PHYS418 – Physics of Atmospheres

Department of Physics and Astronomy

Dr. Adrian McDonald
Room 822
Email:Adrian.mcdonald@canterbury.ac.nz

1. Atmospheric structure and composition

1.1 Atmospheric Temperature, Pressure, and Density

Pressure decreases approximately exponentially with altitude, Figure 1 displays the true variation of atmospheric pressure and density with altitude. In comparison, if we take 273K as a reference temperature, it can be shown that the atmospheric temperatures at different altitudes vary from the reference by only 25% (see Figure 3). Thus, the pressure decrease of the atmosphere with altitude can be modelled with very crude approximations of temperature.

The simplest approximation is that of an isothermal atmosphere (constant temperature), but the result obtained is a good approximation to the general case. For two vertical surfaces in the atmosphere separated by a vertical distance dz. The pressure at the lower surface is P and the pressure at the upper surface is P-δP. Atmospheric pressure results from the weight per unit area of the atmosphere above that level. As the level moves up through the atmosphere, the quantity of the atmosphere above the level decreases, and as a result, the pressure decreases. This means that the difference in pressure δP between z and z+δz comes from the weight of the atmosphere between the two levels. The mass per unit area is $\rho\delta z$ (where ρ is the atmospheric density) and the weight is the mass multiplied by the acceleration due to gravity-

$$\delta P = -\rho g \delta z$$

Equation 1

Using the ideal gas law, we can write –

$$\frac{\delta P}{P} = -g\frac{\delta z}{RT}$$

Equation 2

The variation of pressure with altitude can then be obtained by integrating this relationship between z and infinity where the pressures are P_o and 0, respectively.

$$P = P_0 \,\mathbf{exp}\left(-\frac{gz}{RT}\right) = P_0 \,\mathbf{exp}\left(-\frac{z}{H}\right)$$

Equation 3

H is referred to as the scale height and for an isothermal atmosphere is roughly 8km. In spite of assuming an isothermal atmosphere, this is still a

valid result since atmospheric pressure does decrease very close to exponentially with altitude. The value of the scale height is even approximately correct.

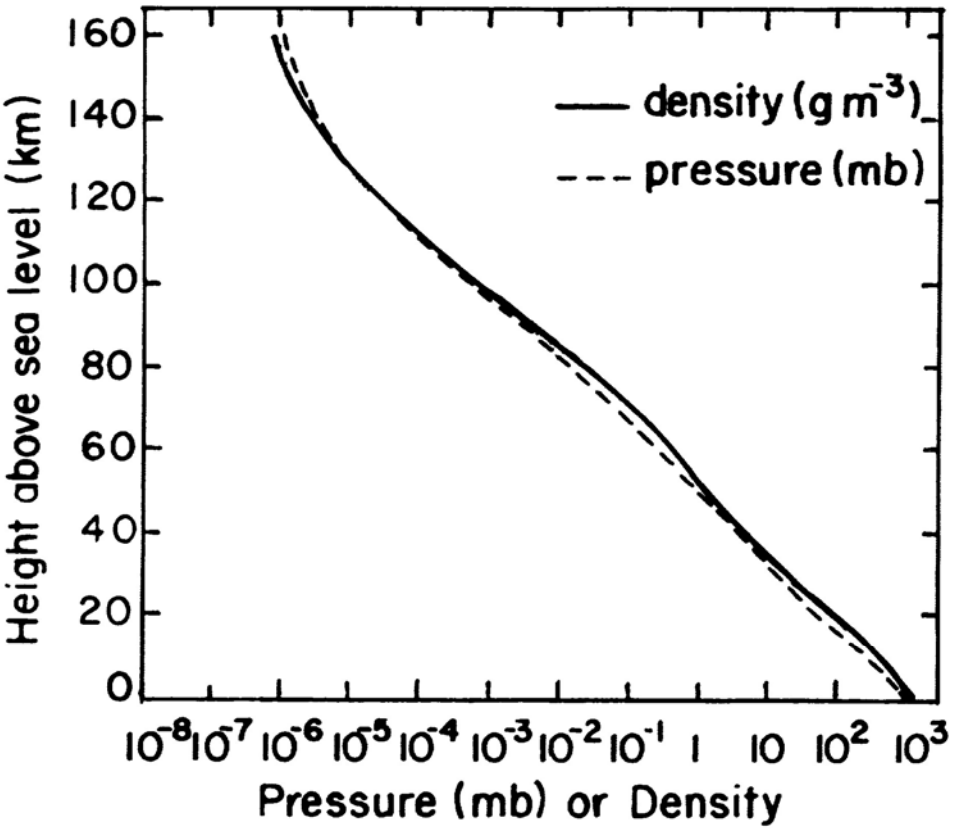

Figure 1 Vertical profile of pressure and density.

Figure 2 shows a typical example of the vertical structure of atmospheric temperature up to 100km (taken from climatological data collated in the CIRA86 standard atmosphere). The atmosphere is conventionally divided into layers according to the variation of temperature with height (also referred to as the lapse rate). The layer from the ground to 10-15km is called the troposphere and is bounded by the tropopause. The layer from the tropopause to about 45-50km in which temperature rises with increasing altitude is called the stratosphere, this layer being bounded by the stratopause. The layer from the stratopause to about 100km is called the mesosphere.

The troposphere is defined by the decrease of temperature with height at a rate of approximately 6.5K/km. The temperature structure of the troposphere varies with altitude such that the temperature increases towards the equator. The altitude of the tropopause also varies with latitude and is higher and colder over the tropics than the poles. The troposphere is the region in which the weather phenomena occur, contains 75% of the total molecular mass of the atmosphere and is characterised by strong vertical mixing. According to the World Meteorological Organization (WMO), the tropopause is defined as

the lowest level at which the temperature lapse rate decreases to 2K/km or less and the lapse rate averaged between this level and any level within the next 2km does not exceed 2K/km.

As well as the change in sign of the temperature gradient (or lapse rate), the transition from the troposphere to the stratosphere is marked by a sharp change in the concentrations of certain important atmospheric species. This abrupt change reflects the fact that there is relatively little mixing of tropospheric and stratospheric air. The increase of temperature with height in the stratosphere means that this region is more stable with respect to vertical mixing than is the troposphere. Thus, there can be large vertical gradients in the trace constituent compositions there. The stratosphere also has a latitudinal temperature distribution, where the summer hemisphere is characterised by a cold equator and a warm pole, while the winter hemisphere displays a distinct temperature maximum over mid-latitudes and cold air over the pole. At the stratopause level the temperature has a nearly constant latitudinal gradient between the warm summer pole and the cold winter pole.

At the mesopause the latitudinal variation of temperature is opposite to that observed at the stratopause, in that a relatively constant temperature gradient is observed from the cold summer pole to the warm winter pole.

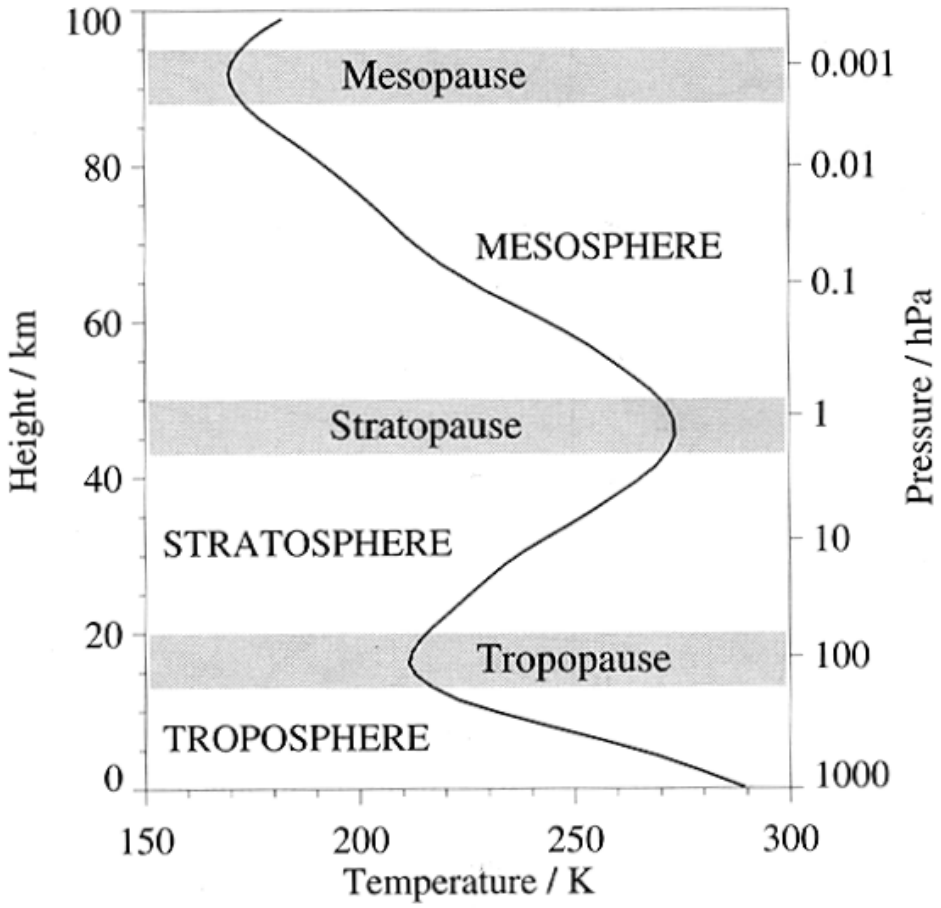

Figure 2 Typical vertical structure of the atmospheric temperature in the lowest 100km of the atmosphere.

An earlier average temperature profile taken from the 1976 US standard atmosphere is displayed in Figure 3. This profile also displays the standard deviations of the temperatures from this profile and the 1% extreme values of temperature observed.

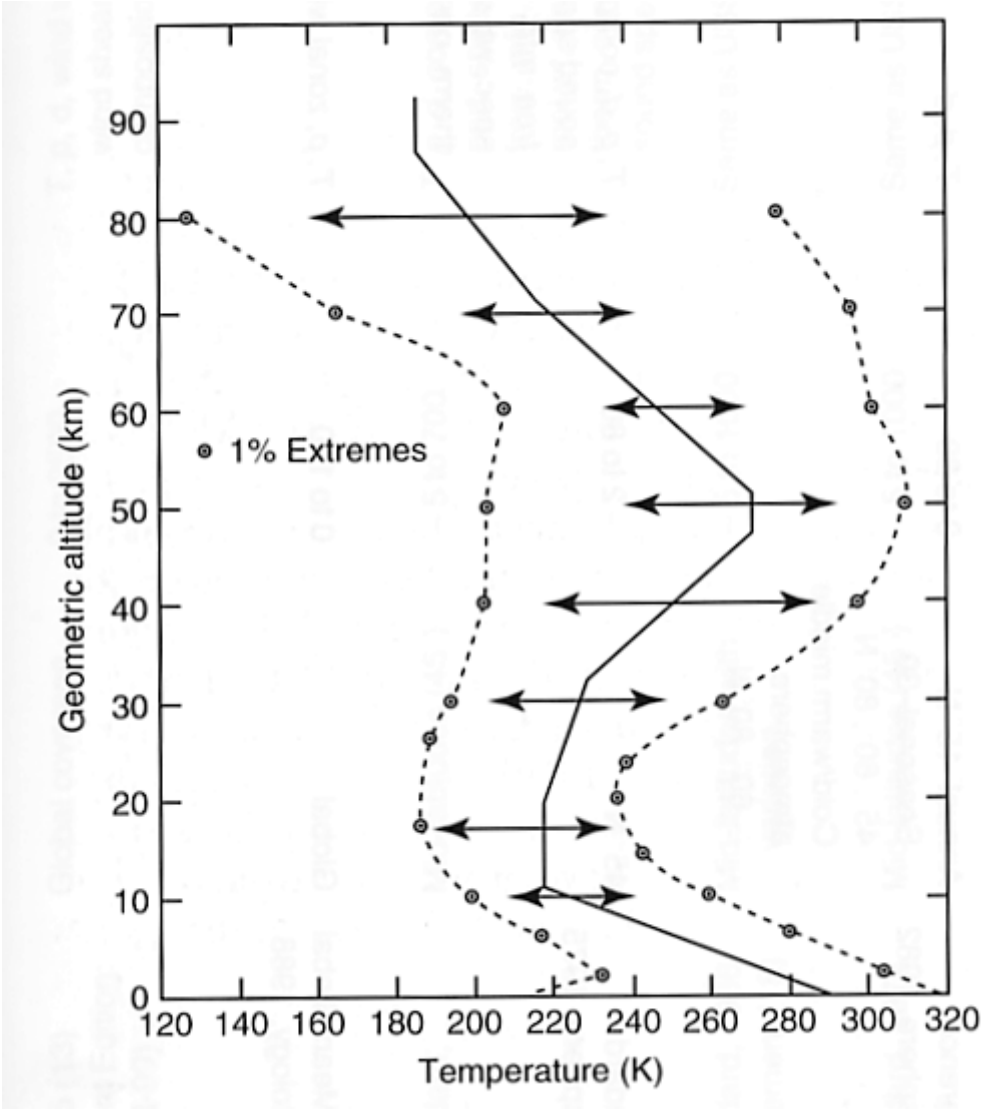

Figure 3 The vertical profile of temperature from the 1976 US standard atmosphere.

As mentioned previously, a clear variation in the atmospheric temperature can be observed with latitude. The CIRA model atmospheric temperature is shown in Figure 4. The mean temperature structure shown is derived from tropospheric data from radiosonde measurements (Oort, 1983), and stratospheric and mesospheric data obtained from satellite measurements (Barnett and Corney, 1985). Within the troposphere temperature decreases with latitude, with the latitudinal gradient being about twice as steep in the winter hemisphere as in the summer hemisphere. The tropopause is also considerably higher and colder over the tropics than over the polar regions. In the lower stratosphere the latitudinal temperature distribution is complicated. The summer hemisphere is characterized by a cold equator and a warm pole,

while the winter hemisphere displays a distinct temperature maximum over mid-latitudes. The cold pool of stratospheric air over the winter pole is highly variable, particularly in the Northern hemisphere where it has been known to disappear completely for periods of a few weeks in midwinter, these periods being related to a phenomenon called a stratospheric sudden warming, the mechanism behind these phenomenon is discussed in detail in this lecture series.

The troposphere is also often referred to as the lower atmosphere. It is here that most “weather” phenomena such as cyclones, fronts, hurricanes, rain, snow, thunder and lightning occur. The stratosphere and mesosphere are also referred to as the middle atmosphere and above this is the upper atmosphere. In the upper atmosphere the effects of ionisation become dominant in determining the atmospheric structure and the air becomes so rarefied that the assumption that it can be treated as a continuous fluid breaks down.

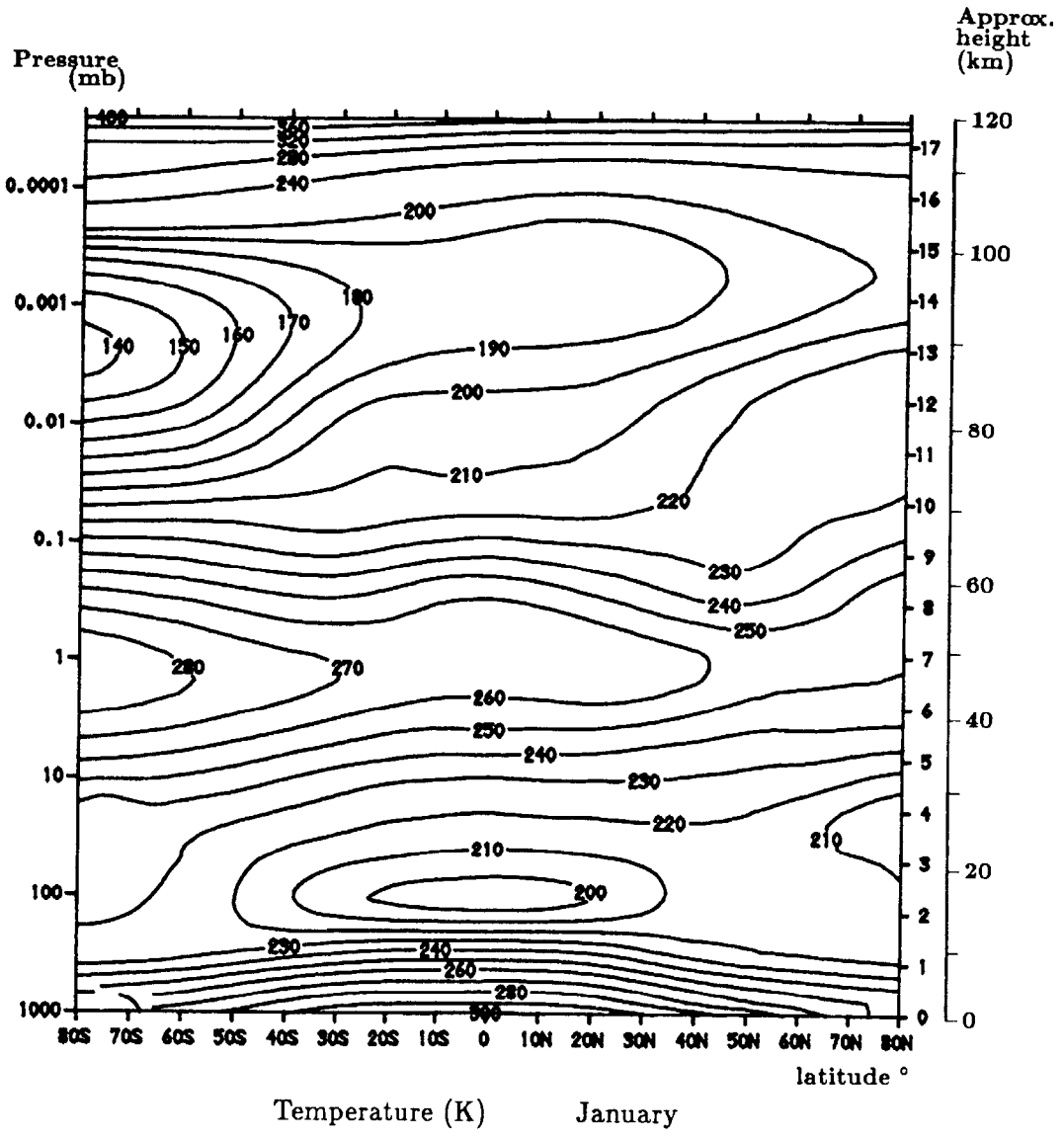

Figure 4 CIRA86 atmospheric temperature contour plot. Taken from Fleming et al. (1990).

Atmospheric temperature measurements made at Aberystwyth using radiosondes which clearly show the variability of temperature are shown in Figure 5. In 1927, the radiosonde, which transmits atmospheric data back to the ground as it is collected, was designed by Pavel Molchanov. A radiosonde is a balloon-borne measurement package that makes measurements of temperature, humidity and pressure. These systems are capable of making observations up to 40km. Radiosondes can also be tracked using ground-based radar to determine the horizontal wind speed and direction with which the background flow advects them. More recently, radiosondes have begun to carry GPS receivers and transmit their position information to the ground from which their velocity can be determined.

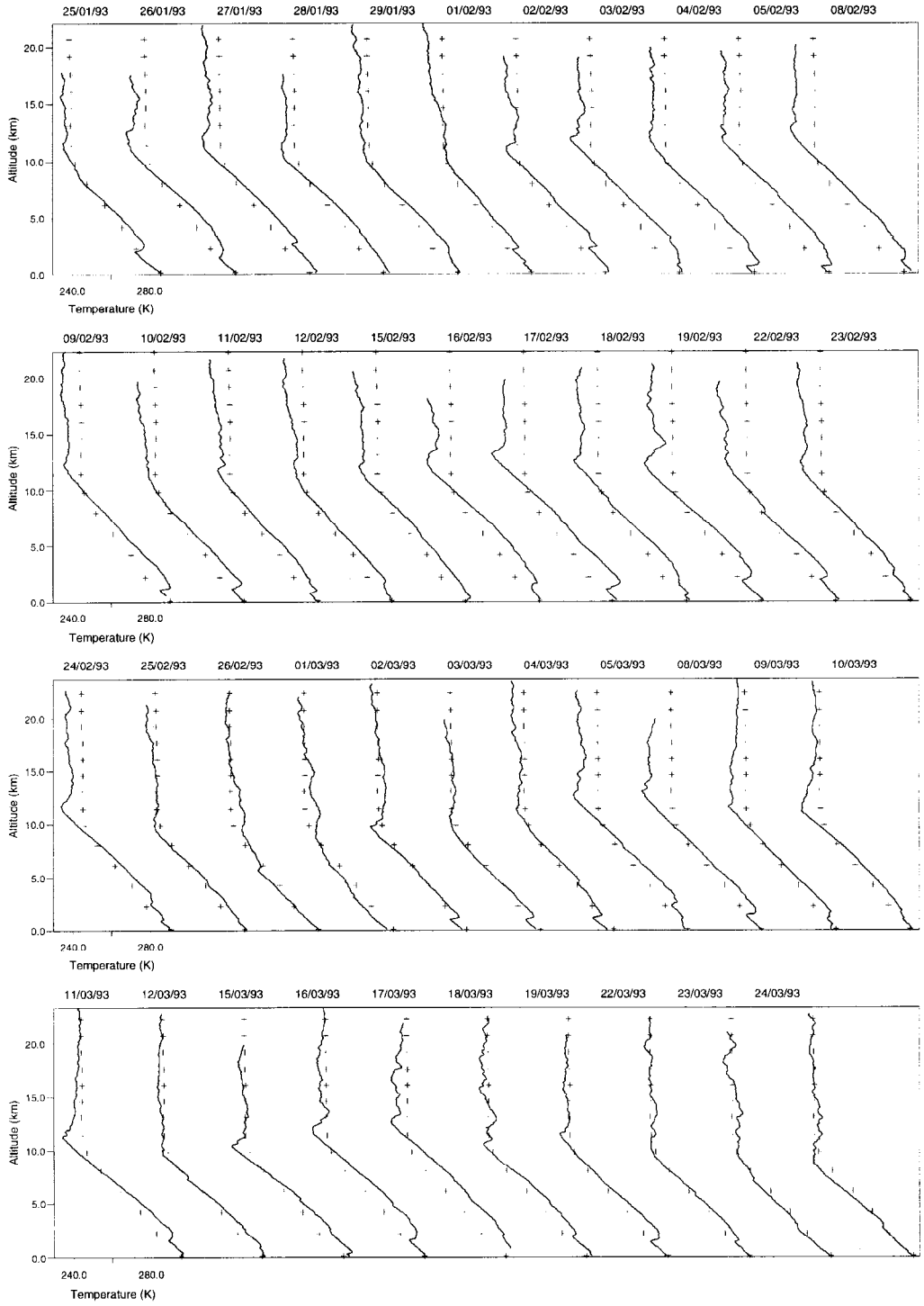

Figure 5 Time series of temperature measurements (full line) in the Troposphere and Stratosphere observed by radiosondes. The labels on the x-axis refer to the first temperature profile, while subsequent profiles have been offset by 40K. The mean temperatures for each month from the CIRA86 model are shown for comparison (crosses). Taken from McDonald (1996)

1.2 General circulation

The basis of geophysical fluid dynamics lies in the principles of momentum, mass and energy conservation. These are expressed mathematically in Newton's 2nd law of motion, the equation of continuity and the thermodynamic energy equation. The eastward, northward and vertical components of the momentum equations, which are derived from Newton's 2nd law, can be written as:

$$\frac{Du}{Dt} - \frac{uv\tan\phi}{a} + \frac{uw}{a} - 2\Omega v\sin\phi + 2\Omega w\cos\phi - F_x = -\frac{1}{\rho}\frac{\partial p}{\partial x}$$

$$\frac{Dv}{Dt} + \frac{u^2\tan\phi}{a} + \frac{wv}{a} + 2\Omega u\sin\phi - F_y = -\frac{1}{\rho}\frac{\partial p}{\partial y}$$

$$\frac{Dw}{Dt} - \frac{u^2+v^2}{a} + g - 2\Omega u\cos\phi - F_z = -\frac{1}{\rho}\frac{\partial p}{\partial z}$$

Equation 4

where the D/Dt is the total derivative which can be expanded out to give something of the form:

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} + \frac{\partial T}{\partial z}\frac{dz}{dt} = \frac{\partial T}{\partial t} + V\cdot\nabla T$$

Equation 5

The formulae in Equation 4 relate to the components of the acceleration in the eastward, northward and vertical directions and are based on using spherical polar coordinate geometry. The right hand term in each of the three formulae in Equation 4 is equal to the pressure gradient force per unit mass for that component. It should be noted that usually at larger scales the atmospheric pressure gradient is very nearly vertical ($\partial p/\partial z$ is very much greater than $\partial p/\partial x$ or $\partial p/\partial y$ since pressure falls exponentially with increasing altitude). It thus follows that the largest component of the pressure gradient force per unit mass acts vertically upward. However, it should be noted that thanks to the independence of the vertical and horizontal components of the equation of motion, the small horizontal pressure gradient forces are still very important. The terms which include the magnitude of the angular velocity of the Earth, Ω, represent the Coriolis force per unit mass. The terms in Equation 4 which are proportional to the inverse of a, which is the Earth's radius, are called the curvature terms. They arise owing to the curvature of the Earth and result from the fact that the natural coordinate system to use is the spherical polar coordinate system. F_x, F_y, and F_z represent the viscous force per unit mass (or frictional force per unit mass) along a latitude circle (eastward), along a longitude circle (northward) and in the vertical direction. In Equation 4 the zonal, meridional and vertical velocity components are defined by u, v and w, respectively, and ϕ is the latitude.

The eastward, northward and vertical component momentum equations are complicated. However, the magnitudes of the different terms in this set of equations are very different depending on the scale of the motion under study. Thus, it is usual to eliminate various terms in the momentum equations based on scale considerations. For example, in order to simplify the momentum equations (Equation 4) for synoptic scale motions, we are basically just considering large scales, we define the following characteristic scales of the field variables in Table 1 based on observed values for mid-latitude synoptic systems.

Scale variable	Symbol	Value
Horizontal velocity scale	U	$10 ms^{-1}$
Vertical velocity scale	W	$0.01 ms^{-1}$
Length scale	L	$10^6 m$
Depth scale	D	$10^4 m$
Horizontal pressure fluctuation scale	$\Delta P/\rho$	$10^3 m^2 s^{-2}$
Time scale	L/u	$10^5 s$

Table 1 Approximate values of field variables for synoptic scale motions.

In Equation 4 it is convenient to use the Coriolis parameter, f, which can be written as:

$$f = 2\Omega \sin\phi$$

Equation 6

where Ω is the angular velocity of the Earth and ϕ is the latitude. For a latitude of 45 degrees we introduce the notation:

$$f_o = 2\Omega \sin\phi_o = 2\Omega \cos\phi_o \approx 10^{-4} s^{-1}$$

Equation 7

which is used in Table 2.

x-component momentum equation	$\frac{Du}{Dt}$	$-\frac{uv\tan\phi}{a}$	$\frac{uw}{a}$	$-2\Omega v\sin\phi$	$2\Omega w\cos\phi$	$=-\frac{1}{\rho}\frac{\partial p}{\partial x}$
y-component momentum equation	$\frac{Dv}{Dt}$	$\frac{u^2\tan\phi}{a}$	$\frac{wv}{a}$	$2\Omega u\sin\phi$		$=-\frac{1}{\rho}\frac{\partial p}{\partial y}$
Scales of individual terms	$\frac{u^2}{L}$	$\frac{u^2}{a}$	$\frac{uw}{a}$	$f_0 u$	$f_0 w$	$\frac{\Delta p}{\rho L}$
Magnitude of terms (ms^{-2})	10^{-4}	10^{-5}	10^{-8}	10^{-3}	10^{-6}	10^{-3}

Table 2 Scale analysis of the Horizontal momentum equations.

Table 2 indicates the characteristic magnitude of each of the terms in the horizontal component terms of the momentum equation (Equation 4) based on the scaling conditions displayed in Table 1. Note frictional terms are not included because for the synoptic time scale frictional dissipation is of secondary importance above about 1km (often referred to as the boundary layer).

1.2.1 Geostrophic Approximation

It is apparent that for midlatitude synoptic scale disturbances the Coriolis term in the horizontal momentum equations and the pressure gradient force are in approximate balance. Therefore, retaining only these two terms, we obtain as a first approximation the geostrophic relationship:

$$-fv=-\frac{1}{\rho}\frac{\partial p}{\partial x}$$

$$fu=-\frac{1}{\rho}\frac{\partial p}{\partial y}$$

Equation 8

where $f=2\Omega\sin\phi$ is the Coriolis parameter. The geostrophic balance is a diagnostic expression which gives the approximate relationship between the pressure field and the horizontal velocity in synoptic scale systems. It is possible to define a horizontal velocity field, $\mathbf{V_G}=\mathbf{i}u_G+\mathbf{j}v_G$, called the geostophic wind. Thus, in vectorial form Equation 8 can be rewritten as:

$$\mathbf{V_G}=\mathbf{k}\times\frac{1}{\rho f}\nabla p$$

Equation 9

Thus, knowledge of the pressure distribution at any time determines the geostrophic wind. The geostrophic approximation describes the familiar situation in which the wind blows parallel to the isobars, for the Northern hemisphere in a clockwise direction around centres of high pressure or anticyclones (see Figure 6) and anticlockwise around centres of low pressure or cyclones. In the Southern hemisphere the opposite relation holds, flow is clockwise around low pressure and anticlockwise around high pressure.

It should be kept in mind that Equation 9 always defines the geostrophic wind. But, that only for large scale motions should the geostrophic wind be used as an approximation of the actual horizontal wind field. It should be noted that in general the wind velocity is not exactly geostrophic. If it were the velocity would have no time dependency and weather could not change.

However, the geostrophic approximation works quite well at heights above 1km (at lower levels friction becomes important) and for latitudes greater than 20 degrees. At latitudes less than the 20 degrees the Coriolis term is not much larger than the other terms. Under these conditions the wind velocity and direction may be deduced from a chart of isobars. At lower levels where friction is not negligible, since friction acts in a direction approximately opposite to **V**, the velocity can no longer be precisely parallel to the isobars. For a high pressure centre the balancing velocity has a component outward from the centre (see Figure 7) and for a low pressure centre an inward component.

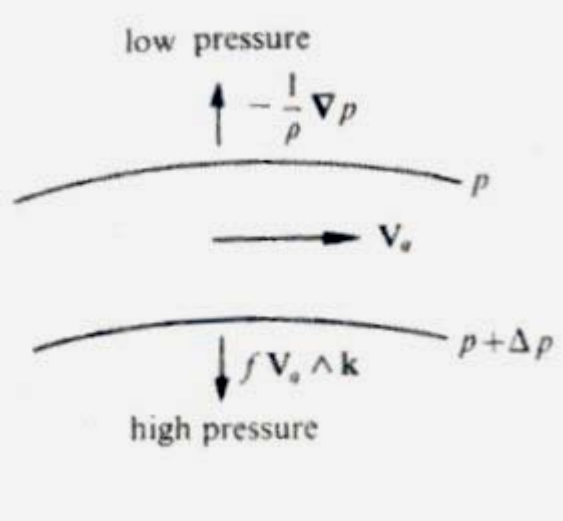

Figure 6 Balance of forces that produces the geostrophic wind around an anticyclone in the Northern hemisphere. Taken from Houghton (2001).

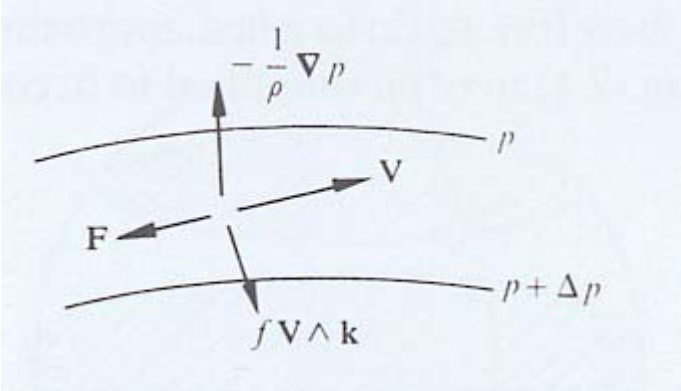

Figure 7 Balance of forces that produces the wind around an anticyclone in the Northern hemisphere when friction must be taken into account. Taken from Houghton (2001).

1.2.2 Hydrostatic approximation

A similar analysis can be applied to the vertical component of the momentum equation (Equation 4). Since pressure decreases by about an order of magnitude from the ground to the tropopause, the vertical pressure gradient may be scaled by P_o/H where P_o is the surface pressure and H is the scale height (see Equation 3). The different terms in the vertical component of the momentum equation, using the scale values specified in Table 1, are displayed in Table 3. As previously frictional terms are neglected and the motions are centred at 45 degrees latitude. Examination of Table 3 indicates that to a high degree of accuracy the pressure field is in hydrostatic equilibrium, that is, the pressure at any point is simply equal to the weight of a unit cross section column of air above that point. Thus, to a good approximation:

$$\frac{\partial p}{\partial z} = -g\rho$$

Equation 10

z-component momentum equation terms	$\frac{Dw}{Dt}$	$\frac{u^2+v^2}{a}$	g	$-2\Omega u\cos\phi$	$-\frac{1}{\rho}\frac{\partial p}{\partial z}$
Scales of individual terms	$\frac{uw}{L}$	$\frac{u^2}{a}$	G	$f_0 u$	$\frac{P_o}{\rho H}$
Magnitude of terms (ms^{-2})	10^{-7}	10^{-5}	10^{1}	10^{-3}	10^{1}

Table 3 Scale analysis of the Vertical momentum equation.

1.2.3 Thermal wind shear balance equation

The geostrophic equation (Equation 9) can be rearranged using the ideal gas law per unit mass ($p = RT\rho$) to give:

$$fu = -\frac{RT}{p}\frac{\partial p}{\partial y} = -RT\frac{\partial \ln p}{\partial y}$$

Equation 11

Similarly, the hydrostatic approximation (Equation 10) becomes:

$$-g = \frac{1}{\rho}\frac{\partial p}{\partial z} = \frac{RT}{p}\frac{\partial p}{\partial z} = RT\frac{\partial \ln p}{\partial z}$$

Equation 12

By taking the vertical derivative of Equation 11 and the zonal derivative of Equation 12 we obtain:

$$f\frac{\partial u}{\partial z} = -RT\frac{\partial \ln p}{\partial z\,\partial y}$$

Equation 13

$$\frac{\partial}{\partial y}\left(-\frac{g}{RT}\right) = \frac{g}{RT^2}\frac{\partial T}{\partial y} = \frac{\partial \ln p}{\partial z\,\partial y}$$

Equation 14

By inserting Equation 13 into Equation 14, we get the relationship between the horizontal temperature gradient and the vertical gradient of the horizontal wind:

$$f\frac{\partial u}{\partial z} = -RT\frac{g}{RT^2}\frac{\partial T}{\partial y}$$

Equation 15

which in turn can be simplified and rewritten as:

$$\frac{\partial u}{\partial z} = -\frac{g}{fT}\frac{\partial T}{\partial y}$$

Equation 16

Equation 16 is generally referred to as the thermal wind shear balance equation. It should be noted that a similar expression for the temperature gradient as a function of longitude and the meridional wind can be derived.

Thus, the zonal-mean zonal winds are related to the temperature field in Figure 4 by the thermal wind shear balance equation. The zonal-mean zonal

(west-to-east) winds for January are shown in Figure 8. Thus, the wind field in the Northern hemisphere is representative of the winter circulation and the Southern hemisphere is representative of the summer circulation. The pattern observed is roughly reversed for winter in the Southern hemisphere. Note that when the atmospheric circulation is averaged with respect to longitude, the zonal (east-west) wind component turns out to be about an order of magnitude larger than the meridional (north-south) component at most locations.

Before going any further it is worth noting that the meteorological term 'easterlies' describes winds that blow from the east, while 'westerlies' blow from the west. Unfortunately, the terms 'eastward' (to the east) and westward (to the west) are also used commonly. Zonal is the wind along lines of longitude and by convention, the zonal wind is positive when from the east. The meridional wind is the wind along lines of latitude and by convention, the meridional wind is positive when from the north.

In the troposphere the mean zonal winds are generally eastward (or westerly) at midlatitudes with two prominent ‘jet streams’, and westward at low latitude. In the stratosphere and mesosphere the mean zonal winds are generally eastward (westerly) in winter and westward (easterly) in summer. The winds at low latitudes shown in Figure 8 are not representative of every January, partly because the winds do not follow a simple annual cycle in the equatorial lower stratosphere; in this region, there is a prominent Quasi-Biennial Oscillation (QBO), with a period of approximately 28 months (see Figure 9).

The QBO dominates the variability of the equatorial stratosphere and is easily seen as downward propagating easterly and westerly wind regimes, with a variable period of approximately 28 months. The QBO is a fascinating example of a coherent, oscillating mean flow that is driven by propagating waves with periods unrelated to that of the resulting oscillation. We will discuss the QBO in some detail later in this set of lectures. Although the QBO is a tropical phenomenon it affects the stratospheric flow from pole to pole by modulating the effects of extra-tropical waves. Chemical constituents, such as ozone, water and methane are also affected by circulation changes induced by the QBO. Through modulation of extra-tropical wave propagation, the QBO also has an effect on the breakdown of the wintertime stratospheric polar vortices and the severity of high-latitude ozone depletion.

During the winter season the high latitude stratosphere cools, forming a deep, strong westerly vortex. The strong westerlies are replaced by easterlies with increasing solar heating in the spring and summer. In both hemispheres the smoothly varying seasonal cycle described is modified by the effects of planetary Rossby waves. If the mean flow of the tropical stratosphere is westerly, planetary waves are able to penetrate into the tropics and even across the equator without encountering a critical line, a phenomenon discussed in detail later in this course. By contrast, when the mean flow in the tropical stratosphere is easterly, planetary waves encounter a critical line on the winter side of the equator. Thus, the effective wave guide for planetary wave propagation is narrower, and as a result, wave activity at middle and

high latitudes in the winter hemisphere tends to be stronger. Stronger wave activity leads to greater wave-induced drag on the mean flow, reduced westerly winds, and hence a warmer polar stratosphere.

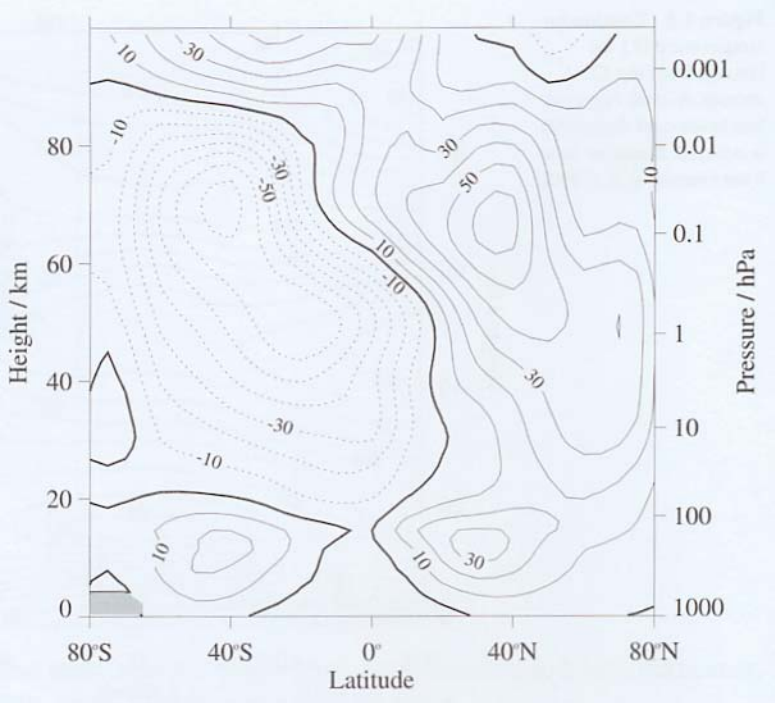

Figure 8 CIRA86 zonal wind contour plot for January. Taken from Fleming et al. (1990).

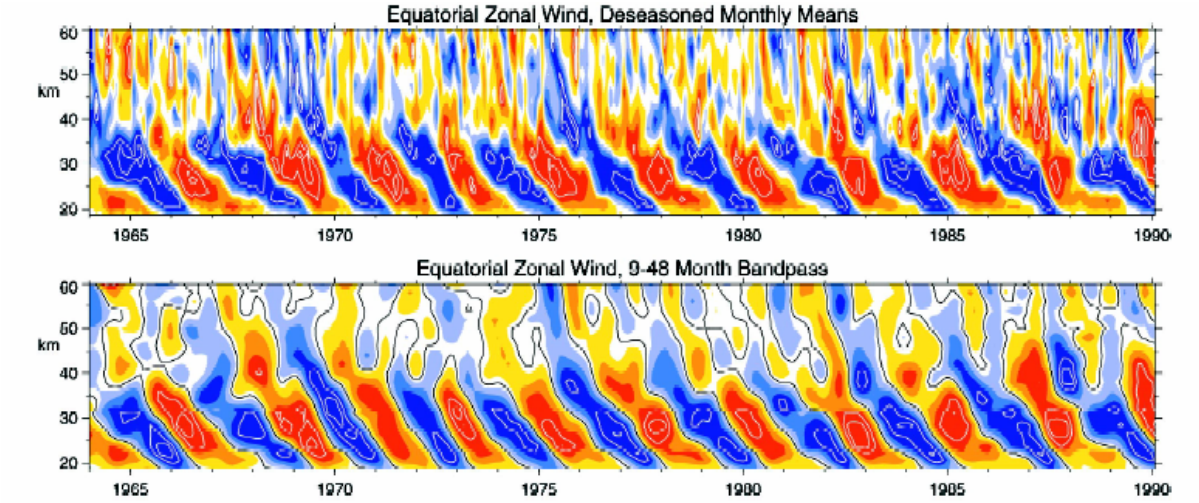

Figure 9 Time-height section of the monthly-mean zonal wind component, with the seasonal cycle removed, for 1964–1990. Below 31 km, equatorial radiosonde data are used from Canton Island (2.8°N, January 1964 to August 1967), Gan/Maledive Islands (0.7° S, September 1967 to December 1975), and Singapore (1.4°N, January 1976 to February 1990). Above 31 km, rocketsonde data from Kwajalein (8.7°N) and Ascension Island (8.0°S) are shown. The contour interval is 6 m/s, with the band between +3 and -3m/s unshaded. Red represents positive (westerly) winds. In the bottom panel the data are band-pass filtered to retain periods between 9 and 48 months. Taken from Baldwin et al. (2001)

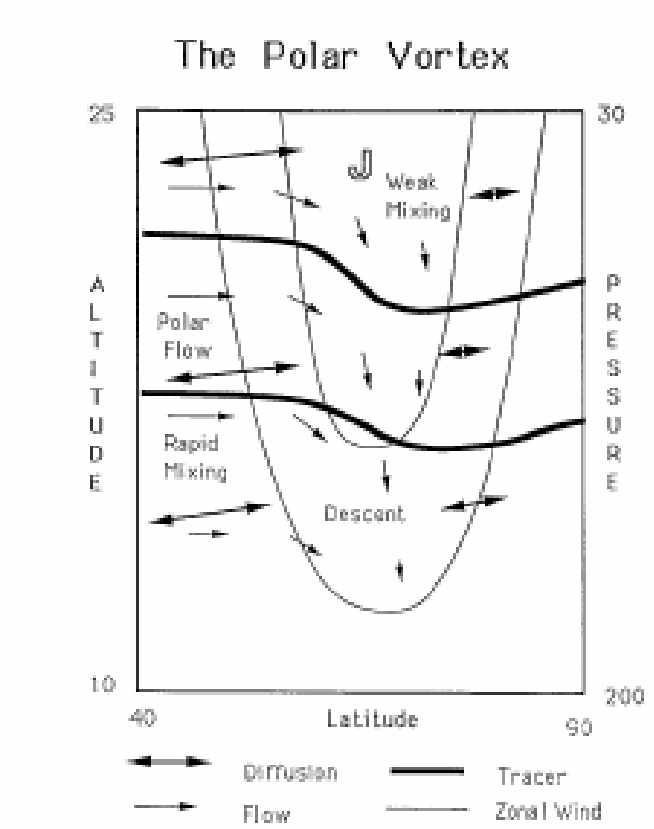

Figure 10 Schematic of the polar night jet. Schoeberl et al. (1992).

The circulation of the winter stratosphere is dominated by a large cyclonic vortex centred near the winter pole. Changes in circulation are generally related to changes in shape or location of the polar vortex. The vortex is encircled by a strong eastward flowing jet, the polar night jet (see Figure 10). This jet is disturbed by planetary-scale Rossby waves that propagate upward from the troposphere. The Rossby waves break in the stratosphere, leading to a well-mixed stratospheric “surf zone” in the mid-latitude region exterior to the vortex. Despite this strong mixing mechanism, the vortex itself appears quite robust and resistant to transport across its boundary.

Although both the Northern and Southern hemisphere vortices form in autumn to early winter and demise in late-winter to spring, with the formation and maximum size occurring earlier at upper levels, there are two significant differences between the hemispheres. The Arctic vortex is much smaller and breaks down earlier. There is also a difference in the vertical variation of the size of the vortices, with the increase in area with altitude larger for the Antarctic vortex. The displacement off the pole and elongation of the vortices is also dramatically different. The arctic vortex is displaced further off the pole and more elongated than the Antarctic vortex. The interannual variability of the Antarctic vortex is small except during the vortex breakdown and in the upper stratosphere in July to August. This is consistent with the variation in the location of the winter jet at 1hPa. The interannual variability of the Arctic vortex is much larger than the Antarctic vortex. The large interannual variability is related to the occurrence in the Northern hemisphere of events (usually stratospheric warmings) where the vortex is highly distorted.

Stratospheric warmings and in particular the unusual Southern hemisphere major stratospheric warming observed in 2002 are discussed in detail.

Geller (1983) demonstrated by means of model calculations that while the general configuration of middle atmosphere dynamics (winter westerlies and summer easterlies) is determined by differential solar heating, deviations of the zonally-averaged temperature field from radiative equilibrium and the closure of the jet structures with increasing altitude result from the action of zonal mean momentum deposition processes. He argued that the primary candidate for these momentum depositions are wave motions in the atmosphere.

Comparison of computed temperatures for radiative equilibrium and observations indicate several regions of disagreement (see Figure 11). For example, the observed tropical cold tropopause temperatures are not modelled when Figure 4 and Figure 11 are observed. The warm summer stratopause is too warm by about 10K, but the winter stratopause is too cold by about 40K, and the summer mesosphere is warmer than the winter mesosphere as must be the case for radiative equilibrium. Comparison shows that the middle atmosphere zonal winds calculated for the radiative equilibrium state are too strong and the jets do not close as observed (see Figure 12). Thus, while the broad configuration of the middle atmosphere circulation is driven by the distribution of solar heating. In order to explain the closed jet configurations and those aspects of the mean temperature distribution that are at variance with radiative considerations, one must include the effects of zonal momentum deposition. A discussion of these mechanisms will be left to later in this lecture series.

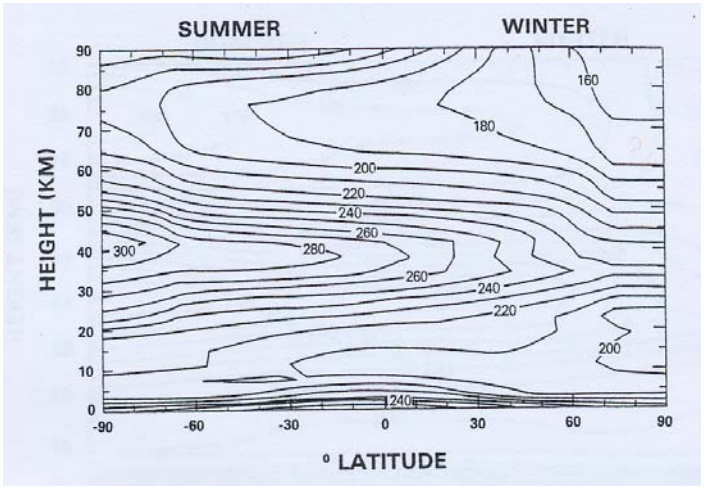

Figure 11 The model temperature determined using only radiative equilibrium. Taken from Geller (1983).

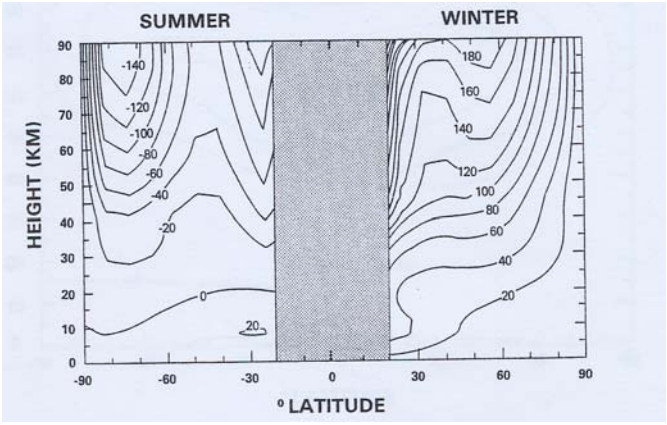

Figure 12 Geostophic mean zonal winds calculated from the radiative equilibirium temperatures shown in Figure 11. Note no values are shown near the Equator because of the inapplicability of the geostrophic formula. Taken from Geller (1983).

1.3: Atmospheric composition and Ozone

The relative abundance of various chemical species in the Earth's atmosphere is shown in Table 4. Chemical species concentrations are usually given as mixing ratios by volume, these mixing ratios are in turn generally given in fractions (e.g. parts per million by volume or ppmv). For example, the ozone mixing ratio is the ratio of the number of 0_3 molecules per unit volume to the total number of molecules in that volume. Figure 13 shows representative profiles of the species listed in Table 4. It is particularly important to note that both molecular oxygen and carbon dioxide are uniformly distributed below about 100km.

Molecule	Volume fraction	Comments
N_2	0.7808	Nitrogen molecules
O_2	0.2095	Oxygen molecules
H_2O	<0.04	Water vapour
A	$9.34X10^{-3}$	Argon
CO_2	$3.45X10^{-4}$	Carbon Dioxide
CH_4	$1.6X10^{-6}$	Methane
N_2O	$3.5X10^{-7}$	Nitrous Oxide
CO	$7X10^{-8}$	Carbon monoxide
O_3	Approx. 10^{-8}	Ozone

Table 4 Atmospheric composition.

Probably the most important gas in the troposphere, at least from the point of view of its interaction with electromagnetic radiation, is water vapour. Water vapour is especially important in the troposphere because of its role in cloud formation and precipitation and in transporting significant amounts of energy in the form of latent heat and infrared radiation. Water vapour is also one of the most variable components of the atmosphere. In the troposphere, the average relative humidity is close to 50% and the water vapour pressure varies over a wide range.

Clouds regularly cover about 65% of the Earth, and occur in various types. Some, such as cirrus in the tropics and stratus near the coastal areas and in the Arctic are climatologically persistent. The total amount of sunlight absorbed by the Earth's surface and the amount of thermal radiation that escapes it are determined, in large part, by the properties of the clouds. By reflecting incoming solar radiation back to space, they cool the Earth-atmosphere system the so-called cloud albedo effect. Clouds also absorb solar radiation in the near-infrared region. The cooling of the Earth-atmosphere system by the cloud albedo effect occurs primarily at the surface. The solar albedo of clouds depends substantially on cloud type and cloud form, as well as the solar zenith angle.

The most straightforward and simple diagnostic measure of the impact of clouds on solar radiation is the short-wave cloud forcing, which is defined as the difference of the net solar irradiances at the top of the atmosphere

between all-sky and cloudless conditions. Here the net irradiance is the incoming solar radiation minus the reflected radiation. Satellite measurements suggest that the global short-wave cloud forcing is about -45Wm^{-2}. Short-wave cloud forcings are maximized (about -120 W m^{-2}) in the summer hemisphere at about latitude 60 degrees where solar input is large and low clouds are abundant.

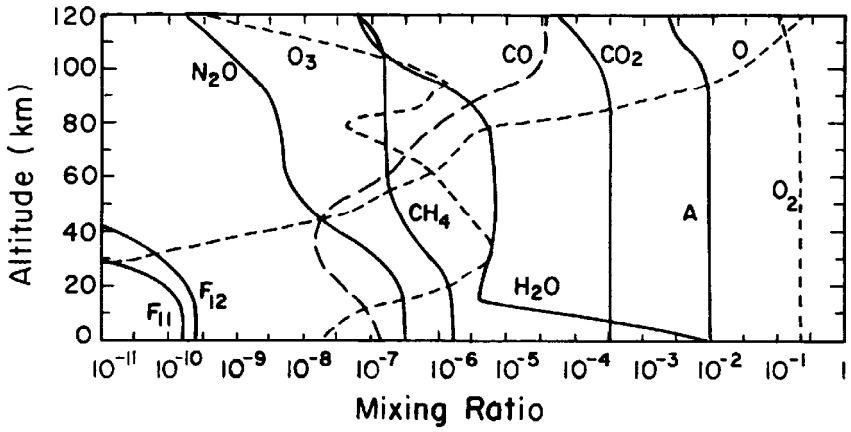

Figure 13 Vertical profiles of mixing ratios of selected chemical species in the atmosphere.

1.3.1 Ozone

At higher altitudes, ozone becomes important. This is mainly due to the fact that ozone absorbs the bulk of solar ultraviolet radiation in the 290-320 nm wavelength range in this region. Ozone is formed from molecular oxygen (O_2) by ultraviolet and extreme ultraviolet photolysis followed by recombination of atomic oxygen (O) with O_2:

$$O_2 + UV \rightarrow O + O$$

Equation 17

$$O + O_2 + M \rightarrow O_3 + M$$

Equation 18

where the third body (M) is present to allow for conservation of energy and momentum. O_3 amounts are small in the stratosphere, with its mixing ratio (the ratio of the number of O_3 molecules per unit volume to the total number of molecules in that volume) rarely exceeding 10 ppmv. The total amount of O_3 in the stratosphere is such that if brought to the surface it would form a layer between 1 and 5 mm thick. Stratospheric O_3 comprises some 90% of the O_3 in the atmosphere. These amounts, though small, are highly important.

In fact, it is these small amounts of O_3 that help define the temperature structure of the stratosphere. Figure 14 shows that the two major characteristics of the stratosphere are a temperature that increases with

altitude and large amounts of O_3. These characteristics are related because the absorption of ultraviolet radiation by O_3 leads to heating of the stratosphere when the energy of the ultraviolet radiation is converted to kinetic energy of atmospheric molecules:

$$O_3 + UV, \text{visible light} \rightarrow O + O_2$$

Equation 19

The net result of processes Equation 18 and Equation 19 is only radiative (there is no net chemical effect). The increase of temperature with altitude in the stratosphere has important consequences for the meteorology there. Areas of increasing temperature with height are more stable with respect to vertical mixing than areas of decreasing temperature. Thus, instead of being mixed vertically by overturning or other convective motions, trace constituents can only be mixed vertically by either the large-scale dynamics of the stratosphere (for example, rising vertical motion in the tropics balanced by descent at high latitudes), by features associated with large planetary-scale wave disturbances, or by smaller-scale gravity waves. The planetary wave activity is located primarily in the winter hemisphere. In the summer hemisphere, mixing by random motion of air parcels (eddy diffusion) becomes increasingly important.

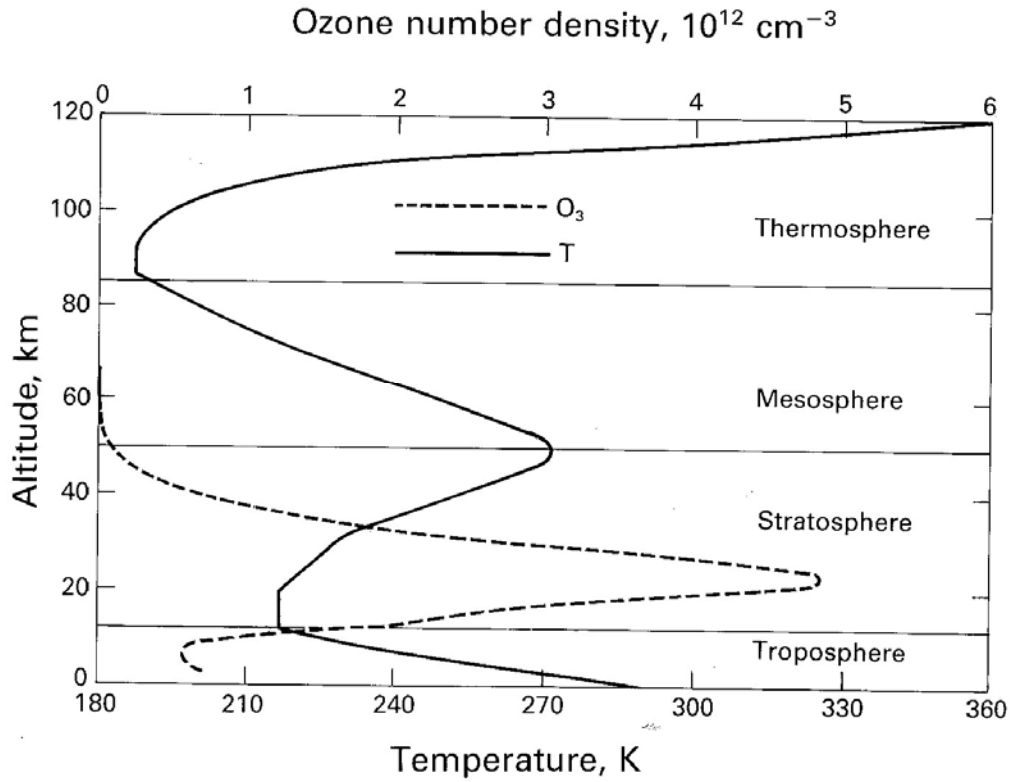

Figure 14 Vertical profile of temperature and ozone number density in the atmosphere.

The distribution of O_3 in the stratosphere is strongly affected by the meteorology there. Most production and destruction of O_3 occurs in the

tropical upper stratosphere, where ultraviolet amounts are largest (see the distribution of solar insolation shown in Figure 25). Dissociation of O_3, however, extends lower in the stratosphere and to higher latitudes than does production; transport processes must obviously play an important role in controlling the distribution of a species whose production and destruction are not in local balance everywhere.

In fact early observations of ozone gave the first clues about the meridional circulation in the stratosphere, this circulation being necessary in order to bring the ozone created in the tropical stratosphere to higher latitudes. The meridional circulation in the stratosphere is shown schematically in Figure 15. This circulation is often called the Brewer-Dobson circulation and is responsible for the transport of all the long-lived trace elements, for instance all the CFCs. These elements manage through the rising branch of the Hadley circulation to enter the tropical stratosphere and from there to reach the polar regions in winter. It is in this way that the CFCs produced in the Northern hemisphere reach Antarctica where they destroy the ozone in spring.

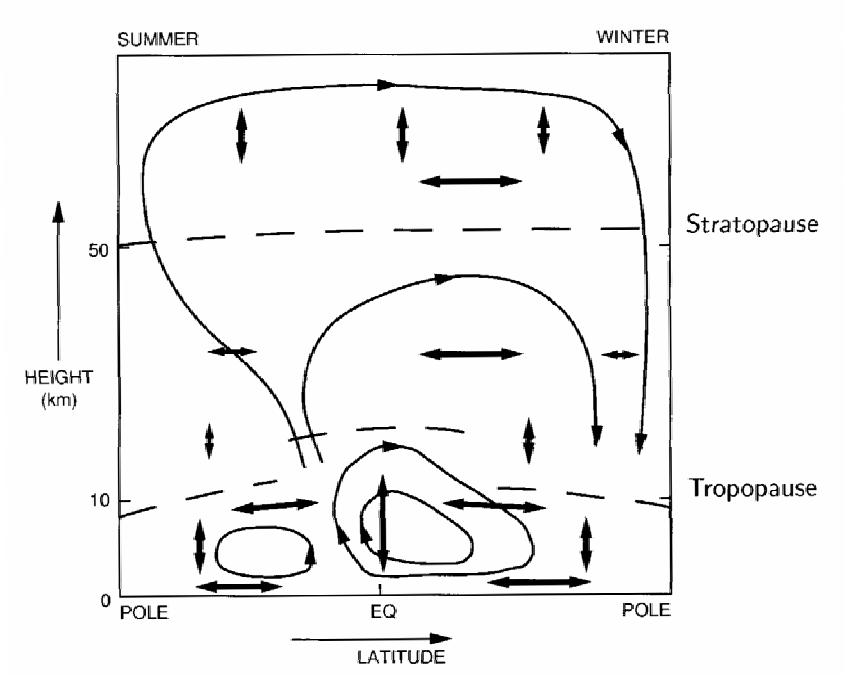

Figure 15 Schematic picture of the meridional circulation in the atmosphere. Taken from Labitzke and Van Loon (1999).

The reason for the separation between the production and destruction regions of O_3 is that the processes resulting in O_3 destruction differ markedly from those resulting in its production. O_3 production is driven by ultraviolet and extreme ultraviolet photolysis of O_2 (wavelengths below 240 nm), while O_3 loss typically involves production of atomic oxygen (see Equation 19) which occurs at longer wavelengths in the ultraviolet (up to 320 nm) and, to a lesser extent, in the visible. Since longer wavelength photons penetrate deeper into the atmosphere, this difference between production and destruction regions of O_3 can occur. Most O_3 destruction occurs by catalytic processes involving free

radical species containing either hydrogen, nitrogen, chlorine, or bromine atoms. These may usually be represented by a two-step cyclic pathway involving the free radical X:

$$O_3 + X \rightarrow XO + O_2$$

Equation 20

$$O + OX \rightarrow X + O_2$$

Equation 21

The most important radicals (X) include hydroxyl (OH), nitric oxide (NO), chlorine (Cl), and bromine (Br). These are themselves formed in the stratosphere by the decomposition of various atmospheric source gases, which can be of either natural or anthropogenic origin. For example, OH comes from decomposition of water (H_2O) or methane (CH_4), NO comes from nitrous oxide (N_2O), Cl comes from naturally occurring methyl chloride (CH_3Cl) or anthropogenically emitted chlorofluorocarbon (CFC) molecules, such as dichlorodifluoromethane (CF_2Cl_2).

Ignoring transport, the rate of change of ozone can be written schematically as:

$$\frac{d[O_3]}{dt} = 2J[O_2] - 2k_a[O][O_3] - 2k_b[NO_2][O] - 2k_c[HO_2][O] - 2k_d[ClO][O] - 2k_e[BrO][O] - etc.$$

Equation 22

J represents a photodissociation rate and the k's are rate constants for individual reactions. Equation 22 includes a number of the more important cycles; others are ignored.

The key point about the catalytic cycles is that radical speices act as catalysts in that their concentration is unaffected by the ozone destruction cycle, reactions (Equation 20 and Equation 21), allowing the cycle to repeat many times. The cycle is terminated only when the radicals are converted to reservoir species. That means that despite having concentrations much less than that of ozone, the radicals can have an enormous influence on the ozone distribution.

Stratospheric chemistry is quite complicated because the chemistry of the various compounds of all these chemical elements (oxygen, hydrogen, nitrogen, chlorine, and bromine) is occurring simultaneously. In addition, the different families' chemistries interact. For example, catalytically active constituents in two different families may recombine to form a so-called 'reservoir' molecule, which are unreactive towards O_3. Thus, in the case of the hydrogen and nitrogen families, the reservoir species nitric acid (HNO_3) can form. Because these reservoir molecules are fairly stable (being only slowly photolysed by ultraviolet radiation and having a lifetime on the order of days to weeks), they can build up in large concentrations. For example, for HNO_3, concentrations of up to 10 parts per billion by volume (ppbv) are not uncommon in the winter; the largest concentrations of OH (a very reactive radical) observed are approximately 1 ppbv. Since these reservoir molecules are long-lived, they can also undergo appreciable transport. They therefore provide a mechanism by which the 'raw materials' for O_3 destruction can be transported from one region of the stratosphere to another.

The stratospheric system, then, is a complex one, whose composition, winds, and temperature are all related. If the amounts of O_3 in the stratosphere changed, ultraviolet heating rates could as well, which would lead to changed stratospheric temperatures. Since winds form in response to temperature gradients, changes in wind fields would also result, and these would then, in turn, affect the transport of O_3 and the chemical source and reservoir species in the stratosphere. The concentration of radical species derived from these species would then change, and these would affect the O_3 distribution, completing a cycle of interconnected physical processes.

The recognition that mankind is dramatically affecting the composition of the Earth's atmosphere has led to substantial interest in stratospheric composition

and meteorology in the last 20 years. In particular, observations of ozone amounts over the last two decades suggest that there have been reductions in ozone amounts. Ground-based measurement showed decreases in the total column ozone amount of 1.7 to 3.0% over the period from 1969 to 1986 between 30° and 64° N, with larger decreases occurring in the winter than in the summer. Much larger decreases in ozone have occurred over the Antarctic in Austral spring. In order to understand better the future of the stratosphere, it is important that its current state be understood. Remote sensing from ground-based and satellite-based instruments provides an important tool for the study of the stratosphere.

The distribution of total ozone is now well known through the measurements made by the satellite-based Total Ozone Mapping Spectrometer (TOMS) instrument and is shown in Figure 16. In tropical latitudes, there is a broad minimum with values below 290 Dobson Units (DU). The Dobson unit is used as a measure of the total ozone amount, and is named after Dobson. One hundred DU corresponds to a layer of ozone 1mm thick at a pressure of 1013hPA and a temperature of 15^{o}C.

The ozone is at a minimum in the tropics because the prevailing stratospheric winds transport the ozone into middle and high latitudes. In middle Northern latitudes the annual mean of total ozone is between 320 and 400DU. However, this value varies through the year for natural reasons between 200 and 500DU, with especially large variability in winter and spring. Figure 16 also shows that the ozone maximum in the Southern hemisphere has less ozone than in the corresponding maximum in the Northern hemisphere. This difference can be explained by the difference in the atmospheric circulation and temperature levels in the two hemispheres. The regions of the ozone maximum are in appreciably colder in the Southern hemisphere, and lower temperatures are associated with lower ozone amounts.

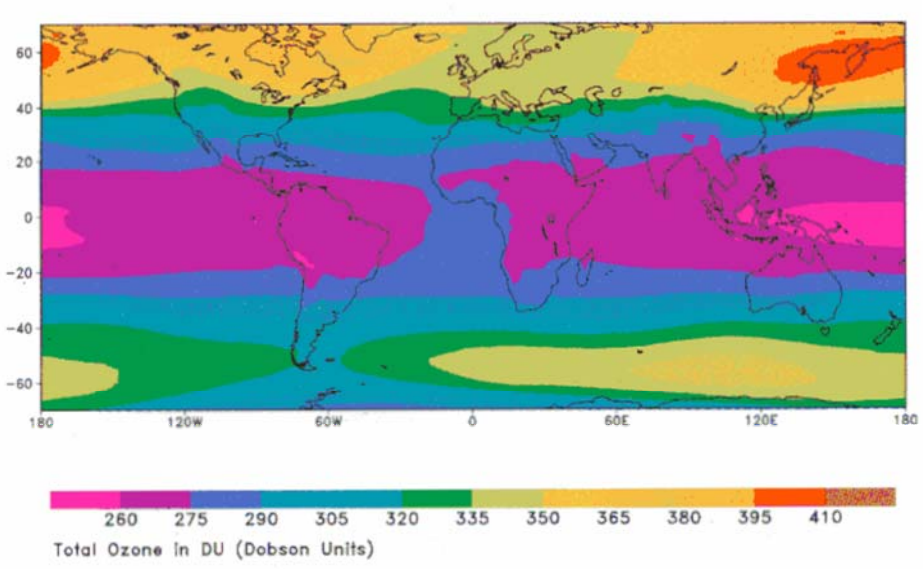

Figure 16 The distribution of annual mean total ozone in Dobson units for 1979 to 1992. Taken from Labitzke and Van Loon (1999).

The variation of atmospheric ozone with latitude and season is shown in Figure 17 and Figure 18 for measurements made in 1980 and 1995, respectively. Figure 17 and Figure 18 display the total column ozone amount in Dobson units. A Dobson unit is the equivalent vertical thickness of atmospheres, in thousandths of cm's, that is occupied by the ozone when concentrated into a uniform layer of pure gas at STP. Comparison shows rather obvious changes can be observed near the summer hemisphere pole in the Antarctic spring which is a clear manifestation of the ozone hole. The zonal mean volume mixing ratio of ozone in ppmv as a function of latitude and height for January is displayed in Figure 19.

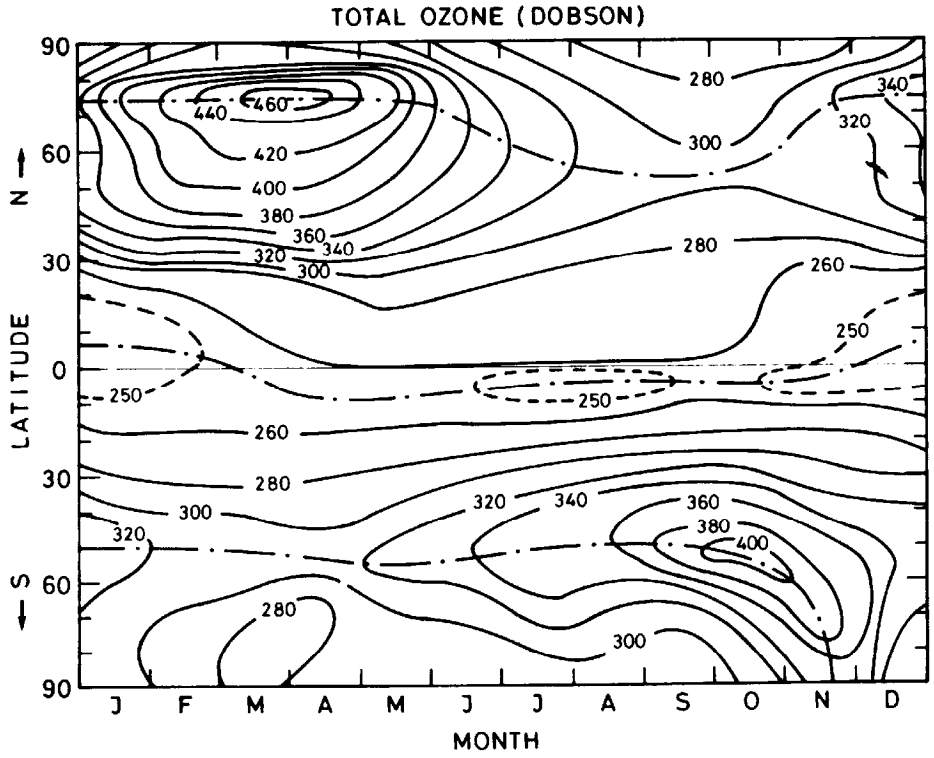

Figure 17 Variation of total ozone column with latitude and season based on data first published in 1980.

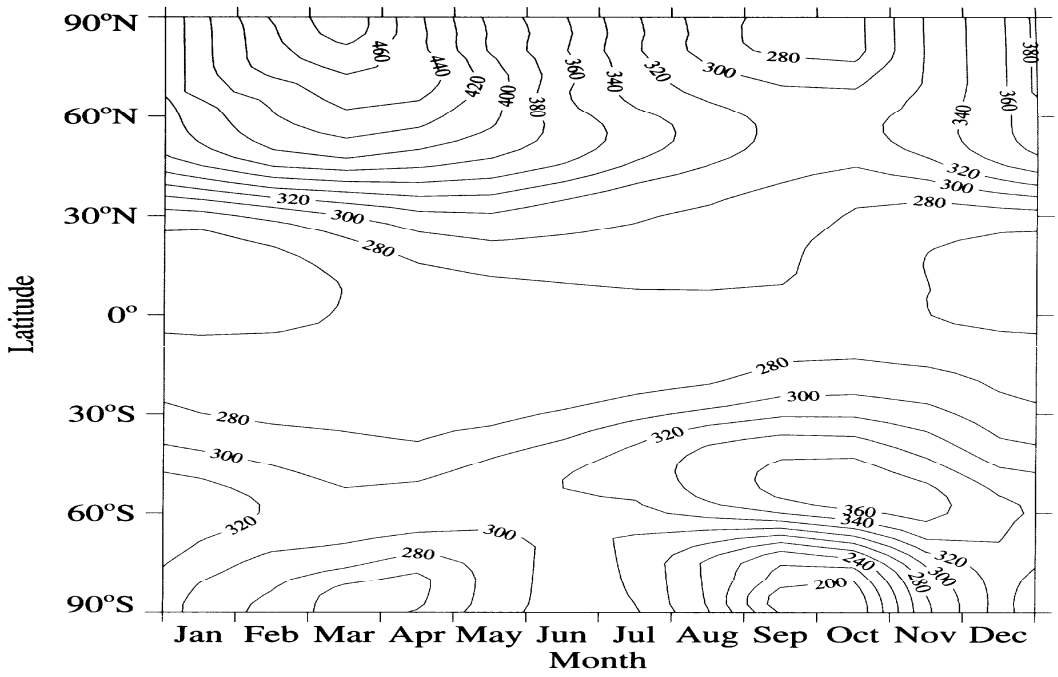

Figure 18 Variation of total ozone column with latitude and season based on data first published in 1995.

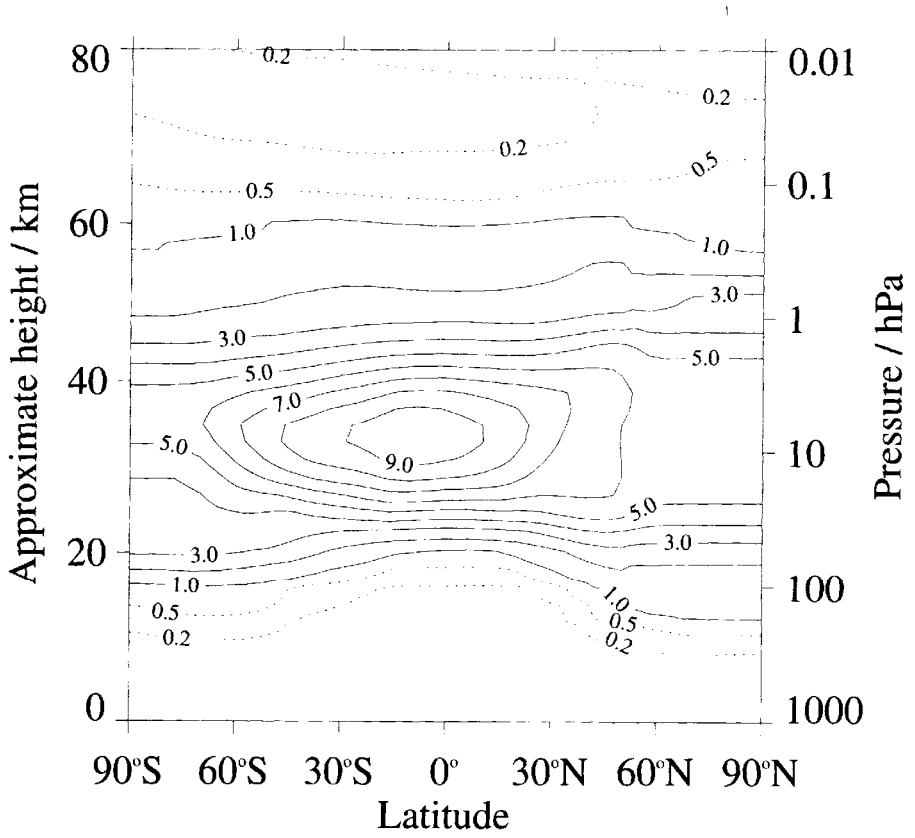

Figure 19 The Zonal mean volume mixing ratio of ozone in ppmv as a function of latitude and height for January based on data from Li and Shine (1995).

1.3.2 Aerosols

Aerosols are suspensions of liquid and solid particles in the atmosphere, excluding clouds and precipitation. The aerosol particle sizes range from 10^{-4} to 10μm, falling under the following broad categories: sulphates, black carbon, organic carbon, dust, and sea salt. Aerosol concentrations and compositions vary significantly with time and location. Thirty years of in situ optical particle counter measured aerosol mixing ratios for particles with radii greater than 0.15 µm and greater than 0:25µm measured at Laramie, Wyoming (41^{o}N) are shown in Figure 20. Figure 20 (a) shows the peak mixing ratio measured and has the times of major volcanic eruptions indicated to show the relationship between these phenomena and the aerosol load of the atmosphere.

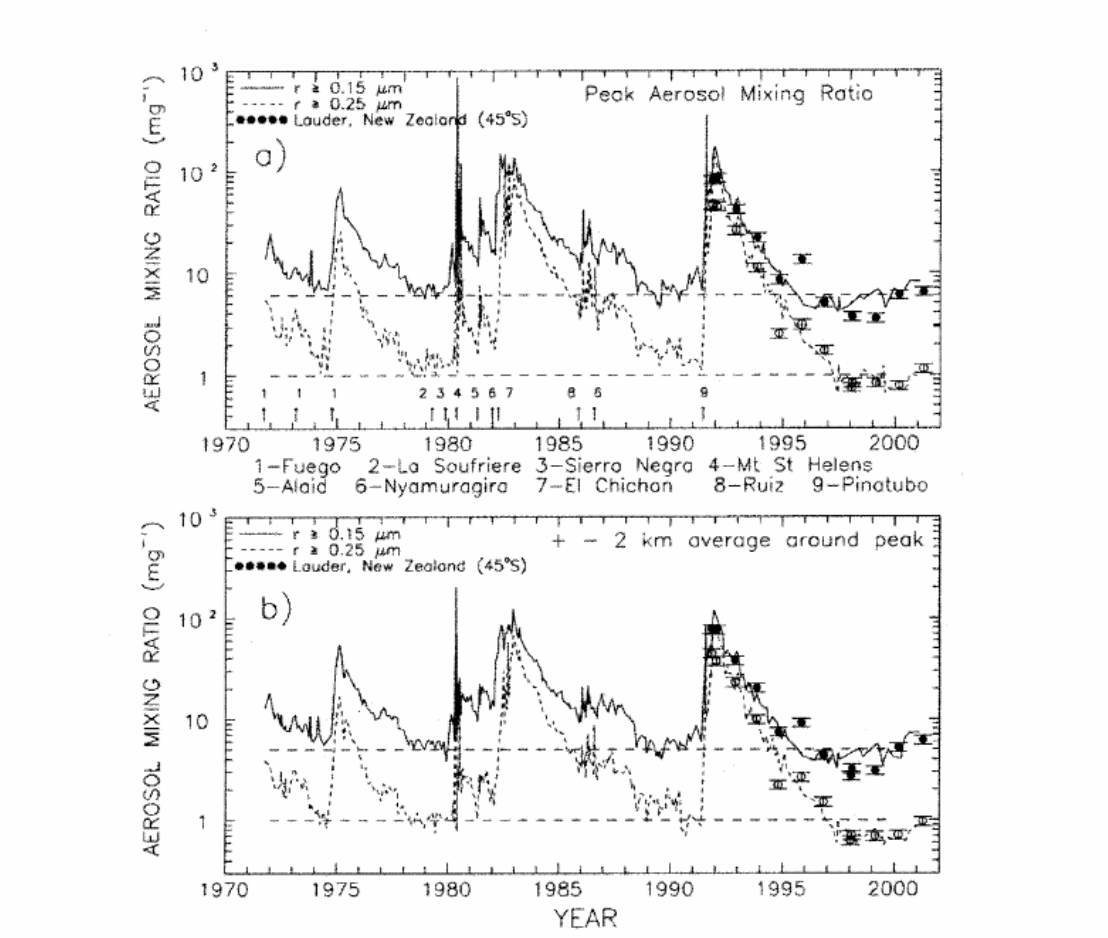

Figure 20 Thirty years of in situ optical particle counter measured aerosol mixing ratios for particles with radii greater than 0.15 µm and greater than 0:25µm measured at Laramie, Wyoming (41^{o}N). Also shown is the 10 year record of measurements made using the same equipment at Lauder, New Zealand (45^{o}S). The error bars on the Lauder data are also indicative of those on the Wyoming data and the horizontal dashed lines provide a reference. Plot (a) shows the peak mixing ratio measured and has the times of major volcanic eruptions indicated. Plot (b) shows an average mixing ratio taken over 4 km centred on the peak mixing ratio. Taken from Thomas (2003).

Aerosols are important directly and indirectly to the radiation budget of the atmosphere and, thus, affect climate since they absorb and scatter radiation. Since the discovery of the stratospheric aerosol layer about 30 years ago, the characteristics and radiative and chemical effects of stratospheric aerosols have been an area of increasing research. Their direct radiative effect is due to their scattering, absorption, and emission properties. Their indirect effect is the result of their ability to act as condensation nuclei in the formation of clouds; thus, aerosols affect the resulting concentration, size distribution and optical properties of cloud particles.

This is particularly important in the polar winter, where the stratosphere cools appreciably (see Figure 4) and pre-existing stratospheric aerosol takes up water and nitric acid. These swollen particles create clouds at temperatures 6°C above the ice point often referred to as polar stratospheric clouds (PSCs). It is the heterogeneous chemistry possible on these clouds which plays a major role in the activation of chlorine over Antarctica. The chemical reaction that produces the catalytic Chlorine is defined by:

$$HCL\,(\text{solid}) + ClONO_2\,(\text{gas}) \rightarrow Cl_2\,(\text{gas}) + HNO_3\,(\text{solid})$$

Equation 23

$$Cl_2 + \text{light} \rightarrow Cl + Cl$$

Equation 24

$$O_3 + Cl \rightarrow ClO + O_2$$

Equation 25

The presence of an ice surface (denoted by solid in Equation 23) allows the reaction of HCl with another chlorine reservoir species, $ClONO_2$, to take place at a much higher rate than it can in the gas phase (Molina, 1991). It should be noted that reactions are catalytic once chlorine is activated (thus once the reaction occurs the chlorine is free to break up more ozone). It should be noted that the nitric acid remains in the solid phase but the Cl_2 is released into the gas phase where it is easily photolysed to liberate chlorine atoms and hence ClO.

ClO reacts with NO_2 to form the reservoir species $ClONO_2$. This reaction could then return active chlorine to its reservoir. However, the removal in the reaction (Equation 23) of nitric acid into the solid phase also effectively reduces the concentration of NO_x (NO and NO_2), a process called denoxification, leading to a longer lifetime for the active chlorine. Indeed, in the Antarctic the very low temperatures usually lead to the growth of large particles being predominantly composed of ice, but incorporating HNO_3. These particles can grow so large that they eventually fall out of the lower stratosphere, reducing irreversibly the concentrations of the nitrogen species (‘denitrification’).

It should be noted that these clouds can only be formed at temperatures less than -80°C and thus only occur in the polar stratosphere. In addition the

hemispheric difference in temperatures means that temperatures less than -80°C occur most frequently over Antarctica and thus ozone loss is most severe there (see Figure 21).

While stratospheric aerosol is vital in the development of PSC's, aerosol also has other important impacts. For example, the large amount of aerosols produced during volcanic eruptions (see Figure 20), can produce artefacts in remotely sensed signals of key atmospheric constituents like H_2O and O_3, rendering their quantitative measurement impossible.

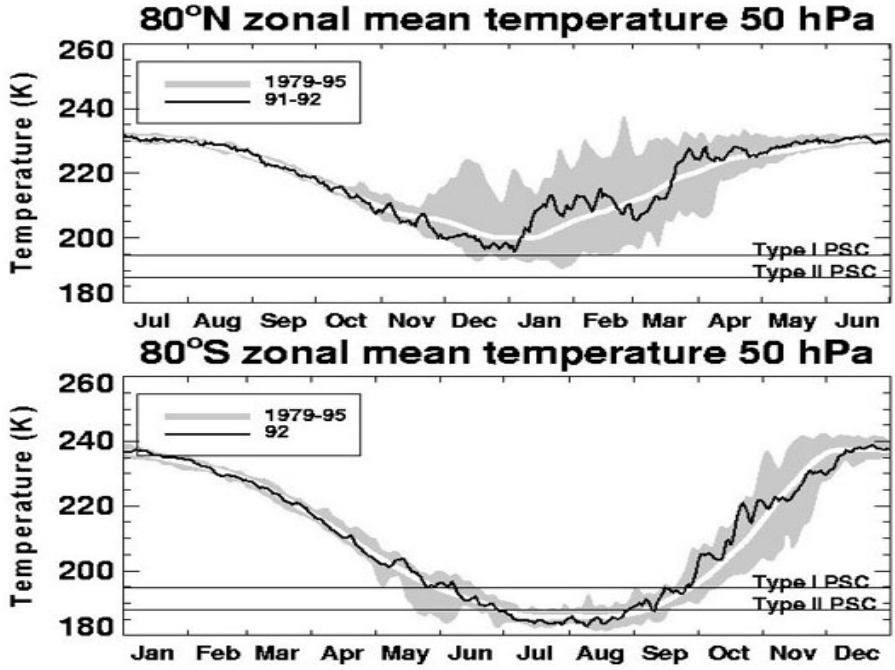

Figure 21 Average temperatures at 50hPa for 80 degrees North and South. Comparison clear indicates that the average temperatures needed to from PSC are much more likely to occur in the Southern hemisphere. Taken from Deshler (2003).

1.3.3 Antarctic Ozone Depletion

It is known that the following catalytic cycle, dominates the destruction of ozone over Antarctica:

$$ClO + CLO + M \rightarrow Cl_2O_2 + M$$

Equation 26

$$Cl_2O_2 + M \rightarrow ClO + ClO + M$$

Equation 27

$$Cl_2O_2 + h\nu \rightarrow Cl + ClO_2$$

Equation 28

$$ClO_2 + M \rightarrow Cl + O_2 + M$$

Equation 29

$$2 \times \left(Cl + O_3 \rightarrow ClO + O_2\right)$$

Equation 30

When reaction (Equation 26) is followed by reaction (Equation 27) the no ozone loss occurs. However, when the dimmer of chlorine monoxide (Cl_2O_2) is stabilized at low temperatures, typical of the Antarctic lower stratosphere, when the rate of reaction (Equation 27) is low, then the dimmer is lost predominantly by photolysis, leading eventually to the formation of two chlorine atoms which destroy ozone in the reaction (Equation 30). The net effect of this catalytic cycle is thus:

$$2O_3 + h\upsilon \rightarrow 3O_2$$

Equation 31

However, this cycle is not sufficiently rapid to explain all of the ozone loss observed and a second catalytic cycle is important, namely:

$$ClO + BrO \rightarrow Cl + Br + O_2$$

Equation 32

$$Cl + O_3 \rightarrow ClO + O_2$$

Equation 33

$$Br + O_3 \rightarrow BrO + O_2$$

Equation 34

The net effect of which can be written as:

$$2O_3 + h\upsilon \rightarrow 3O_2$$

Equation 35

Inclusion of these two catalytic cycles into numerical models is sufficient to explain the observed rate of decline of ozone in the Antarctic spring. In Antarctica, the dimer cycle is responsible for more than 50% of the observed loss.

The chemical processes outlined above which involve both chlorine and bromine, and which lead to the destruction of ozone, are today the accepted explanation of the ozone hole. Measurements of ClO were made during various research flights into the polar vortex. The data shown in Figure 22 clearly shows the inverse relationship between ozone and ClO.

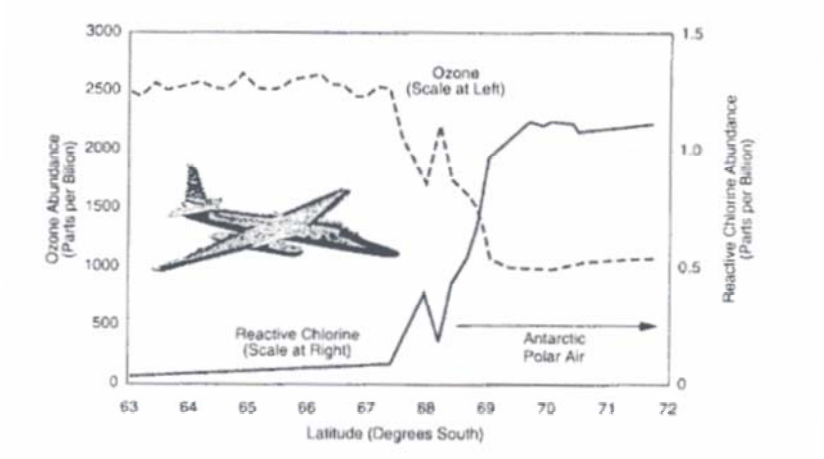

Figure 22 The ozone and ClO content in the stratosphere as measured on a southward flight leg in September 1987. Taken from Labitzke and Van Loon (1999).

In addition to the colder temperatures in the Antarctic the transition from the winter (cyclonic) to the summer (anticyclonic) circulation in the stratosphere takes place about two months earlier in the Arctic than in the Antarctic. The later interseasonal transition in the Antarctic, after the sun has returned to the polar regions (see Figure 25), produces the meteorological conditions that are necessary for the formation of the ozone hole in spring (see Figure 10). A comparison of the temperature and ozone distribution in the two hemispheres in spring (March and September for the Northern and Southern hemisphere, respectively) shows the marked differences in the polar regions. In Figure 23 the temperature is still extremely low over the Antarctic when the Sun returns and enables the ozone-destroying chemistry to take place. In contrast, the mean temperature above the Arctic in March is high and the waves, which are responsible for the higher temperatures, have also transported ozone in to the polar region.

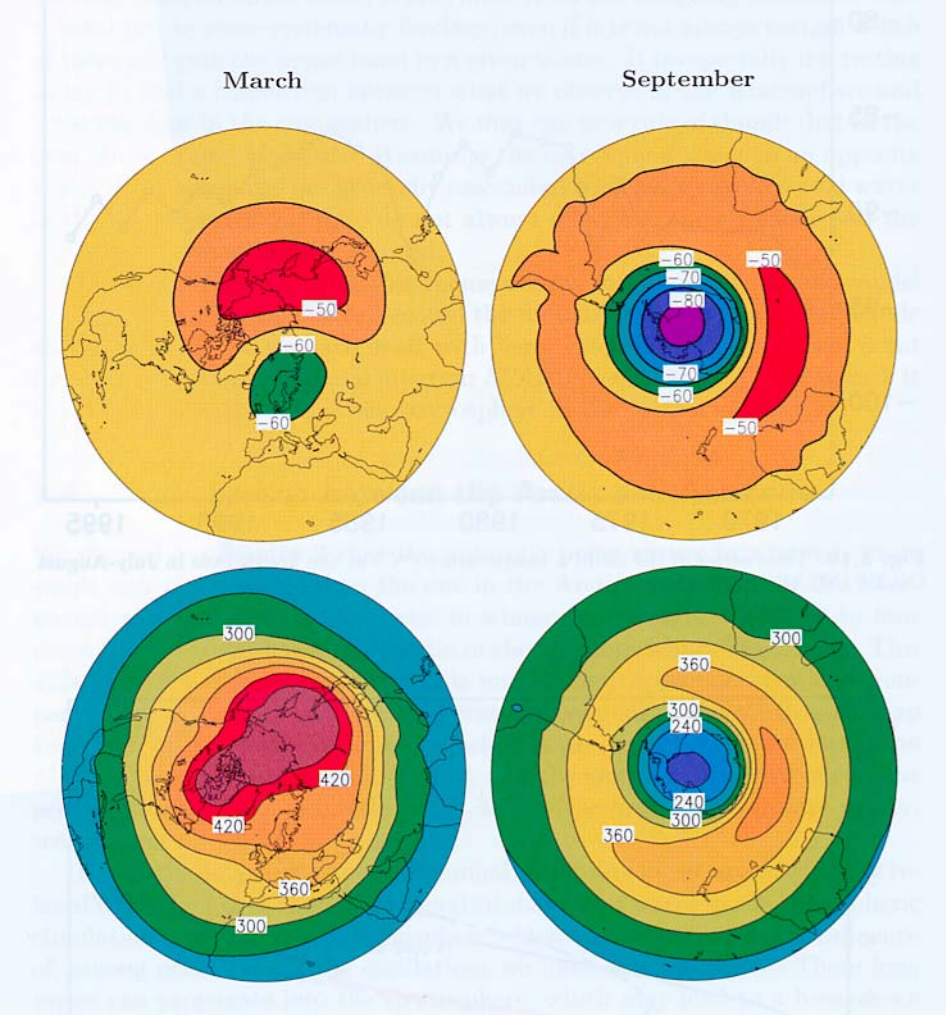

Figure 23 The monthly mean maps of the 30hPa temperature (above) and total ozone in Dobson units (below) on the two hemispheres in March, Northern hemisphere (left), and September, Southern hemisphere (right). Taken from Labitzke and Van Loon (1999).

It should be noted that in the winter and springtime stratosphere, the flow is dominated by the presence of a large vortex. The vortex is encircled by a strong eastward flowing jet, the polar night jet. This jet is disturbed by planetary-scale waves in the Northern hemisphere and is generally undisturbed in the Southern hemisphere. Last year for the first time observed a major stratospheric warming in the Southern hemisphere associated with planetary wave breaking and this produced significant changes to the pattern of ozone depletion observed above Antarctica (see Figure 24).

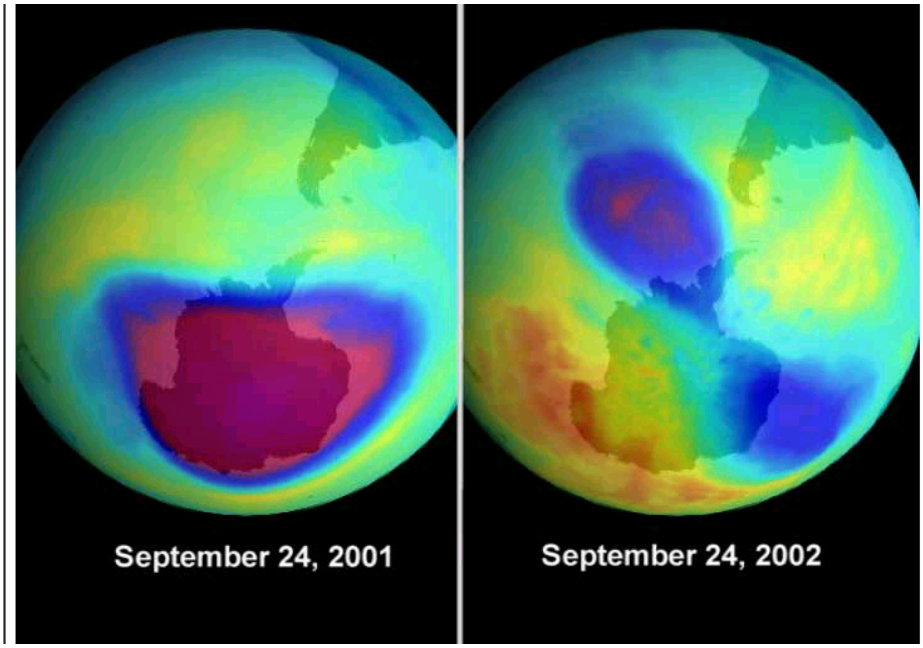

Figure 24 Total column abundance of ozone measured on September 24th 2001 and September 24th 2002. The striking difference observed is associated with the first stratospheric warming observed in the Southern hemisphere which led to the early break up of the Antarctic ozone hole. Taken from Deshler (2003).

2 Radiation in the Atmosphere

2.1: Solar radiation at the Top of the Atmosphere

The distribution of solar radiation as a function of the wavelength is called the solar spectrum, which consists of a continuous emission with some superimposed line structures. The Sun's total radiation output is approximately equivalent to that of a blackbody at 5776 K. The solar radiation in the visible and infrared spectrum fits closely with the blackbody emission at this temperature. However, the ultraviolet (UV) region (< 0.4μm) of solar radiation deviates greatly from the visible and infrared regions in terms of the equivalent blackbody temperature of the Sun. In the interval 0.1-0.4μm, the equivalent blackbody temperature of the sun is generally less than 5776 K with a minimum of about 4500 K at about 0.16μm. The deviations seen in the solar spectrum are a result of emission from the non-isothermal solar atmosphere.

The solar insolation is the actual amount of solar radiation incident upon a horizontal surface over a specified period of time for a given location. It depends strongly on the solar zenith angle χ and also on the ratio of the actual distance to the mean distance of the Earth from the Sun. The solar irradiance at the top of the atmosphere may be expressed by –

$$F = S\left(\frac{d}{d_m}\right)^2 \cos\chi$$

Equation 36

where S is the solar constant. The solar zenith angle depends on latitude, day of year, and time of day, and is given by-

$$\cos\chi = \sin\lambda\sin\delta + \cos\lambda\cos\delta\cos h$$

Equation 37

where λ is the latitude, δ is the solar declination, and h is the hour angle. The hour angle is zero at solar noon and increases by 15 degrees for every hour. For reference, the solar zenith angle is at 90 degrees at sunset and sunrise. The solar insolation for a specified period of time between t_1 and t_2 is-

$$Q = \int_{t_1}^{t_2} F(t)dt$$

Equation 38

The values of the daily insolation at the top of the atmosphere are presented in Figure 25.

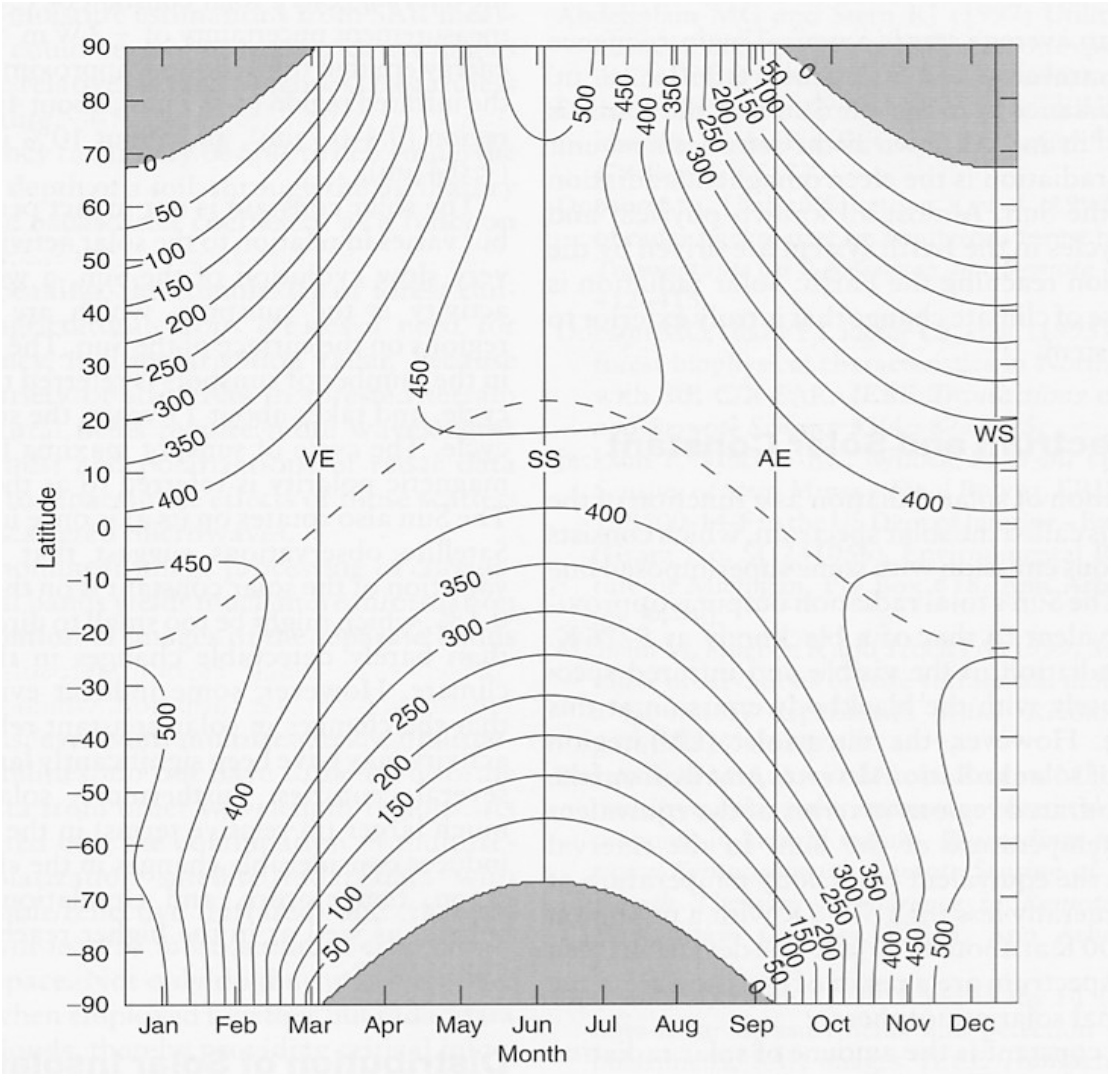

Figure 25 Daily solar insolation at the top of the atmosphere as a function of latitude and day of year. The shaded areas denote zero insolation. Taken from Liou (2002).

2.2: Atmospheric Absorption

As solar photons penetrate into the Earth's atmosphere, they undergo collisions with atmospheric molecules and are progressively absorbed and scattered. The probability of absorption by a molecule depends on the nature of the molecule and the wavelength of the incoming radiation. An effective absorption cross section, $\sigma(\lambda)$, can be defined for each species. It should be noted that this quantity is independent of the concentration of the atmospheric species under consideration, but may depend on temperature and pressure.

A simplified picture of a one-stage process which conveys the general nature of the interaction between a photon and an air molecule is shown in Figure 26. A molecule can be assumed to exist in one of two quantized states, M and M* (M represents the state before interaction with the photon and M* represents an excited molecule). We will assume that in the initial state (M)

the molecule is in thermodynamic equilibrium (i.e., the populations of energy levels are determined by Boltzmann's law). The first stage in the interaction is for the photon to disappear and the molecule to be raised to the excited state, M* (which will not be in Boltzmann equilibrium).

In the second stage of the interaction, several things may happen to the excited molecule, M*, including decomposition and chemical reactions. Leaving chemistry aside, there are two generic possibilities. First, collisions with other molecules may be so rapid that the excess energy in M* is thermalized and the molecule returns to an equilibrium state, but with slightly more internal energy in the thermal reservoir. This process is absorption. The alternative process is for the excited molecule to decay spontaneously, as all excited molecules will if left undisturbed for long enough. This again returns the molecule to an equilibrium state, but now a photon is emitted that is the same as the incident photon, except with different direction and polarization. The internal energy of the thermal reservoir is unchanged. This is an elastic scattering process. The branching between absorption and scattering depends upon the ratio of the times for collisional thermalisation and spontaneous decay. For the wavelength bands discussed, this ratio is very small in the troposphere and stratosphere, and absorption is the dominant mode of interaction between molecules and radiation in the lower atmosphere.

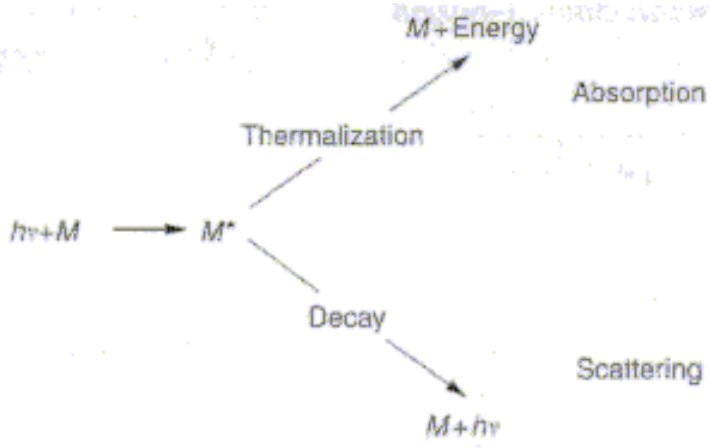

Figure 26 A molecule-photon interaction schematic. The branching depends on the ratio of the collisional thermalisation rate to the spontaneous decay rate. Taken from Goody and Hu (2002).

2.2.1: Absorption by atmospheric gases

The absorption of infrared radiation by the six most significant gaseous absorbers is summarised in Figure 27. This figure shows the transmittance for a vertical beam passing through the whole atmosphere, as a function of wavelength. The gases shown are all minor constituents (that is, they are present in much smaller quantities than molecular nitrogen and oxygen and all but ozone (0_3) are concentrated mainly in the troposphere. The major constituents, N_2 and O_2, are not strongly radiatively active in the infrared

because they possess no permanent electric dipole moment, on account of their symmetry. Since, on the scale of the diagram, much of the fine structure associated with individual spectral lines is not shown, the diagram can be regarded as plots of the band transmittance between the top of the atmosphere and the ground, corresponding to band widths associated with wavelength differences of 0.1 μm.

The bottom panel of Figure 27 shows the total long-wave absorptance due to all gases. There is a broad region from 8 to 12μm, called the *atmospheric window,* within which absorption is weak, except for a band near 9.6μm associated with 0_3. Water vapour absorbs strongly over a wide band of wavelengths near 6.3 μm and over a narrower band near 2.7 μm. At longer wavelengths, especially beyond 16 μm, rotational transitions of water become important, leading to strong absorption. Carbon dioxide is also a strong absorber in a broad band near 15 μm, associated with a vibrational mode, and in a narrower band near 4.3 μm.

Ozone absorbs strongly near 9.6 μm in the atmospheric window. Since the other gases do not absorb significantly in this spectral region, ozone which is mainly concentrated in the stratosphere (see Figure 14) can therefore exchange radiation with the lower atmosphere.

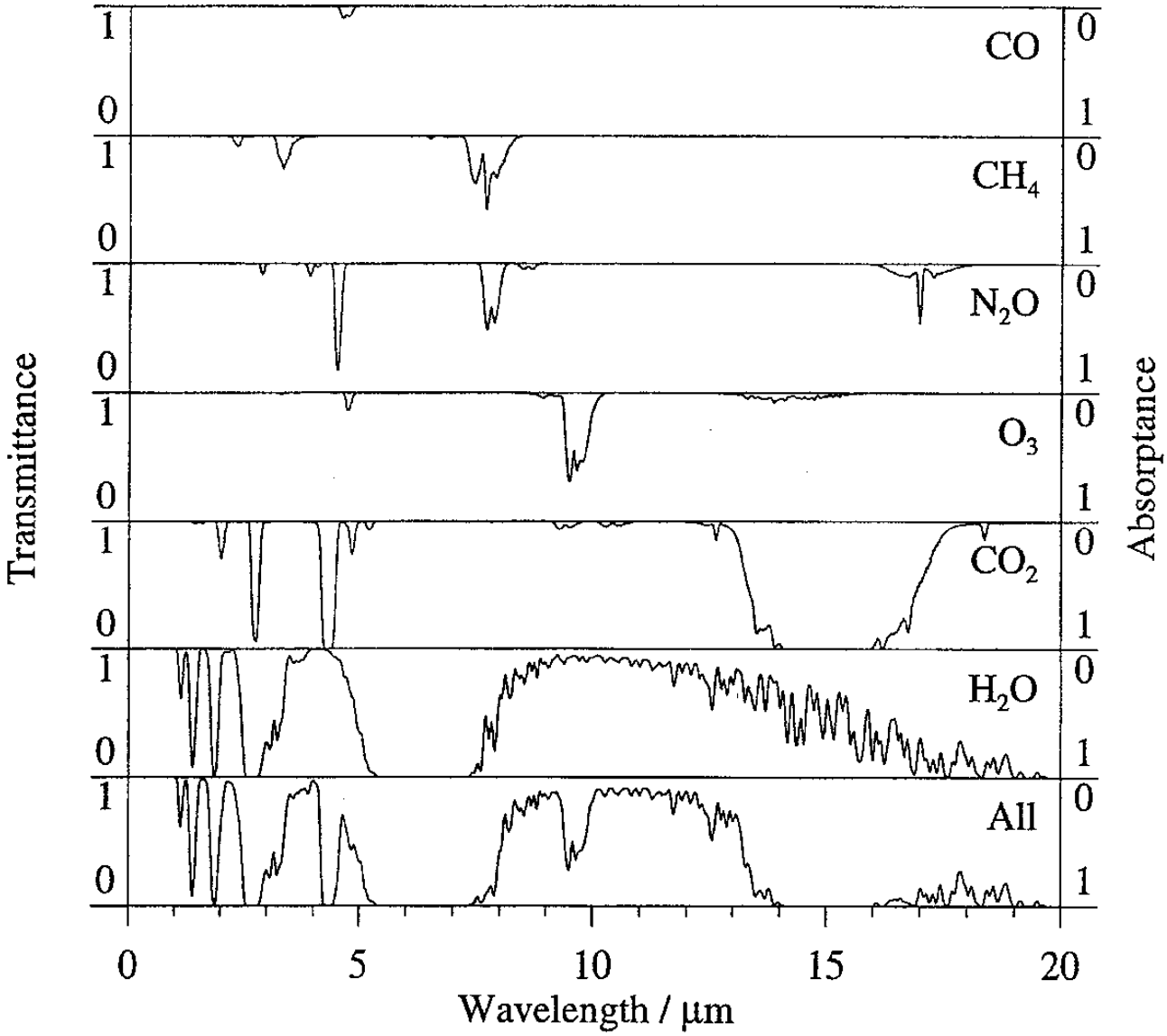

Figure 27 Infrared absorption spectra for six strongly absorbing gases and for the six gases combined. Taken from Andrews (2000).

It should be noted that Figure 27 gives information on the absorption of infrared radiation over the total depth of the atmosphere, but not about the way in which the absorption varies with altitude.

In the ultraviolet, the main absorbers are molecular oxygen and ozone. Absorption at these wavelengths is often depicted in terms of the *absorption cross-section*, which is equal to the absorption coefficient times the molecular mass. Unlike in the infrared, we must take account of electronic transitions, as well as vibrational and rotational transitions, when considering absorption at discrete wavelengths; moreover, photodissociation and photoionization lead to important *continuum absorption (i.e*. absorption over a continuous range of wavelengths).

The absorption cross-section for oxygen molecules (Figure 28) has large values due to ionization at wavelengths below 100 nm; in the range 100-130 nm there are irregular bands of unknown origin. The Schumann-Runge continuum, in the range 130-175 nm, is due to dissociation where one oxygen atom remains in the ground state and the other goes into an excited state. The Schumann-Runge bands, in the range 175-200 nm, are associated with an electronic transition and superimposed vibrational transitions. The Herzberg continuum is found in the range 200-242 nm. At 242 nm dissociation into two ground-state oxygen atoms occurs; although this is an insignificant absorption feature, it is very important in the formation of ozone (see Equation 17). A further electronic transition gives rise to the weak Herzberg bands in the range 242-260nm.

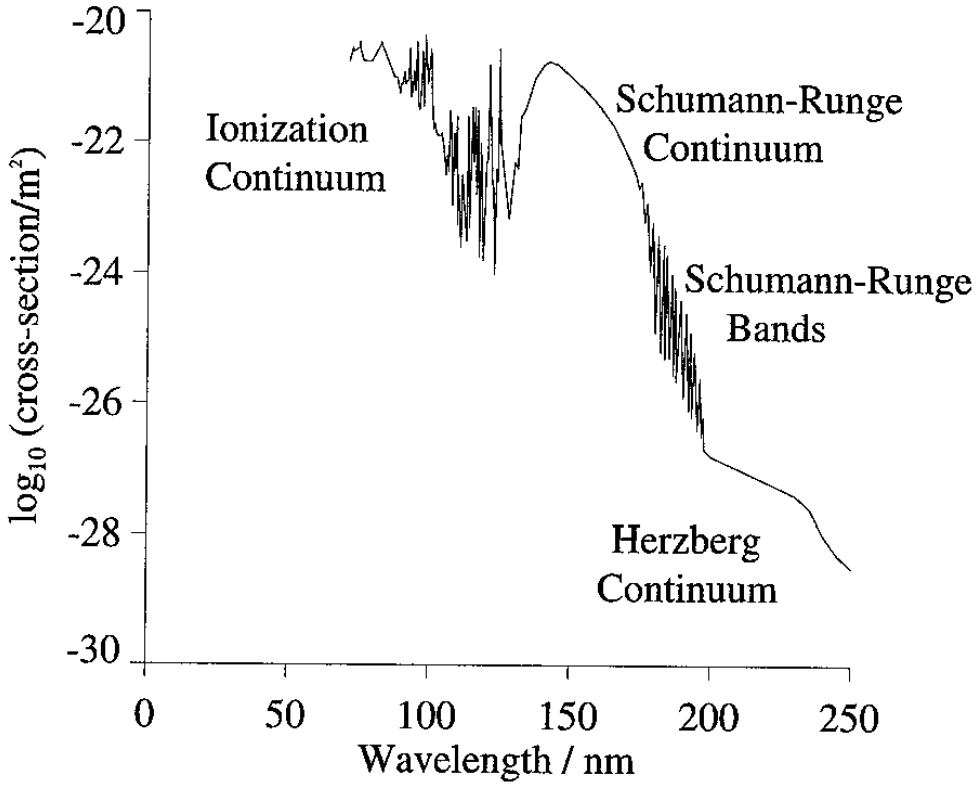

Figure 28 Absorption cross section as a function of wavelength for oxygen molecules. Taken from Brasseur and Solomon (1986).

However, the principal absorber of ultraviolet light in the Earth's atmosphere is ozone. Absorption in the *UV* by ozone is strong enough that a few parts per million of ozone remove all of the sunlight at wavelengths shorter than about 300 nm before they can reach the ground. The ozone absorption cross-section (Figure 29) exhibits two continua in the ultraviolet and one in the visible and near infrared, all due to photodissociation: the Hartley band, in the range 200-310nm, the Huggins bands, in the range 310-350nm and, in the

visible and near infrared, the Chappuis bands, in the range 450- 750 nm. Although the absorption cross-section for the Chappuis bands is much smaller than those for the Hartley and Huggins bands, the Chappuis bands are important since they occur near the peak of the solar spectrum and absorb in the troposphere and lower stratosphere. Shorter-wavelength (more energetic) radiation is almost absent at these levels, since it is mostly absorbed at higher altitudes.

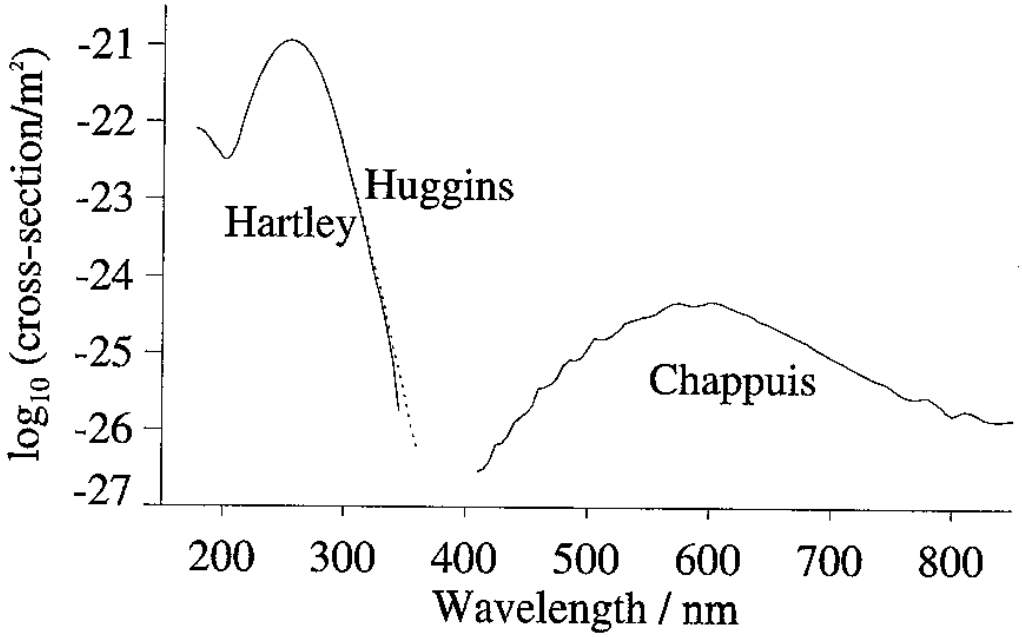

Figure 29 Absorption cross section as a function of wavelength for ozone. Taken from Andrews (2000).

The relative contributions of absorption by oxygen molecules and ozone for different wavelengths are shown in Figure 30. The cross sections of each of these gases must be multiplied by their column abundances to compare their absorbing capacities. The effect of molecular oxygen dominates for wavelengths between 164 and 190nm, while ozone absorption dominates near 250nm.

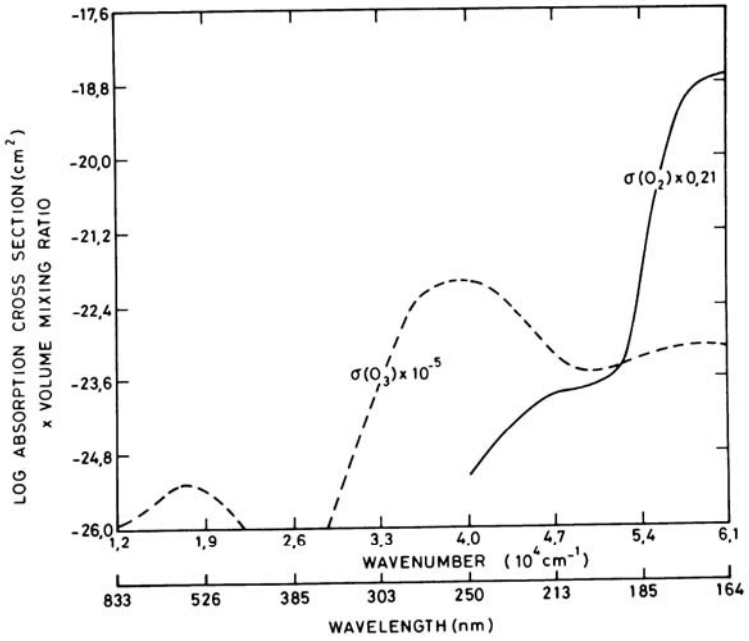

Figure 30 Relative contributions of ozone and molecular oxygen to absorption by the atmosphere.

2.3: Emission

Thermal emission is the complementary process to absorption: internal energy is transferred spontaneously from a molecules thermal reservoir to photons in the radiation field. In Boltzmann equilibrium, states above the ground state have finite populations and there is a probability that decay by spontaneous emission will take place, emitting a photon. As a consequence, there is a steady state of photons associated with the interior of a constant-temperature enclosure or cavity. Cavity or blackbody radiation is, according to Kirchhoff's law, isotropic and the same as that emitted by a perfectly black body at the same temperature. Most meteorological and climatological studies assume that air molecules are in a state of local thermodynamic equilibrium, whereby Boltzmann equilibrium is obeyed on a local basis. Consequently, equilibrium radiation laws can be employed.

Emission differs from absorption by being direction dependent. The transmission is the same in either direction along a path, but the radiation reaching the opposite ends of a path from gases along the path is generally different. The difference comes from the nonsymmetric thermal structure along the path.

Transmission refers to what happens to radiation as it propagates between two points. Emission occurs at a single point and normally has a random directional distribution that emits equal energy in all directions. Emission is a physical process whereas transmission is a property of the atmosphere that exists whether or not a signal is propagating along the path.

Emission can be difficult to analyse and to calculate in scattering atmospheres. The difficulty occurs because, in principle, each scattering centre can redirect the radiation, and the redirected radiation is again redirected by the next scattering centre. This redirection must be done even if the emission in a single direction is desired because some of the redirected radiation may finally be redirected again into the direction of interest. Scattering occurs at discrete points along the path, so until the number of scatterings is very large, the scattering cannot easily be treated as a continuous function. In the treatment of atmospheric emission, scattering will be neglected. Scattering in the Earth's atmosphere is not too important at the wavelengths where emission occurs. Microwave emission and water drops are the exception to this generalization.

2.3.1: Atmospheric Emission

The generation of electromagnetic waves occurs as a general result of accelerating electric charges. In general, any object is composed of vast numbers of molecules. These molecules oscillate over a continuous range of frequencies and therefore emit radiation at all frequencies. However, this radiation is not emitted equally at all frequencies rather it is distributed according to the emission spectrum which depends strongly on the temperature of the object.

To describe the emission spectrum and its relationship to the temperature of the body the concept of a blackbody is necessary. A blackbody can be defined as a body whose surface absorbs all radiation incident upon it. From this definition it follows that any two blackbodies at the same temperature emit precisely the same radiation and that a black body emits more radiation than any other type of object at the same temperature.

If we consider an isolated cavity with walls opaque to all radiation. The cavity walls constantly emit, absorb, and reflect radiation until a steady state is reached. This equilibrium radiation fills the cavity uniformly and is just the same as the radiation emitted by a hypothetical blackbody at the same temperature as the cavity. It can then be shown that the ability of a body to radiate is closely related to its ability to absorb radiation. The mathematical formulation of this statement is known as Kirchoff's Law, which may be written as -

$$E_\lambda = a_\lambda B_\lambda(T)$$

Equation 39

where E_λ. is the emitted radiation and B_λ. *(T)* is the radiation (expressed in intensity units) of the hypothetical blackbody. The proportionality constant is the absorption coefficient which varies between 0 and 1. If a_λ. = 0, then Equation 39 states that a body neither absorbs radiation at the given wavelength nor emits radiation at the same wavelength. For a_λ. = 1, the

emitted radiation is the blackbody radiation. The wavelength dependence of the absorption coefficient varies dramatically according to the nature of the matter emitting the radiation and the portion of the electromagnetic spectrum under consideration.

It is through the statement of Kirchhoff's Law that the whole point of blackbody radiation becomes apparent. All blackbodies at some temperature behave identically and the radiation emitted by such bodies at a given λ depends only on the temperature of the body. Thus the emission of radiation at some chosen wavelength is solely determined by the characteristics of the emitting matter (through a_{λ}.) and temperature (through B_{λ}.).

Planck's blackbody function, which can be written as:

$$B_{\lambda}(T) = \frac{2hc^2}{\lambda^5\left(e^{hc/k\lambda T} - 1\right)}$$

Equation 40

describes blackbody radiation where *k* is Boltzmann's constant and *T* is the absolute temperature of the cavity walls. Planck's function for different temperatures is given in Figure 31. The examples demonstrate an obvious relationship the hotter the object the shorter the wavelength of the maximum intensity. This observation can be mathematically stated by an expression known as *Wien's displacement law* which establishes a connection between the wavelength of maximum emission (λ_{max}) and the temperature of the radiator. This law is simply derived from the derivative of Equation 40 with respect to wavelength from which it follows that:

$$T\lambda_{\max} = 5.098 \times 10^{-3}$$

Equation 41

Wien's displacement law is indicated on Figure 31 as the solid diagonal line joining the maxima of the three Planck functions. For example, at 6000 K, the maximum emission is in the blue region of the visible spectrum and substantial amounts of radiation are emitted across the entire visible spectrum. This distribution is very similar to the wavelength distribution of radiation emitted by the sun.

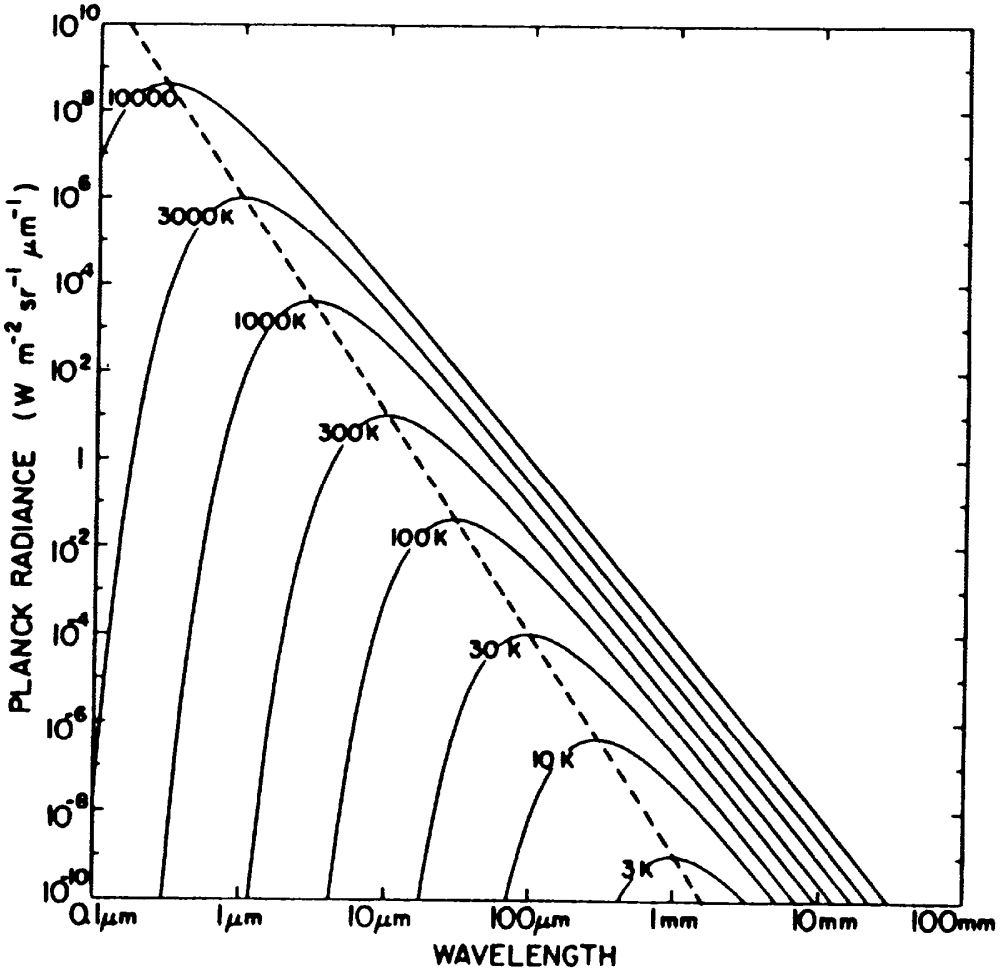

Figure 31 Planck's blackbody curves at a number of temperatures. The dotted diagonal line depicts Wien's displacement law.

Another obvious characteristic of blackbody radiation is the hotter the object, the greater the total amount of radiation emitted from a given surface area. This is just a statement of *Stefan-Boltzmann's law,* which derives from integration of B_λ over the entire wavelength domain:

$$B(T) = \int_0^\infty B_\lambda(T) d\lambda = \frac{\sigma}{\pi} T^4$$

Equation 42

where σ is the *Stefan-Boltzmann constant.* The radiation emitted by a 6000 K blackbody, for instance, is 160000 times that emitted by a 300 K blackbody.

Figure 32 shows the spectral irradiance (spectral irradiance will be defined later) of solar radiation at the top of the atmosphere (TOA) and at sea level. It is seen that the TOA spectral irradiance is fairly close to the black-body

spectral irradiance at most wavelengths. However, absorption and scattering in the atmosphere cause significant deviations between the TOA spectral irradiance and the sea-level spectral irradiance, with especially large deviations (apparent in the sharp dips in the sea-level curve) at certain wavelengths, which are identified with particular absorbing gases, notably ozone (0_3) in the ultra-violet and visible and carbon dioxide *(CO_2)* and water vapour (H_2O) in the infrared.

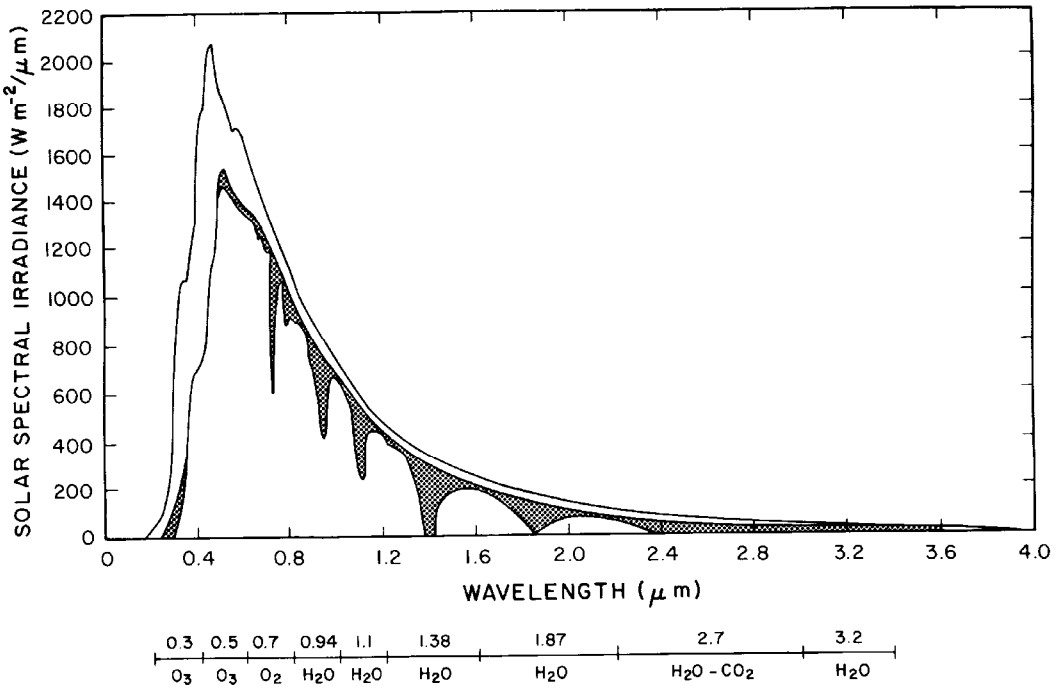

Figure 32 Spectral irradiance distribution curves related to the sun. The upper line shows the observed solar irradiance at the top of the atmosphere. The lower line shows the solar spectral irradiance observed at sea-level. The shaded area represents absorption due to various gases in a clear atmosphere. The upper envelope of the shaded area denotes the reduction of solar spectral irradiance due to scattering.

Fundamental to an understanding of atmospheric emission is the relation of emission and absorption. Emission is proportional to absorption with the proportionality factor being Planck's black body function. Consider a layer of gas in front of a blackbody at some temperature. The gas absorbs a fraction of the radiation and is heated. The amount of radiation absorbed is proportional to the absorptance of the gas. The absorption increases the gas temperature until it becomes as warm as the black body. For the gas to become warmer than the black body would violate thermodynamic principles. For the gas to always remain cooler than the black body would imply that the gas and the black body could be in radiative equilibrium at different temperatures. Instead, the gas and black body must come to the same temperature. This means the gas emits the same amount of energy as it absorbs. But the amount of energy absorbed is directly proportional to the absorptance. Thus, the emission by the gas must also be directly proportional to the absorptance.

2.4: Radiative transfer

The preceding discussions have described different aspects of how electromagnetic radiation interacts with matter. Most practical applications, however, rely on measurements of the accumulated effect of many interactions not a single interaction. The actual path along which these interactions occur might be just a few meters in some cases or hundreds of kilometres for instruments on meteorological satellites. Some way of describing the accumulated effects of all processes as radiation is transferred from one volume of atmosphere to another along these paths is needed. The theory of radiative transfer provides such a description.

2.4.1: Definitions

Several different, but related, quantities are used in the description and measurement of radiation. The most important definitions are described below.

2.4.1.1 Spectral radiance

The *spectral radiance* (or monochromatic radiance) $L_\nu(r,s)$ is the power per unit area, per unit solid angle, per unit frequency interval in the neighbourhood of the frequency ν, at a point *r,* in the direction of the unit vector *s.* The spectral radiance can be visualised in terms of the photons emerging from a small area ΔA with unit normal *S,* centred at a point *r (see* Figure 33*)*. Consider those photons whose momentum vectors lie within a cone of small solid angle $\Delta\Omega$ centred on the direction *s* and whose frequencies lie between ν and $\nu + d\nu$. Then $L_\nu \Delta A \Delta\Omega \Delta\nu$ is the energy transferred by these photons, per unit time, from 'below' the area ΔA to 'above'.

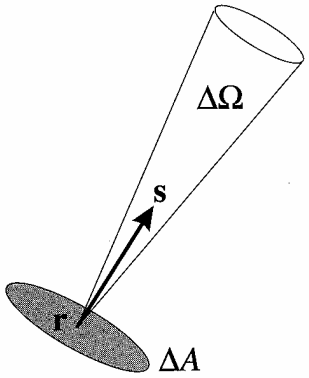

Figure 33 Spectral radiance geometry. Taken from Andrews (2000).

We have already encountered a special case of the spectral radiance, for isotropic black-body radiation in an isothermal cavity, when $L_\nu = B_\nu(T)$, the Planck function; this depends only on the temperature of the cavity and is independent of position and direction.

2.4.1.2 Radiance

The *radiance L(r, s)* is the power per unit area, per unit solid angle at a point *r* in the direction of the unit vector *s;* in other words it is the integral of L_ν over frequency:

$$L(r,s)=\int_0^\infty L_\nu(r,s)d\nu$$

Equation 43

2.4.1.3 Spectral Irradiance.

The *spectral irradiance* (or monochromatic irradiance) $F_\nu(r,n)$ is the power per unit area, per unit frequency interval in the neighbourhood of the frequency ν, at a point *r* through a surface of normal *n*.. It is obtained from the spectral radiance by integration over a hemisphere on one side of the surface:

$$F_\nu(r,n)=\int_{2\pi} L_\nu(r,s)n\cdot sd\Omega(s)$$

Equation 44

where $d\Omega(s)$ is the element of solid angle in the direction *s (*see Figure 34*)*. We therefore integrate over *all* photons in the frequency interval that emerge into the region above the surface. As with the Planck function, the spectral radiance and spectral irradiance can alternatively be expressed per unit *wavelength* interval.

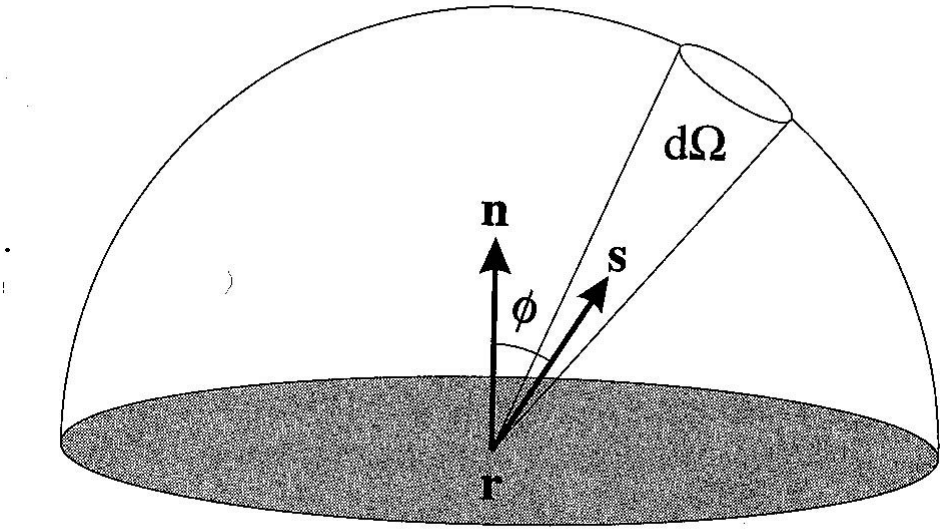

Figure 34 Spectral irradiance geometry. Taken from Andrews (2000).

2.4.1.4 Irradiance

The *irradiance* (or flux density) *F(r, n)* is the power per unit area at a point *r* through a surface of normal *n,* i.e., the integral of F_ν over frequency, and also the integral of the radiance *L* over a hemisphere:

$$F(r,n)=\int_0^\infty F_\nu(r,n)d\nu=\int_{2\pi} L(r,s)$$

Equation 45

It must be borne in mind that the irradiance has a specific direction associated with it; for example, if the surface in question is horizontal and the normal *n* points upwards, then the irradiance under consideration (denoted by F↑) is associated with upward-moving photons. Conversely, the irradiance *F*↓ = *F(r, -n)* is associated with downward-moving photons. If we require the *net* upward power per unit area, F_z, then we must take the difference:

$$F_z = F\uparrow - F\downarrow$$

Equation 46

There is a simple relationship between the radiance and irradiance from an isothermal plane surface, at temperature *T,* that emits black-body radiation. Since the black-body radiation is isotropic, $L_\nu = B_\nu(T)$ is independent *of s* and

r, so the hemispheric integral is straightforward. Equation 44 for the spectral irradiance becomes -

$$F_\nu(r,n) = \int_{2\pi} L_\nu(r,s) n \cdot s d\Omega(s) = 2\pi B_\nu \int_0^{\pi/2} \cos\phi \sin\phi d\phi = \pi B_\nu(T)$$

Equation 47

where ϕ is the angle between *s* and the normal *n* (see Figure 34) so that $d\Omega = 2\pi \sin\phi d\phi$, since there is axisymmetry around the normal. Integrating over all ν we obtain the Stefan-Boltzmann law for the irradiance (see Equation 42).

2.4.2: The Equation of Transfer and Schwartzchild's Equation

Consider a beam of radiation of unit cross-sectional area, moving in a small range of solid angles $\Delta\Omega$ about the direction *s (*see Figure 35). If the photons experience absorption or scattering in a small distance *ds* along the beam, due to the presence of a radiatively active gas, then the spectral radiance L_ν will be reduced.

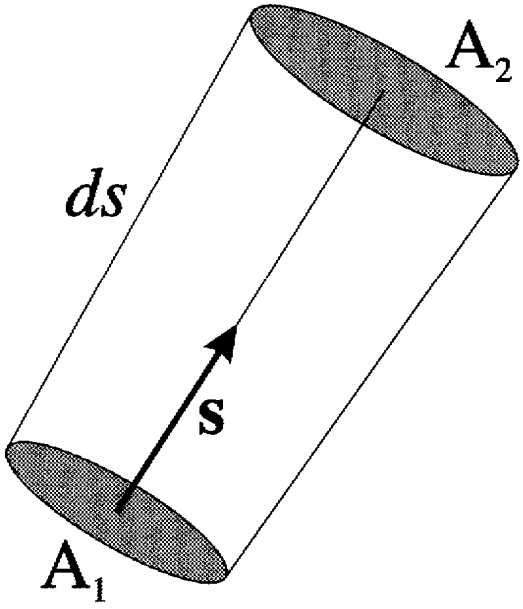

Figure 35 A beam of radiation travelling a distance ds from surface A_1 to a surface A_2. It should be noted that the area of A_2 is slightly greater than unity, owing to the divergence of photons into a solid angle $\Delta\Omega$ from each point of A_1. Taken from Andrews (2000).

The physics of the process is complex; however, it may be summed up by the *Beer-Lambert's Law,* which states that the fractional decrease of the spectral radiance is proportional to the mass of absorbing or scattering material encountered by the beam in a distance *ds.* Since the beam has unit cross-sectional area, this mass is ρ_a *ds,* where ρ_a is the density of the radiatively active gas, so -

$$dL_\nu = -k_\nu(s)\rho_a(s)L_\nu(s)ds$$

Equation 48

The quantity k_ν is called the *mass extinction coefficient;* it is the sum of an absorption coefficient a_ν and a scattering coefficient s_ν, defined in an obvious manner in terms of the contributions to dL_ν from absorption and scattering, respectively -

$$k_\nu = a_\nu + s_\nu$$

Equation 49

The extinction coefficient (k_ν) generally depends on temperature and pressure, and can be regarded as calculable from detailed quantum mechanics or as an empirical quantity, to be derived from measurements. (Note that ρ_a is *not* generally the same as the total gas density ρ). If the gas is also emitting photons of frequency, ν, an extra term must be added to the right-hand side of Equation 48, to represent the additional power per unit area introduced into the beam. This term will also be proportional to the mass ρ_a ds, so it is convenient to write it as $k_\nu \rho_a J_\nu ds$, where $J_\nu(s)$ is called the source function. The source function defines the contribution to the beam by scattering and emission. Including both extinction and emission we obtain the *radiative-transfer equation* (also called the general equation of transfer).

$$\frac{dL_\nu}{ds} = -k_\nu(s)\rho_a(L_\nu - J_\nu)$$

Equation 50

The terms making up the general equation of transfer are indicated schematically in Figure 36. It should be noted that the fundamental equation describing the propagation of electromagnetic radiation is the equation of transfer.

If k_ν, ρ_a and $J_\nu(s)$ are given as functions of distance s, a formal solution of the radiative-transfer equation can be obtained as follows. First introduce the *optical path* χ_ν, defined by:

$$\chi_\nu(s) = \int_{s_o}^{s} k_\nu(s')\rho_a(s')ds'$$

Equation 51

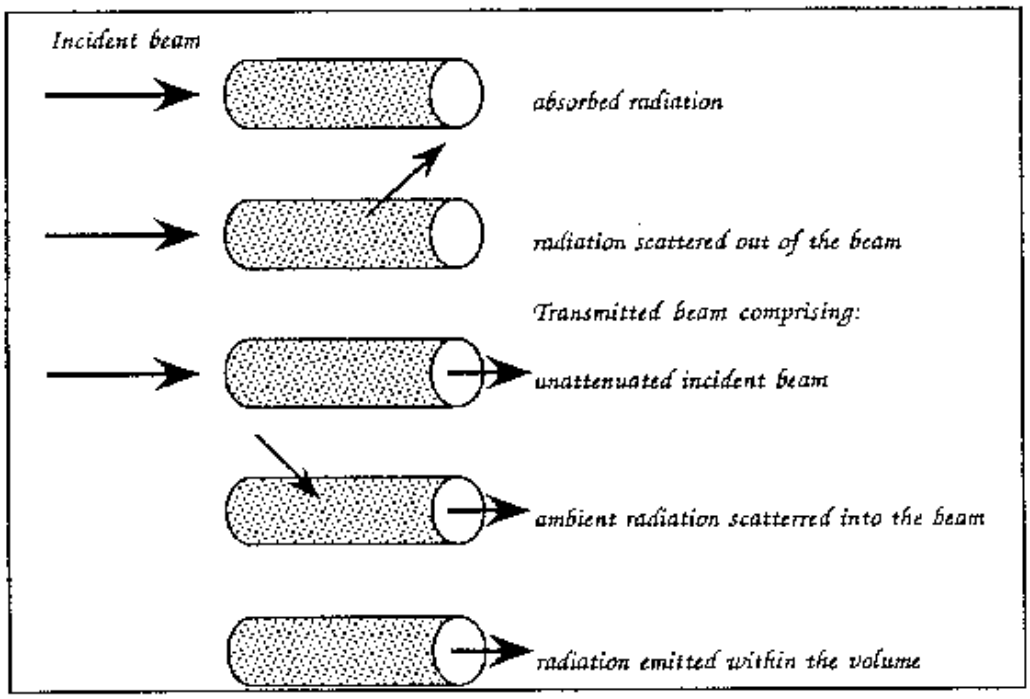

Figure 36 Elemental radiative transfer processes. Taken from Andrews (2000).

where s_o is the start of the path; then Equation 50 can be written as

$$\frac{dL_\nu}{d\chi_\nu} + L_\nu = J_\nu$$

Equation 52

We can then integrate Equation 52 to obtain

$$L_\nu e^{\chi_\nu} = \int J_\nu e^{\chi_\nu'} d\chi_\nu' + constant$$

Equation 53

If the spectral radiance equals $L_{\nu 0}$ at the point s_0 then -

$$L_\nu(s) = \int_0^{\chi_\nu} J_\nu e^{(\chi_\nu - \chi')} d\chi' + L_{\nu 0} e^{-\chi_\nu}$$

Equation 54

Note that, in the absence of emission, the spectral radiance falls exponentially, decreasing by a factor of *e* over a distance corresponding to unit optical path. A region is said to be *optically thick* at a frequency if the total optical path χ_ν through the region is greater than 1 and *optically thin* if the total optical path is less than 1. A photon is likely to be absorbed or scattered within an optically thick region, but is likely to traverse an optically thin region without absorption or scattering.

As a simple example, suppose that the mass extinction coefficient k_ν and the density ρ_a of the radiatively active gas are both constant and take $s_0 = 0$. Then from Equation 51 the optical path is proportional to the distance *s,* $\chi_\nu = k_\nu \rho_a s$, so -

$$L_\nu(s) = k_\nu \rho_a \int_0^s J_\nu(s') e^{-k_\nu \rho_a (s-s')} ds' + L_{\nu 0} e^{-k_\nu \rho_a s}$$

Equation 55

The radiance $L_\nu(s)$ reaching *s* thus has a simple interpretation, as follows. The second term on the right-hand side of Equation 55 represents the radiance at the starting point *s* = 0, attenuated by an exponential factor due to extinction over the distance *s,* while the integral represents the sum of contributions emitted from elements *ds'* at different distances *s'* along the path, each attenuated by the factor $e^{-k_\nu \rho_a (s-s')}$ due to extinction over the remaining distance *s-s'.* A schematic of this process is shown in Figure 37.

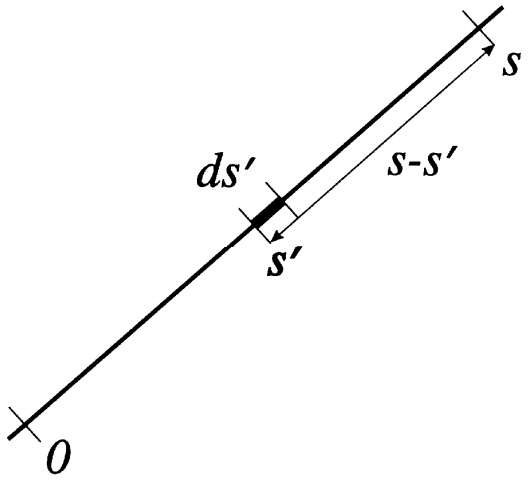

Figure 37 Schematic of the simple example process. Taken from Andrews (2000).

Under local thermodynamic equilibrium (LTE) conditions, in the absence of scattering, the source function equals the blackbody spectral radiance, or the Planck function (Equation 40). This can be shown by using Kirchhoff's Law, which holds under LTE conditions. The radiance emitted from a mass ρ_a *ds* of gas in the beam is $k_\nu \rho_a J_\nu ds$ and the radiance absorbed is $k_\nu \rho_a L_\nu ds$. Hence the spectral emittance, the ratio of the emitted radiance to the radiance emitted by a black body is $\varepsilon_\nu = \frac{k_\nu \rho_a ds J_\nu}{B_\nu}$. The spectral absorptance, the fraction of the incident radiation that is absorbed is $\alpha_\nu = \frac{k_\nu \rho_a ds L_\nu}{L_\nu} = k_\nu \rho_a ds$ neglecting scattering. However, Kirchoff's law states that $\varepsilon_\nu = \alpha_\nu$ and hence $J_\nu = B_\nu$.

2.5 Radiative heating

One of the main goals of radiative calculations is to obtain radiative-heating rates throughout the atmosphere. For this one requires knowledge of the heating due to absorption of solar (short-wave or UV) photons and the heating and cooling due to absorption and emission of thermal (long-wave or infrared) photons.

Consider a horizontal slab of atmosphere, of horizontal area A, at height z and of thickness Δz and make the plane-parallel atmosphere assumption. (i.e. the model assumption that the atmosphere is made up of layers which are horizontally infinite and uniform at each level of the atmosphere). The upward power entering the bottom of the slab is $AF_z(z)$ and the upward power emerging at the top is $AF_z(z + \Delta z)$. The loss of radiative power within the volume $A\Delta z$ of the slab is therefore:

$$A\left[F_Z\left(z - F_Z\left(z + \Delta z\right)\right)\right] \approx -\left(A\Delta z\right)\frac{dF_Z}{dz}$$

Equation 56

This loss of radiative power implies that radiative diabatic heating of the slab is occurring at a rate $-\frac{dF_Z}{dz}$ per unit volume. The diabatic heating per unit mass can be written as:

$$Q = -\frac{1}{\rho(z)}\frac{dF_Z}{dz}$$

Equation 57

per unit mass, where ρ is the density of air. The units of Q are W kg^{-1}; often the quantity Q/c_p arises in calculations of the dynamical effects of radiative

heating, where c_p is the specific heat capacity at constant pressure, and this quantity has units Ks^{-1}. Since $F\uparrow$ and $F\downarrow$ both involve integration of spectral irradiances over frequency, F_z and Q also comprise contributions from different frequency bands.

2.5.1 Short-wave heating / Chapman layers

Consider the diabatic heating rate per unit volume, ρQ_ν^{SW}, produced by absorption of short-wave solar radiation of frequency ν by a gas of density $\rho_a(z)$ and extinction coefficient $k_\nu(z)$; scattering will be neglected. Assuming that the Sun is directly overhead, the appropriate optical path for each frequency is the optical depth, measured vertically downwards from the top of the atmosphere (taken to be $z=\infty$) :

$$\chi_\nu(z)=\int_z^\infty k_\nu(z')\rho_a(z')dz'$$

Equation 58

By analogy with the second term of Equation 54, the downward irradiance of the solar radiation is given by:

$$F_\nu^\downarrow(z)=F_{\nu\infty}^\downarrow e^{-\chi_\nu(z)}$$

Equation 59

where $F_{\nu\infty}^\downarrow$ is the downward solar irradiance at the top of the atmosphere. (Some care is needed to ensure that correct signs are obtained in Equation 58 and Equation 59). Since scattering is neglected, the upward solar irradiance $F_\nu^\uparrow(z)$ must be zero and the net vertical irradiance at frequency ν is:

$$F_{Z\nu}(z)=-F_{\nu\infty}^\downarrow e^{-\chi_\nu(z)}$$

Equation 60

Note that the exponential term is equal to the transmittance between height z and the top of the atmosphere. The contribution to the heating rate per unit volume from this frequency is therefore:

$$\rho Q_V^{SW}=\frac{d}{dz}\left(F_{\nu\infty}^\downarrow e^{-\chi_\nu(z)}\right)=F_{\nu\infty}^\downarrow\left(-\frac{d\chi_\nu}{dz}\right)e^{-\chi_\nu(z)}=F_{\nu\infty}^\downarrow k_\nu(z)\rho_a(z)e^{-\chi_\nu(z)}$$

Equation 61

Suppose now that the extinction coefficient is independent of z and that the density of the absorber decays exponentially with height, $\rho_a(z)=\rho_a(0)e^{-z/H_a}$, where H_a is a constant. Then the integral of Equation 58 can be evaluated to give:

$$\chi_\nu(z)=H_a k_\nu \rho_a(0)e^{-z/H_a}=\chi_\nu(0)e^{-z/H_a}$$

Equation 62

this shows how the optical depth increases as the solar radiation penetrates downwards, i.e., as z decreases. Substitution into Equation 60 then gives the vertical irradiance:

$$F_{Z\nu} = -F_{\nu\infty}^{\downarrow} \exp\left[-\chi_\nu(0)e^{-z/H_a}\right]$$

Equation 63

and differentiation gives the monochromatic volume heating rate:

$$\rho Q_V^{SW}(z) = F_{\nu\infty}^{\downarrow} k_\nu \rho_a(0) \exp\left(-\frac{z}{H_a} - \chi_\nu(0)e^{-z/H_a}\right)$$

Equation 64

Figure 38 shows the variation of the optical depth, the vertical irradiance and the volume heating rate, as functions of height in this simple example; note that the volume heating rate has a broad single peak. Differentiation of Equation 64 with respect to z shows that the monochromatic volume heating rate has a maximum at the height where the optical depth equals unity. We assume that the absorber is sufficiently 'optically thick' for this level to occur above the ground.

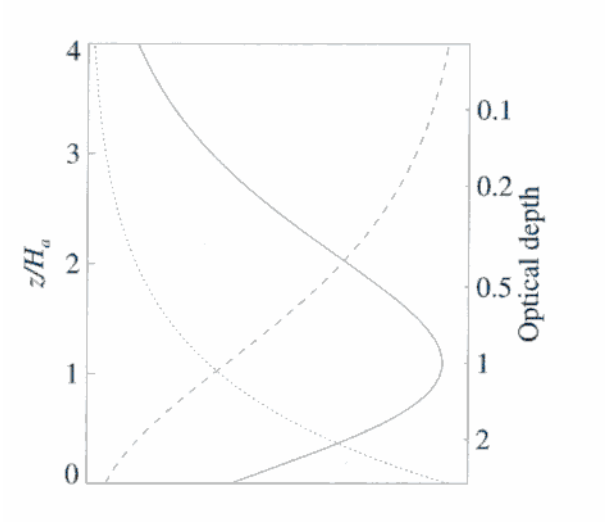

Figure 38 Vertical profiles of the short-wave volume heating, $\rho Q_V^{SW}(z)$, (full line) the negative of the vertical irradiance, $-F_{z\nu}(z)$, (dashed line) and the absorber density, $\rho_a(z)$, (dotted line) for solar radiation at frequency ν. The horizontal scales are arbitrary. The left-hand ordinate shows the height z, divided by H_a; the right-hand ordinate shows the optical depth. The optical depth of the ground is arbitrarily chosen to be 3. The sun is overhead. The absorber gas has an extinction coefficient and an exponentially decaying density. Taken from Andrews (2000).

A vertical structure of the type given by Equation 64 (and shown in Figure 38) is said to exhibit a Chapman layer. The peaked shape of the Chapman layer in the heating rate can be interpreted as follows. At high levels, above the peak there is a large vertical irradiance (since little absorption of the solar beam has yet occurred), but few absorber molecules; at low levels, below the peak, there is a small vertical irradiance (since much absorption has occurred) but many absorber molecules. In each case the heating rate is small. However, near the level of unit optical depth both the irradiance and the absorber density are significant and so the heating rate is comparatively large. Chapman layers also occur in other processes determined primarily by the absorption of radiation; an example is the photo-dissociation that contributes to the formation of the 'ozone layer'.

2.5.2 Chapman layers as a function of solar zenith angle

It is also possible to determine the form of a Chapman layer in a more realistic situation. For example, solar radiation penetrates the atmosphere at an angle of incidence which depends on the local time, season and latitude (See Figure 25). Using the Beer-Lambert law (Equation 48) which describes the absorption of a ray of incident light passing through a thin layer ds. The variation of radiance is given by:

$$dL_\nu = -k_\nu(s)L_\nu(s)ds$$

Equation 65

where k_ν is the extinction coefficient. A schematic is shown in Figure 39. Neglecting scattering as before this coefficient is proportional to the concentration (n) of the absorbing particles and is related to the effective cross section (σ) by the expression:

$$k_\nu = \sigma_\nu n$$

Equation 66

Integrating Equation 65 and substituting into Equation 66 gives –

$$L_\nu = L_{\nu 0}\exp\left(-\int_s \sigma_\nu(s)n(s)ds\right)$$

Equation 67

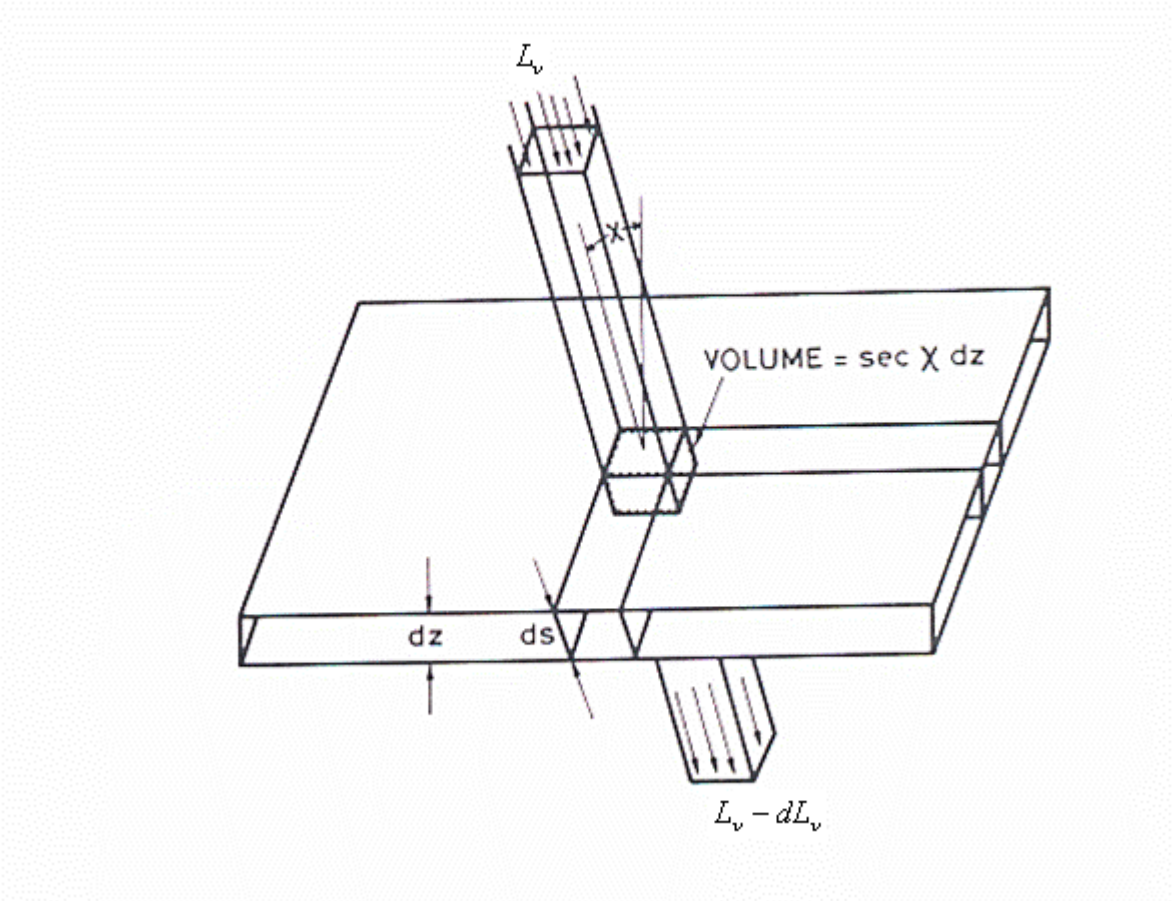

Figure 39 Absorption of solar radiation in an atmospheric layer of unit area. Adapted from Brasseur and Solomon (1986).

In reality, solar radiation penetrates the atmosphere at an angle of incidence which depends on the local time, season and latitude. The cosine of the local solar zenith angle (χ) is given by Equation 37. Assuming a flat earth –

$$ds = dz\sec(\chi)$$

Equation 68

The concentration n(z) is often assumed to vary exponentially with altitude according to a scale height H. Thus, for a medium containing only one absorbing gas, the variation of solar radiation with altitude can be written as-

$$L_\nu = L_\nu(\infty)\exp\left(-\sec(\chi)\int_z^\infty \sigma_\nu(z) n_o e^{-z/H} dz\right)$$

Equation 69

where $L_\nu(\infty)$ represents the solar radiance outside the Earth's atmosphere. After integration Equation 69 can be written as –

$$L_\nu = L_\nu(\infty)\exp\left(-\sec(\chi)\sigma_\nu n_o H e^{-z/H}\right)$$

Equation 70

The rate of photodissociation and production of heat, are all directly linked to the rate of energy deposition in the atmosphere by absorption. The rate of energy deposition in the atmosphere due to absorption is given by-

$$r = \frac{dL_\nu}{ds} = \frac{dL_\nu}{dz}\cos(\chi) = \sigma_\nu n_0 L_\nu(\infty)\exp\left(-\left(\frac{z}{H} + \sec(\chi)\sigma_\nu n_o He^{-z/H}\right)\right)$$

Equation 71

It can be shown that the rate of energy deposition in the atmosphere exhibits a maximum at the altitude –

$$z_m = H\ln(\sigma_\nu n_0 H\sec(\chi))$$

Equation 72

The rate of energy deposited at this altitude can be derived by substituting Equation 72 into Equation 71 to give –

$$r_m = \sigma_\nu n_0 L_\nu(\infty)\cos(\chi)\exp\left(-1-\frac{z_o}{H}\right)$$

Equation 73

When χ=0 (sun overhead), Equation 72 and Equation 73 can be simplified to:

$$z_{\chi=0} = H\ln(\sigma_\nu n_0 H)$$

Equation 74

and

$$r_{\chi=0} = \sigma_\nu n_0 L_\nu(\infty)\exp\left(-1-\frac{z_0}{H}\right)$$

Equation 75

Using the substitution –

$$x = \frac{z - z_{\chi=0}}{H}$$

Equation 76

The rate of energy deposition can be written as –

$$\frac{r}{r_{\chi=0}} = \exp(1 - x - \sec(\chi)\exp(-x))$$

Equation 77

The theoretical discussion presented above was suggested by Chapman in 1931 and is thus sometimes referred to as Chapman theory. The appearance of a characteristic layer due to absorption of solar energy is shown in Figure 40; this distribution is obtained by the Chapman theory for different solar zenith angles.

In the theory outlined above the attenuation of solar radiation in the atmosphere depends on the solar zenith angle because of the sec(χ) term. Thus, the radiative heating varies as a function of latitude, and time of year. However, when the sun is near the horizon, the solar zenith angle must be replaced by the Chapman function, in order to account for the Earth's curvature.

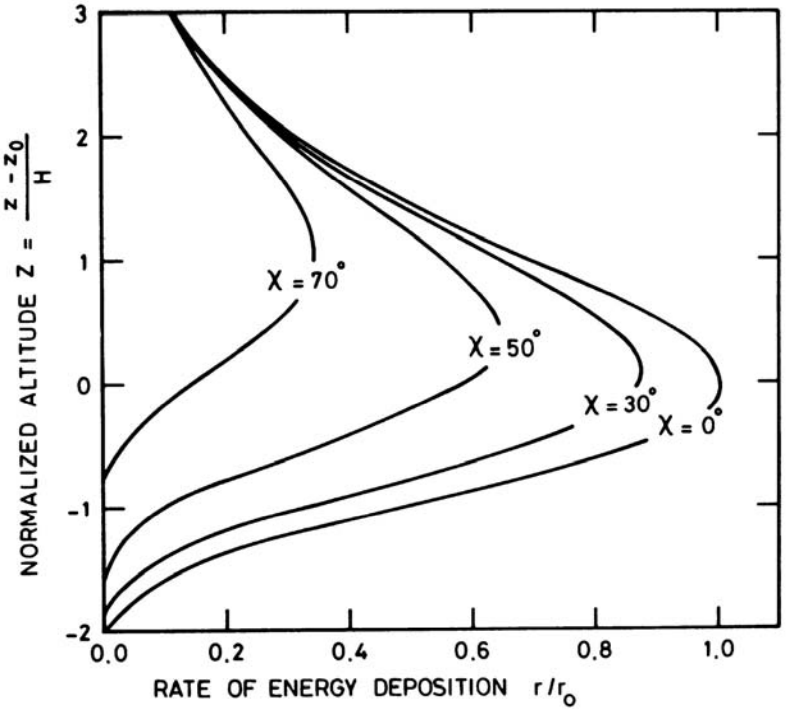

Figure 40 Schematic of a layer of absorption due to absorption of solar energy using the Chapman theory. Taken from Brasseur and Solomon (1986).

In reality, the atmosphere can be separated into many regions associated with absorption by different species. The altitude of the peak of the Chapman layer as a function of wavelength is shown in Figure 41. The various atmospheric species which dominate the absorption at different wavelengths are also displayed.

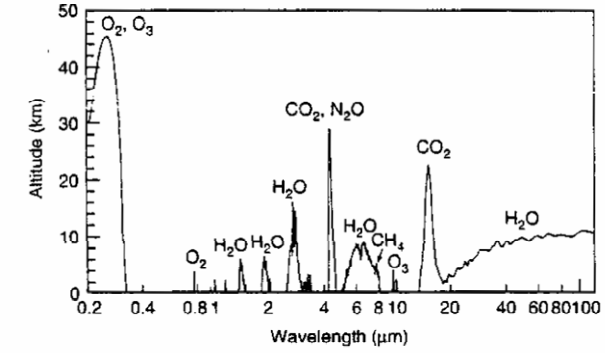

Figure 41 Altitudes of Solar absorption and thermal emission. Heights of Chapman layers and primary absorbers for vertical radiances. Note the gap between 0.4 and 1.0 microns associated with the visible region. Taken from Goody and Hu (2002).

2.5.3 Long-wave heating and cooling

We now consider the effects of thermal (long-wave or infrared) photons, allowing for both downward and upward paths. It can be shown, by solving the radiative transfer equation and integrating over solid angle, that the upward thermal irradiance at frequency ν and height z is:

$$F_\nu^\uparrow(z) = \pi \int_0^z B_\nu(z') \frac{\partial T_\nu^*(z', z)}{\partial z'} dz' + \pi B_\nu(0-) T_\nu^*(0, z)$$

Equation 78

Here $T_\nu^*(z',z)$ is the spectral transmittance, averaged over the upward hemisphere to take account of all slanting paths between heights z' and z, and $B_\nu(0-)$ is the Planck function (see Equation 46) evaluated at the temperature of the Earth's surface. (Note that this temperature may be different from that of the air just above the surface.) The assumption of LTE (local thermodynamic equilibrium) has been made, so that the source function $J_\nu = B_\nu$, and the Earth's surface has been assumed to radiate as a black body. Similarly, the downward irradiance is given by:

$$F_\nu^\downarrow(z) = -\pi \int_z^\infty B_\nu(z') \frac{\partial T_\nu^*(z', z)}{\partial z'} dz'$$

Equation 79

there is no 'boundary' term here since the downward thermal irradiance at the top of the atmosphere is zero.

Noting that the net upward long-wave spectral irradiance is:

$$F_{Z\nu}(z) = F_\nu^\uparrow(z) - F_\nu^\downarrow(z)$$

Equation 80

and from this the net long-wave diabatic heating rate per unit mass, Q_ν^{LW}, can be calculated using Equation 57. The resulting expression for the net long-wave diabatic heating rate per unit mass is quite complicated, but has a simple physical interpretation; namely the net heating or cooling at a given level is due to the difference between the energy gained per unit time by absorption of photons from neighbouring levels and the Earth's surface and the energy lost per unit time by emission of photons to neighbouring levels, the Earth's surface and space.

These heating and cooling terms can also be obtained directly. Consider for example the rate of loss of energy by a horizontal slab of atmosphere of thickness Δz and horizontal area A at height z by emission of photons to space (often called the cooling-to-space term). The derivation of the radiative transfer equation (Equation 50) shows that the spectral power emitted in a vertical direction from this slab is:

$$= k_\nu \rho_a J_\nu A \Delta z$$

Equation 81

Where the source function is J_ν, which is equal to $J_\nu = B_\nu$ under LTE. The fraction of this power that escapes to space is given by the transmittance:

$$T_\nu(z,\infty) = \exp\left(-\int_z^\infty k_\nu \rho_a dz'\right)$$

Equation 82

Noting that

$$\frac{\partial T_\nu(z,\infty)}{\partial z} = k_\nu(z)\rho_a(z)T_\nu(z,\infty)$$

Equation 83

we find that the power escaping to space from the slab in a purely vertical beam is:

$$= B_\nu(z)\frac{\partial T_\nu(z,\infty)}{\partial z} A\Delta z$$

Equation 84

Now, integrating over all slanting paths as above and replacing T_ν by T_ν^*, we obtain a contribution to the heating rate per unit mass:

$$Q_\nu^{CTS}(z) = -\frac{\pi B_\nu(z)}{\rho(z)}\frac{\partial T_\nu^*(z,\infty)}{\partial z}$$

Equation 85

The factor π in this equation comes from integration over the upward hemisphere and the minus sign from the fact that the power loss to space implies a negative heating at height z. The other contributions to the heating rate at height z can be calculated in similar, but more complicated, ways; these must include the heating of the slab due to absorption of photons emitted at other levels. A useful simplification for some purposes is the cooling-to-space approximation, in which the loss of photon energy to space dominates the other contributions; therefore, under this approximation:

$$Q_\nu^{LW} \approx Q_\nu^{CTS}$$

Equation 86

All gases that absorb and emit at frequency ν must in principle be included in T_ν^*. Then Q_ν^{LW} must be integrated over all relevant frequencies to obtain the total long-wave cooling $Q^{LW}(z)$.

2.5.4 Net radiative-heating rates

The short-wave and long-wave contributions to the diabatic heating rate Q can be computed using the principles described previously, together with information on the atmospheric temperature structure and the concentration and spectroscopic characteristics of absorber gases. The heat budget is governed primarily by the absorption of solar unltarviolet radiation by ozone and emission of infrared radiation by carbon dioxide, ozone and water vapour. Central to the suceessful evaluation of the heat budget are accurate radiative transfer calculations at wavelengths spanning the far-infrared to the ultraviolet.

Figure 42 shows the spectral cooling rate for all atmospheric gases as a function of wavenumber and pressure. It should be noted that pressure is used as the vertical coordinate in Figure 42. The cooling rate profiles generated by integrating the spectral cooling profiles shown in Figure 42 are shown in Figure 43. Calculation shows that 96% of middle atmospheric cooling is associated with three spectral bands. Carbon dioxide at 15 microns, the ozone band at 9.6 microns and the water vapour rotational bands contribute 72, 17 and 7%, respectively. Heating rates are dominated by absorption by molecular oxygen in the Schumann-Runge bands, the Schumann-Runge continum, the Herzberg badns, and at Lyman alpha wavelengths. In general, the UV heating by oxygen is small in the stratosphere compared with that of ozone which we include in the Hartley, Huggins and Chappius bands (see Figure 44).

Figure 45 shows the vertical profiles of the global-mean short-wave heating rate Q_ν^{SW}/c_p and the long-wave cooling rate $-Q_\nu^{LW}/c_p$, in convenient units of K day^{-1}. The corresponding profiles of heating and cooling due to the most important atmospheric gases are also shown. It is clear that the total heating is approximately equal to the total cooling over much of the profile, except in the troposphere and above about 90 km. The net effect is shown in Figure 48 for the troposphere and stratosphere. In the global mean, the middle atmosphere is therefore close to being in radiative equilibrium.

Below 80 km the short-wave heating rate is dominated by a Chapman-layer like structure, centred near 50 km, due to absorption of solar radiation by ozone. The peak of the heating rate, at over 10 K day^{-1}, lies above the maxima both in the ozone density, near 22 km, and in the ozone mixing ratio near 37km (see Figure 19). The fact that this peak is above the maximum ozone density is consistent with Figure 38. Below 15 km, in the troposphere, the main contribution to the heating rate is from water vapour, at about 1 K day^{-1}. Heating due to absorption by ozone and molecular oxygen is important between 80 and 100km.

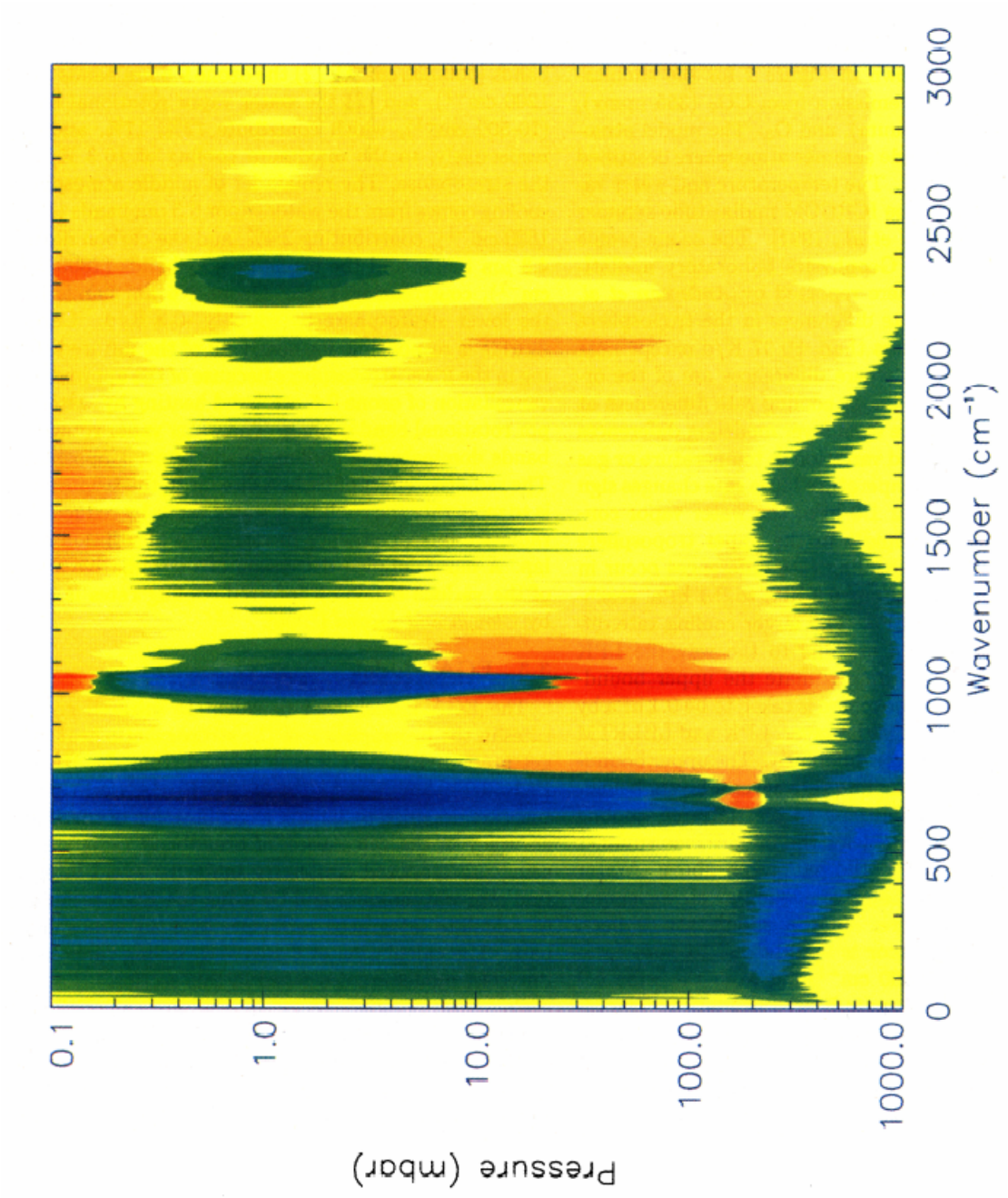

Figure 42 Spectral cooling rate for all atmospheric gases as a function of wavenumber and pressure. Taken from Mertens et al. (1999).

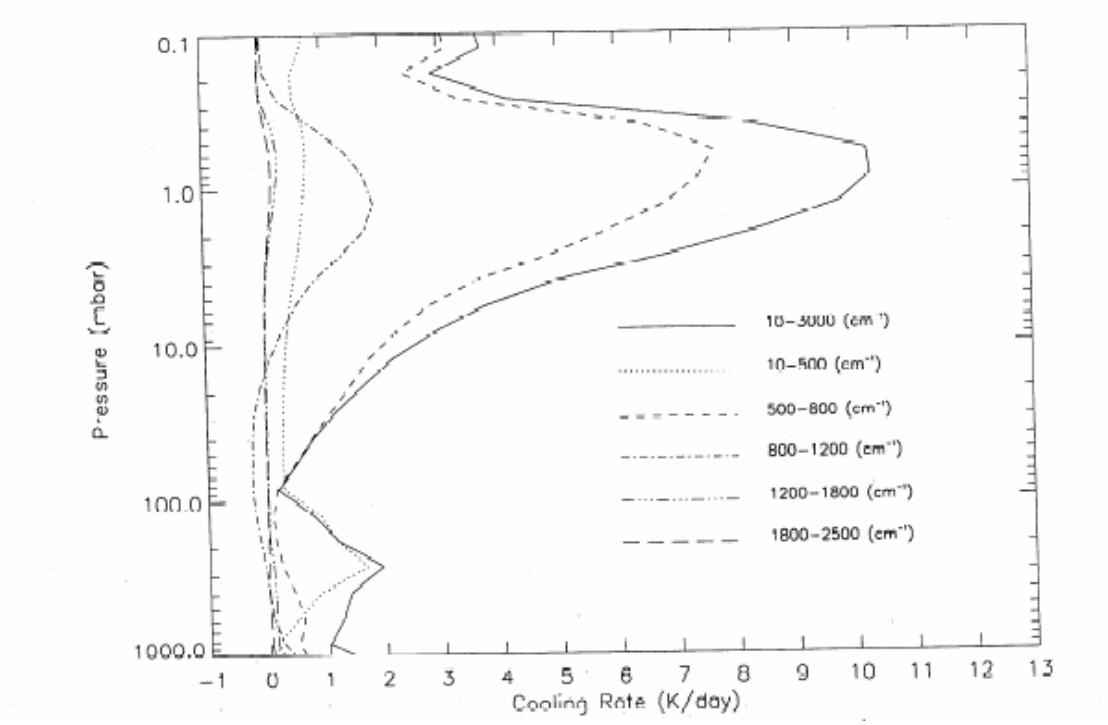

Figure 43 Cooling rate profiles generated by integrating the spectral cooling profiles shown in Figure 42. Taken from Mertens et al. (1999).

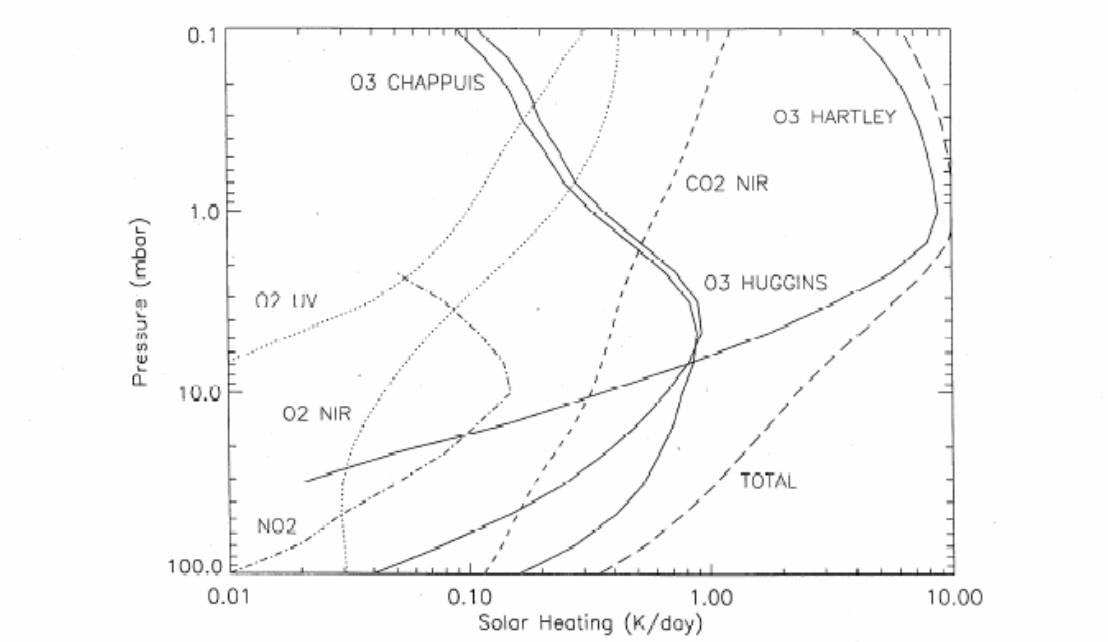

Figure 44 Solar heating rates at Equinox in the stratosphere. Taken from Mertens et al. (1999).

The peak in long-wave cooling near 50 km has significant contributions due to carbon dioxide and ozone, both cooling to space. The dominant wavelength bands involved are 15μm for $C0_2$ and 9.6μm for 0_3. A small amount of long-wave heating appears near 20 km, in the lower stratosphere, due to

absorption by ozone of upwelling radiation from the troposphere at wavelengths near 9.6μm, in the atmospheric window. Tropospheric cooling is dominated by water vapour, at about 2 K day^{-1}.

Figure 45 omits complications due to clouds and aerosols. The radiative cooling rates associated with the increased aerosol loading in the stratosphere associated with the eruption of Pinatubo is shown in Figure 46. Figure 46 shows that the radiative cooling observed in the atmosphere was primarily observed around the equatorial region, this is because the vast majority of the aerosol loading occurred over the tropics.

A close up of the heating rates in the troposphere and stratosphere is shown in Figure 48. This second representation of the heating and cooling rates taken from a separate model also indicates that the models show the same general form. But, that the magnitude of the cooling and heating rates is a little different in the two models.

The temperature structure associated with radiative equilibrium is compared to the climatological mean represented in the US standard atmosphere in Figure 47. It should be noted that the approximate radiative equilibrium in the global mean does not apply to individual latitudes and seasons (see Section 2.5.2). For example, in the winter stratosphere and the summer upper mesosphere there are big differences between Q_v^{SW} and Q^{LW}. In such regions dynamical heat transport is also significant and this explains the differences between the two temperature profiles. It must be emphasised that the net heating rate $Q = Q^{SW} + Q^{LW}$ should not be thought of as a pre-ordained heating, to which the atmospheric temperature and wind fields respond. One reason is that Q itself depends strongly on the temperature, especially through Q^{LW}.

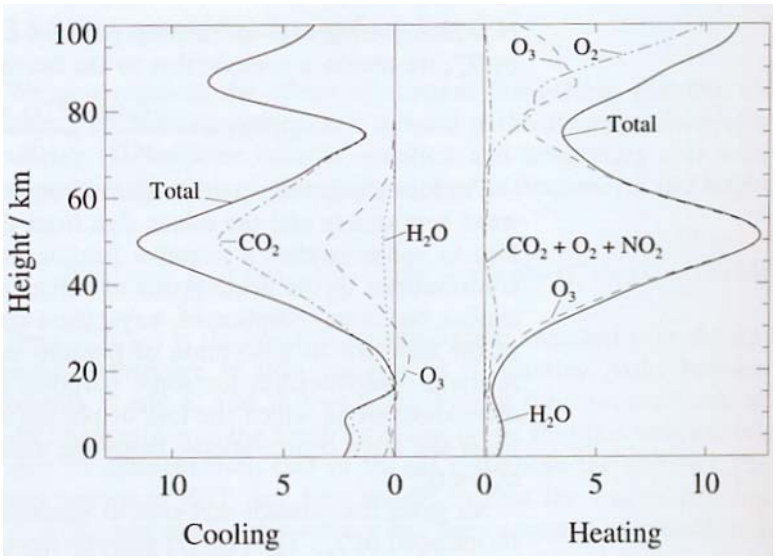

Figure 45 Global-mean vertical profiles of the short-wave heating rate and the long-wave cooling rate, in K per day. Taken from Andrews (2000).

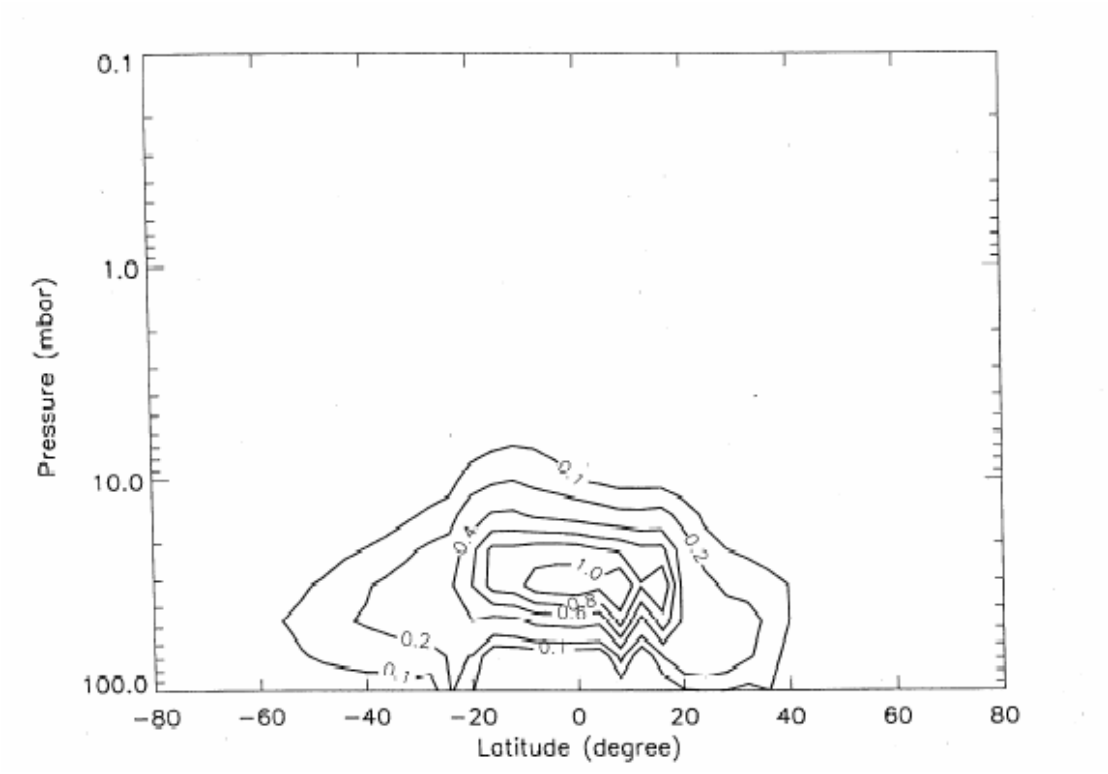

Figure 46 Infrared cooling rates by Pinatubo aerosols for 30th December 1991 to 29th January 1992. Taken from Mertens et al. (1999).

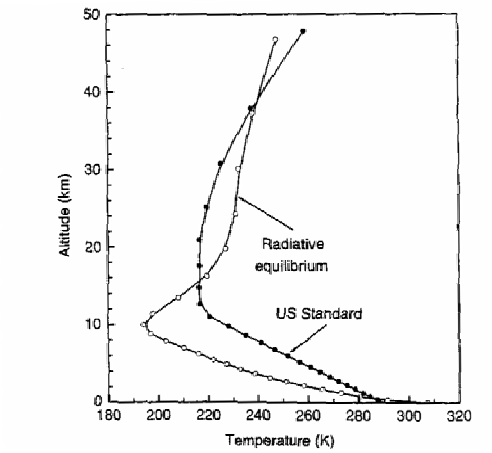

Figure 47 Radiative equilibrium temperatures compared to the US standard atmosphere. Taken from Goody and Hu (2002).

The application of the theory presented to the long-wave radiation component shows that this component generates a net cooling on the order of 2.5 K per day in the troposphere. The maintenance of a steady state in the atmosphere is only possible through the transfer of sensible heat (enthalpy) and latent heat (evaporation-condensation) from the earth's surface to the atmosphere.

In the daytime and with clear skies, the net radiation balance is dominated by the short-wave solar radiation but at night the balance is, of course, entirely due to the long-wave radiation. Because the concentrations of some of the

absorbers (e.g., H_2O and 0_3) vary strongly with height the solar and terrestrial radiation fluxes are also strongly height dependent. In general, in clear and calm nights the long-wave radiative flux increases with height, leading to a flux divergence and a cooling of the atmosphere.

In addition. clouds have a strong effect on the transfer of long-wave radiation because they modify the emissivity of the atmosphere at certain wavelengths. Clouds are almost completely opaque for infrared radiation. They act as if they close the atmospheric window, preventing the escape of long-wave radiation into space. This effect is large enough to strongly influence the surface temperature. For example, under cloudy conditions the nocturnal temperature drop will be much reduced as compared with clear night conditions. Even the presence of a thin layer of cirrus clouds can be enough to cause an increase in surface air temperature because of the additional long-wave radiation emitted by the clouds.

Manabe and Strickler (1964) used a simple one-dimensional climatology to study the contributions by H_2O, $C0_2$, and 0_3 to the atmospheric heating and cooling rates. Their results for the case that the cloudiness is equal to the global mean value are shown in Figure 48.

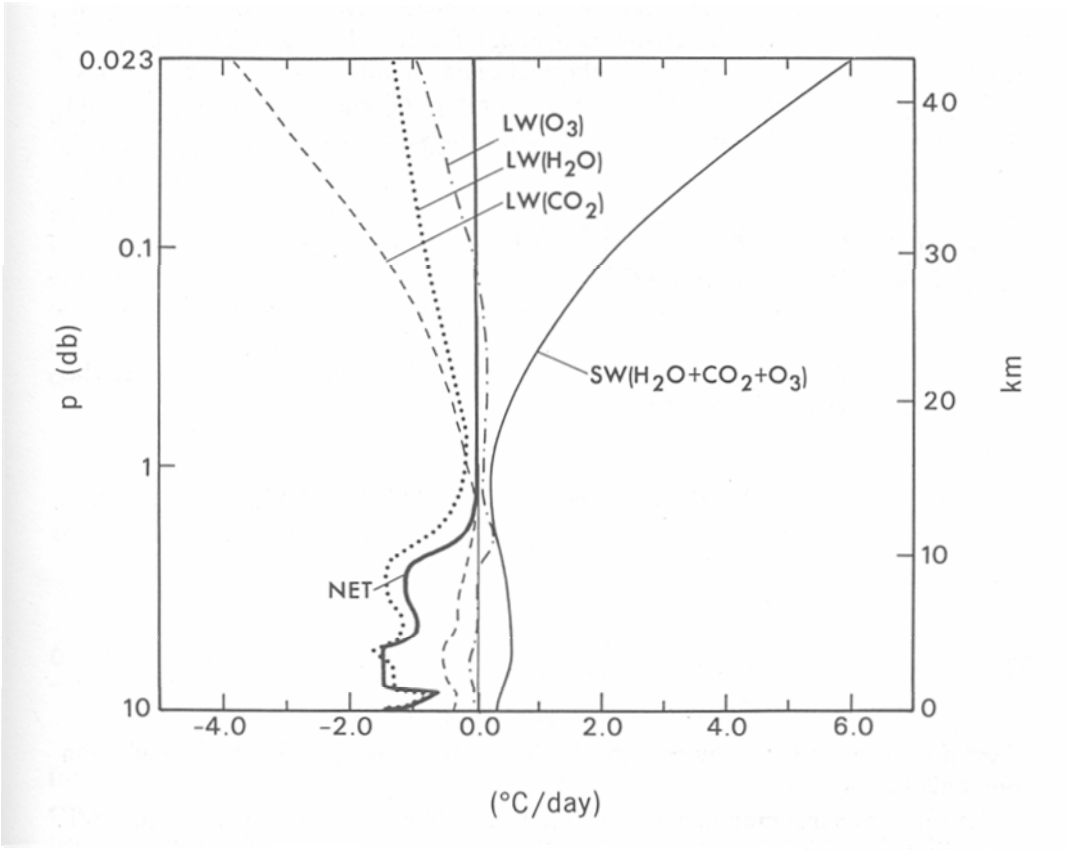

Figure 48 Vertical distribution of the computed rate of temperature change in the atmosphere for thermal equilibrium due to various absorbers. LW (H_2O), LW(CO_2), and LW(O_3) show the rate of temperature change due to long-wave radiation emission and absorption by water vapour, carbon dioxide, and ozone, respectively. SW(H_2O+CO_2+O_3) shows the rate of temperature change due to the absorption of solar radiation by these three gases. The think line represents the net rate of temperature change due to all components. Taken from Piexoto and Oort (1992).

The troposphere shows a net radiative cooling mainly due to water vapour, which is compensated by the latent and sensible heating associated with moist convection. In the stratosphere, there is a strong heating due to the absorption of ultraviolet solar radiation by ozone and to a much smaller extent, due to the absorption of long-wave terrestrial radiation in the 9.6μm band. We may note that the long-wave heating is only possible because of the low concentrations of 0_3 in the troposphere. The cooling in the stratosphere is due to long-wave emission mainly by $C0_2$ and to a lesser extent by water vapour and ozone.

In the 8-12μm region the atmosphere is almost transparent for long-wave terrestrial radiation, with the exception of the absorption at 9.6 μm by ozone. This region of the spectrum is therefore known as the atmospheric spectral window. It is also the range where the long-wave radiation for the atmosphere is a maximum. It is interesting to note that, through absorption in the spectral window, small increases of CO_2 or of other trace gases may have a large impact on the climate. In addition, Ramanathan et at. (1985) have brought attention to the fact that some minor trace gases, such as

chlorofluorocarbons, methane (CH_4) and nitrous oxides (N_2O), also have absorption lines in the spectral window. The recent increase in the concentrations of these trace gases can lead to additional trapping of the radiation emitted by the earth's surface and thereby to heating of the troposphere and cooling of the stratosphere.

Several other studies have been made using observed distributions of the absorbers in the atmosphere. Results show a net radiative cooling throughout the year at almost all latitudes in the troposphere. However, in the stratosphere he finds heating at low latitudes and cooling at mid and high latitudes. All the studies mentioned so far are limited by an incomplete knowledge of the concentrations of the various absorbers in the upper troposphere and stratosphere and by the difficulties of incorporating the effects of clouds.

As previously indicated the radiative heating varies as a function of latitude, and time of year (see Figure 49). The absorption of ultraviolet radiation by ozone in the Huggins and Hartley bands constitutes the principal source of heat in the stratosphere and mesosphere. The heating rate reaches as much as 12K/day near the stratopause, with a maximum of about 18K/day near the summer pole. The effect of the Huggins bands in the visible region becomes important in the lower stratosphere, where the resulting heating rate is not quite 1K/day. These numbers are obviously related to the amount of ozone present, and an increase in ozone density would lead to an increase in the stratospheric and mesospheric temperatures, as well as possible changes in the locations of the stratopause and mesopause.

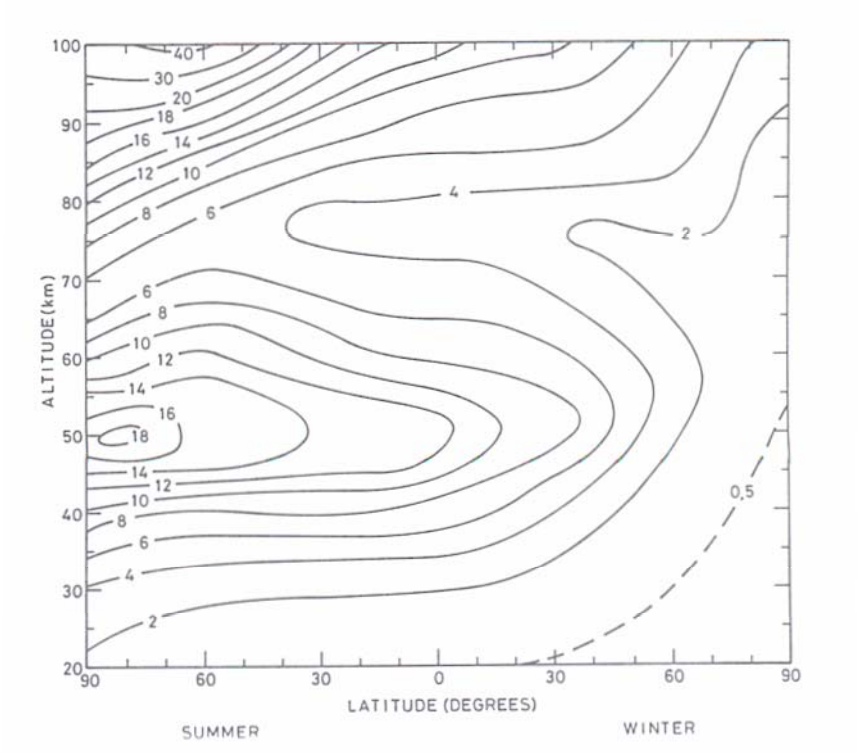

Figure 49 Distribution of the heating rate associated with the effects of absorption by ozone and oxygen as a function of latitude and altitude. Taken from Brasseur and Solomon (1986).

3 Planetary waves

A planetary wave is a large-scale perturbation of the atmospheric dynamical structure that extends coherently around a full longitude circle. The perturbations have wave-like form in the longitudinal and vertical directions and often also in the altitudinal direction. These large-scale waves are a dominant part of the spatial and temporal variability in the stratosphere, and also make significant contributions at higher altitudes in the mesosphere. In their simplest form they occur because of the variation of the Coriolis parameter with latitude and are known as Rossby waves. Most important are quasi-stationary mid-latitude Rossby waves, which propagate upward from the troposphere and are ubiquitous but quite variable in the middle atmosphere during winter. They are important because they have significant influence on the wind speeds, temperature, distribution of ozone, and structure.

Where they occur planetary waves are evident in all dynamical fields (i.e. temperature, wind, pressure and density). However, observations of wave structure are most commonly viewed using the geopotential height.

3.1 Geopotential height and Thickness

The geopotential ϕ at any point in the atmosphere is defined as the work done against the Earth's gravitational field in order to raise a mass of 1kg from sea level to that point. In other words, ϕ is the gravitational potential for unit mass. The work in raising 1kg from z to z+δz is gδz, therefore:

$$\delta\phi = g\delta z$$

Equation 87

The geopotential $\phi(z)$ at height z is thus given by:

$$\phi(z) = \int_0^z g\,dz$$

Equation 88

where the geopotential at sea level, is by convention taken as zero. It should be noted that the geopotential at a particular point depends only on the height and not the path taken to the point. We may also define a quantity called the geopotential height Φ as:

$$\Phi \equiv \frac{\phi(z)}{g_o} = \frac{1}{g_o}\int_0^z g\,dz$$

Equation 89

where g_o is taken as 9.8ms^{-2}. The geopotential height and the geometric height are approximately the same in the lower atmosphere where $g_o \approx g$. The difference in geopotential height Φ_2-Φ_1 between any two layers in the

atmosphere is called the thickness of the intervening layer. It is possible to show that thickness charts are effectively charts of the mean temperature between different pressure levels using the derivation indicated below. Integrating the hydrostatic equation (Equation 10) from a height z to the top of the atmosphere gives:

$$p(z) = \int_z^\infty \rho g dz$$

Equation 90

Using Equation 90 and the ideal gas law ($p = RT\rho$, where R is the gas constant per unit mass) we may write the thickness as:

$$\Delta\Phi = \Phi_2 - \Phi_1 = \frac{R}{g_o}\int_{p_2}^{p_1} T d\ln p$$

Equation 91

Defining the mean temperature for a pressure layer as:

$$\overline{T} = \frac{\int_{p_2}^{p_1} T d\ln p}{\int_{p_2}^{p_1} d\ln p}$$

Equation 92

We can rewrite Equation 91 as:

$$\Delta\Phi = \frac{R\overline{T}}{g_o}\ln\left(\frac{p_1}{p_2}\right)$$

Equation 93

Thus, the thickness of a layer is proportional to the mean temperature of the layer.

3.2 Planetary wave climatology

Figure 50 shows the geopotential height for the 10hPa pressure level in the middle stratosphere of the Northern hemisphere derived from satellite data. The contours indicate the height in kilometres of the 10hPA surface, which has an average altitude of about 28 to 32km. The planetary waves are the deviations from symmetry around a longitude circle. These can readily be isolated by performing Fourier analysis in longitude at a fixed latitude and pressure level (see Figure 51). The wave field shown is dominated by zonal wavenumber 1, but wavenumber 2 also has significant amplitude. Wavenumber 1 waves relate to a wave with a single wavelength around a latitude circle (at 45^o latitude this corresponds to a wavelength of approximately 28,300km). Similarly, wavenumber 2 waves relate to a wave with two wavelengths around a latitude circle (at 45^o latitude this corresponds

to a wavelength of approximately 14,150km). On the polar plot shown in Figure 50, a wavenumber 1 pattern tends to be evident as a displacement of the lowest heights off of the pole and sometimes, as in this case, a distinct region of high heights. Wavenumber 2 is evident as an elongation of the contours of low height.

From geostrophic balance, air motion on a pressure surface tends to follow lines of constant geopotential height with low values to the left in the Northern hemisphere (anti-clockwise flow around the low pressure). One can see that following such a trajectory will move an air parcel alternately closer to the pole and farther from the pole. The time taken to complete a circuit of the pole depends on the latitude and wind speed, but is normally several days to greater than a week. The air parcel will also move vertically as it follows the circumpolar path, although the vertical excursions are much smaller than those in latitude, being typically on the order of 1km.

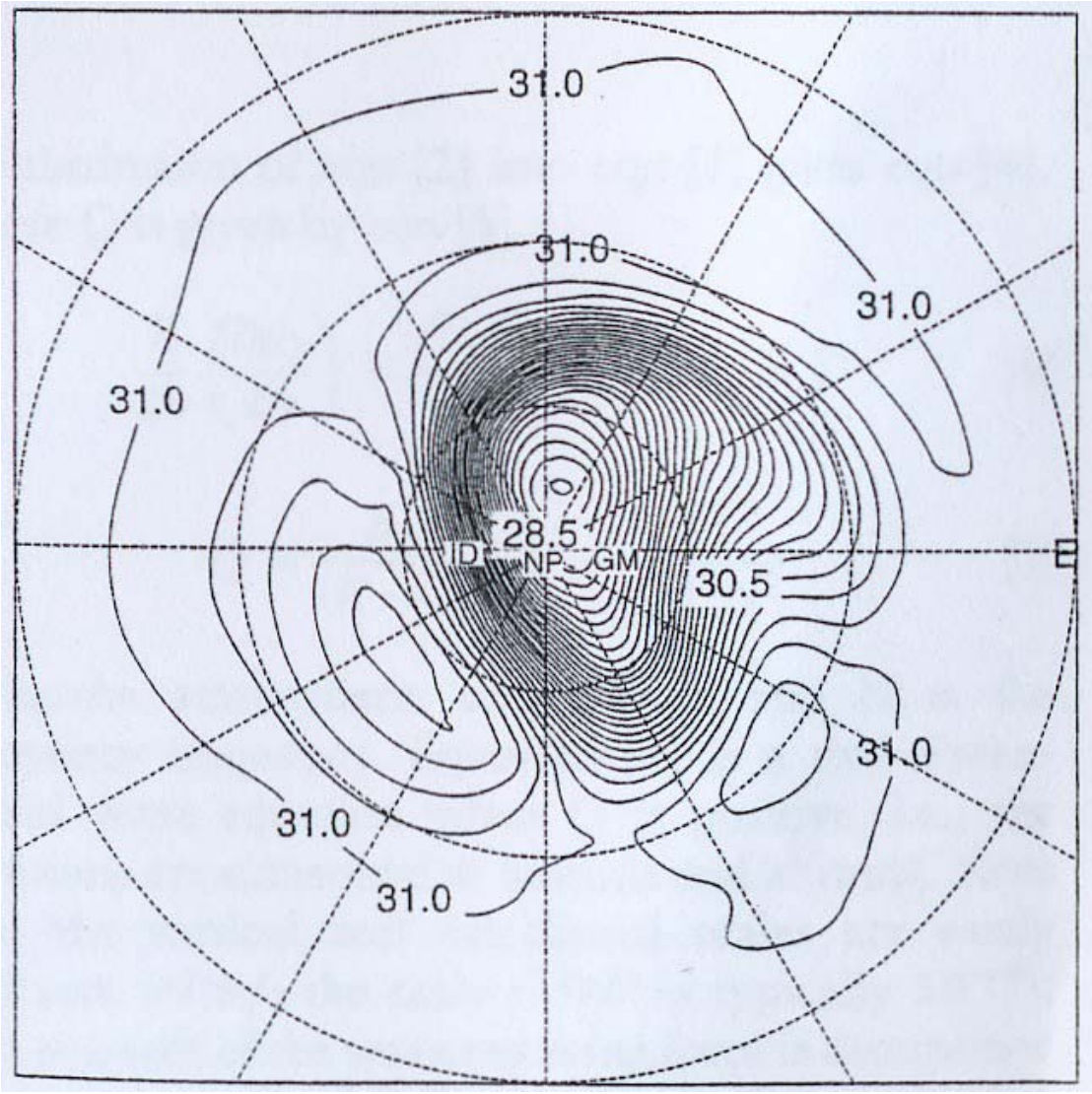

Figure 50 Polar stereographic projection of the geopotential height for the Northern hemisphere at 10hPa on 20th January 1979. Contours indicate the height in kilometres. Taken from Smith (2003).

The amplitude of the wavenumber 1 and 2 Rossby waves as a function of pressure (or altitude) and latitude for the case illustrated in Figure 50 is displayed in Figure 52. Typical mid-latitude Rossby waves vary in amplitude daily, but the longitude of the maximum geopotential usually stays in the same quadrant. Figure 52 shows several typical features of wave amplitudes in the winter hemisphere:

- The amplitudes are largest in middle to high latitudes near 60^o.
- Maximum amplitudes in the vertical are reached in the stratosphere between 10hPa and 0.3hPa.
- Wavenumber 1 waves have larger amplitude than wavenumber 2 waves.

It should be noted that this case is slightly atypical because the wavenumber 1 wave has a very large magnitude. It should be noted that some of the Rossby waves generated in the troposphere by thermal contrasts between land and ocean and by flow over mountains propagate vertically and are the source of mid-latitude waves in the middle atmosphere. A substantial part of the variation observed is a result of the variability of the large-scale weather patterns in the troposphere. Since some of these large scale waves are forced by features at the surface, by orography and thermal contrast (e.g. between land and sea) they are stationary with respect to the surface to produce, for example, the Aleutian low (or cyclone).

There are several reasons that constitute the substantial interest in these phenomena. For example, planetary waves drive the circulation out of radiative equilibrium, they are known to be important mechanisms for transport processes in the stratosphere, are responsible for the intermittent breakdowns of the polar vortex structure ("sudden stratospheric warmings") and are involved in vortex erosion processes. Especially, the latter point is of substantial interest since the dramatic springtime depletions of ozone in polar regions require that polar stratospheric air has a high degree of dynamical isolation and extremely cold temperatures necessary for the formation of polar stratospheric clouds. Both of these conditions are produced within the undisturbed stratospheric polar vortex.

It should be noted that there is a clear seasonal variation in planetary wave activity. A minimum in planetary wave activity is observed in the summer and a maximum in the winter. This can be clearly observed when the climatological mean temperature field at various pressure levels for January in both hemispheres is examined (see Figure 53). This difference is demonstrated by the fact that the fields are nearly zonally symmetric in the summer hemisphere, but in the winter hemisphere a large degree of longitudinal variation is observed. Comparison of Figure 53 and Figure 54 also shows another interesting point, namely, that the amplitude of planetary wave activity in the Northern hemisphere is significantly larger than those observed in the Southern hemisphere.

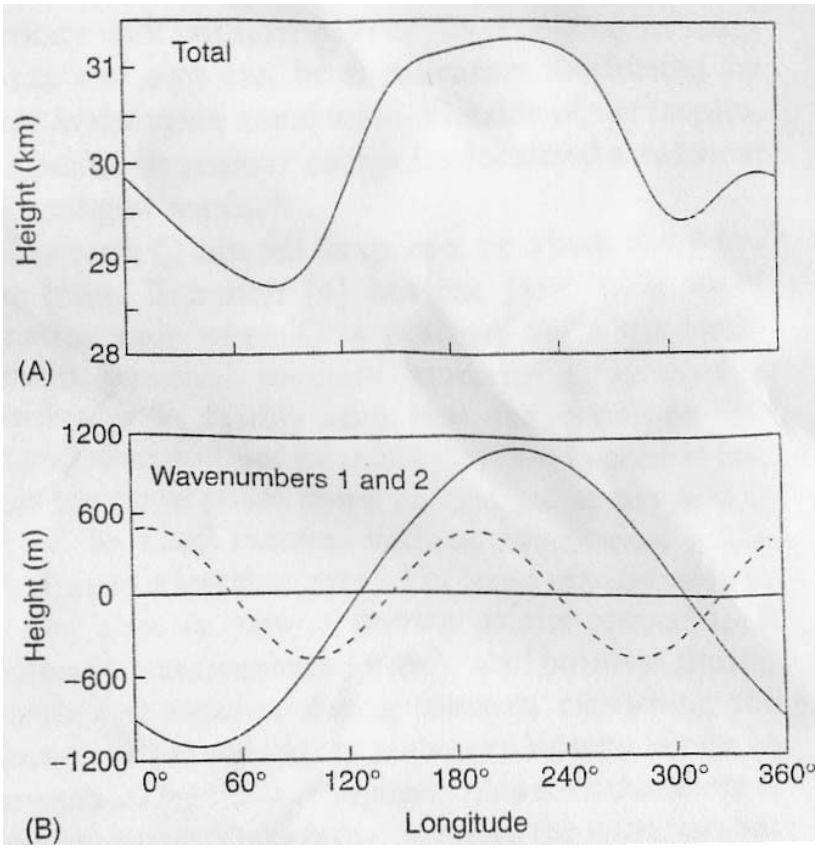

Figure 51 (A) Variation of geopotential height with longitude and (B) Fourier analysis showing the contributions of wavenumbers 1 (solid) and 2 (dashed) at 60°N. Taken from Smith (2003).

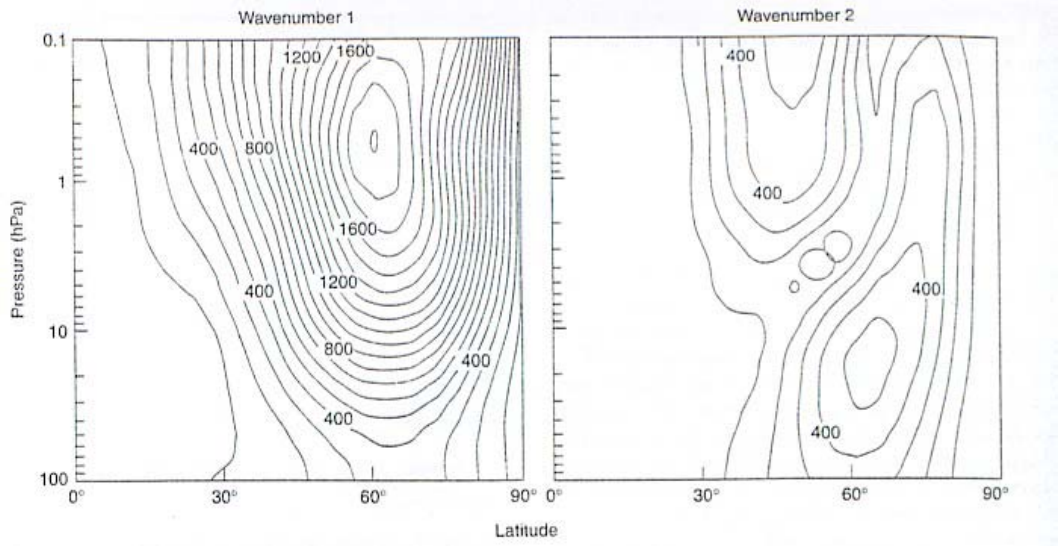

Figure 52 The amplitude of wavenumber 1 and 2 as a function of latitude and pressure. Taken from Smith (2003).

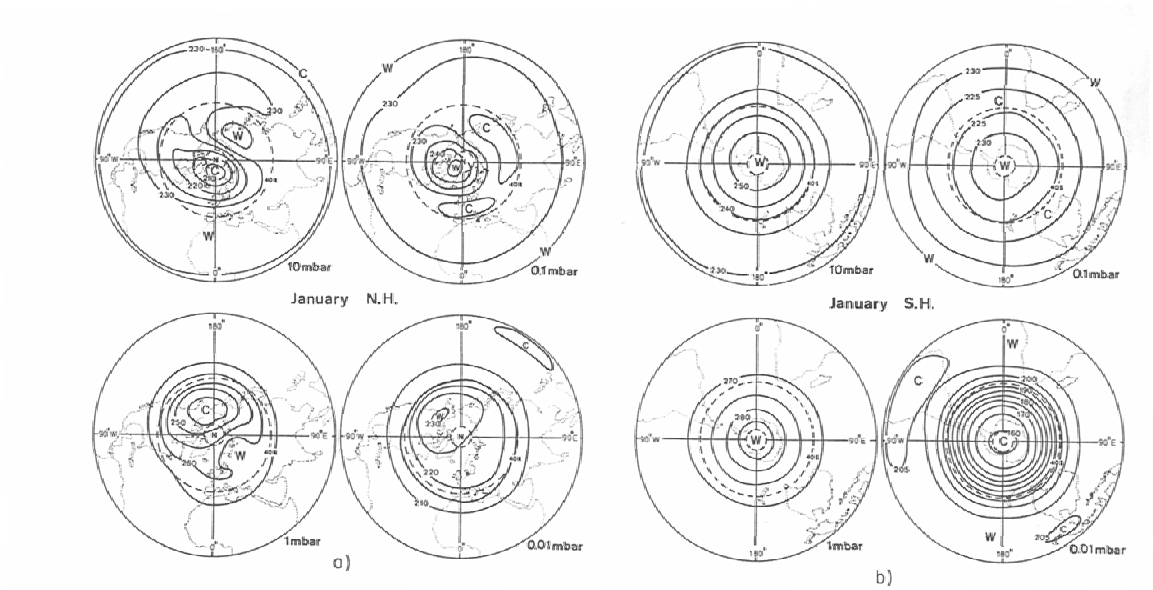

Figure 53 The climatological mean temperature field at various pressure levels for January in both hemispheres. Northern hemisphere values are associated with (a) and Southern hemisphere values are associated with (b). Taken from Barnett and Corney (1985).

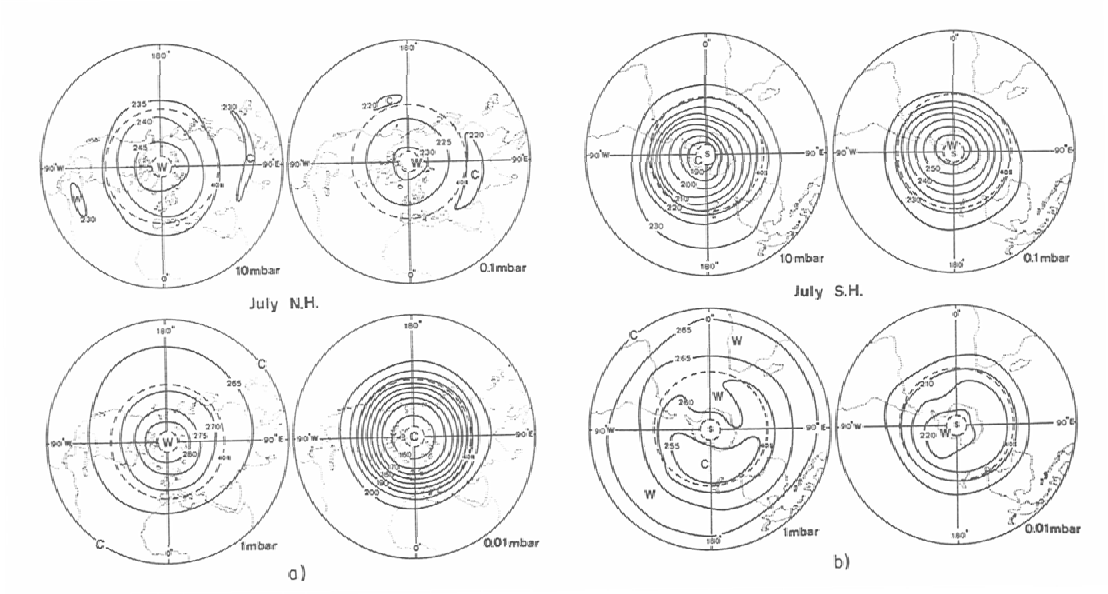

Figure 54 The climatological mean temperature field at various pressure levels for July in both hemispheres. Northern hemisphere values are associated with (a) and Southern hemisphere values are associated with (b). Taken from Barnett and Corney (1985).

3.3 Rossby waves

The simplest Rossby wave solution is obtained for an atmosphere of constant density under the assumption of uniform zonal flow ($\bar{u}$) and no vertical motion. The appropriate momentum equations (see Equation 4) and the

continuity equation (which states that the net flow of mass into unit volume per unit time is equal to the rate of change of density and is thus related to mass conservation) can be simplified to:

$$\frac{Du}{Dt} + \frac{1}{\rho}\frac{\partial p}{\partial x} - fv = 0$$

Equation 94

$$\frac{Dv}{Dt} + \frac{1}{\rho}\frac{\partial p}{\partial x} + fu = 0$$

Equation 95

and

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$

Equation 96

Operating on Equation 94 with $\partial/\partial y$ and Equation 95 with $\partial/\partial x$ and subtracting gives:

$$\frac{D}{Dt}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) + f\left(\frac{\partial v}{\partial y} + \frac{\partial u}{\partial x}\right) + v\frac{\partial f}{\partial y} = 0$$

Equation 97

The quantity $\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$ is one component of the $\nabla \times V$ which is known as the vorticity ξ; it may be considered as a vector equal to twice the local angular velocity of fluid elements. The second term in Equation 97 is zero using Equation 96. Because the Coriolis parameter f only varies with latitude it is possible to write Equation 97 in the form:

$$\frac{D(\xi + f)}{Dt} = 0$$

Equation 98

The quantity $(\xi + f)$ is known as the absolute vorticity; it is the vorticity due to the rotation of the fluid itself combined with that due to the Earth's rotation. Equation 98 shows that under the conditions imposed (i.e. non-divergent, frictionless flow) absolute vorticity is conserved.

To solve Equation 98 we assume a linear relationship between f and y (i.e. we let $f = f_o + \beta y$ where β is a constant (this is known as the β-plane approximation).

We assume a uniform zonal flow ($\bar{u}$) in the unperturbed situation and introduce perturbations (i.e. $u = \bar{u} + u'$ and v=v') so that Equation 98 becomes:

$$\left(\frac{\partial v}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right)\left(\frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}\right) + \beta v' = 0$$

Equation 99

Because the flow is non-divergent in the horizontal (that is air is not flowing out of the region in question) a stream function ψ can be introduced such that by using it, the equation of continuity (Equation 96) is automatically satisfied, that is:

$$u' = -\frac{\partial \psi}{\partial y}$$

and

$$v' = \frac{\partial \psi}{\partial x}$$

Equation 100

Substituting Equation 100 into Equation 99 we obtain:

$$\left(\frac{\partial v}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right)\nabla^2\psi + \beta\frac{\partial \psi}{\partial x} = 0$$

Equation 101

For wave solutions:

$$\psi = \mathrm{Re}\{\psi_o \exp[i(\omega t + kx + ly)]\}$$

Equation 102

which to obtain real answers the following relationship must be satisfied:

$$c = \frac{-\omega}{k} = \bar{u} - \frac{\beta}{k^2 + l^2}$$

Equation 103

this relationship being known as the dispersion relation. The velocity relative to the zonal flow is $c - \bar{u}$ where c is the phase velocity in the x direction. Rossby waves, therefore, drift to the west relative to the basic flow, at typical speeds of a few metres per second. It should be noted that the phase speed of the wave increases with wavelength.

Under the assumptions of two-dimensional flow in an atmosphere of uniform density, Equation 98 is a statement of the conservation of absolute vorticity. Before dealing with the propagation of Rossby-type waves in three dimensions we need to find an equation to describe vorticity changes under less stringent assumptions. For motions of synoptic scale this can be done by starting with the approximate horizontal momentum equations (Equation 94 and Equation 95). Operating on Equation 94 with $\partial/\partial y$ and Equation 95 with $\partial/\partial x$ and subtracting, and also noting that:

$$\frac{Df}{Dt} = v\frac{\partial f}{\partial y}$$

Equation 104

We can obtain for the rate of change of absolute vorticity:

$$\frac{D(\xi + f)}{Dt} = -(\xi + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) - \left(\frac{\partial w}{\partial x}\frac{\partial v}{\partial z} - \frac{\partial w}{\partial y}\frac{\partial u}{\partial z}\right) + \frac{1}{\rho^2}\left(\frac{\partial p}{\partial x}\frac{\partial p}{\partial y} - \frac{\partial p}{\partial y}\frac{\partial p}{\partial x}\right)$$

Equation 105

This is known as the vorticity equation. The first term on the right hand side is the most important since it arises because of the horizontal diveregence. If there is positive horizontal divergence, air is flowing out of the region in question, and the vorticity decreases. This is the same effect as occurs with a rotating body whose angular velocity decreases because of angular momentum conservation if its moment of inertia increases.

Scale analysis of Equation 105 shows that for synoptic scale motions the last two terms are considerably smaller than the others and that to a first approximation:

$$\frac{D_h(\xi + f)}{Dt} = -(\xi + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)$$

Equation 106

where $\frac{D_h}{Dt}$ denotes:

$$\frac{D_h}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$$

Equation 107

(i.e. the total derivative in two dimensions).

It is instructive to apply Equation 106 to an a atmosphere of constant density and temperature for which the continuity equation can be used to show that Equation 106 becomes:

$$\frac{D_h(\xi + f)}{Dt} = -(\xi + f)\frac{\partial w}{\partial z}$$

Equation 108

Because of the constant temperature the geostrophic wind is independent of height z. Further, because to a first approximation the vorticity is equal to the vorticity of the geostrophic wind the vorticity will not vary with altitude. Therefore, integrating Equation 107 between levels z_1 and z_2 where $z_2 - z_1 = h$ gives:

$$\frac{1}{(\xi + f)}\frac{D_h(\xi + f)}{Dt} = \frac{w(z_2) - w(z_1)}{h}$$

Equation 109

Now considering the fluid which at one time is confined between the level distance h apart, we obtain:

$$\frac{Dh}{Dt} = w(z_2) - w(z_1)$$

Equation 110

Thus, Equation 109 may now be rewritten as:

$$\frac{D_h}{Dt}\left(\frac{\xi + f}{h}\right) = 0$$

Equation 111

Equation 111 is a simplified statement of the conservation of potential vorticity. It has important consequences for atmospheric flow,. Consider, for example, adiabatic flow over a mountain barrier. As a column of air flows over the barrier its vertical extent decreases so that ξ must also decrease. A westward moving airstream will therefore, move equatorwards as it passes over the barrier.

To examine how Rossby waves propagate vertically we need to look for constraints imposed on three dimensional solutions. To simplify the treatment we need to work at nearly constant density and so we use the *Boussinesq approximaton* which allows us to neglect changes of density except where they are coupled with gravity to produce buoyancy forces. With this approximation the equation of continuity is as for an incompressible gas applies, so that the appropriate vorticity equation is Equation 108.

With a uniform unperturbed zonal flow $\bar{u}$, the perturbation form of Equation 108 can be written as:

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\xi'+v'\frac{\partial f}{\partial y}-f\frac{\partial w'}{\partial z}=0$$

Equation 112

where ξ has been neglected compared with f in the right hand side of Equation 108. We also require the perturbation form of the hydrostatic equation, which may be written as:

$$\frac{1}{\rho}\frac{\partial \rho}{\partial z}+\frac{\rho'}{\rho}g=0$$

Equation 113

And the thermodynamic energy equation which under the Boussienesq approximation becomes:

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\frac{\rho'}{\rho}-\frac{N^2}{g}w'=0$$

Equation 114

Eliminating ρ' and w' from Equation 112, Equation 113 and Equation 114 and putting:

$$\beta=\frac{\partial f}{\partial y}=\text{constant}$$

Equation 115

which is the β-plane approximation (that is we assume that the variation of the Coriolis parameter with latitude is constant) we obtain:

$$\left(\frac{\partial}{\partial t}+\bar{u}\frac{\partial}{\partial x}\right)\left(\xi'+\frac{f_o}{N^2\bar{\rho}}\frac{\partial^2 p'}{\partial z^2}\right)+\beta v'=0$$

Equation 116

Now most of the vorticity perturbation arises from the vorticity perturbation of the geostrophic wind whose components are:

$$u_g'=-\frac{1}{f_o\bar{\rho}}\frac{\partial p'}{\partial y}$$

and

$$v_g'=\frac{1}{f_o\bar{\rho}}\frac{\partial p'}{\partial x}$$

Equation 117

For the geostrophic winds, as in Equation 100, we can define a stream function ψ which, comparing with Equation 117, will be equal to:

$$\psi = \frac{p'}{f_o \bar{\rho}}$$

Equation 118

Introducing this term into Equation 116 and also approximating v' by its geostrophic value in the last term we obtain:

$$\left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right)\left(\nabla^2\psi + \frac{f_o^{\,2}}{N^2}\frac{\partial^2\psi}{\partial z^2}\right) + \beta\frac{\partial\psi}{\partial x} = 0$$

Equation 119

These substitutions involve the quasi-geostrophic approximation in which the wind is replaced by its geostrophic value except in the divergence term. For wave solutions:

$$\psi = \mathrm{Re}\{\psi_o \exp[i(\omega t + kx + ly + mz)]\}$$

Equation 120

to be possible for Equation 119, the dispersion relation:

$$c = -\frac{\omega}{k} = \bar{u} - \frac{\beta}{\left(k^2 + l^2 + m^2\frac{f_o^2}{N^2}\right)}$$

Equation 121

must be satisfied. Since $\frac{f_o^2}{N^2} << 1$, vertical wavelengths are of the order of 1% of horizontal wavelengths (see Figure 55 and Figure 56). Figure 55 and Figure 56 display temperature of a layer approximately 10km thick centred at roughly 45km and 15km, respectively. Comparison indicates that about half a wavelength of the wave in the vertical is observed between these two altitudes.

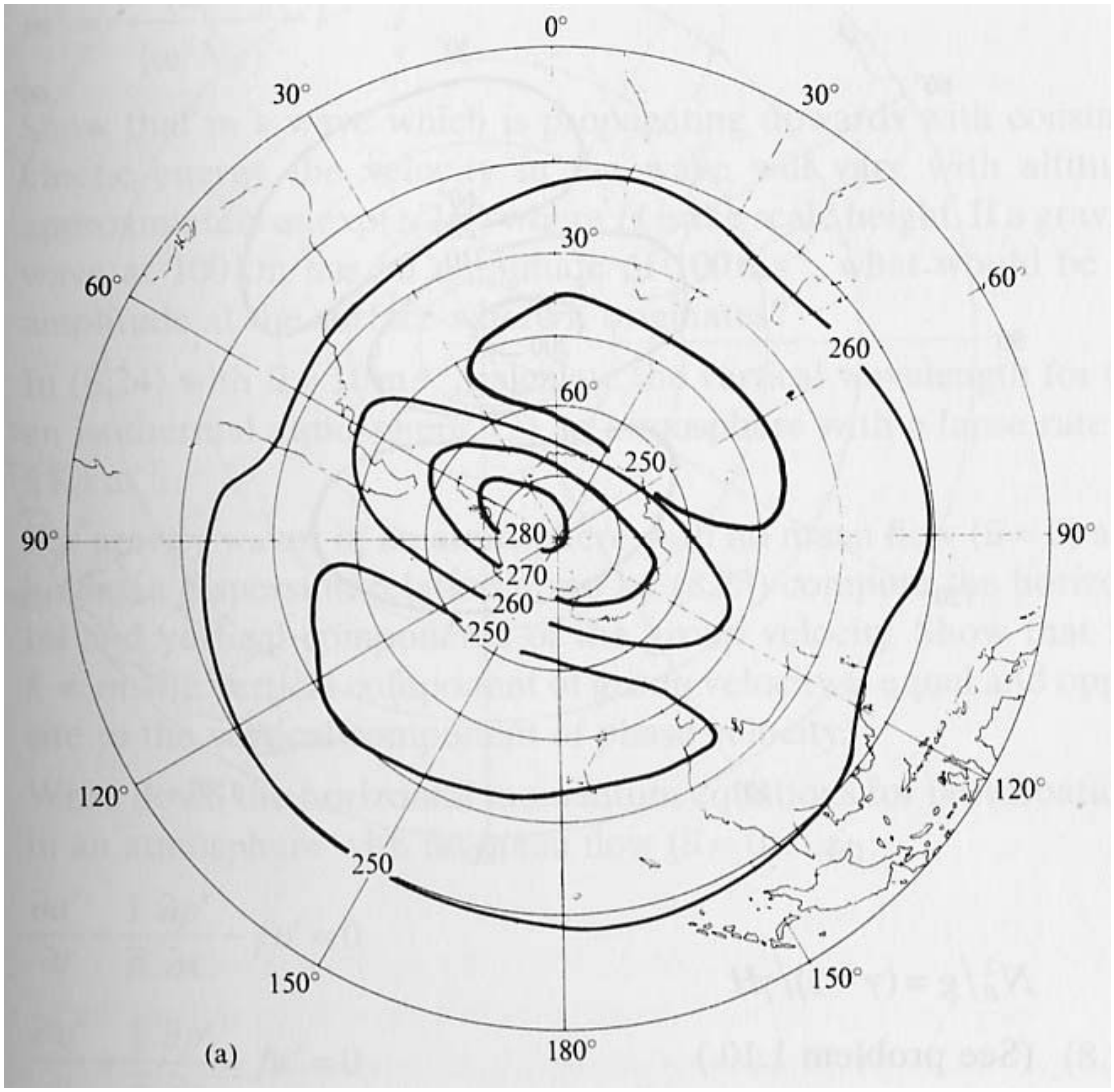

Figure 55 A quasi-stationary wave in the Southern hemisphere stratosphere of wavenumber 2 observed by the Nimbus 4 satellite. The quantity plotted is temperature of a layer approximately 10km thick centred at roughly 45km. Taken from Houghton (2001).

We might expect some of these large scale waves to be forced by features at the surface, by mountains or large land masses, in which case they would be stationary with respect to the surface. For c to be zero from Equation 121:

$$m^2 = \frac{N^2}{f_o^2}\left(\frac{\beta}{\bar{u}} - \left(k^2 - l^2\right)\right)$$

Equation 122

For vertical propagation to occur m^2 must be positive, which sets the condition:

$$0 < \bar{u} < \frac{\beta^2}{\left(k^2 + l^2\right)}$$

Equation 123

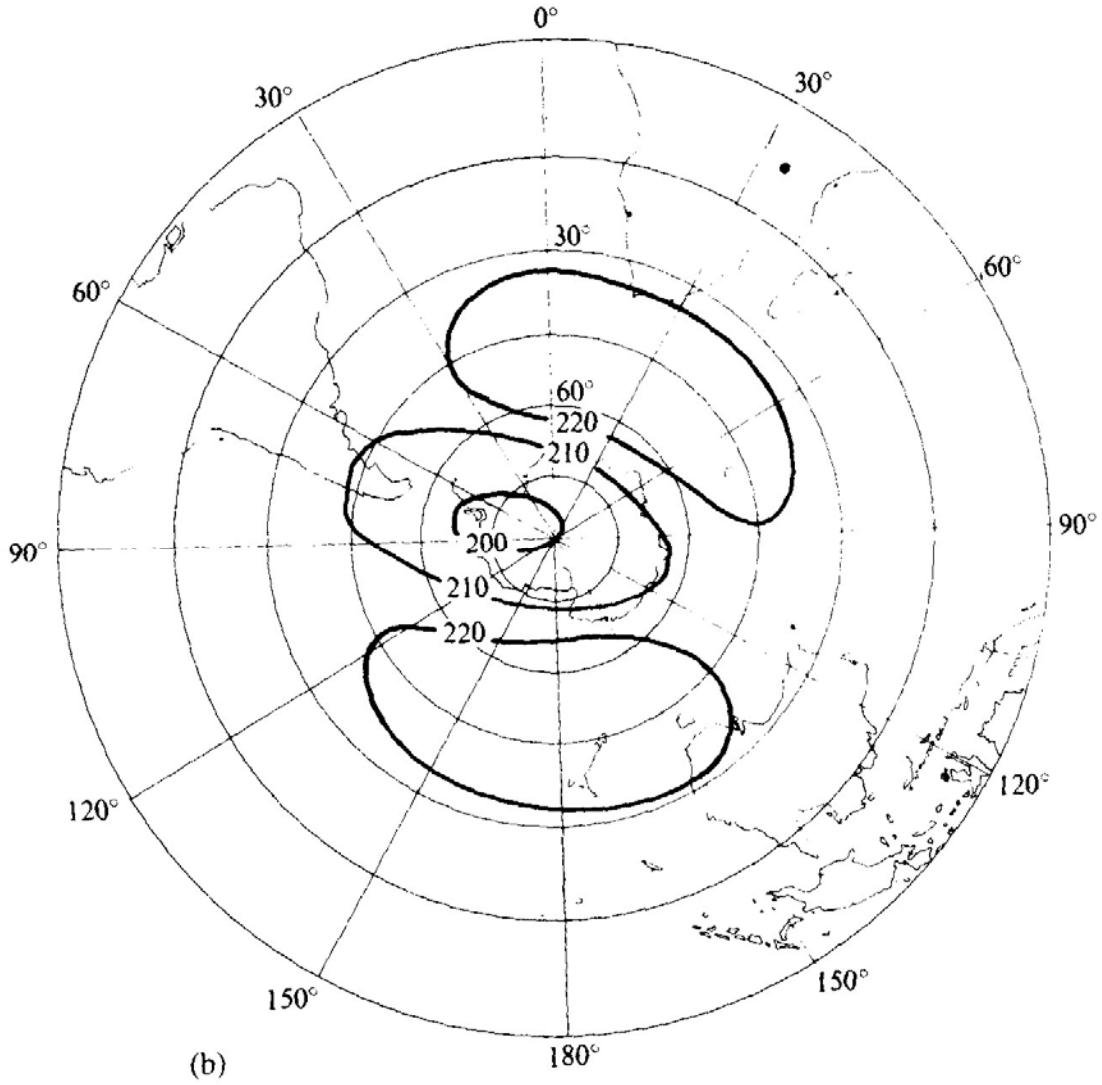

Figure 56 A quasi-stationary wave in the Southern hemisphere stratosphere of wavenumber 2 observed by the Nimbus 4 satellite. The quantity plotted is temperature of a layer approximately 10km thick centred at roughly 15km. Taken from Houghton (2001).

Under conditions of easterly flow, therefore, or very large westerly flow no vertical propagation will occur. This result is confirmed by the lack of planetary wave activity in the stratosphere during the summer months when the stratospheric flow is easterly (this is valid for both hemispheres). This can be observed when Figure 53 and Figure 54 are compared. Further the waves of largest horizontal wavelength propagate most readily.

The wave type that is of most importance for large-scale meteological processes is the Rossby wave. In an inviscid barotropic the Rossby wave is an absolute vorticity-conserving motion that owes its existence to the variation of the Coriolis force with latitude. More generally, in a baroclinic atmosphere the Rossby wave is a potential vorticity-conserving motion that owes its existence to the isentropic gradient of potential vorticity. Rossby wave

propagation can be understood in a qualitative fashion by considering a closed chain of fluid parcels initially aligned along a circle of latitude. For absolute vorticity which is related to the relative vorticity (ξ) and the Coriolis parameter (f) by $(\xi + \mathrm{f})$. Assume that $\xi = 0$ at time t_0. Now suppose that at t_1, δy is the meridional displacement of a fluid parcel from the original latitude. Then at time t_1 we obtain:

$$(\xi + f)_{t_1} = f_{t_0}$$

Equation 124

Which can in turn be rewritten as:

$$\xi_{t_1} = f_{t_0} - f_{t_1} = -\beta \delta y$$

Equation 125

where β is defined by Equation 115 and is the planetary vorticity gradient at the original latitude.

From Equation 125 it is evident that if the chain of parcels is subject to a sinusoidal meridional displacement under absolute vorticity conservation, the resulting perturbation vorticity will be positive (i.e. cyclonic) for a southward displacement in the Northern hemisphere and negative (anticyclonic) for a northward displacement in the Northern hemisphere. These values will be opposite in the Southern hemisphere.

The perturbation vorticity field will induce a meridional velocity field, which advects the chain of fluid parcels southward west of the vorticity maximum and northward west of the vortcity minimum, as shown in Figure 57. Thus, the fluid parcels oscillate back and forth about their equilibrium latitude, and the pattern of vorticity maxima and minima propogates to the west. This wetward propagating vorticity field constitutes a Rossby wave. Just as a positive vertical gradient of potential temperature resits vertical fluid displacements. The meridional gradient of absolute vorticity resists meridional displacements and provides the restoring mechanism for Rossby waves.

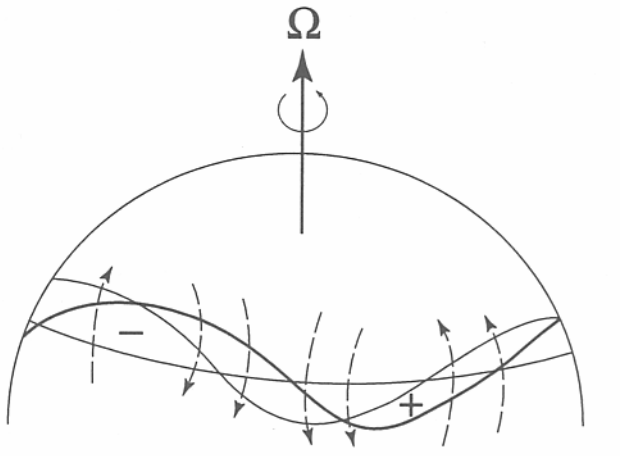

Figure 57 Perturbation vorticity field and induced velocity field (dashed arrows) for a sinusoidal displacement of a chain of fluid parcels from its mean latitude. Heavy wavy lines show the westward displacement caused by advection of the pattern by the induced meridional velocity field. Taken from Holton (2000).

As previously indicated, in the summer hemisphere the flow in the stratosphere is nearly parallel to latitude circles. Large-scale wave motions have very small amplitudes. In the winter (especially in the Northern hemisphere) the flow is disturbed by planetary-scale waves of zonal wavenumbers 1 and 2. Thus, the flow distribution for the Northern hemisphere features a cyclonic polar vortex that is distorted by quasi-stationary planetary waves that from an anticyclonic disturbance called the Aleutian high.

At times the planetary waves in the Northern winter stratosphere may amplify dramatically over a short span of time, and produce rapid meridional transport, which leads to rapid deceleration of the mean flow, and an accompanying sudden stratospheric warming in the polar region. The zonal flow deceleration associated with sudden warming of the polar stratosphere represents a temporary enhancement of the normal westward-darg force in the winter stratosphere. Breaking of planetary (Rossby) waves primarily produces this wetward zonal force. These waves produce a westward force that is hemispheric in scale. Owing to the rapid rotation of the Earth, the response to a westward force is to produce poleward drift. The Coriolis force associated with the polewrad drift approximately balances the wave-drag force, and maintains the zonal winds weaker than they would be in the absence of wavebreaking. Further, by mass continuity , the meridional drift leads to upward motion in low latitudes and downward motion in high latitudes. Thus, the extra-tropical wave-drag acta as a sort of “pump”. It pulls air upward in the tropics , and pushes air downward at high latitudes. Through adiabatic cooling accompanying the rising motion, the low latitudes are maintained at temperatures below radiative equilibrium. While through adiabatic warming accompanying the sinking motion, the high latitudes are maintained at temperatures above radiative equilibrium in the winter hemisphere. The meridional and vertical circulation produced by wavebreaking is shown in Figure 58.

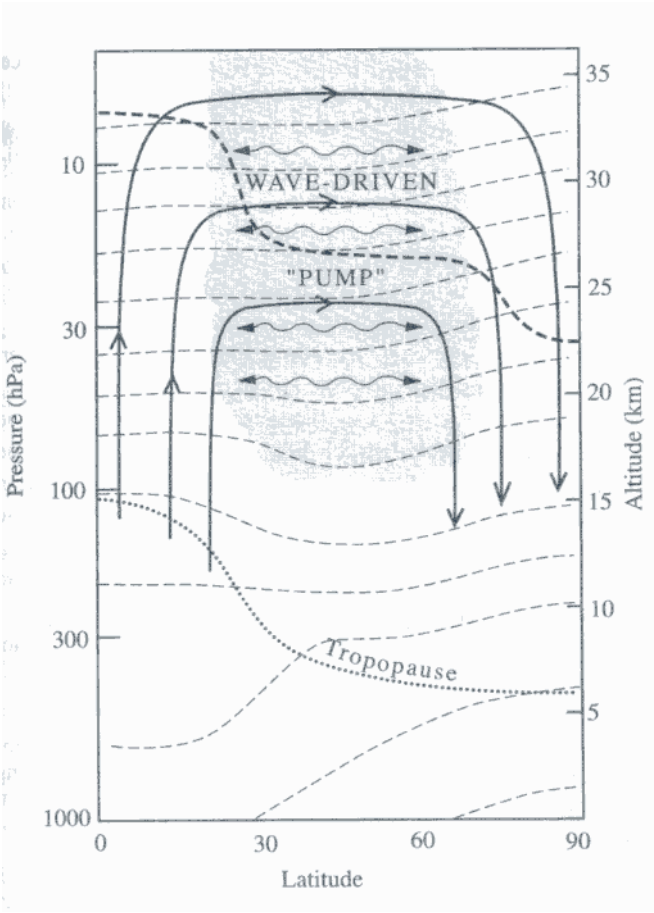

Figure 58 Schematic latitude-height cross-section showing the wave-driven circulation (solid lines with arrows) of the winter stratosphere. Shading indicates regions of planetary wave-breaking called the surf zone. Light dashed lines are isentropes; dotted line indicates the tropopause. Wavy double-headed arrows indicate meridional mixing by wavebreaking. Heavy dashed line shows a constant mixing ratio surface for a long-lived tracer. Taken from Holton (2000).

The observed annual cycle of temperature in the tropical lower stratosphere provides strong support for this hypothesis of upwelling driven non-locally by extra-tropical wave-induced forces. Owing to adiabatic cooling, the lowest temperatures should occur at the time of strongest upwelling, and highest temperatures at the time of weakest upwelling. In fact, tropical lower stratospheric temperatures are observed to vary annually, with lowest temperatures over the entire tropics during Northern hemisphere summer. This annual cycle is consistent with the annual cycle in the extratropical wave-induced stratospheric forcing, which is stronger in the Northern winter than in the Southern winter owing to the strong Rossby-wave activity of the Northern hemisphere in winter.

The prevailing zonal winds in the equatorial stratosphere are known to undergo a remarkable oscillation from strong westerlies to strong easterlies with a slightly irregular period that averages about 28 months (see Figure 9). The QBO influences on the extra-tropical Northern hemisphere and Southern hemisphere is very apparent (see Figure 59). The basis explanation for the observed QBO effect in the extratropics was advanced by Holton and Tan (1980,1982). Typically the dominant direction of wave activity propagation for

Quasi-stationary planetary waves (QSPW) forced in the extra-tropical troposphere is upward and equatorward. If the mean flow in the tropical stratosphere is westerly, QPSW's will be able to penetrate into the tropics and even across the equator without encountering a critical surface. By contrast, when the mean flow in the tropical stratosphere is easterly, QPSW's will encounter a critical surface on the winter side of the equator. It is believed that the critical surface should act to partially reflect the QSPW wave activity. Then, when there are easterlies in the tropics, the effective wave guide for QSPW propagation should be narrower and the wave activity at mid and high latitudes of the winter hemisphere should be stronger. Stronger QSPWs in high latitudes should lead to greater wave-induced drag on mean-flow, and hence a warmer winter stratospheric pole. All this is consistent with the observed Holton-Tan effect shown in Figure 59. A schematic diagram showing the zonal wind structure in the winter hemisphere and the resultant propagation direction of planetary waves for different phases of the QBO is shown in Figure 60.

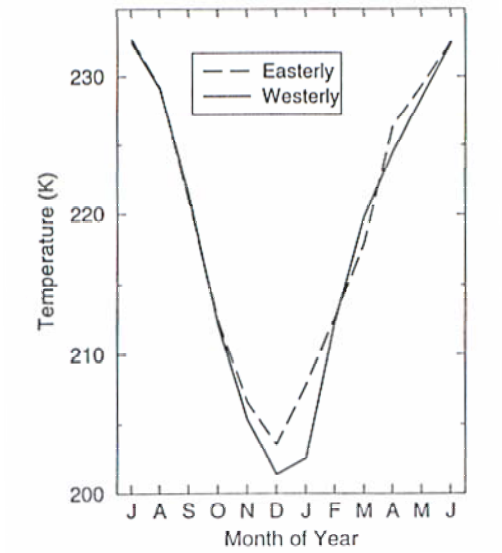

Figure 59 The zonal mean 30hPa temperature at 78°N composited for each calendar month for times when the 40hPa Singapore winds were westerly (solid) and easterly (dashed). Taken from Hamilton (2000).

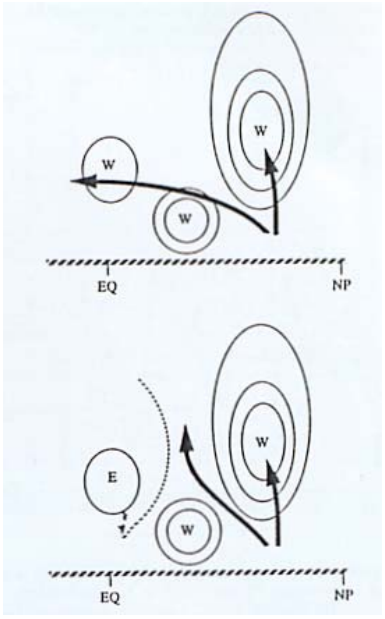

Figure 60 A schematic diagram showing the zonal wind structure in the winter hemisphere when the phase of the QBO in the equatorial lower stratosphere is westerly (top) and easterly (bottom). The dashed line in the bottom panel shows the locations of the zero mean wind surface. The thick arrow denotes the dominant paths of wave activity propagation associated with topographically-forced quasi-stationary planetary waves. Taken from Hamilton (2000).

3.4 Stratospheric Sudden warmings

The stratospheric sudden warming (SSW) is the most dramatic meteorological phenomena to occur in the stratosphere. SSWs occur in the stratosphere of the Northern hemisphere during winter. Temperatures rise dramatically by as much as 80^{o}C or more in a few days. Figure 61 shows temperature changes recorded at a high-latitude location during a strong sudden warming. In addition to the large increase in the stratosphere it is also interesting to note the cooling observed in the mesospheric regions (see also Figure 62). Accompanying these changes, the stratopause descends over some locations by as much as 20km, see Figure 62. Before the onset of the warming, the stratospheric circulation is dominated by a cold and strong westerly polar vortex, which lies over the North pole, covering most of the Northern hemisphere outside the tropics. During a major warming, this polar vortex is almost entirely broken down in a matter of days. The stratospheric circulation undergoes a dramatic change. Westerly winds are replaced by easterly winds throughout much of the stratosphere at high latitudes. The occurrence of a major warming can be detected in the stratosphere over the whole globe.

In addition, Major Warmings occur mostly in January-February. In addition to warming of the north polar region and reversal of the meridional temperature gradient, they are also associated with a breakdown of the polar vortex, which is replaced by a high. That is, the definition of a Major Midwinter Warming requires not only the warming but also a total change of circulation. The definition of a breakdown of the polar vortex is that the usual westerlies in the

Arctic at 10 hPa are replaced by easterlies so that the centre of the vortex moves south of 60-65°N. The vortex is either displaced entirely or split into two. This type of warming had not been observed in the Antarctic until the austral spring of 2002.

Major warmings tend to occur every three to four years in the Northern hemisphere. However, it should be noted that Minor warmings occur almost every winter. Although there are large increases in temperature during such events, the polar vortex is not broken down. However, it should be noted that Minor Warmings can indeed be intense and sometimes also reverse the temperature gradient, but they do not result in a reversal of the circulation at the 10 hPa level. They are found in the Antarctic as well.

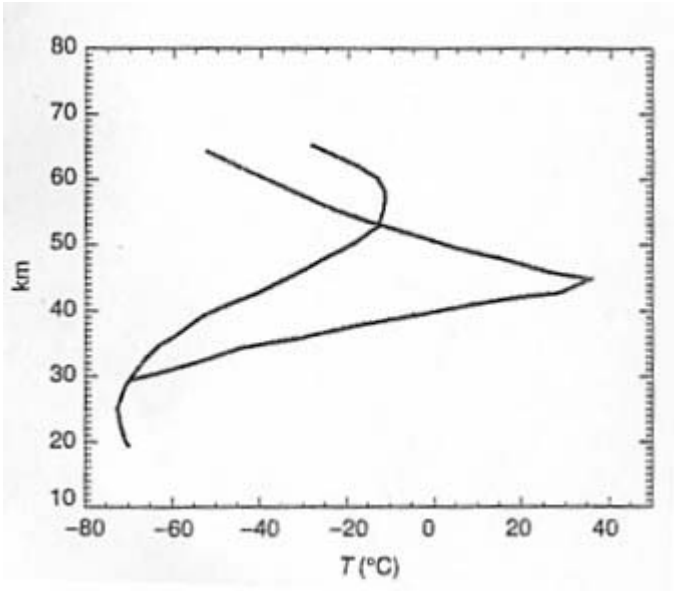

Figure 61 Typical rocketsonde observations before and after a major stratospheric warming. Taken from O'Neill (2003).

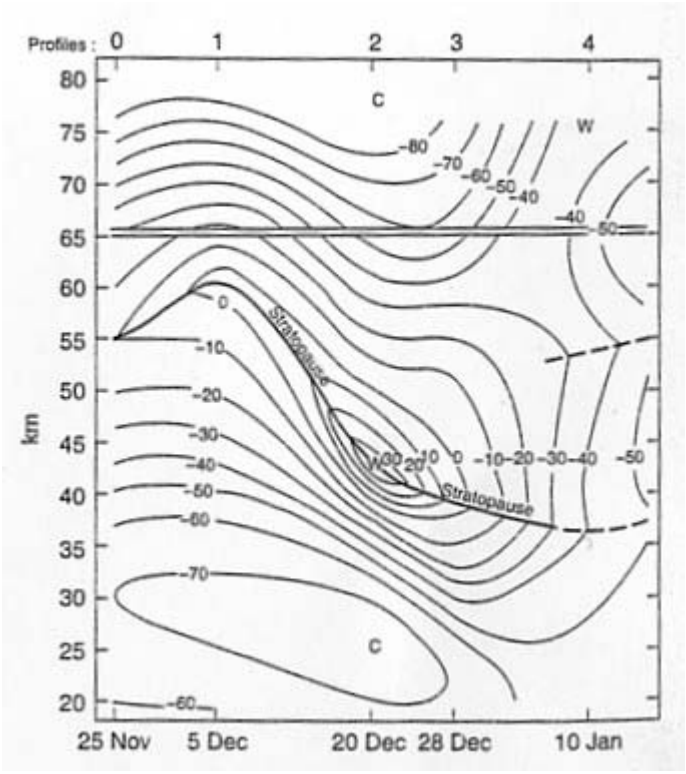

Figure 62 Schematic evolution of temperature at 60°N during the development of a major midwinter warming. Note the height of maximum temperature defining the stratopause descends with time. Taken from O'Neill(2003).

Final Warmings mark the transition from the cold cyclonic vortex in winter to the warm high centred on the pole in summer. Their intensity varies much and they can be divided into major and minor Final Warmings. The time when the Final Warmings take place - when the westerlies of winter are replaced by the easterlies of summer - also varies a good deal, so they are further divided into early and late Final Warmings. Naturally, there are also Final Warmings in the Southern Hemisphere.

In addition, major warmings may be separated into wave-1 type warmings and wave-2 type warmings. A classical example of a wave-1 type warming occurred in the northern winter of 1986/187. The Aleutain high has intensified, displacing the polar vortex and the associated cold pool of air from the pole. Such intensification is often the result of merger between the quasi-stationary Aleutian high and an eastward travelling anticyclone that forms to the west. Temperatures rise sharply in the strengthening jet stream between the polar vortex and the Aleutian high. This temperature rise is a result of adiabatic compression of the air as it descends (up to 2km) on entering the jet stream.

The most dramatic major warmings (wave-2 type warmings) involve a complete split in the polar vortex, followed by a rapid breakdown of one or both of the two cyclonic vortices that result from the split. This type of warming tends to occur less frequently than major warmings of the wave-1 type. An

unusually symmetrical example of a wave-2 major warming occurred in Northern winter 1984/1985.

Matsuno(1980) developed the theory of SSWs around a theoretical model of stratospheric dynamics that divides the circulation into zonal-mean and wave components. The work postulated that the mean flow changes observed during sudden wamings, including the deceleration in zonal mean wind and the mean temeperature rise near the pole, are attributable to the effects of vertically propagating planetary waves forced in the troposphere by large scale disturbances therein. Matsuno used a simplified numerical model to simulate the interaction between a single planetary wave and the zonal-mean flow. Wave amplitudes could be specified at a lower boundary, near the tropopause, so that the effects of forced transient planetary waves could be studied. Discovered that the interaction between travelling and stationary planetary waves can result in resonance leading to a strong and rapid increase of wave amplitude. If the amplitude gets too high the wave becomes unstable and breaks. Energy is transferred out of the wave field into the surrounding atmosphere and the amplitude decays quickly. Such a feature describes the stratospheric warming event.

The strong association between SSWs and planetary waves can also be used to explain the lack of major warmings in midwinter in the Southern hemisphere. Planetary waves are generated by mountains and by contrasts between the temperatures of land and sea. Because much of the Earth's surface in the Southern hemisphere is oceanic, such waves are weaker than in the Northern hemisphere. The polar vortex in the Southern hemisphere is correspondingly stronger and more resilient to short-term increases in planetary wave activity.

The 2002 Southern hemisphere Major Stratospheric warming

The southern hemisphere stratospheric winter of 2002 was the most unusual winter yet observed in the southern hemisphere climate record. Temperatures near the edge of the Antarctic polar vortex were considerably warmer than normal over the entire course of the winter. The polar night jet was considerably weaker than normal, and was displaced more poleward than has been observed in previous winters. These record high temperatures and weak jet resulted from a series of wave events that took place over the course of the winter. The first large event occurred on 15 May, and the final warming occurred on 25 October. The wave events tended to occur irregularly over the course of the winter, and pre-conditioned the polar night jet for an extremely large wave event on 22 September. This large wave event resulted in the first ever observed major stratospheric warming in the Southern hemisphere. This wave event split the Antarctic ozone hole. The combined effect of the wave events of the 2002 winter resulted in the smallest ozone hole observed since 1988.

During the fall of 2002, the Antarctic ozone hole was unusually disturbed. First, the ozone hole was considerably smaller than has been observed during the early September period. Second, the ozone hole split into two parts on 22 September 2002 (see Figure 63). It has been shown that the reason for this unusual behaviour was not a result of chlorine chemistry, but because of the stratospheric temperature and dynamics that occurred over the winter period. There are two necessary conditions for causing the Antarctic ozone hole: high levels of halogens (specifically chlorine and bromine), and temperatures cold enough to form polar stratospheric clouds (PSCs). The impact of temperature on Antarctic ozone losses has always been considered to be of secondary importance because temperatures are always cold enough to form extensive PSCs, and the southern hemisphere (SH) has small interannual variability during winter. In contrast, Arctic temperatures are both warmer and more variable. This variability will occasionally results in large Arctic ozone losses (e.g., 1997), or virtually no ozone loss at all (e.g., 1999).

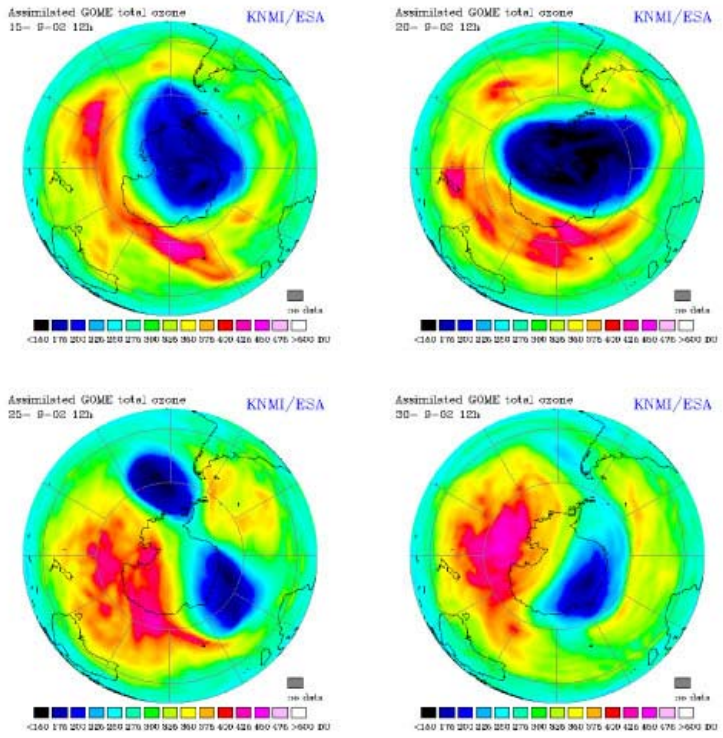

Figure 63 Time evolution of the total ozone field assimilated at KNMI: September 15, 20, 25, and 30 in 2002. Taken from Baldwin et al. (2003).

Figure 64 shows the daily evolution of three quantities which characterise the polar vortex from 1979 to 2002: temperature at the South Pole and geopotential height amplitudes of zonal wavenumber 1 and 2 in high latitudes, at the 20 hPa level. The polar temperature in late September 2002 was extremely high, indicating this is an unprecedented event at least in the past 24 years studied. Three minor warming events are also observed in late August and early to mid September (these can also be observed clearly in Figure 65 and Figure 66), which are comparable to that in 1988. These warming events are associated with amplification of planetary waves. The daily amplitude of wavenumber 1 is very large in August 2002, while that of wavenumber 2 is large in July and early September. It should be noted that note that the largest amplitudes of wave 1 and wave 2 ever observed occurred during the winter of 2002. Examination of the year-to-year variations of the monthly mean of the daily amplitudes of zonal wavenumber 1 and 2 also indicates that the planetary waves in the polar region are more active than normal through the winter 2002.

In order to discuss the rarity of the stratospheric sudden warming event in September in the Southern Hemisphere it should be noted that recent work by Taguchi and Yoden (2002) with a simple global circulation model investigated interannual variations of the troposphere-stratosphere coupled system based

on 1,000 year integrations. In a run corresponding to the Southern Hemisphere, a highly skewed distribution of the polar temperature in the upper stratosphere is obtained for the months from April to September. The probability distribution is far from the normal (Gaussian) distribution, and about 0.5 % of the 1,000 years are extremely warm with monthly temperature anomalies exceeding 6 standard deviations. It should also be noted that the possibility of major warmings in the Southern Hemisphere is also supported by some GCM experiments. However, prior to September 2002 the occasional major warmings in the Southern Hemisphere were considered a model "defect."

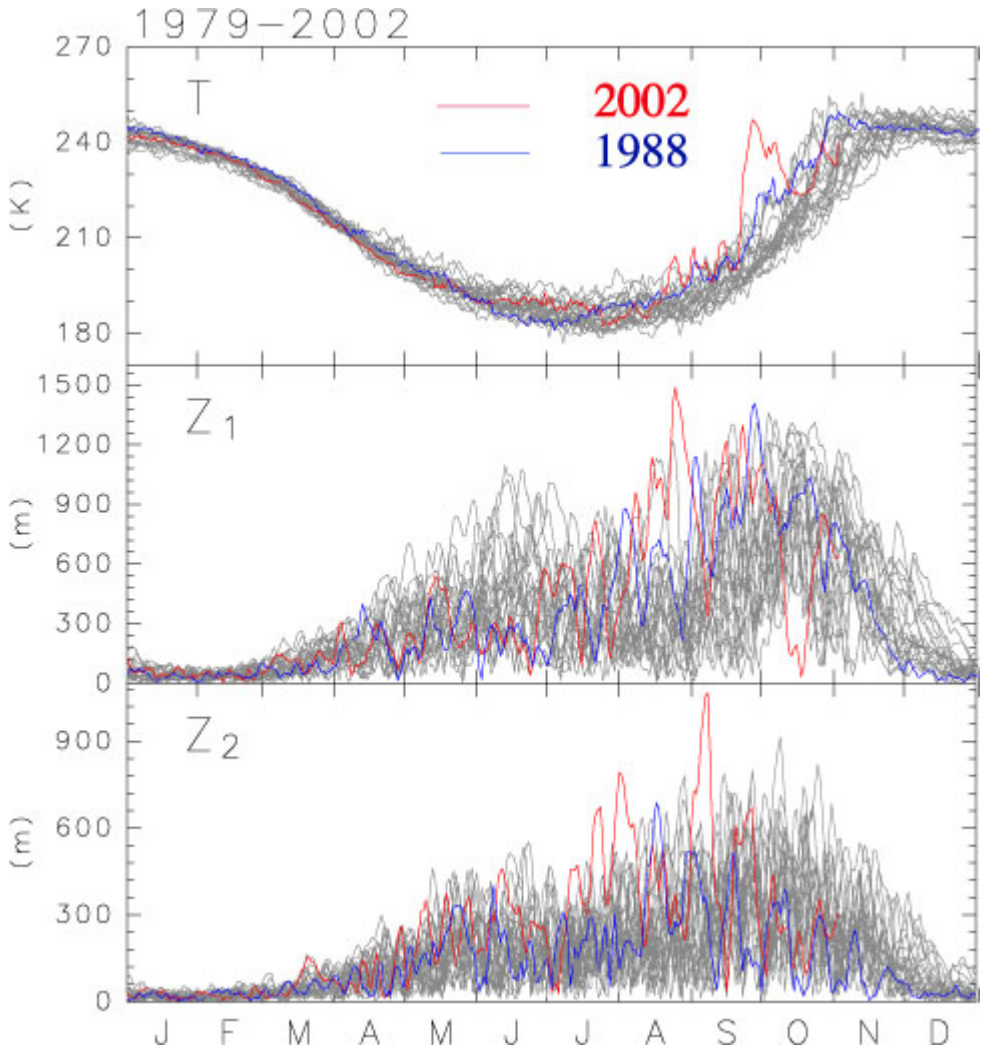

Figure 64 24-year records of the daily evolution of the polar vortex temperature at the South Pole (T), geopotential-height amplitudes of zonal wavenumber 1 (Z1) and 2 (Z2) averaged over the latitudes between 60°S and 70°S, at the 20 hPa level. Taken from Baldwin et al. (2003).

Figure 65 shows time-height sections of the mean zonal wind at 60°S for the period from July to October in 2002 and 2001. In 2002, we can see a regularly oscillatory change of the polar night jet with a typical time scale of about 10 days throughout the winter season prior to the sudden warming. Eventually,

the sudden warming occurred on 26 September, with winds reversing to easterly at 10 hPa, 60°S and a reversal of the temperature gradient between 60°S and the pole, meeting the WMO criteria for a major stratospheric warming.

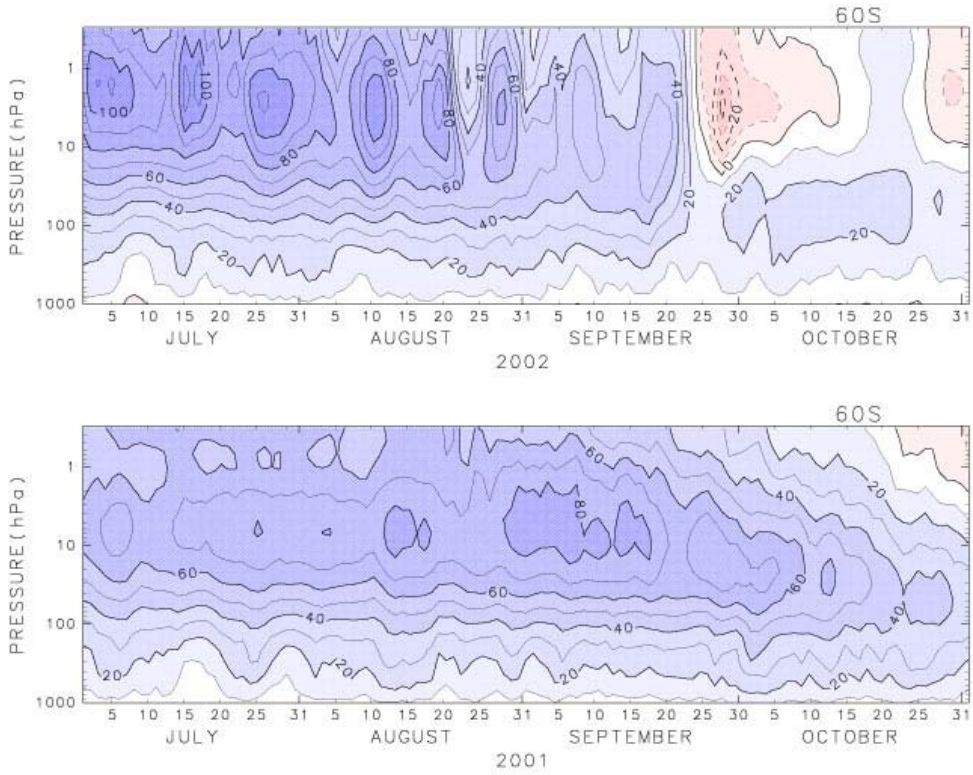

Figure 65 Time-height sections of the zonal mean zonal wind at 60°S for the period from July to October in 2002 (upper panel) and 2001 (lower panel). Taken from Baldwin et al. (2003).

The warming of the polar vortex collar region (displayed in Figure 66 (A)) is directly controlled by the planetary waves that propagate upward from the troposphere into the stratosphere. Figure 66 (B) displays a 2002 time series of mid-latitude eddy heat flux at 100 hPa for waves 1–3. This heat flux can be thought of as representing the vertical flux of wave energy. Over the course of the May–October period, 11 significant wave events can be identified. Vertical lines are superimposed between panels of Figure 66 to show the connection between the wave events and the warmings. The wave events are shown to be irregularly spaced with about a 1–3 week periodicity.

The waves not only impact temperature, but also the zonal-mean wind. Figure 66 (C) displays the height-time plot of the zonal-mean zonal wind difference between 2002 and a1979–2001 average in the 20-90^{o}S zone. This figure effectively represents the deviation of the average jet strength from climatology. Each wave noted in Figure 66 (B) impacts the strength of the jet in Figure 66 (C), with the 22 September wave completely reversing the westerlies to weak easterlies.

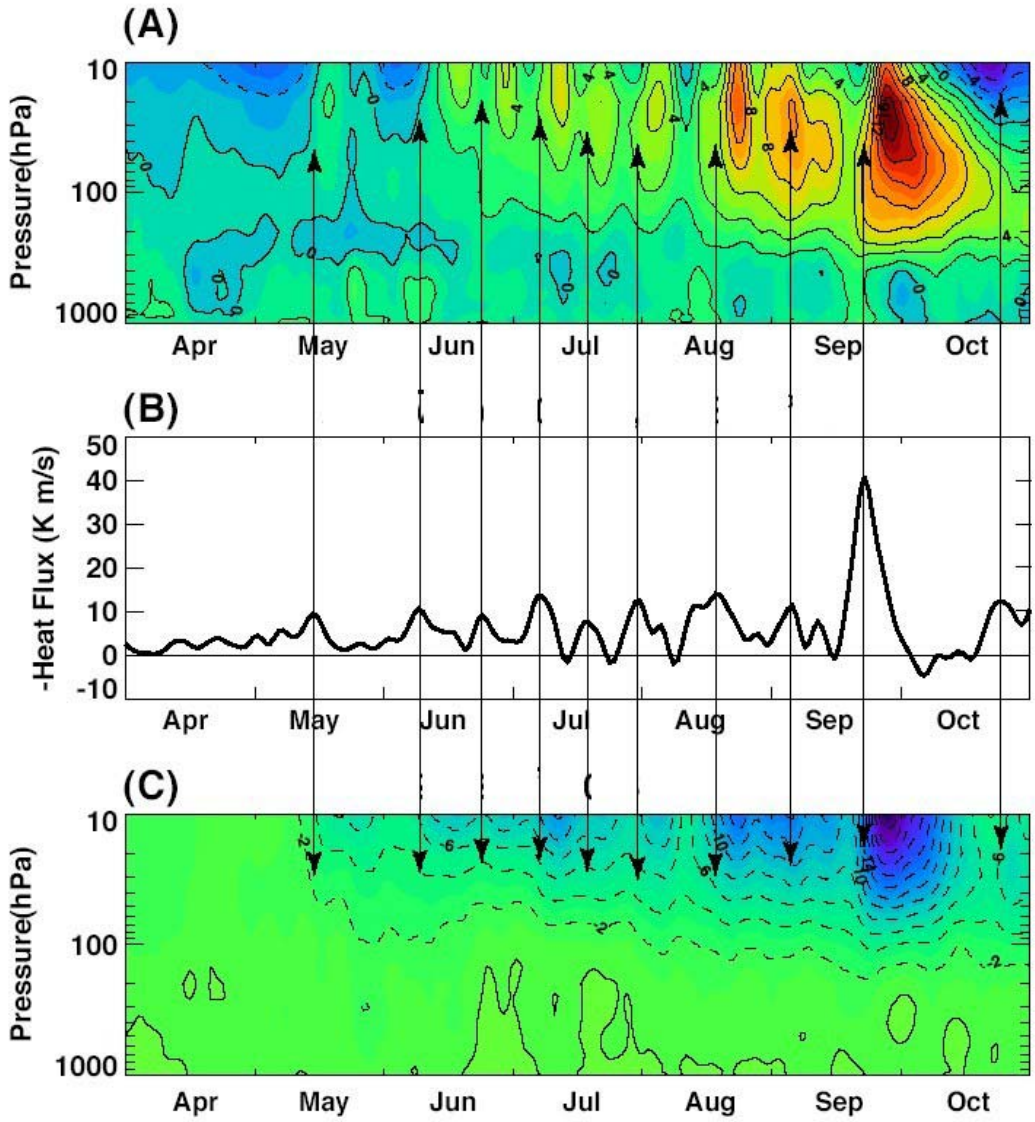

Figure 66 (A) Daily temperatures for 1 April to 31 October 2002, averaged for 55-75^{o}S at 50hPa. Contour intervals are 2 K. (B) Daily eddy heat flux at 100 hPa, averaged for 40-70^{o}S. (C) Daily zonal-mean zonal winds, averaged for 20–90^{o}S. Contour intervals are 2 m s^{-1} and negative contours are dashed. The vertical lines passing through all panels indicate the days of maximum eddy heat flux at 100 hPa - which corresponds to planetary wave activity. Taken from Newman and Nash (2003)