PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Satellite Measurements of Middle Atmosphere Temperature Structure

J. J. Barnett

Phil. Trans. R. Soc. Lond. A 1980 **296**, 41-57
doi: 10.1098/rsta.1980.0154

Phil. Trans. R. Soc. Lond. A **296**, 41–57 (1980) [41]
Printed in Great Britain

Satellite measurements of middle atmosphere temperature structure

By J. J. Barnett
Department of Atmospheric Physics, Clarendon Laboratory, Oxford, U.K.

Sixteen years have elapsed since the first satellite measurements of atmospheric temperature. These were observations of the lower stratosphere. Techniques have developed rapidly, and observations now extend from the surface to the mesopause. The instruments and techniques are briefly described and a review is given of the wide range of middle atmosphere research that has been based upon these measurements. The Nimbus 6 pressure modulator radiometer has made over 3 years' observations of upper stratospheric and mesospheric temperature, with weighting functions peaking at up to 80 km. The main results from this instrument and their relation to variations at lower levels are discussed. Temperature variations are generally smaller in the upper mesosphere than in the stratosphere. Planetary waves penetrate to this level in winter. There is a strong negative correlation between zonal mean temperature near the mesopause and in the upper stratosphere on both long and short timescales.

1. Spectroscopic remote sounding

Soon after the first artificial satellite was launched it was realized that measurements of infrared radiation emitted by the atmosphere could be used to deduce its temperature structure (Kaplan 1959). At wavelengths sufficiently long (> 2 μm) the radiant energy from the atmosphere is mainly thermal emission by atmospheric constituents, with little or no contributions from airglow, fluorescence or reflected solar radiation. For nadir observations from space, i.e. observations looking vertically down at a single wavelength at which just one gas has significant absorption, the radiation observed originates from a layer. Radiation from below this layer is absorbed by the atmosphere above, so little reaches space, while radiation from above the layer is emitted by so much smaller a quantity of gas because the density decreases so rapidly with height that it is insignificant. Where Lorentz line broadening is dominant, the variation of linewidth with pressure causes the emitting layer to be still thinner. This varying response to emission from different levels is represented by the weighting function K, such that the radiation observed, I, is given by

$$I = \int_0^\infty K(z)\, J(z)\, \mathrm{d}z, \tag{1}$$

where $J(z)$ is the source function and z is a height or pressure-like variable, most conveniently ln (pressure). Figure 1 gives nadir view weighting functions for various instruments viewing carbon dioxide emission at 15 μm. These weighting functions are integrated over a range of frequencies, since no measurement can be monochromatic, and so are broader than single frequency weighting functions. When given in pressure coordinates, weighting functions depend mainly upon the mixing ratio profile of the emitting gas, but also to a varying extent upon the temperature profile because of the temperature dependence of Lorentz and Doppler line broadening. When given in height coordinates the temperature dependence is greater because the hydrostatic equation introduces an additional component of variation. For carbon dioxide the mixing ratio is sufficiently constant up to the thermosphere, and the weighting functions are, to a good extent, fixed except for a little temperature dependence in the mesosphere.

Whereas the weighting function describes the pressure range of the observation, the measured intensity is determined by the source function at these pressures. When possible, measurements are made of spectral bands where there is local thermodynamic equilibrium between the translational energy states of gas molecules and the vibrational and rotational states responsible for the emission. In this case the source function equals the Planck function, which is a simple

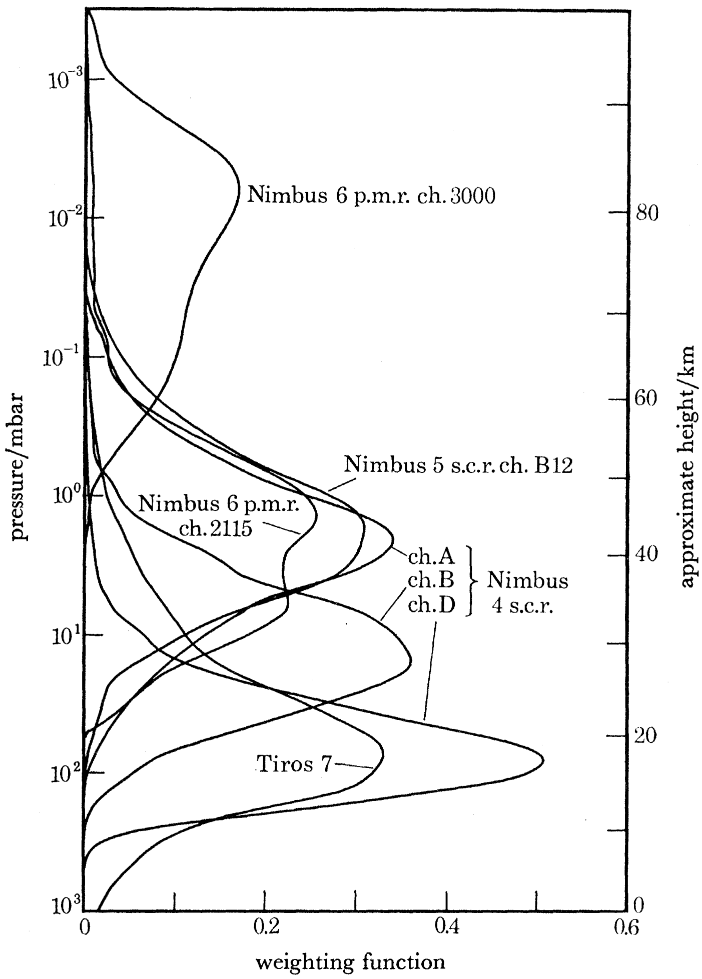

FIGURE 1. Weighting functions for various temperature sounders measuring infrared emission from carbon dioxide. The Tiros 7 weighting function is an average for views within 40° of nadir. The curves are functions of pressure; the height scale applies for a climatological mean temperature profile.

function of temperature and wavelength. However, at sufficiently low pressure (10^{-2} mbar† for the 15 μm v_2 carbon dioxide band), collisions between molecules are too infrequent to maintain local thermodynamic equilibrium and the source function is less than the Planck function.

Thus with a gas whose mixing ratio is known, it is possible to determine source function, and hence temperature, for various layers by measuring at appropriate wavelengths. This is the basis of temperature sounding in most instruments launched so far. Alternatively, given the temperature profile, observations of a gas at various wavelengths can be used to determine the vertical distribution of that gas except in regions where it is isothermal.

† 1 mbar = 10^{-2} Pa.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

2. Temperature sounding instruments

The first instrument to use these principles was the medium resolution infrared radiometer (m.r.i.r.) carried by the Tiros 7 satellite launched in 1963. One channel measured emission from carbon dioxide in the 14.8–15 μm region. The satellite was spinning; this caused the instrument to scan the Earth at varying angles to the nadir. By averaging observations within 40° of the nadir over regions about 5° latitude by 5° longitude, and over a few days, almost complete analyses were obtained for an emitting layer in the lower stratosphere (see figure 1). With this instrument many important and now well known phenomena were seen for the first time in such detail, and they were described by Kennedy & Nordberg (1967) in what is surely a classic paper.

The m.r.i.r. was carried by several subsequent Tiros satellites, and an improved version but with a broader filter and a slightly lower weighting function was carried by Nimbus 2 in 1967 (Nordberg *et al.* 1967). The Nimbus series of satellites are more convenient than the early Tiros were, having three-axis stabilization relative to the vertical and the direction of travel. Orbital heights are normally about 1000 km. The orbits are Sun-synchronous, passing over the tropics and mid-latitudes near local noon and midnight, and reach 80° N and 80° S. Orbital periods are about 106 min. Each successive orbit is 26° longitude further west, and as the Earth rotates beneath the orbital path, the satellite passes within 13° longitude of each place twice per day. The data are recorded within the satellite and replayed at high speed when passing ground stations, so that data for much of the globe are routinely obtained. The recorders normally degrade faster than other parts of the satellite, and after about a year of operation, coverage starts to deteriorate. Nevertheless, Nimbus satellites have an excellent record of long life: Nimbus 4 launched in 1970 is still used, and Nimbus 5 launched in 1972 can still give almost complete coverage each day.

The m.r.i.rs had only one carbon dioxide channel. It is fortuitous for stratospheric research that this, the simplest temperature sounder to construct, which measures a broad part of the 15 μm carbon dioxide emission, receives most of its radiation from the lower stratosphere. Developments since then fall into three main categories:

(*a*) improvement of spectral resolution, by several orders of magnitude in some cases, to allow a choice of the height observed, from the surface to the mesopause;

(*b*) finer horizontal resolution, which is particularly important in the troposphere where it is often necessary to view through gaps between clouds;

(*c*) improved accuracy.

The satellite infrared spectrometer (s.i.r.s.) carried by Nimbus 3 (1969) and Nimbus 4 (1970) was the first instrument to have channels sounding at more than one level and so be able to resolve vertical structure. The instrument used eight channels 5 cm^{-1} wide in the 15 μm (668 cm^{-1}) carbon dioxide band. The highest two channels measure emission from layers 15–30 km thick covering the middle and lower stratosphere. More elaborate instruments with approximately the same weighting functions, but using filters, were carried on later satellites, namely the v.t.p.r. on the N.O.A.A. 2 to N.O.A.A. 5 series of satellites (1972–6), the i.t.p.r. on Nimbus 5 (1972) and the h.i.r.s. on Nimbus 6 (1975). These instruments were intended primarily for tropospheric sounding, the stratospheric weighting functions being necessary to eliminate from the tropospheric weighting functions any contribution from the stratosphere. However, the data are used extensively for stratospheric studies, including operational monitoring of sudden winter warming events, by the Upper Air Branch of the U.S. National Oceanic and Atmospheric Administration.

The main problem of sounding above the troposphere is obtaining sufficient spectral resolution. Conventional infrared spectrometers or filter radiometers can give broad weighting functions in the mid and low stratosphere, but nothing higher, since spectral lines becoming narrower and non-overlapping cause transmission to vary much more rapidly with wavelength. Furthermore, the energy available in any sufficiently small spectral interval would be so small as to require long observation times. These problems are overcome by a method used by Oxford and Heriot-Watt Universities in the selective chopper radiometer (s.c.r.) on Nimbus 4 and extensively developed by them since. The radiation is passed through a filter to select radiation only from the 15 μm carbon dioxide band, and is then passed through a cell containing carbon dioxide before falling onto the detector. The carbon dioxide acts as a filter to attenuate radiation near the centre of each line, the overall attenuation depending upon the cell length and carbon dioxide pressure. The Nimbus 4 s.c.r. selected the radiation by chopping the beam between a cell containing carbon dioxide and an identical but empty cell, the difference signal being the intensity near the line centres summed over all of the lines. This gave a weighting function peaking at about 2 mbar (43 km) with a 15 km half-width. Another channel chopped between two cells containing carbon dioxide at different pressures to obtain the radiation in narrow intervals either side of each line, and had a weighting function peaking near 20 mbar. The Nimbus 5 s.c.r. (1972) used the same basic method but, instead of chopping between cells, the radiation was measured in turn through four identical cells, one being empty and the other containing carbon dioxide at different pressures. By subtracting radiances during data analysis, four different weighting functions were obtained, peaking at levels up to 2 mbar. These signals typically differed by only 10%, so care was necessary to obtain an accurate difference signal, and for this reason these designs were unsuitable for higher weighting functions where much smaller differences would occur. The main problems arise from lack of 'balance' between the two paths because of (*a*) inevitable small differences between cells, and (*b*) the imprecision of chopping. These problems were overcome by using one cell containing carbon dioxide whose pressure could be modulated by a piston. This technique is called pressure modulation and was first used in the pressure modulator radiometer (p.m.r.) carried by Nimbus 6 (1975). A refinement was to include a molecular sieve and heater in the cell to allow the mean pressure to be varied if necessary to obtain various mean weighting function heights from a single cell. The p.m.r. has a top weighting function peak near 80 km (see figure 1), for which the spectral resolution is about 1 part in 10^6. A potential problem of sounding at such heights is that the Doppler shift due to just 5% of the satellite velocity is comparable with the Doppler line width. Thus at heights where Doppler broadening is dominant (in this case above about 50 km), moving the direction of view away from nadir along the direction of travel moves the atmospheric emission spectrum relative to the lines in the cell, causing the weighting function to move down. This was used to advantage on the Nimbus 6 p.m.r., where the instrument scans to 15° either side of nadir along the direction of travel, causing the weighting functions to scan vertically. A similar method, without Doppler scanning, is used on the Tiros and N.O.A.A. series of operational satellites, the first being carried by Tiros N (launched in 1978). Three pressure modulator cells have weighting functions peaking in the stratosphere. A similar instrument is carried by the Pioneer Venus orbiter.

In the instruments described so far, selection of the height of observation is carried out spectroscopically. The limb sounder views the limb of the Earth with a very narrow field of view and the height viewed is determined geometrically, most of the radiation originating near the tangent height, with a weighting function that may be as thin as 5 km. However, the height viewed must be known accurately, since an error of 1 km (0.02° angular deviation for a Nimbus orbit where the horizon is 3000 km away) leads to a radiance error of up to 15 % for viewing carbon dioxide at 15 μm. Fortunately, the viewing level can be determined sufficiently accurately from radiance measurements alone, but there remains a difficult requirement that the spacecraft remains stable to seconds of arc during a complete observation. The method has two main advantages:

(*a*) thinner weighting functions can be obtained than with nadir sounding;

(*b*) constituents can be sounded more precisely and unambiguously, partly because of the greater optical thickness of the atmosphere for a limb path.

Three limb sounders have been carried on satellites: the Nimbus 6 l.r.i.r., the Nimbus 7 l.i.m.s. (both basically the same design) and the Nimbus 7 s.a.m.s. which uses pressure modulation. They each measure temperature in the stratosphere and mesosphere and the concentrations of various constituents. They are fully described elsewhere in this volume, and so will not be referred to here again.

Data from other instruments have been used for stratospheric studies. The Nimbus 5 n.e.m.s. and Nimbus 6 s.c.a.m.s. sounders measured microwave thermal emission from oxygen, and thereby obtained narrow (*ca.* 10 km wide) weighting functions in the troposphere and lower stratosphere. The Eole constant density balloon experiment (1971–2) provided temperature and wind fields for the Southern Hemisphere at about 200 mbar. A large number of balloons were tracked by satellite, which was used to relay their measurements and to determine their velocities. These data have been used in several stratospheric studies since they provide a reference from which winds in the stratosphere can be determined. A similar experiment known as TWERLE was performed with Nimbus 6.

3. Interpretation of data

One of the shortcomings of satellite temperature sounding is undoubtedly the limited vertical resolution. With a vertical sounding instrument the weighting functions have a half-width of 10 km or more, and although measurements from different channels can be linearly combined to give a narrower weighting function, a width of about 10 km is a practical limit in the middle atmosphere. Thus disturbances with a vertical wavelength of 20 km or less are severely attenuated or are invisible. Fortunately the gross features of the stratospheric and mesospheric temperature structure, i.e. zonal mean variations and planetary waves, have vertical distance scales greater than 10 km. However, some wave motions that determine these features have wavelengths that are too short to resolve, including gravity waves which may be important in the energy, heat and momentum budgets above the middle mesosphere, and equatorial modes which are believed to produce the quasi-biennial oscillation.

Approaches to data interpretation fall into three categories:

(*a*) direct use of radiance data, possibly converted to brightness temperature by using the inverse Planck function;

(*b*) derivation of thickness for layers well described by the measurements;

(*c*) derivation of temperature profiles as continuous functions of height by a process known as retrieval.

There are several methods of temperature retrieval. Insufficient information is measured to determine the temperature at each level which might be desired, so there is an infinity of temperature profiles that will fit a given set of observations, each differing in detail over a distance scale less than the resolution and also in ways allowed by the measurement uncertainty. Retrieval methods attempt to derive the most reasonable profile, although different retrievals from the same data may differ substantially, depending upon how reasonableness is built into the method. A common retrieval method is to determine the most probable profile, given the climatological mean and covariance of deviations from the mean for the appropriate latitude, day of the year, and the probable error in the measurements. The solution is given by a simple linear combination of measured radiances. Mathematically this method is similar to a linear regression between temperature and the observed radiance. A review of retrieval methods was given by Rodgers (1976*b*).

Temperature retrievals are almost essential for some applications such as budget calculations, but they suffer from the problem that more information is present than is actually measured. Thus it is never clear whether medium scale features on the borderline of the vertical resolution are present in the measurement or are artefacts of the retrieval process or originate from other information used by it, such as atmospheric statistics. Hence where temperature retrievals are not essential, radiances or thicknesses are often preferred because they display no more information than is measured. Thickness may be determined by the same type of retrieval methods as temperature. The simplest method is to regress, for a wide range of temperature profiles, thickness against radiance or brightness temperature measured by one channel whose weighting function averages over approximately the same layer. A high correlation usually results, so allowing radiance analyses to be relabelled as thickness for this layer. A more sophisticated approach is to use multilinear regression with several channels, and possibly different coefficients for different seasons and latitude bands.

One of the great advantages of satellite data is their completeness and uniformity of data coverage. This greatly simplifies the mapping task and there is no vital need for objective analysis methods able to assimilate data for arbitrary places, although such methods are frequently used. In Oxford we have used linear interpolation along the orbit for filling small gaps and zonally between adjacent orbits but not over larger gaps, to obtain data on a grid of intervals 10° longitude and 4° latitude. This is done separately for north- and south-bound parts of orbits and the two are combined yielding as a final product a grid which may still have gaps. Only where gaps are unacceptable, such as when zonal Fourier components are to be derived, are they filled by linear interpolation and then only if sufficiently small. Temperature structures observed by satellites in the stratosphere and mesosphere have long distance scales such that they are adequately represented by the lowest four zonal waves, possibly because shorter-scale horizontal variations have vertical scales too small to be observed. Temperature fields can change substantially during a day, but rarely is it necessary to use analyses at more than daily intervals. However, it is frequently a cause of concern that simple analyses as described above are not for any single synoptic time, so that in principle there should be a discontinuity representing the change during a day along the line between the first and last orbits of the day. Rodgers (1977*b*) overcomes this problem by using Kalman filtering to give statistically rigorous time interpolation and smoothing with a long data series, and thereby obtains analyses at any desired synoptic times.

4. Accuracy of measurements

The problems of calibrating an infrared radiometer have been discussed by Houghton (1977). Errors fall into four main categories:

(*a*) Errors that are random from one observation to the next, usually detector noise; although they limit the accuracy they are readily measured and can be taken into account.

(*b*) Errors that vary more slowly but cannot be corrected. For example, it may be necessary to correct calibration parameters for the temperature of various parts of the instrument, but this is likely to fail where rapid temperature changes cause thermal gradients across those parts; since the pattern of temperature variation is likely to repeat each orbit, this problem is liable to produce a pattern of errors which also repeats from orbit to orbit.

(*c*) Constant or slowly varying errors such as those caused by imperfect mirrors, black bodies or filters, or stray radiation, or errors in transmission data used to calculate weighting functions. Pre-launch testing should provide information to eliminate these errors, but there are bound to be residual components.

(*d*) Errors that arise from numerical short cuts when processing the data, such as taking a spectral interval to be monochromatic when integration over a range of wavelengths would be desirable, or neglecting the effect of weighting function variation with temperature, or non-local thermodynamic equilibrium effects. Most of these sources of error can, in principle, be made as small as desired, so there results a compromise between the effort required and the resulting gain of accuracy.

Observation of changes of atmospheric temperature from place to place are of much greater importance than measurement of absolute temperature. For observations from a single instrument, constant errors (category (*c*)) cancel when taking temperature differences from place to place at the same pressure and to some extent when comparing measurements of different pressure levels. However, this does not mean that such errors can be ignored because different instruments, even when using the same basic techniques, may have quite different forms of error and results can only be combined if they have absolute accuracy or are adjusted to be compatible.

Detailed comparisons between s.c.r. and rocket measurements were given by Barnett *et al.* (1975). For the Nimbus 5 s.c.r. they found mean differences between the radiances observed by the s.c.r. and those calculated from coincident rocket temperature profiles equivalent to about 2 K in the upper stratosphere and less than 1 K in the lower stratosphere. Standard deviations from these mean differences were of similar size, and correlation coefficients between the two forms of measurement were typically 0.95. The set of comparisons included a variety of stations and seasons. An indication of the random error was given by the standard deviation of radiance at one location (0° N, 0° E) during July 1973. This is composed of atmospheric plus instrumental variation, so gives an upper limit to either. Values corresponded to approximately 0.1 K for a channel measuring in the lower stratosphere, rising by an order of magnitude towards the upper stratosphere. Such accuracy is impressive and shows that the instruments can provide excellent observations of waves; however, it should be remembered that retrieval methods are able to improve vertical resolution by combining weighting functions, but at the expense of noise, so that final retrievals may be considerably less accurate than this. Miller in an accompanying paper gives an account of the errors in Tiros N s.s.u. measurements.

5. Studies using satellite data

Satellites have extended our measurement of atmospheric temperature in three main ways:

(*a*) Daily radiosonde analyses were previously possible up to 10 mbar for the Northern Hemisphere and weekly analyses up to 0.2 mbar, although the latter depended upon a network of rocket stations which is very sparse. Satellites now give us complete daily analyses throughout the middle atmosphere.

(*b*) Satellites give us complete Southern Hemisphere coverage of the middle atmosphere where previously there were just radiosonde observations at relatively few stations on land.

(*c*) Satellites provide homogeneous data from the same instrument, so allowing waves of as little as 0.3 K amplitude to be detected.

Most studies have dealt with the winter disturbed stratosphere and mesosphere, which is dominated by planetary waves and stratospheric warmings.

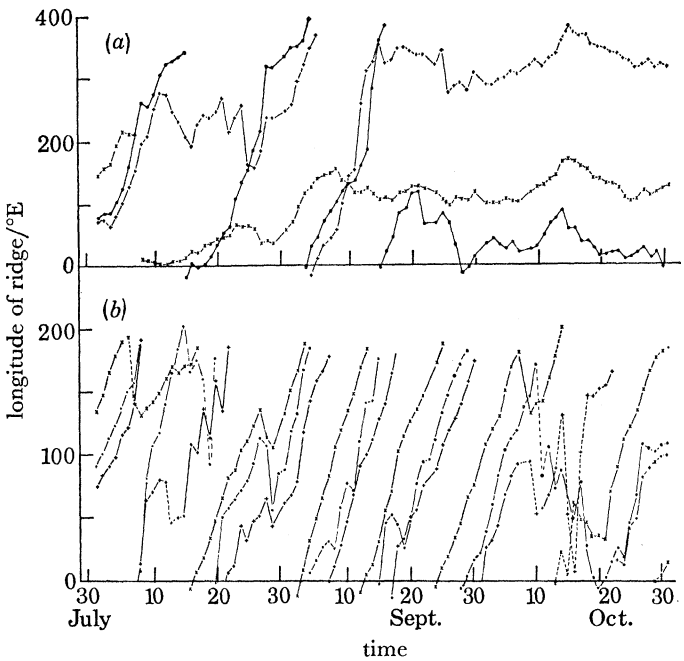

Figure 2. Phases (longitude of the ridge) of temperature waves at 52° S during 1971 measured by the Nimbus 4 s.c.r. +, channel A; •, channel B; ×, channel D (see figure 1 for their weighting functions). (*a*) wavenumber one – mainly stationary with a westward tilt; (*b*) wavenumber two – eastward travelling waves with a westward tilt with height. From Harwood (1974).

5.1. *Planetary waves*

The general behaviour of planetary waves has been investigated by several workers. Harwood (1975) studied the 1971 Southern Hemisphere winter and observed a marked wavenumber 2 pattern of up to 12 K amplitude in the stratosphere, which rotated uniformly eastwards about the pole through five cycles in 3 months (figure 2). He also noted an inverse relation between zonal and eddy available potential energy in the lower stratosphere. During this time wavenumber 1 phase was relatively stationary. Leovy & Webster (1976) studied the same data set but also Eole derived temperatures in the upper troposphere, and found excellent coherence between the 200 mbar level and the upper stratosphere. They noted that wavenumber 2 amplitudes have their peak in the late winter in both hemispheres, and that, at least in the Northern Hemisphere, this follows tropospheric behaviour. They asked what determines the meridional extent of planetary waves and pointed out that, although it seems to follow the strength of the zonal mean westerly wind field, the eddies are more alike in the two hemispheres during their respective winters than the wind fields, suggesting that another mechanism, possibly photochemical damping at low latitudes, might be important. Deland (1973, 1977) studied s.i.r.s. and s.c.r. data and found westward travelling planetary waves in the tropics and mid-latitudes, apparently coupled and forced from outside the tropics, and eastward travelling

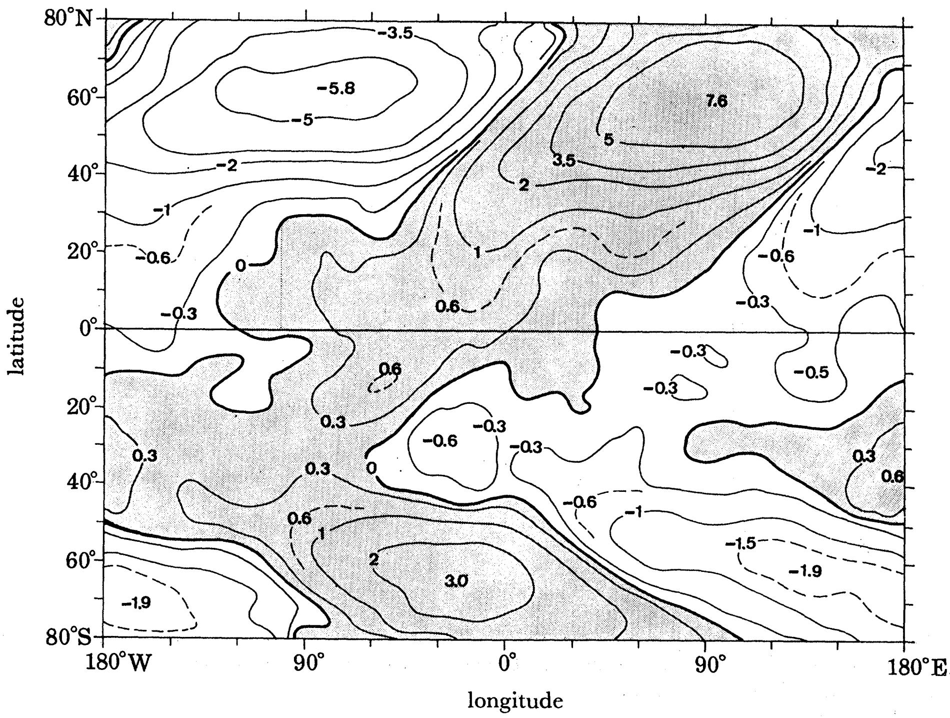

FIGURE 3. Deviation of Nimbus 5 s.c.r. channel B12 brightness temperature (/K) from the zonal mean averaged over 9–23 March 1973. Regions of positive deviation are dotted. Large stationary waves, mainly of wave-number one, were present in both hemispheres, but that of the Northern Hemisphere appeared to penetrate into the Southern as indicated by the westward tilt with latitude. From Barnett (1975).

waves in the Southern Hemisphere winter. He was able to separate them into two classes, one clearly that seen by Harwood, Leovy and Webster (but for different years) which tilted westward with increasing height suggesting forcing from below, and a class of faster moving waves which tilted eastward with height, suggesting downward energy propagation. The temperature and geopotential height structure of Southern Hemisphere travelling waves have also been described by Hartmann (1976) in an extensive study using a data set retrieved by N.O.A.A. combining the Nimbus 5 n.e.m.s., s.c.r. and i.t.p.r. instruments. Similar results were obtained by Stanford & Dunkerton (1978), although their method of analysis differed by using spherical harmonics.

Hirota (1976, 1978) has made extensive use of s.c.r. and p.m.r. data to study travelling waves. He has shown that transient planetary waves penetrate into the tropics near the equinoxes (although amplitudes are less than a degree), a result similar to a finding by Barnett (1975*a*) that stationary planetary waves appear to propagate across the equator at such times to mid-latitudes of the other hemisphere (see figure 3). It is not clear whether this behaviour relates to the semiannual zonal wind oscillation in the tropics, and which is the controlling process. Hirota (1976) used power spectral analysis to obtain the spectrum of phase velocities for wave-number one during various seasons. This gave convincing support to the simple theory that planetary waves can only propagate westwards relative to the zonal flow (see figure 4).

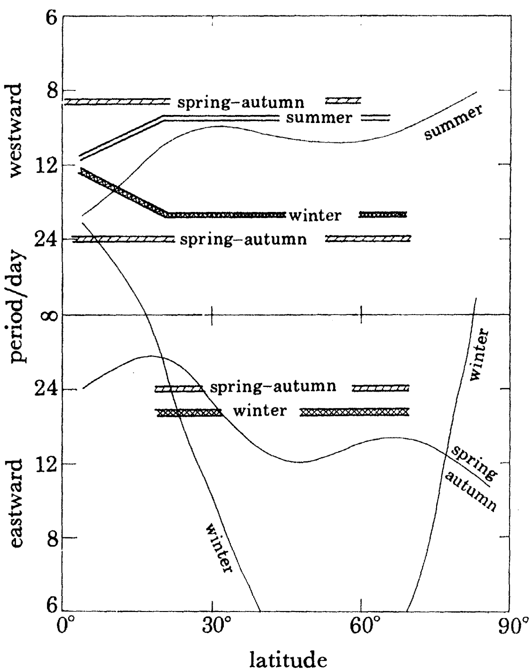

FIGURE 4. Schematic diagram of seasonal characteristics of transient waves for wavenumber one observed by Nimbus 5 s.c.r. channel B12. Thin full lines give the zonal wind velocity at 45 km height. Bold horizontal lines mark regions of high wave amplitude, which are seen always to be at phase velocities westward relative to the zonal flow. (From Hirota 1976.)

Peckham and colleagues at Heriot-Watt University have also made extensive studies of travelling waves using spectral analysis methods with s.c.r. data (see, for example, Chapman *et al.* 1974). Their work is left to Dr Peckham to describe in the accompanying paper.

Rodgers (1976*a*) discovered, by using s.c.r. data, westward travelling wavenumber 1 waves of global extent, period about 5 days, amplitude 0.2–0.6 K and symmetric about the equator with phase independent of latitude. This he identified as a global Rossby mode which had been predicted theoretically and observed in surface pressure data.

Stationary planetary waves have been studied by Labitzke (1977) and many other authors. When averaged over a month or more the temperature field shows little or no eddy activity during summer. The winter field is mainly composed of wavenumber 1, which is in a similar phase for different months and different years. The phase tilts westwards with increasing height, having a vertical wavelength of about 100 km, and towards the equator, with a change of 180° between low and high latitudes in the stratosphere. This is illustrated by figure 2, where wave-number 1 was fairly stationary for the whole period, and tilted westward with height, while there was little stationary component to wavenumber 2.

5.2. *Stratospheric and mesospheric warmings*

Planetary waves vary in amplitude by an order of magnitude during winter, and sometimes large amplifications are closely followed by rises of the polar cap temperature (an alternative description is that the hot centre or centres of the wave move to the pole). In the Southern Hemisphere these perturbations rarely exceed a few degrees and occur mainly after midwinter, but in the Northern Hemisphere at the stratopause, polar temperature sometimes rises by more than 50 K accompanying planetary waves of 100 K peak to peak. Events of this magnitude are described as major stratospheric warmings. For a few days winter polar temperature may exceed that of midsummer, and the normal winter westerly mean flow is replaced by easterlies. They occur in late December or January, in which case they are midwinter warmings, and the high polar temperature is followed by rapid cooling and very low temperatures, or they occur at the end of winter (March) in which case the circulation and temperature fields remain summerlike. These features can be seen on figure 5 (I). They are very dramatic events and have received much study with satellite and conventional data, notably by Labitzke and colleagues in Berlin and the Upper Air Branch of N.O.A.A.

Fritz & Soules (1972) used s.i.r.s. data to study the winter behaviour and noted cooling in the tropics and summer hemisphere during warmings at the winter pole. Barnett (1974, 1975*b*) saw similar behaviour at higher levels including a notable example for the Southern Hemisphere where the largest zonal mean temperature rise was at mid-latitudes, and the cooling was visible to almost the North Pole. Quiroz (1975) gave a detailed account of the movement and development of temperature anomalies during four winters with the use of s.i.r.s. and v.t.p.r. data. The major warmings during this period seemed to arise from a superposition of a standing wave with a travelling wave. Although links between winter activity and the tropical quasi-biennial oscillation had been suspected, none was evident in these data. Labitzke (1972, 1974, 1977) has made many studies of the winter stratosphere and mesosphere, by using all types of data including s.i.r.s. and s.c.r. She has placed much emphasis on comparison between the hemispheres, on timescales ranging from a few days to the seasonal, and by amassing data from many years has given the climatology of winter behaviour.

5.3. *Seasonal variations and climatology*

This section covers variations on a monthly or longer time-scale, which, although often studied for only a single year, can be reasonably taken to show the climatology. Professor Labitzke's paper in this volume should also be referred to.

Studies of the annual and semiannual cycles have been made by Fritz & Soules (1972) and Barnett (1974), both using 1 year of data, and McGregor & Chapman (1978) using 2 years. Each study used data from a different time and a different instrument, but the major findings were in agreement. The annual cycle has amplitude maxima at the poles near the stratopause of about 20 K in the Northern Hemisphere and 30 K in the Southern Hemisphere. The times of maxima are close to the summer solstice. The Southern Hemisphere amplitude is larger, probably because the eccentricity of the Earth orbit around the Sun causes the Southern Hemisphere to experience colder winters and warmer summers than the Northern, and because the Northern Hemisphere has a warmer winter because of larger stratospheric warmings. Small separate maxima of the annual wave of about 2 K amplitude occur near the Equator in the tropical lower and upper stratosphere. The mean temperature of the whole planet for

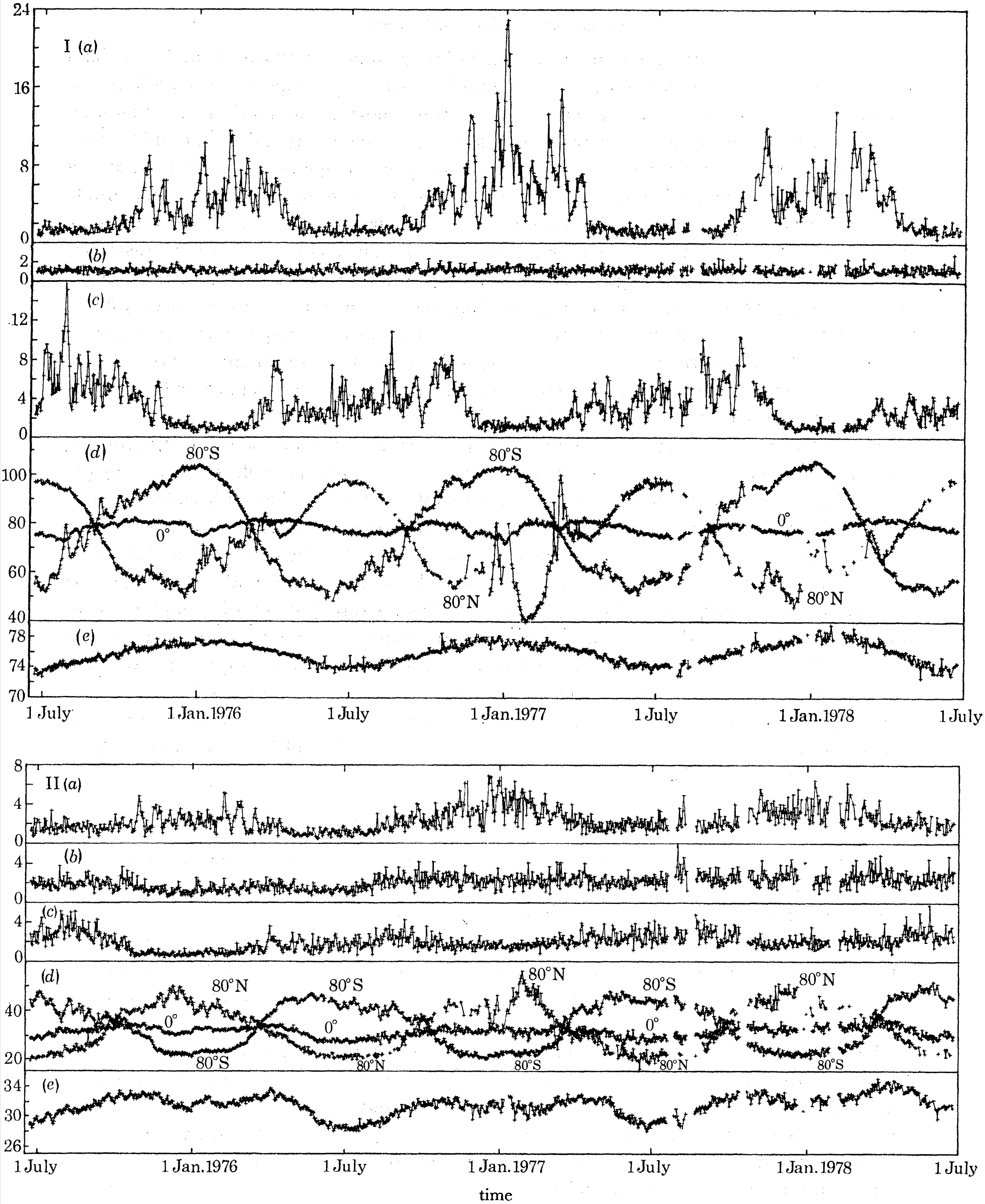

FIGURE 5. Results from the Nimbus 6 p.m.r. for 21 June 1975–22 June 1978. I. channel 2115 (near 45 km – see figure 1 for weighting functions); II. channel 3000 (near 80 km). (*a*), (*b*), (*c*) Standard deviation of radiance around the globe at 64° N, the equator, 64° S respectively; (*d*) zonal mean around globe at 80° N, equator, 80° S; (*e*) global mean.

stratospheric layers undergoes a substantial annual variation, apparently because of the ellipticity of the Earth orbit about the Sun (see figure 5). Insolation is about 7 % greater in December than June; whereas the global mean has this phase, the variation is less than expected from very simple considerations, while a detailed study does not appear to have been published. For the semiannual wave there are amplitude maxima near the pole which apparently are just a result of midwinter warmings giving a warm period in the middle of winter. For this reason the semiannual wave can be expected to vary substantially from year to year. There is also a pronounced 4 K amplitude maximum in the tropical upper stratosphere which dominates the temperature variation there. Maxima occur near the equinoxes, and being out of phase with the polar cycles, the cycle may be forced dynamically from the winter hemisphere through the out-of-phase relationships discussed in § 5.2. The twice yearly passage of the Sun overhead is in correct phase and may also play a part. Fritz (1974) discusses these mechanisms in detail for the lower stratosphere.

Crane (1979*b*) has extended these analyses to the mesopause by using p.m.r. data for 18 months, some of which are shown in figure 5. The annual cycle near 80 km is much as below except for a 180° phase change in mid and high latitudes (maximum temperature in mid-winter), as is well known. The global mean in the upper mesosphere is in phase with lower levels and with the solar forcing. The semiannual wave dominates in the tropics (2 K amplitude) with maxima near the equinoxes, but an important difference is that the semi-annual wave has this phase at all latitudes; consequently the global mean has a strong semi-annual component.

Retrieved zonal mean cross sections have been given by various authors (usually for individual days and not seasonal means) e.g. Barnett (1974), Rodgers (1977*a*), Labitzke & Barnett (1979). They show substantial deviations from the climatologies based on rocket data, and substantial differences between the hemispheres when data six months apart are compared.

5.4. *Budget calculations*

Quite distinct from work described so far are calculations from retrieved temperature fields of such quantities as meridional and vertical wind velocities, eddy fluxes, energy transformations, etc. Several studies have used the measurements in this way and in some cases have attempted to derive substantially complete budgets of energy, heat or momentum. These studies require high data quality, since extensive manipulation of the temperature field is required, and for some products the errors are greatly magnified. Errors are difficult to assess even for an experimenter who has intimate knowledge of the instrument design, operation, calibration, retrieval and final calculations, so error bounds are rarely given and the reader must assess the reliability of a study by comparison with others. A height analysis is normally required at one level to enable the geopotential field to be derived from the temperature field. A lower stratospheric height is often used, but in the Southern Hemisphere the only sufficiently complete analysis may be at the surface and tropospheric retrievals must be used, despite problems caused by cloud, to build up to greater heights.

Adler (1975) calculated meridional circulations in the lower stratosphere of both hemispheres by using s.i.r.s. data. Northern Hemisphere measurements were compared with fields derived by radiosondes and this gave credence to the new Southern Hemisphere results. A 1000 mbar height field was used as a basis for the geopotential field. There were significant hemispheric differences, the most notable being that the centre of ascending motion in the

Southern Hemisphere was at about 70° S compared with at the pole in the Northern Hemisphere.

Hartmann (1976, 1977) used a data set derived by N.O.A.A. from the Nimbus 5 n.e.m.s., s.i.r.s. and i.t.p.r., and made extensive calculations of various fluxes including potential vorticity, with special emphasis on the Southern Hemisphere and comparison between the hemispheres. In particular, he found similarities between the energy exchanges of the two hemispheres during minor warmings, despite the greater total energy involved in Northern Hemisphere events. Crane (1979*a*) made a detailed study of the energetics of the early 1973 major warming by using s.c.r. data. There was no evidence of vertical phase propagation, the warming occurring simultaneously at all levels. Near the pole the eddy kinetic energy maxima were mainly caused by vertical energy flux convergence, but at mid-latitudes by barotropic conversion. A large part of the work was to develop a satisfactory retrieval method, which illustrates how important this problem is thought to be. Johnson & Rodenhuis (1977) studied the vorticity budget for the same warmings with the N.O.A.A. n.e.m.s., s.c.r., i.t.p.r. data set. In particular they noted extensive counter-gradient flux of potential vorticity before and during the warming, in agreement with Hartmann's (1977) Southern Hemisphere study. They also noted that maximum eddy amplitudes tend to follow the maximum zonal wind westerlies, a common but important observation from the theoretical viewpoint.

6. Satellite studies of the mesosphere

The Nimbus 6 l.r.i.r. and p.m.r. were the first satellite instruments to observe temperature in the mesosphere. The Nimbus 7 s.a.m.s. and l.i.m.s. are now adding to those measurements. Being from a nadir sounder the p.m.r. data are easier to interpret than those of the other instruments, which probably explains why only p.m.r. data have been extensively published. The satellite was launched in June 1975, and 40 months of data were taken until data acquisition was largely discontinued with the launch of Nimbus 7. All of these data have been processed, then reprocessed with optimum calibrations, and calculation of parametrized temperature dependent weighting functions taking non-l.t.e. deviations into account has been recently completed. The instrument was described by Curtis *et al.* (1974).

Figure 5 gives various quantities for the first 3 years of p.m.r. data, 21 June 1975–22 June 1978 for the highest and lowest weighting functions peaking at about 80 km and 43 km respectively (see figure 1). Most of the features of the lower level have already been referred to. The upper level has similarities with these but also differences. The most disturbed times near 80 km are again during winter, but the differences between the winter, summer and tropical eddy amplitudes are less marked than at lower levels. At the mean radiance of the 80 km level, 1 radiance unit change corresponds to about 1.4 K temperature change, compared with 0.84 K change for the lower level. Taking this into account, temperature amplitudes near 80 km are approximately half of those near 43 km during winter. Lower summer and tropical standard deviations are apparent between October 1975 and August 1976 and are caused by the instrument operating in a different mode with one channel not scanning during this period.

Perhaps the most important difference between the levels is the strong negative correlation between the 80° N zonal mean temperatures at the two levels, not only seasonally, which has long been well known, but also in the short term during stratospheric warming events. This is

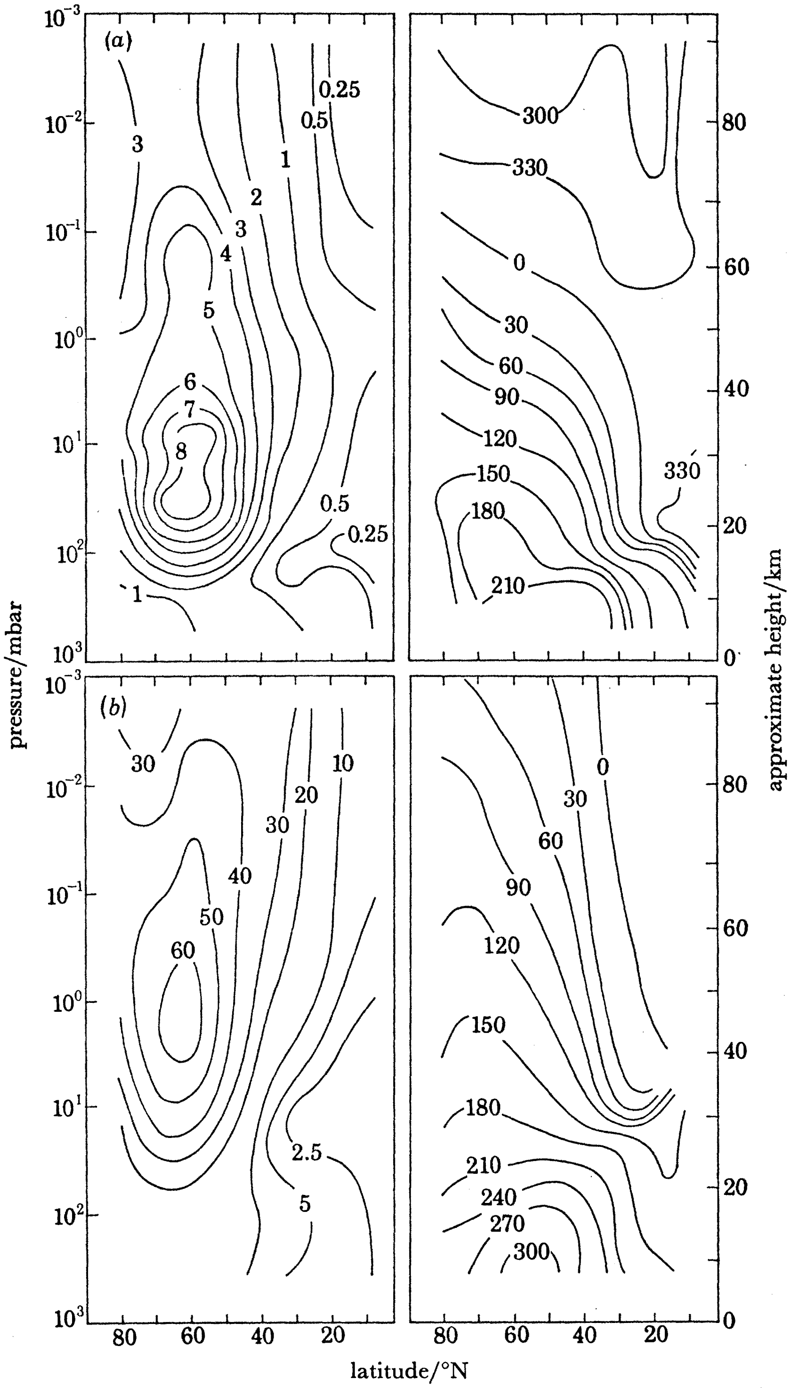

FIGURE 6. (*a*) Amplitude and phase of the wavenumber one component of temperature (K) averaged over December 1975, January and February 1976, derived from observations by the Nimbus 5 s.c.r. and Nimbus 6 p.m.r. (*b*) the geopotential height field (dm) derived from (*a*) by using the climatological 30 mbar height as a reference surface. The phase is the longitude (°E) of the maximum.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

most obvious during the 1976–7 winter when large oscillations of the zonal mean occurred at 43 km and were mirrored at 80 km. Houghton (1978) and Hirota & Barnett (1977) discussed these changes and showed that they fit in well with theory, the upper mesospheric cooling during a warming of the stratosphere and lower mesosphere resulting from the updraught forced over the polar cap by eddy heating. As discussed earlier, in the stratosphere tropical minima coincide with polar maxima due to warmings, and are easily seen at 45 km. At the higher level there is no such clear response, particularly during the large 1976–7 midwinter warming, when it might be expected to be most obvious. However, there are strong suggestions of such negative correlations (tropical minima at 80 km coincident with high latitude maxima at the same level) during other warmings, and this is awaiting further analysis.

The structure of planetary waves has been shown by Austen *et al.* (1976), Houghton (1978), Barnett (1977) and Hirota & Barnett (1977). The westward tilt with increasing height previously seen at lower levels continues up to the mesopause, both in the mean and during large single events, with approximately a 150° phase change between the upper stratosphere and upper mesosphere (figure 6). This is a further example of out-of-phase relationships between these two levels, but there is no suggestion that they are due to the same cause. They were previously observed by Quiroz (1969) and Labitzke (1972) from individual rocket observations. Hirota & Barnett (1977) gave latitude–time sections of wave amplitude and phase and also means for a month in midwinter 1975–6. Temperature variations for wavenumber 1 are approximately twice as large near 45 km as near 80 km, both for short term variations and for the monthly mean. Short term variations were not correlated except that particularly large amplitude maxima at 43 km were accompanied by maxima at 80 km.

The phase variations of the geopotential height (figure 6) have direct physical meaning. Westward tilt with height implies eddy heat flux towards the poles and, if the wind is westerly, upward eddy energy flux. Westward tilt towards the Equator implies eddy momentum flux towards the poles and for westerly wind, energy flux towards the Equator. Since the wind is normally westerly during winter, the phase tilts shown indicate forcing of the wave from the high latitude troposphere and absorption at low latitudes. It is interesting to note that the eddy maxima lie in the upper stratosphere or lower mesosphere, showing that the stationary winter low wave number planetary wave is essentially a middle atmsophere phenomenon.

7. Conclusion

A large number of measurements of the middle atmosphere temperature now exist, and although there have been many studies of the data, they have barely started to explore the information which has been gathered.

Until now instruments have been experimental and we have relied upon the goodwill of N.A.S.A. to operate them longer than the nominal 1–2 years design lifetime, and so to provide an unbroken series of measurements. However, the observations are on a firmer basis with the start of the Tiros N series of operational sounders having temperature sounding channels as high as the upper stratosphere. The importance of a long sequence of observations cannot be overemphasized. One reason is the large year-to-year variations of the winter atmosphere. Major warmings occur on average about once per year, and each seems very different from the others. Whether there is a pattern to these variations, such as a link with the 26-month oscillation or the solar cycle, or whether these cycles appear in other ways are open questions and may need decades of data for their answers.

Mesospheric sounding will still rely on experimental instruments and with greater emphasis on constituent measurements in the stratosphere, mesospheric temperature may not be measured continuously for many more years. Given that the mesosphere is just as important a part of the middle atmosphere as the stratosphere with many processes taking place in both, a case can be made for the next generation of operational sounders after the Tiros N series to carry mesospheric sounders.

References (Barnett)

Adler, R. F. 1975 *J. atmos. Sci.* **32**, 893–898.

Austen, M. D., Barnett, J. J., Curtis, P. D., Morgan, C. G., Houghton, J. T., Peskett, G. D., Rodgers, C. D. & Williamson, E. J. 1976 *Nature, Lond.* **260**, 594–596.

Barnett, J. J. 1974 *Q. Jl R. met. Soc.* **100**, 505–530.

Barnett, J. J. 1975*a* *Q. Jl R. met. Soc.* **101**, 835–845.

Barnett, J. J. 1975*b* *Nature, Lond.* **255**, 387–389.

Barnett, J. J. 1977 *Phil. Trans. R. Soc. Lond.* A **279**, 247–259.

Barnett, J. J., Harwood, R. S., Houghton, J. T., Morgan, C. G., Rodgers, C. D. & Williamson, E. J. 1975 *Q. Jl R. met. Soc.* **101**, 423–436.

Chapman, W. A., Cross, M. J., Flower, D. A., Peckham, G. E. & Smith, S. D. 1974 *Proc. R. Soc. Lond.* A **338**, 57–76.

Crane, A. J. 1979*a* *Q. Jl R. met. Soc.* **105**, 185–206.

Crane, A. J. 1979*b* *Q. Jl R. met. Soc.* **105**, 511–522.

Curtis, P. D., Houghton, J. T., Peskett, G. D. & Rodgers, C. D. 1974 *Proc. R. Soc. Lond.* A **337**, 135–150.

Deland, R. J. 1973 *Mon. Weath. Rev.* **101**, 132–140.

Deland, R. J. 1977 *Proc. IAGA/IAMAP Assembly*, Seattle, 1977.

Fritz, S. & Soules, S. D. 1972 *Mon. Weath. Rev.* **100**, 582–589.

Fritz, S. 1974 *J. atmos. Sci.* **31**, 813–822.

Hartmann, D. L. 1976 *J. atmos. Sci.* **33**, 1141–1154.

Hartmann, D. L. 1977 *Space Res.* **17**, 167–174.

Harwood, R. S. 1975 *Q. Jl R. met. Soc.* **101**, 75–91.

Hirota, I. 1976 *Q. Jl R. met. Soc.* **102**, 757–770.

Hirota, I. & Barnett, J. J. 1977 *Q. Jl R. met. Soc.* **103**, 487–498.

Hirota, I. 1978 *J. atmos. Sci.* **35**, 714–722.

Houghton, J. T. 1977 *Appl. Opt.* **16**, 319–321.

Houghton, J. T. 1978 *Q. Jl R. met. Soc.* **104**, 1–29.

Johnson, K. W. & Rodenhuis, D. R. 1977 *Proc. IAGA/IAMAP Assembly*, Seattle, 1977.

Kaplan, L. D. 1959 *J. opt. Soc. Am.* **49**, 1004–7.

Kennedy, J. S. & Nordberg, W. 1967 *J. atmos. Sci.* **24**, 711–719.

Labitzke, K. 1972 *J. atmos. Sci.* **29**, 756–766.

Labitzke, K. 1974 *J. geophys. Res.* **79**, 2171–2175.

Labitzke, K. 1977 *Cospar Space Research* **17**, 159–165.

Labitzke, K. & Barnett, J. J. 1979 *Cospar Space Research* **19**, 97–106.

Leovy, C. & Webster, P. J. 1976 *J. atmos. Sci.* **33**, 1624–1638.

McGregor, J. & Chapman, W. A. 1978 *J. atmos. terr. Phys.* **40**, 677–684.

Nordberg, W., McCulloch, A. W., Foshee, L. L. & Bandeen, W. R. 1967 *Bull. Am. met. Soc.* **47**, 857–872.

Quiroz, R. S. 1969 *Mon. Weath. Rev.* **97**, 541–552.

Quiroz, R. S. 1975 *J. atmos. Sci.* **32**, 211–224.

Rodgers, C. D. 1976*a* *J. atmos. Sci.* **33**, 710–714.

Rodgers, C. D. 1976*b* *Rev. Geophys. Space Phys.* **14**, 609–624.

Rodgers, C. D. 1977*a* In *Dynamical and chemical coupling* (ed. B. Grandal & J. A. Holtet), pp. 3–16. Dordrecht: Reidel.

Rodgers, C. D. 1977*b* In *Inversion methods in atmospheric remote sounding* (ed. A. Deepak), pp. 117–138. New York: Academic Press.

Stanford, J. L. & Dunkerton, T. J. 1978 *Beitr. Phys. Atmos.* **51**, 174–188.