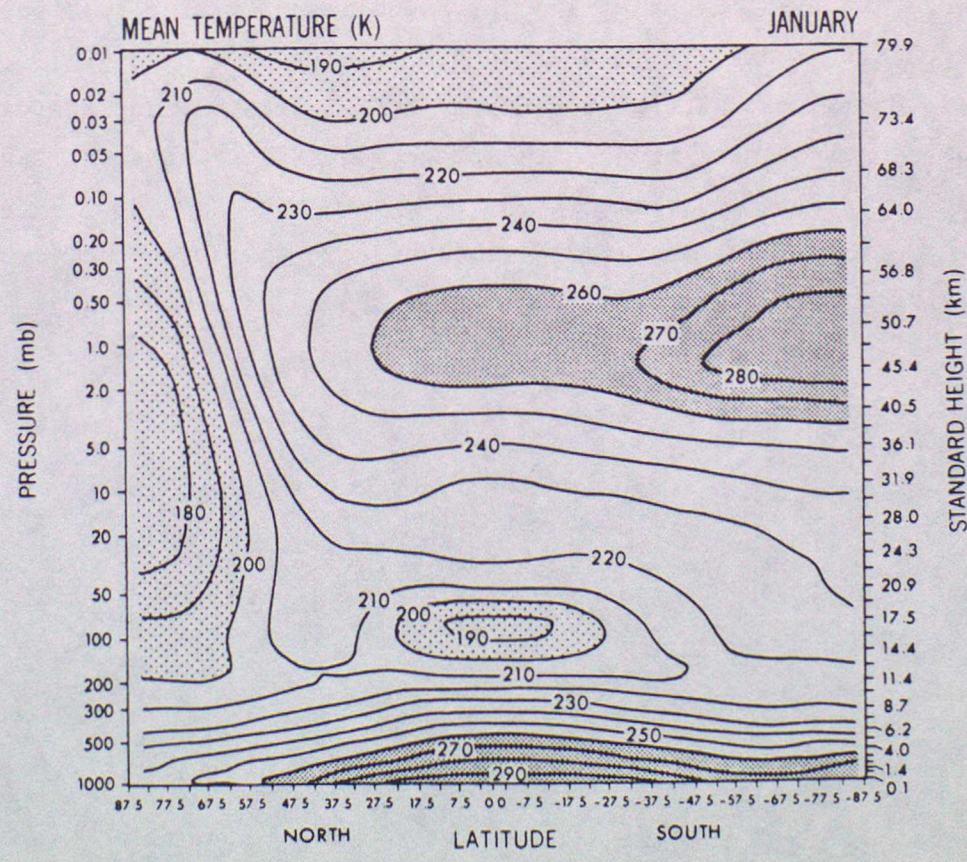


Fig. 10.2. $\overline{T}$ (K) for January "1982" from the SKYHI model. (From Mahlman and Umscheid 1984.)

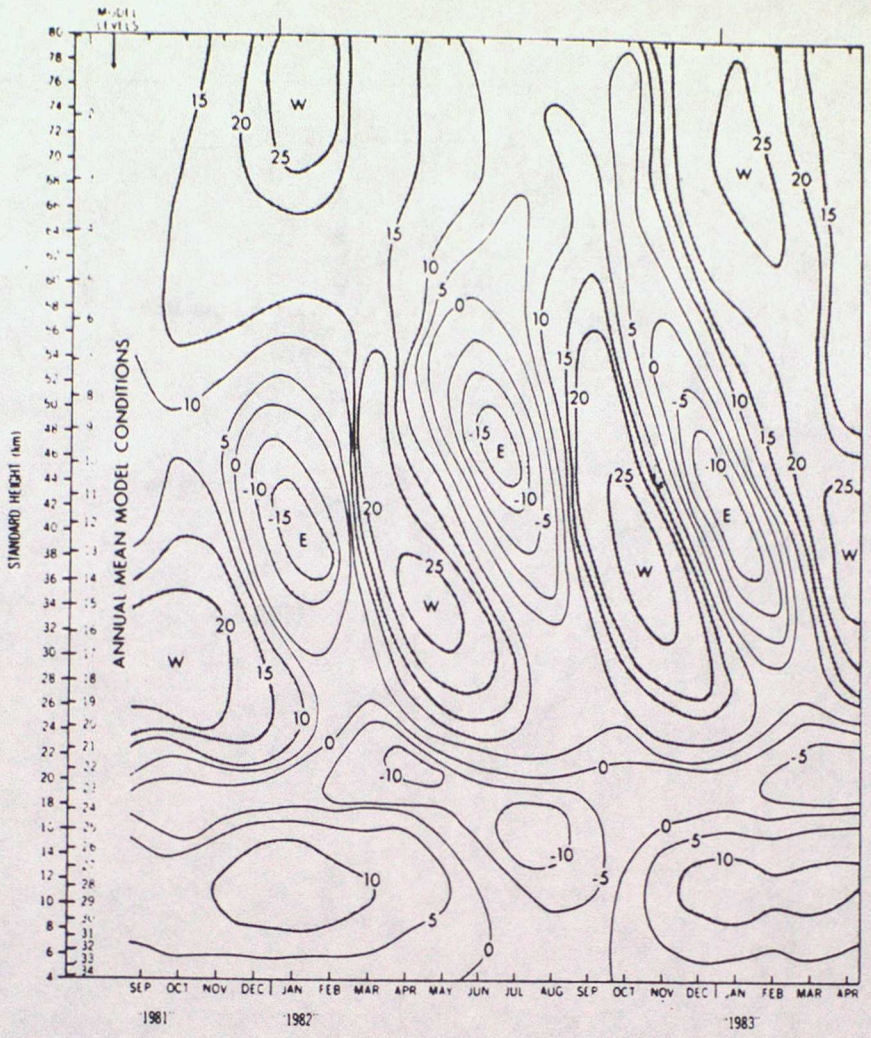


Fig. 10.3. Time-height section of $\bar{u}$ at 2.5°N for the SKYHI model, showing the simulated semiannual oscillation near the stratopause. (From Mahlman and Umscheid 1984.)

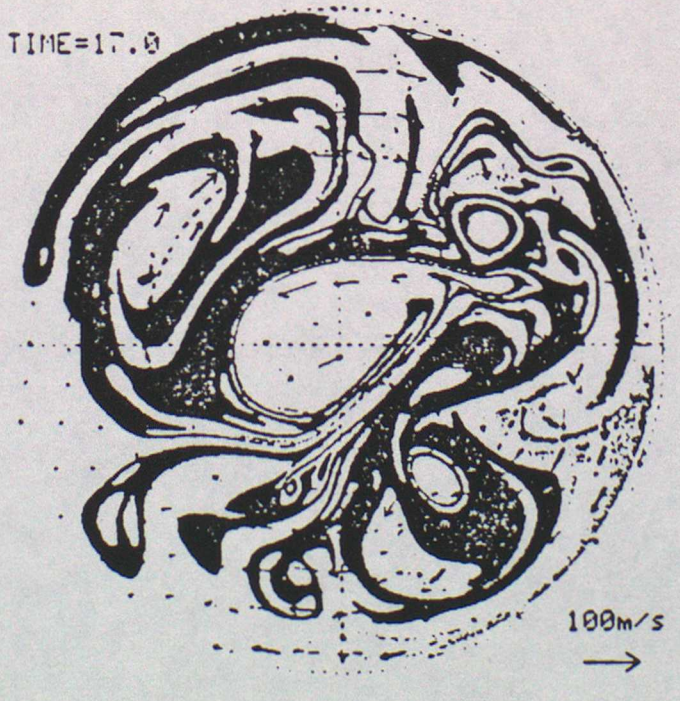


Fig. 10.4. Polar stereographic map indicating the vorticity field on day 17 of an integration of Juckes and McIntyre's high-resolution barotropic model: see text for details.

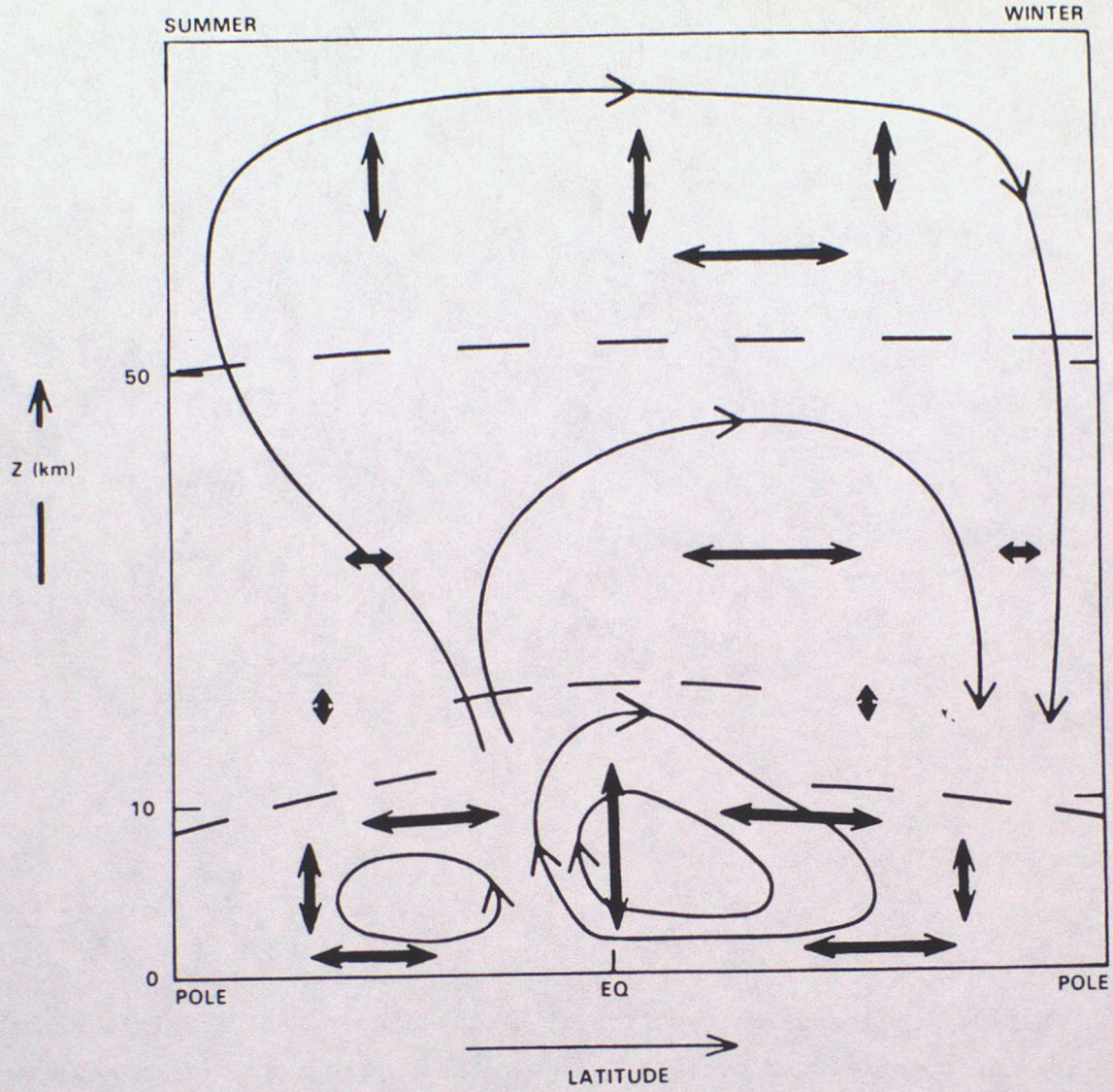


Fig. 10.5. Schematic illustration of zonally-averaged transport processes up to the mesopause. Single arrows denote mean circulation; double arrows indicate quasi-horizontal and vertical diffusion. In the troposphere there is advection by the Hadley circulation, quasi-horizontal mixing by planetary and synoptic eddies and vertical convective mixing. In the stratosphere one finds a mean cell and quasi-horizontal mixing, both mainly driven by planetary waves. In the mesosphere the transport seems to be dominated by a pole-to-pole circulation and vertical mixing, both mainly due to breaking gravity waves. (From WMO_3.)