Tests of Schemes to Infer Stratospheric Temperature from Satellite Measurements

David R Jackson

Doctor of Philosophy

University of Edinburgh

1990

Declaration

I declare that this thesis consists of research performed solely by myself, and that it has been composed solely by myself.

Acknowledgements

I would like to thank my supervisors, Bob Harwood and Eric Renshaw, for their invaluable guidance and encouragement throughout the course of this research. The support received from all the members of Edinburgh University Meteorology Department is also appreciated. The Met Office Middle Atmosphere Group are thanked for their co-operation, especially Mike Bailey and Alan O'Neill. I am grateful to the Natural Environment Research Council and the Met Office for funding this research, and to these organisations and the University of Edinburgh for providing money to help me attend conferences and meetings.

Summary

In this thesis we test two retrieval/analysis schemes for inferring stratospheric temperature from satellite observations of radiance. The first scheme is similar to that used by the UK Meteorological Office. The retrievals are made by using a multiple linear regression model which regresses radiances against Planck function, whilst the analyses are made using a linear time/space interpolation method. The retrieval part of the second retrieval/analysis scheme is the same as above, but the analysis part sequentially estimates Fourier coefficients at fixed latitudes using a version of the Kalman Filter. Analyses made using both methods are compared.

Because of the lack of 'ground truth' observations in the stratosphere, the schemes are tested in simulation experiments. Preliminary tests of the time/space interpolation and sequential estimation analysis schemes are made using idealised radiance fields which resemble observations made by a satellite radiometer in the northern hemisphere winter stratosphere. The regression retrieval scheme and the two analysis schemes are also tested in a more sophisticated experiment in which the 'true' atmosphere is represented by an atmosphere simulated by a numerical model. Simulated observations are calculated by computing the radiance that would be observed from the 'true' atmosphere by a satellite instrument. The radiances are then retrieved and analysed and the resultant analyses compared with the corresponding 'true' fields. Tests are made using output from a day when a sudden warming was present.

The retrieval scheme is seen to perform less well within the area of the sudden warming than outside it. However, this may be expected as the vertical structure within the sudden warming is generally too small to be resolved by a

satellite instrument. The analysis scheme analyses the stratospheric field well, even in the area of a sudden warming. These results, and results from preliminary tests made using idealised radiance fields, suggest that the analysis is generally of better quality when the distance radius used to select observations for the scheme is small. Results of tests of the sequential estimation scheme reveal that this method also produces satisfactory analyses of idealised radiance and model fields. Constraints of time prevented more rigorous testing of the scheme, but suggestions for further research are given.

TABLE OF CONTENTS

1 Introduction 1

2 Climatology of the Stratosphere 7

2.1 Zonal Mean Temperature Structure 7
2.2 Dynamical Processes in the Stratosphere 12
2.2.1 Stationary and Transient Waves 12
2.2.1.1 Quasi-Stationary Waves 14
2.2.1.2 Transient Wave Analysis 16
2.2.2 Stratospheric Sudden Warmings 17
2.2.3 Equatorial Waves 22
2.2.4 Gravity Waves and Tides 25
2.3 Ability to Retrieve and Analyse Stratospheric Wave Motions Using Satellite Measurements 28

3 Principles of the Remote Sounding of Temperature 31

3.1 Radiative Transfer Theory 31
3.2 Instruments used for Stratospheric Temperature Retrievals 34
3.3 Calculation of Instrument Weighting Functions 37
3.4 Current Retrieval and Analysis Methods 39
3.4.1 Retrieval Methods 39
3.4.1.1 Introduction 39
3.4.1.2 Minimum Information Methods 43
3.4.1.3 Regression 46
3.4.1.4 Error Characteristics of Retrievals 47
3.4.2 Analysis 50
3.4.2.1 Introduction 50
3.4.2.2 Time/Space Interpolation 52
3.4.2.3 Sequential Estimation of Fourier Field Components 54
3.5 Summary 58

4 Preliminary Tests of the Time/Space Interpolation Method Using Analytical Radiance Fields 59

4.1 The Model Radiance Field 60
4.1.1 Theory 60
4.1.2 Numerical Details 61
4.2 Calculation of the 'Observed' Radiance Field 65
4.3 Results 67
4.3.1 Field 1 68
4.3.2 Field 2 74
4.3.3 Field 3 82
4.3.4 Analyses Using Different Weights 87
4.3.4.1 Cosine Weights - Results 88
4.3.4.2 Negative Exponential Weights - Results 90
4.3.4.3 Discussion of Results Made Using Negative Exponential Weights 94
4.3.5 Summary of Results 95
4.3.6 Recommendations for Further Tests of the Time/Space Interpolation Scheme 96

5 Tests of Retrieval/Analysis Schemes Using Met. Office Stratosphere/Mesosphere Model fields : Method of Simulation and Data Used 98

5.1 Purpose of the Tests 98
5.2 The Stratosphere/Mesosphere Model 100
5.3 Simulation of Radiances 101

6 Tests of the Regression Retrieval Scheme 105

6.1 The Regression Model 105
6.2 Calculation of Regression Coefficients 107
6.2.1 Relation of The Sonde Dataset to the Numerical Model 108
6.2.2 Testing of the Regression Model 113
6.3 Retrieval Results 116
6.3.1 Cross-Sections 123
6.4 Discussion 127

7 Tests of the Time/Space Interpolation Method 130

7.1 Interpolation of Retrieved Temperatures 131
7.1.1 Examination of the Model Field Used in the Tests 132
7.1.2 R.m.s Errors 138
7.1.3 Comparison of Maps at 5 mb 150
7.2 Conclusions 156

8 Tests of the Sequential Estimation Analysis Method 158

8.1 Tests Using Analytically Calculated Fields - Details and Results 159
8.1.1 Preliminaries - Interpolation and Calculation of Initial 'First Guess' Estimates 161
8.1.2 Results 163
8.2 Tests Using a Met. Office Model Field 175
8.2.1 Results 176
8.3 Conclusions and Pointers to Future Research 191

9 Conclusions 196

I. Proofs to Optimal Solutions of the Retrieval Problem 200

I..1 Combination of Observations 200
I..2 Rearranging Optimal Estimate and Error Covariance Equations 202

II. References 204

III. Publication 213

CHAPTER 1

INTRODUCTION

Electromagnetic radiation leaving the top of the atmosphere carries information about the distribution of temperature and of the emitting gases. If measurements are made at wavelengths at which the emission is by a gas of known mixing ratio such as carbon dioxide, then some details of the temperature distribution may be deduced (Kaplan, 1959); this principle is exploited in several remote sensing satellites. Until remotely sounded measurements of the stratosphere and mesosphere became available in the late 1960s the only routine measurements of the region were made by radiosondes (in the lower stratosphere) and by rocketsondes (in both the stratosphere and mesosphere). Radiosonde observations are made chiefly over continents, whilst rocketsonde ascents are made infrequently from only a small number of stations. Accordingly, it is difficult to form a complete picture of middle atmosphere structure and dynamics from sonde measurements alone. On the other hand, observations from a polar orbiting satellite make it possible to map the temperature structure in three dimensions, and to follow its changes day by day. Consequently, numerous studies of the structure and dynamics of the middle atmosphere have been made using satellite data. Examples include: the construction of a new draft reference middle atmosphere (eg Barnett and Corney, 1985); comparisons between the middle atmosphere dynamics of the southern and northern hemispheres (eg Andrews, 1989); the study of stratospheric sudden warmings using daily maps of isentropic potential vorticity (eg Fairlie and O'Neill, 1988; Clough et al, 1985); the identification of extratropical transient waves (eg Rodgers, 1976a; Hirota and Hirooka, 1984; Hirooka and Hirota, 1985) and of equatorial waves (eg Hirota, 1979; Salby et al 1984); the study of seasonal, inter-seasonal and inter-annual variability in the middle atmosphere (e.g Mechoso et al, 1985; Hirota, 1978; Hirota, 1980). Since

satellite data are widely used in stratospheric studies, it is important to realise that the process of deducing the temperature structure from measurements is not necessarily straightforward. Usually there are two aspects to the problem, although it is possible to devise procedures in which they are combined. The two aspects are: 'retrieval', in which a single temperature profile is deduced from a more-or-less instantaneous set of measurements; and 'analysis' in which the state of the atmosphere at a given instant is deduced on a regularly spaced grid of points from the retrieved profiles which are asynoptic and distributed according to the shifting satellite orbit. Without further information the retrieval problem is under-constrained because in general an infinite number of atmospheric profiles can yield the same finite set of measurements. Moreover, the analysis problem can suffer from aliasing difficulties. Aliasing occurs in all Fourier analyses of discrete data; the time period of the data imposes a limit on the highest resolvable frequency, and hence any higher frequency present will be analysed falsely within the range of the lower, resolvable, frequencies.

The widespread use of satellite observations in stratospheric studies means it is important to estimate how well we can infer stratospheric temperature structure from such observations. Thus the aim of this thesis is to evaluate the performance of retrieval/analysis schemes for obtaining stratospheric temperatures from the TIROS Operational Vertical Sounder (TOVS) instrument (Schwalb, 1978 ; Smith et al, 1979) on the TIROS-N series of polar-orbiting satellites.

The retrieval scheme tested here is based on a regression model similar to that used by the UK Meteorological Office (Pick and Brownscombe, 1981), the main difference being that their scheme uses the measured radiances to give thicknesses of fairly thick layers of atmosphere, whereas the present scheme

gives the temperature profile at 31 pressure levels from 0.2 to 570 mb. Many previous tests of retrieval schemes have compared retrievals with coincident rocketsonde measurements. Nash and Brownscombe (1983) and Pick and Brownscombe (1981) tested the Stratospheric Sounding Unit (SSU) on TOVS; Barnett et al (1975) tested the Selective Chopper Radiometer (SCR) on the Nimbus 5 satellite. Whilst the chief purpose of those tests was to assess the performance of the satellite instrument, our tests are different in that we pay particular attention to the regression/retrieval scheme itself. We examine the representativeness of the datasets used to calculate the regression coefficients, and test the ability to retrieve temperature in various atmospheric conditions.

Two analysis schemes are tested: 1) 'time/space interpolation'; 2) 'sequential estimation'. The time/space interpolation method is that used operationally by the UK Meteorological Office (Pick and Brownscombe, 1979, 1981), so this thesis will help users of such analyses to evaluate the confidence which can be placed in them. The scheme gives each observation a time and distance weight which decreases the further the observation is from the gridpoint or analysis time. Only observations lying within a specified time and distance (called 'search radii') of the gridpoint and analysis time are used in the scheme. Tests of the scheme are initially made using idealised fields, and then on a field simulated by a numerical model. We concentrate chiefly on the way the quality of the analysis changes when the search radii are changed. The other analysis scheme sequentially estimates Fourier field components at fixed latitudes using a version of the Kalman Filter (Kalman, 1960; Kalman and Bucy, 1961). The Kalman Filter is a set of equations which provide an optimal estimate that is changed and updated as fresh observations arrive. This scheme is tested using idealised and model fields similar to those used to test the time/space interpolation method, and thus the two analysis methods can be compared. The sequential estimation technique has previously only been used to analyse

measurements from limb sounding satellites (eg the Limb Infrared Monitor of the Stratosphere (LIMS)), and sequentially estimated analyses of LIMS measurements have been compared with time/space interpolated analyses of TOVS measurements (eg Grose and O'Neill, 1989). However, since these analyses are made using observations from different satellite instruments, differences between them are due both to differences in the methods of observation, and to differences in the analysis methods. Here, since both analysis techniques are tested using the same set of observations, comparisons highlight only differences in the analysis methods.

A difficulty in testing any retrieval/analysis scheme using real observations is the absence of adequate 'ground-truth' observations. Rocketsondes observe temperature in the upper stratosphere but rocket flights are infrequent and badly spaced. Radiosondes observe in the lower stratosphere and give better global coverage, but even so there are few observations made over the oceans or in the southern hemisphere. Moreover, there are a multiplicity of sonde types, leading to calibration problems. Most tests which have been done with real data compare retrieved profiles with coincident rocketsonde measurements (eg Nash and Brownscombe, 1983; Pick and Brownscombe, 1981; Barnett et al, 1975). Furthermore it is generally even more difficult to make comparisons with ground truth for *fields* analysed from satellite measurements than it is for retrieved profiles. Possible difficulties with analyses have been reported by Al-Ajmi et al (1985) and by Clough et al (1985). The former paper gives evidence that there may be temporal variations too rapid for proper resolution, and the latter provides evidence that the vertical temperature structure is not always adequately resolved. Accordingly we have chosen to test the schemes in simulation experiments. Preliminary tests of the time/space interpolation scheme are made using idealised radiance fields which resemble the stratosphere during a northern hemisphere winter. A model of a TIROS-N series

satellite orbit is used to compute the radiance that would be observed from these idealised fields by a satellite radiometer. These radiances are then analysed and the resultant analyses compared with the corresponding idealised fields. The sequential estimation method is similarly tested, and results from both methods are compared. More sophisticated tests of the retrieval and analysis schemes are made using an atmosphere calculated in a numerical model. Simulated observations are calculated by computing the radiances which would be observed from this model atmosphere by a TOVS-like instrument, including the effects of instrumental noise, and temperatures are retrieved from these observations using the regression retrieval scheme. Then the retrieved temperatures are analysed using both the time/space interpolation and sequential estimation techniques, and these analyses are compared with each other and with the corresponding model fields.

The layout of the thesis is as follows. Chapter 2 describes the climatology and time/space variability of the stratosphere, and the ability of satellite observations to detect such variations. Chapter 3 describes the theory of remote temperature sounding, and gives details of the TOVS instruments. A review of methods to retrieve and analyse satellite observations also appears in this chapter, and includes descriptions of the methods to be tested in subsequent chapters. Details of the idealised radiance fields used in tests of the time/space interpolation and sequential estimation methods appear in Chapter 4, together with results of preliminary tests of the time/space interpolation scheme. Chapter 5 contains details of the Met. Office model field used in tests of the regression retrieval scheme and the two analysis schemes, together with the method used to simulate observations from this field. Chapter 6 describes the retrieval scheme and the method of obtaining the regression coefficients, together with results of tests of the retrieval scheme, whilst results of tests of the time/space analysis scheme appear in Chapter 7. Chapter 8

contains results of tests of the sequential estimation analysis scheme, and conclusions appear in Chapter 9.

CHAPTER 2

CLIMATOLOGY OF THE STRATOSPHERE

The aim of this thesis is to test methods of retrieving and analysing stratospheric temperature from satellite measurements. It is therefore important to adequately describe the stratospheric climatology, and in particular to relate the temporal and spatial variability of stratospheric dynamical processes to the ability to retrieve and analyse them from satellite data. Hence, in Section 2.1 the observed zonal mean temperature structure is described. However, there can be large day-to-day departures from this observed zonal mean state caused by quasi-stationary planetary waves, transient waves, stratospheric sudden warmings and other dynamical features. The size and variability of these features is described in Section 2.2, and the ability of satellite measurements to resolve their spatial and temporal scales is discussed in Section 2.3.

2.1. Zonal Mean Temperature Structure

The vertical temperature structure in the stratosphere is mainly determined by differences in the absorption of solar radiation. We shall consider absorption of solar radiation from the top of the stratosphere downwards. In the stratosphere the main contribution to the temperature structure comes from absorption of radiation by ozone. The very strong absorption by O_3 in the Hartley and Huggins bands (0.18 μm to 0.35 μm) gives heating which is most intense at 50 km (the maximum heating is 18 K day^{-1} at the summer pole (London, 1980)). Below 25 km the ozone is protected from these wavelengths and hence the temperature is lower, as evidenced by Figure 2.1, which schematically represents the atmospheric vertical temperature structure. In the lower stratosphere weak absorption by CO_2 and H_2O in the near infrared and by O_3 in the visible occurs. These absorptions give heating rates of only a few

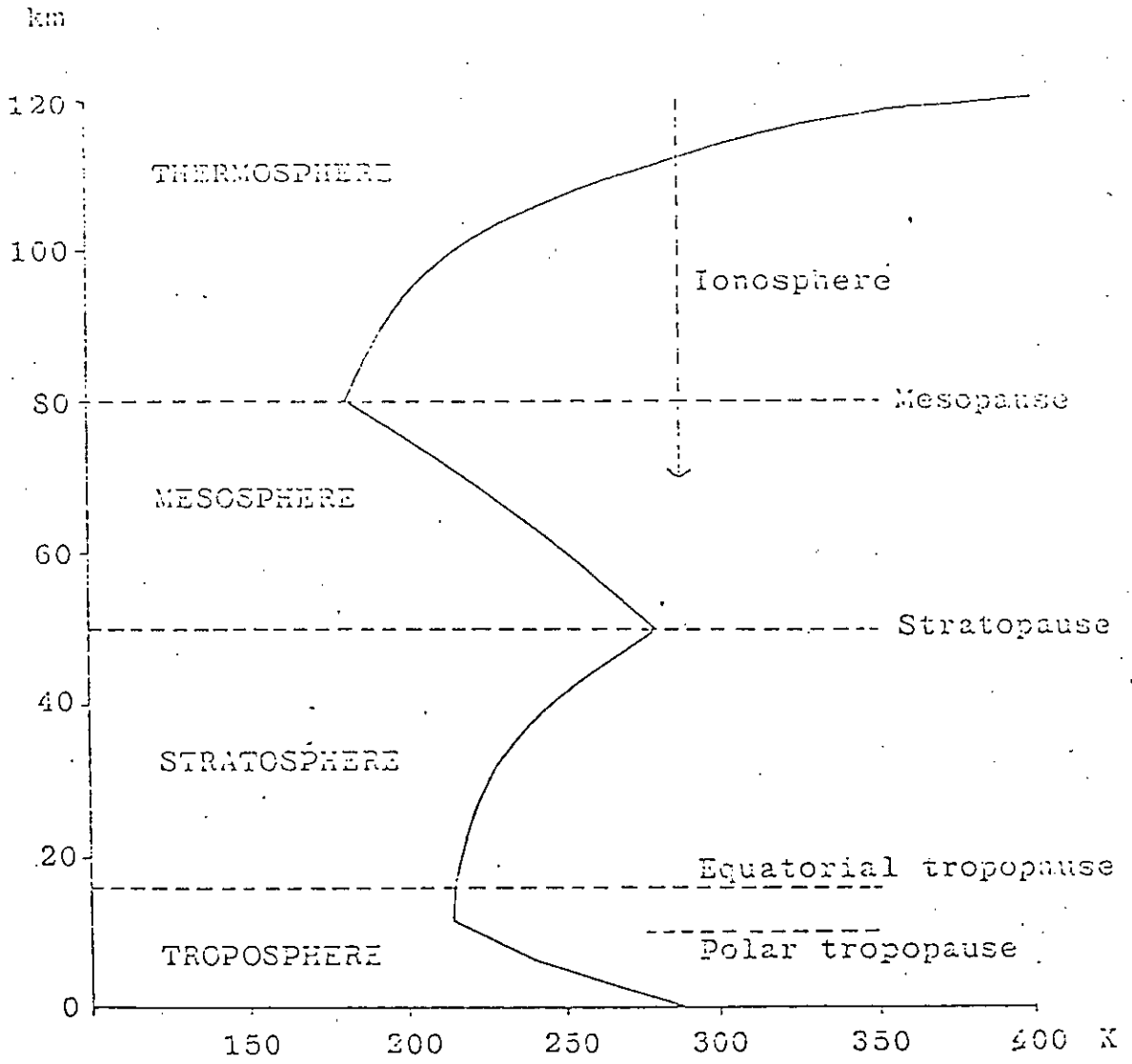
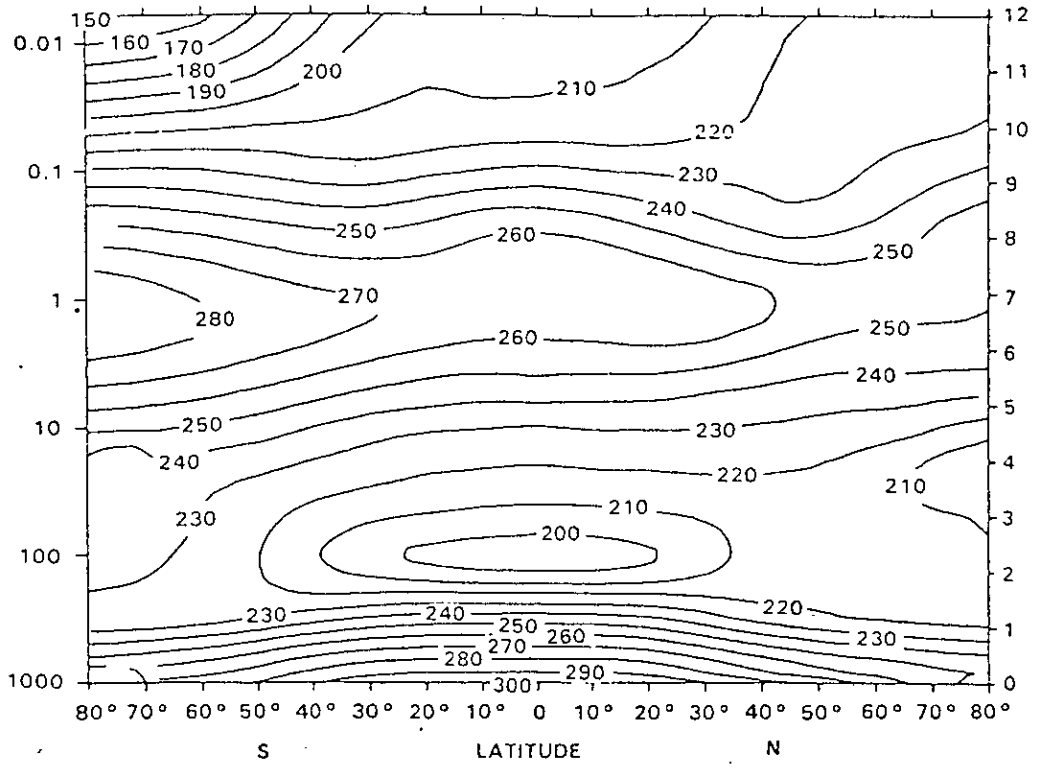


Figure 2.1 A schematic representation of the vertical temperature structure of the atmosphere with the nomenclature used to describe the different layers and their boundaries (from Crane, 1978).

tenths of a degree per day, hence there is a temperature minimum between 10 and 17 km (the tropopause). In the troposphere direct heating by absorption of solar radiation is of little importance and the temperature profile is determined mainly by energy exchange with the Earth's surface. Energy loss due to thermal emission is due mainly to CO_2 at 15 μm and O_3 at 9.6 μm in the stratosphere, and by clouds and H_2O in the troposphere.

A description of the zonal mean climatology of the middle atmosphere appears in a number of publications (see eg W.M.O, 1985). Observed zonal mean temperatures and zonal geostrophic winds of the stratosphere and mesosphere for January and July are shown in Figures 2.2 and 2.3 respectively. The lowest stratospheric temperatures occur in the polar night and the highest occur over the summer pole at 1 mb. Except at high latitudes in the winter lower stratosphere, the temperature increases with height in the stratosphere, whilst in the middle and upper stratosphere the temperature decreases monotonically from the summer to the winter pole. In addition, differences exist between the hemispheres in winter: the winter polar temperature in most of the stratosphere is higher in the northern hemisphere than in the southern, but the northern winter temperature is lower near the stratopause. In the summer there is a close resemblance between the two hemispheres with a cold equatorial tropopause (with temperatures less than 200 K) and a warm near-horizontal stratopause (peaking at over 280 K at the pole). The temperature is slightly higher at the southern hemisphere summer stratopause, probably because the Earth is closer to the sun in January than it is in July. Zonal winds in winter and summer are westerly and easterly, respectively, with maxima in the lower mesosphere. Easterlies occupy equatorial latitudes in the stratosphere at the solstices, except in the lower stratosphere during the westerly phase of an inter-seasonal phenomenon known as the Quasi-Biennial Oscillation. Dramatic changes to this zonal mean structure can occur during stratospheric sudden warmings, which are discussed in Section 2.2.2.

a)



b)

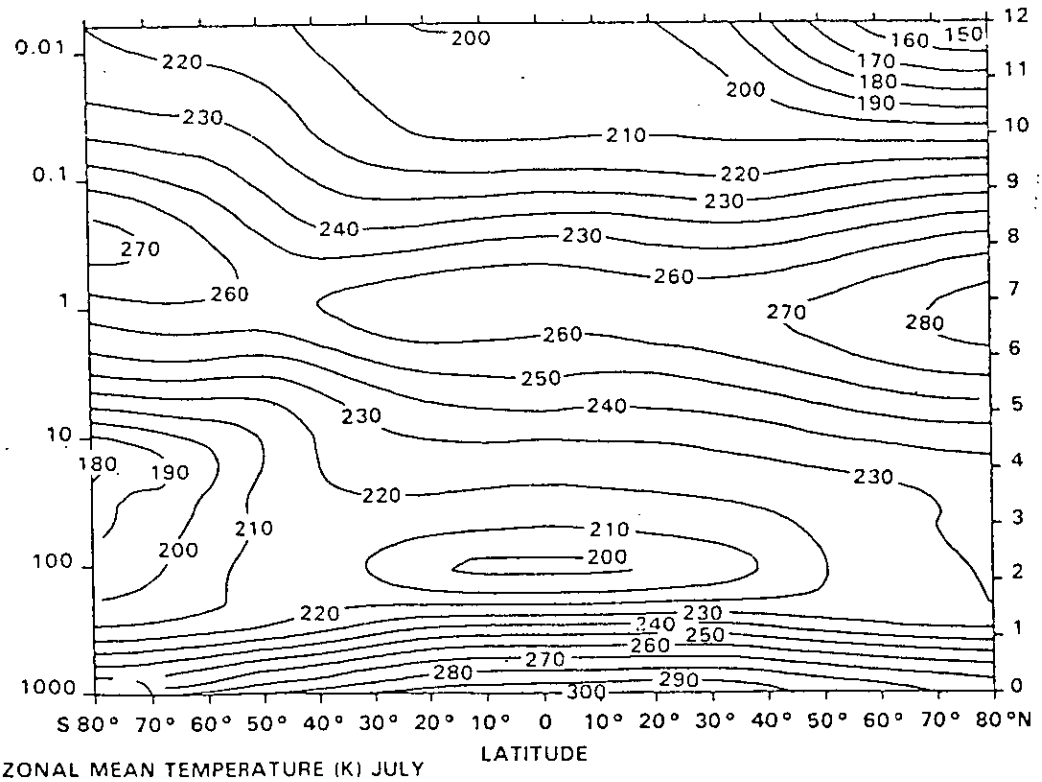
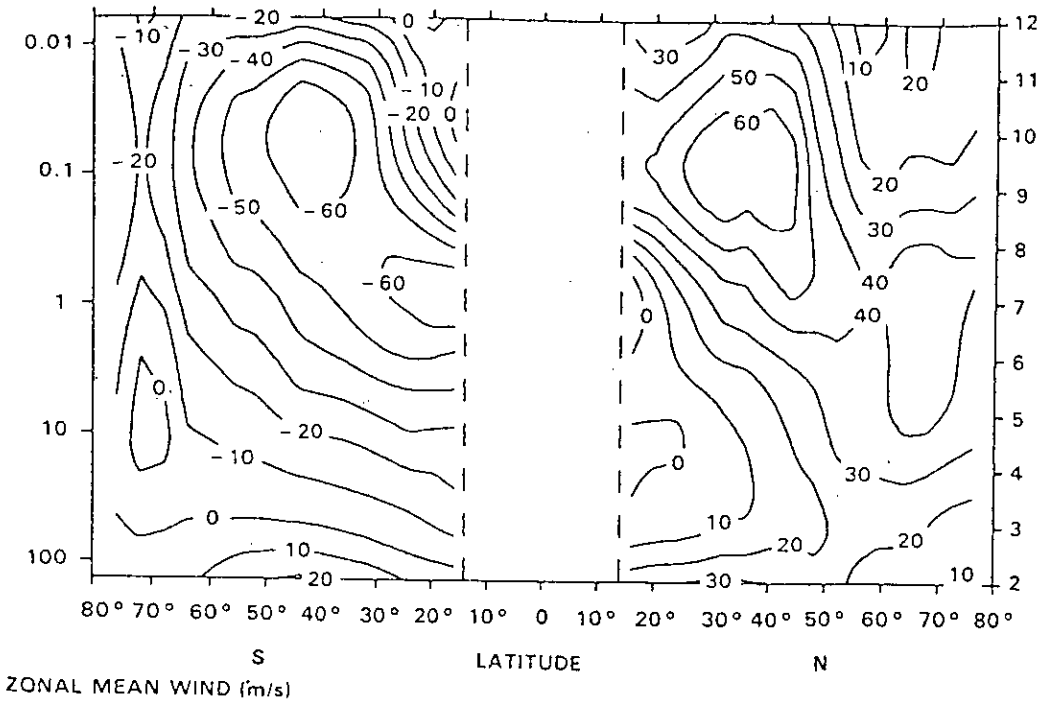


Figure 2.2 Cross-section (pressure (mb)-latitude) of zonal mean temperature (K) for the average over 5 years of the monthly means for a) January; b) July. The data are from the combined SCR/PMR retrieval made at the University of Oxford for the period January 1973 to December 1974 and July 1975 to June 1978 (from (W.M.O, 1985).

a)



b)

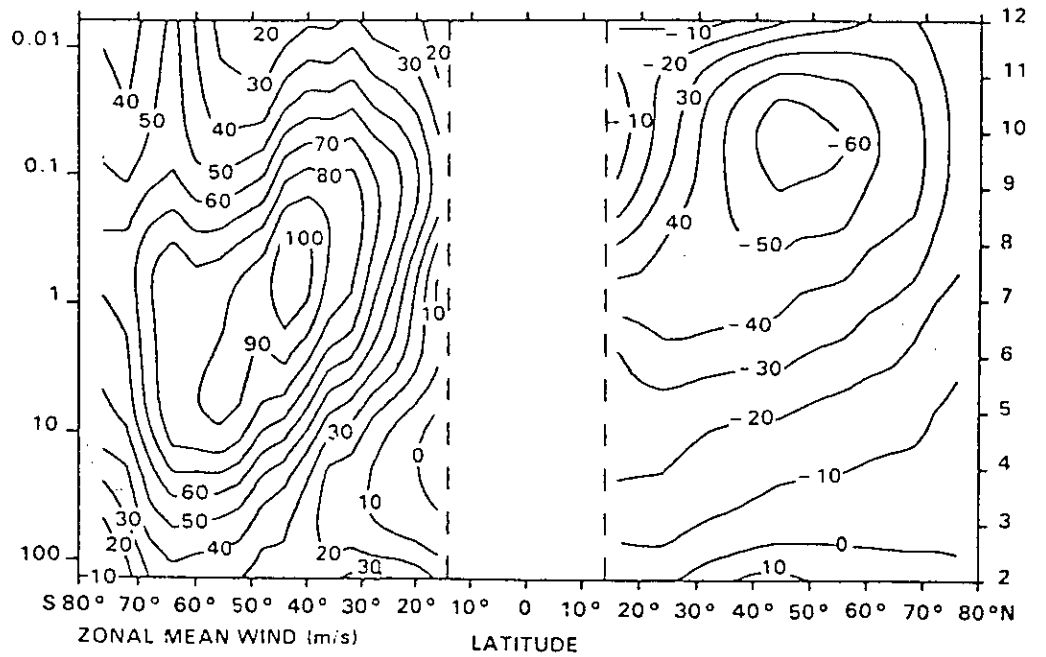


Figure 2.3 Like Figure 2.2, except zonal mean geostrophic zonal winds are plotted. Units are ms^{-1}. a) January; b) July.

2.2. Dynamical Processes in the Stratosphere

The zonal mean structure of the stratosphere is an incomplete description of the three-dimensional state. The winter stratosphere field may depart substantially from zonally uniform flow because of the presence of large-amplitude quasi-stationary waves. In addition, transient waves cause substantial day-to-day variations. Both transient and quasi-stationary waves are discussed in greater detail in Section 2.2.1, whilst stratospheric sudden warmings, which produce the most drastic departures from the regular seasonal cycle, are described in Section 2.2.2. Recent work suggests that planetary waves can 'break', leading to a cascade of energy to smaller scales, and this is also discussed in Section 2.2.2. The waves observed in the tropics are distinct from extra-tropical waves. Details of these equatorial waves appear in Section 2.2.3, whilst Section 2.2.4 contains details of tides and gravity waves.

2.2.1. Stationary and Transient Waves

The extra-tropical planetary wave field may be characterised in terms of two basic contributions to frequency spectra of individual wavenumbers. One contribution is due to baroclinic 'quasi-stationary' waves, which result from the interaction of the mean flow with topographical features and associated thermal effects. They are 'quasi-stationary' in the sense that they fluctuate in amplitude and phase about climatologically preferred values. A satellite in a sun-synchronous orbit will typically make 14 observations around a latitude circle in 24 hrs. If such observations are Fourier analysed, then quasi-stationary waves with zonal wavenumbers of 1, 2, 3 etc will give rise to components in a time-power spectrum with peaks at frequencies of 1, 2, 3 etc cycles per day (c/d). This is demonstrated by Figure 2.4, which shows the power spectra of waves for channels 1 to 5 of the Nimbus 4 Selective Chopper Radiometer (SCR). The dotted lines show the contributions of the quasi-stationary waves. A clear characteristic of these waves is the decrease in the power with increasing zonal wavenumber. Since other waves present in the stratosphere have periods of less than a month, monthly mean fields will chiefly contain information about

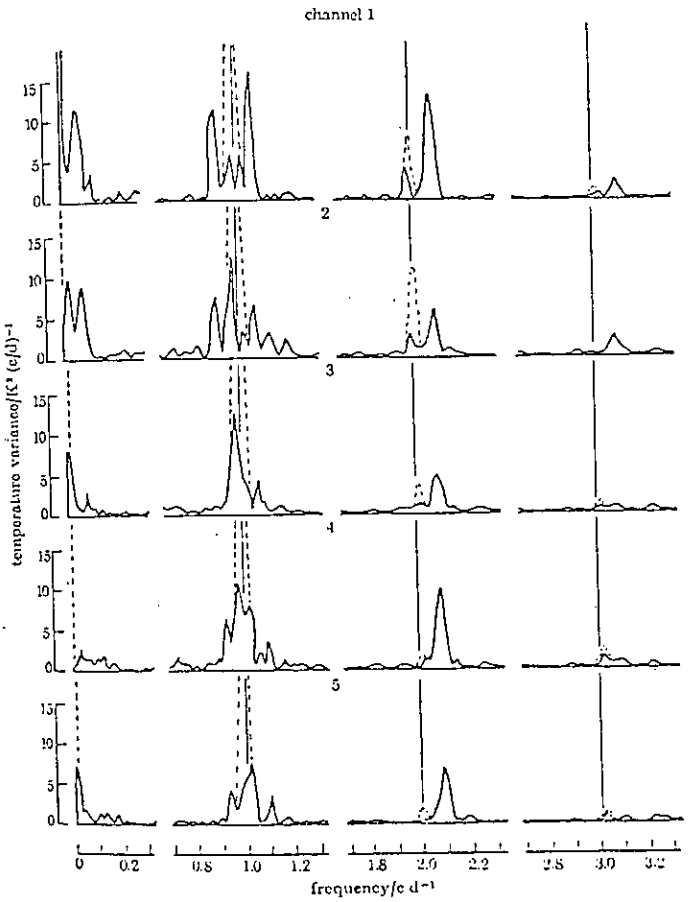


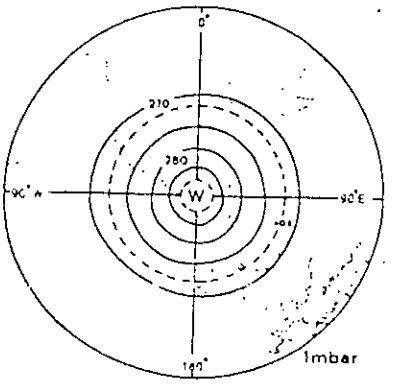
Figure 2.4 Power spectra of travelling waves for Nimbus 4 SCR channels 1 to 5, latitude 60°S, September to October 1970. The broken lines show the spectra without the subtraction of the stationary waves (from Chapman et al, 1978).

quasi-stationary waves. Hence the climatological structure of quasi-stationary waves are presented as monthly mean values in Section 2.2.1.1, although it must be stressed that the departures of the actual wave field from these means can be large.

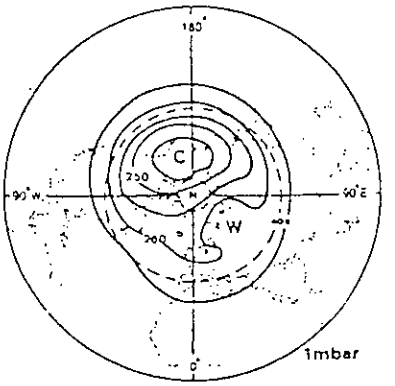
The second contribution to the frequency spectra of individual wavenumbers is from travelling waves. These waves are of global extent (at least at the lower levels), and their vertical structure is mainly barotropic. These waves give rise to Fourier components with non-integral frequencies, and appear in the spectrum displaced from the position of the quasi-stationary waves. Waves travelling eastward will be displaced to higher and those travelling westward to lower frequencies. A typical spectrum of travelling waves is represented by the solid lines in Figure 2.4. In this example, both eastward and westward propagating waves are observed, with eastward travelling waves being more common at wavenumbers 2 and 3. A further discussion of the gravest observed travelling waves appears in Section 2.2.1.2.

2.2.1.1. Quasi-Stationary Waves

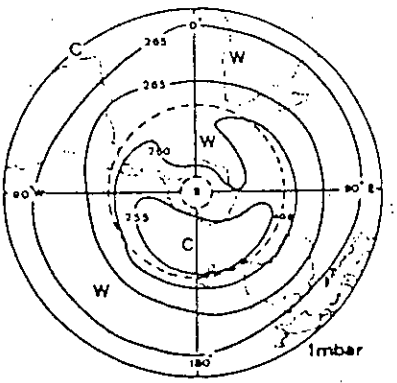
Most of the wave activity in the stratosphere is thought to originate in the troposphere. Middle and high latitude Rossby waves forced in this way play an important role in the atmosphere in transferring heat and momentum from low to high latitudes. The predominant waves in the troposphere are typically of zonal wavenumber 5 to 10. However, an examination of the mean temperature field in summer and winter (for both hemispheres) at a variety of stratospheric pressure levels shows that in winter the predominant zonal wavenumbers of these quasi-stationary waves are 1, 2 and 3 and that in summer these waves are absent. Figure 2.5 shows the monthly mean temperature field at 1 mb in January and July for both hemispheres. In summer (July for the northern hemisphere, January for the southern hemisphere) there is little or no wave



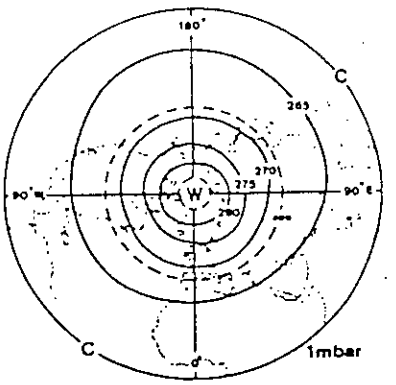
January S.H.



January N.H.



July S.H.



July N.H.

Figure 2.5 Monthly mean temperatures at 1 mb for January and July (from Barnett and Corney, 1985).

activity, whilst in winter large longitudinal variations are evident. These variations are chiefly of low wavenumber. This behaviour is broadly consistent with the Charney-Drazin theory of linear vertical wave propagation (Charney and Drazin, 1961). This states that a quasi-stationary wave can propagate from the troposphere to the stratosphere only if its phase speed is westward relative to the mean zonal flow and less than a critical speed. Since this critical speed decreases with increasing wavenumber, only long waves will propagate in winter, whilst in summer, because the mean flow is easterly, there is no propagation. However, recent work on 'wave breaking', (McIntyre and Palmer, 1983, 1984) however, cautions against an over-reliance on linear theory.

2.2.1.2. Transient Wave Analysis

Day-to-day departures of the actual stratospheric circulation from monthly means can be substantial. Some of the variance of the transient components may be associated with atmospheric normal modes. Unlike quasi-stationary waves, normal modes are predicted to be somewhat independent of forcing details. This property has permitted the identification of a number of these features.

Madden (1978) presented an extensive study of observational reports of travelling planetary-scale waves. This involved a variety of data sources, mainly tropospheric analyses, ranging in length from one month to 73 years. Wavenumber 1 waves are shown to migrate almost exclusively westward, with reported periods clustering about 5 and 16 days. Wavenumber 2 waves propagate predominantly westward, with most frequently reported periods being 3-7 days and 10-18 days, while wavenumber 3 waves are equally divided between eastward and westward propagating waves.

The first normal mode to be convincingly identified was the westward travelling 5-day wave. It has a wavenumber 1 structure and is symmetric about

the equator. Rodgers (1976a) used global temperature retrievals to identify the 5-day wave in the upper stratosphere. A power-spectrum analysis reveals a peak corresponding to a westward travelling wavenumber 1 wave with a period of about 5 days. The wave's typical amplitude is 0.5 K at about 50°N and 50°S, and somewhat smaller in equatorial and polar regions. Figure 2.6 shows the wave amplitude in November. With the increased availability of global satellite measurements, other stratospheric normal modes have been identified. The wavenumber 2 analogue of the 5-day wave is the 4-day wave at the stratopause. Figure 2.7 shows the amplitude (in geopotential height) of this wave as derived from TIROS-N SSU data by Hirota and Hirooka (1984). (At 1 mb a 200m increment of geopotential height is roughly equal to a 1 K increment of the mean temperature of a layer 40 km in depth). These workers have also used SSU data to identify higher-degree modes, such as the 10-day wavenumber 1 wave (Hirooka and Hirota, 1985). Figure 2.8 shows the amplitude of this wave at 1 mb. The amplitude is larger than for the 4-day wave, and there is a marked asymmetry about the equator. Other observed transient features in the stratosphere are attributable to local instability, for example the polar eastward-moving 4 day wave (Venne and Stanford, 1982), and the eastward-travelling wavenumber 2 anomaly in the southern hemisphere (Harwood, 1975).

2.2.2. Stratospheric Sudden Warmings

Stratospheric sudden warmings are the most spectacular large-scale dynamical events to occur in the middle atmosphere. They occur in winter and involve rapid rises of temperature (up to 80 K in a week locally), and wave amplitudes can double. An example is Figure 2.9, which shows the change in radiance observed by channel B12 on the Nimbus 5 satellite during a southern hemisphere warming: a maximum radiance increase of 40 r.u in 5 days is observed in southern hemisphere latitudes near 0°E (1 r.u in increment is

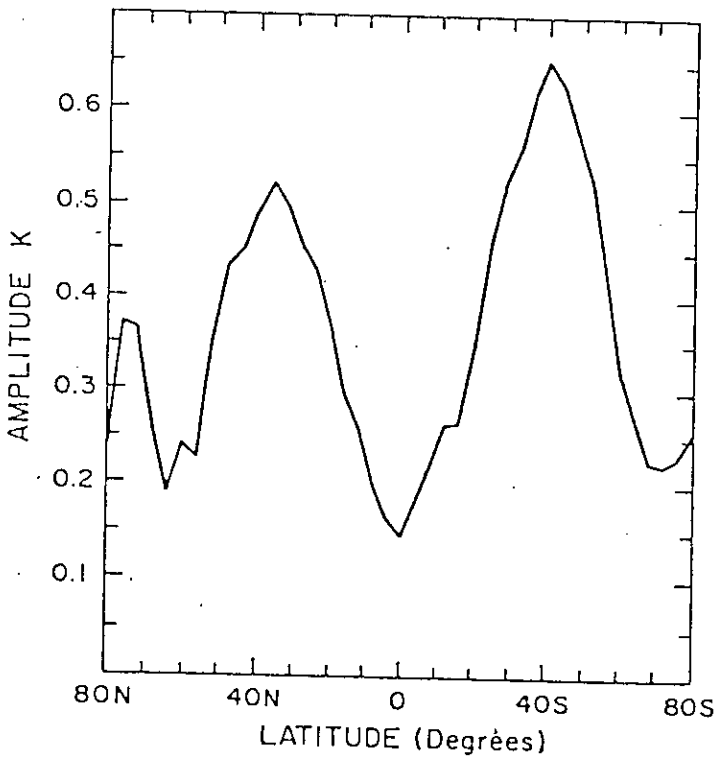


Figure 2.6 Amplitude (K) of the 5-day wave as a function of latitude for November 1973 (from Rodgers, 1976a).

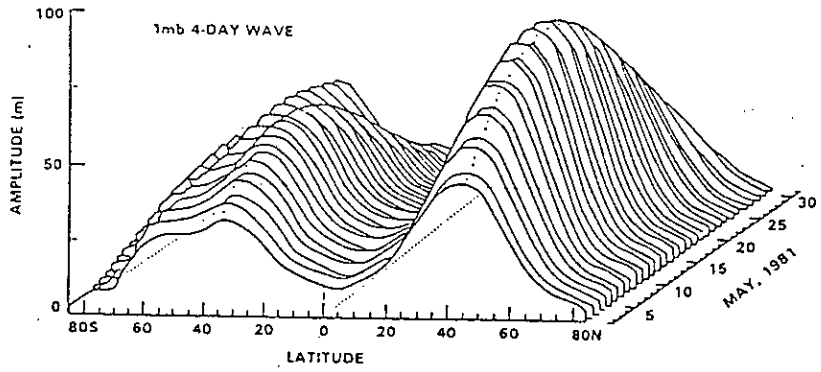


Figure 2.7 Three-dimensional plot of the 4-day wavenumber 2 wave amplitude (in geopotential height) versus latitude and time at 1 mb for May 1981 (from Hirota and Hirooka, 1984)

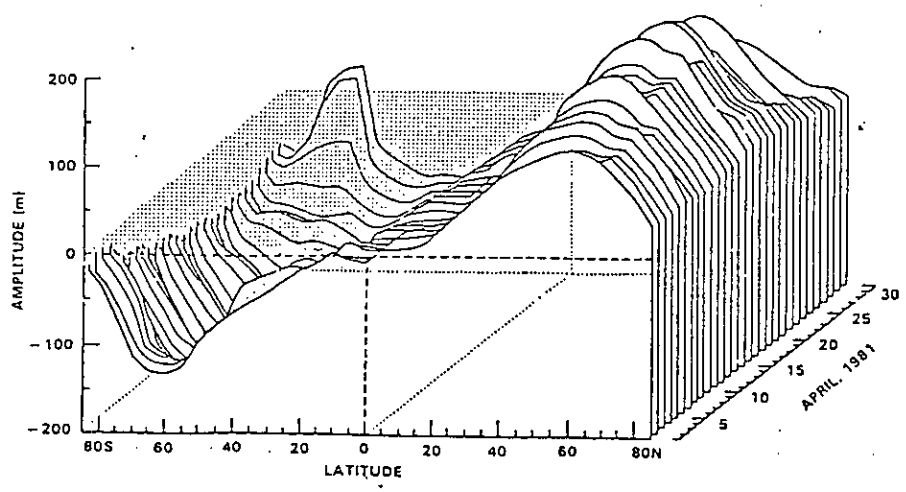


Figure 2.8 Like Figure 2.7, except the amplitude of the 10-day wavenumber 1 wave is shown (from Hirooka and Hirota, 1985).

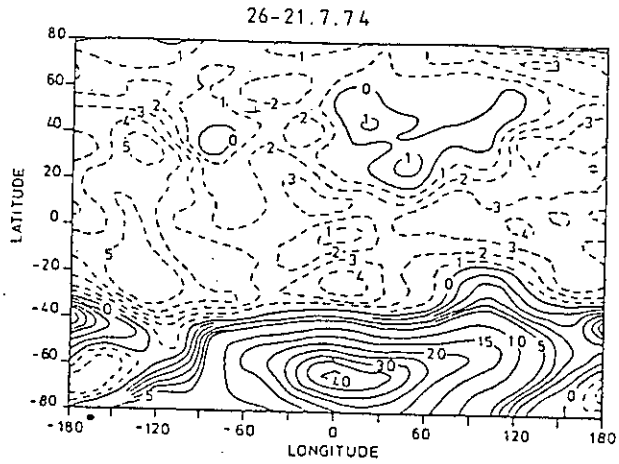


Figure 2.9 Change in radiance observed by channel B12 on Nimbus 5 between 21 and 26 July 1974. This approximately indicates the change in temperature of a 10 to 15 km thick layer of atmosphere situated about the peak of the channel B12 weighting function (1.5 mb). Contours are plotted at values of 0 (bold), ± 1, ± 2, ± 3, ± 4, ± 5 and then at multiples of 5 r.u. Positive values are solid, negative values dashed. (from Al-Ajmi et al, 1985).

approximately equal to 1 K). Sudden warmings in the northern hemisphere are usually even stronger than those in the southern hemisphere.

The temperature changes result in a reversal of the latitudinal gradient of zonal mean temperature, and also in a deceleration of the zonal mean westerly zonal wind. According to W.M.O definitions, the warming is 'major' if the winds reverse to become easterly (below 10 mb and polewards of 60°N) and 'minor' if not. Major warmings occur on average about once every two years in the northern hemisphere, but have never been observed in the southern hemisphere, whilst strong minor warmings occur frequently in both hemispheres.

Prior to a warming, the stratospheric circulation is a strong westerly flow around a cold pole, and highest temperatures exist in mid-latitudes. This is then perturbed, leading to a 'wave-1' warming (where the vortex is displaced off the pole), or a 'wave-2' warming (where the vortex is split, resulting typically in two cyclonic vortices separated by a warm anticyclone on or near the pole). Usually, the maximum warming occurs significantly earlier in the upper stratosphere at high latitudes, and is followed by downward propagation of the perturbation into the lower layers. After some sudden warmings the temperature falls again and the zonal winds accelerate, with the zonal mean structure reverting to roughly its previous form. Towards the end of winter, however, some warmings lead directly into the changeover to summer conditions of warmer polar temperatures and climatological easterlies: such warmings are called 'final warmings'.

During warmings there is a large exchange of material between high and low latitudes, and such poleward advection of air from low latitudes is indicated by maps of Ertel's potential vorticity on isentropic surfaces. Provided that the flow is adequately resolved by measurements, contours of potential vorticity on

such surfaces can be taken as material lines for about a week or so in the middle stratosphere. An example is Figure 2.10: the area of low potential vorticity is drawn around the westerly vortex into the polar cap from low latitudes. These maps are a useful tool for studying sudden warmings, but the ability of satellite observations to resolve small-scale features of the temperature field imposes a constraint on their use. This is discussed further in Section 2.3. Maps of isentropic potential vorticity also demonstrate 'wave breaking' (McIntyre and Palmer, 1983, 1984 and 1985). This name is applied to large-scale large-amplitude disturbances in which non-linear advection causes material lines to be irreversibly deformed, rather than simply undulating back and forth as is assumed in linear wave theory. This deformation and buckling of the material lines leads to a cascade of energy to smaller scales. The long tongue of potential vorticity in Figure 2.10 is an example of wave breaking.

2.2.3. Equatorial Waves

A distinct category of wave is observed in the tropics. These 'equatorial waves' are of large zonal scale, but are confined in latitude about the equator. They have large horizontal phase speeds and short vertical wavelengths (typically 10 to 20 km). Because of their small vertical wavelength these waves are easier to observe by satellite using the limb scanning technique, which has a better vertical resolution than nadir-viewing instruments.

The first equatorial waves to be identified, using radiosonde data, were eastward propagating Kelvin waves and the westward propagating Rossby gravity wave. It is presumed that both these waves are excited by convective heating in the tropical convergence zone. Both of these waves are confined to the tropics and propagate vertically with wavelengths of about 10 km. They have periods of 10 to 20 days, corresponding to a phase speed of 23 to 46

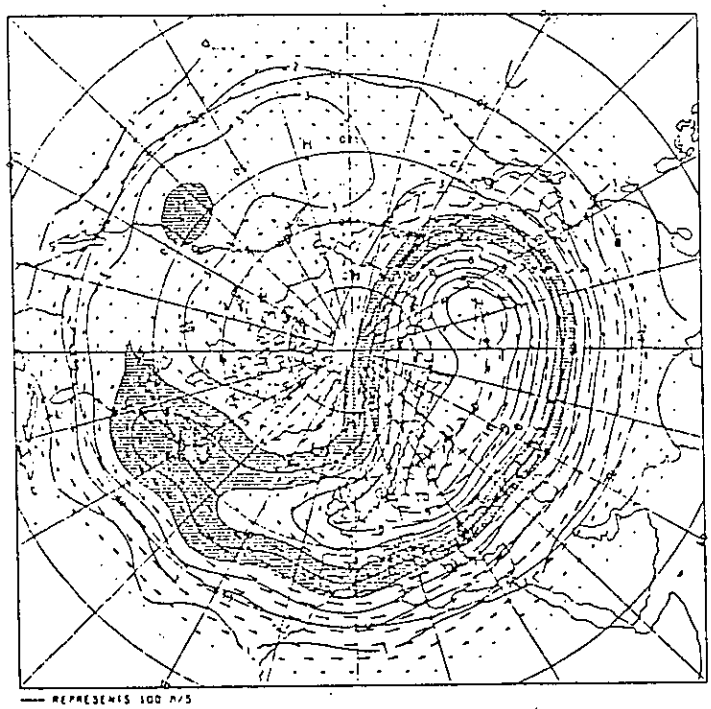


Figure 2.10 Ertel's potential vorticity and geostrophic winds evaluated on the 850 K isentropic surface on 4 December 1981 using data from the NOAA-6 satellite. Units : $Km^2kg^{-1}s^1 \times 10^4$. (from Clough et al, 1985).

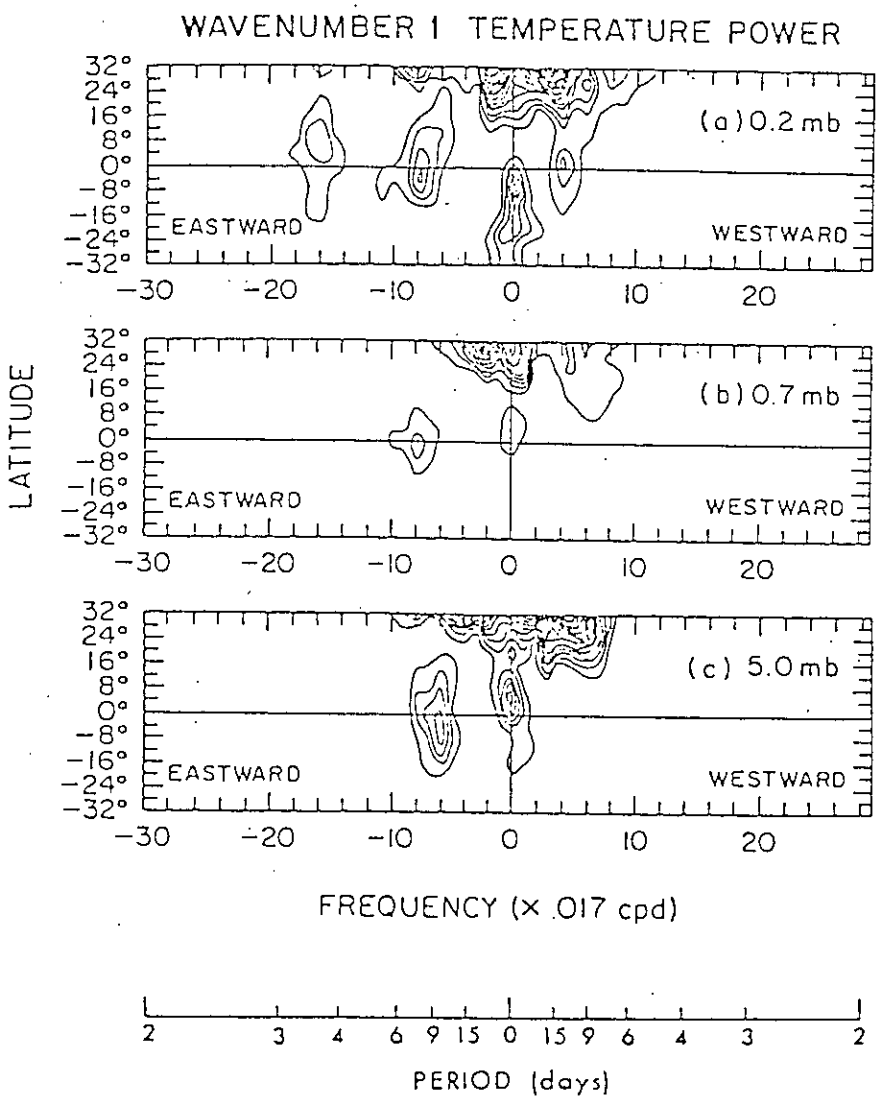


Figure 2.11 Temperature power for wavenumber 1 as a function of frequency and latitude for the January-February sample period at a) 0.2 mb, contour increment = 0.04 K^2; b) 0.7 mb, contour increment = 0.10 K^2; c) 5.0 mb, contour increment = 0.02 K^2. (from Salby et al, 1984)

ms^{-1}. In addition to these waves faster eastward propagating waves have been recorded at upper stratospheric levels. These were first identified from rocketsonde data by Hirota (1978). These waves correspond to Kelvin waves with periods of 5 to 10 days, a wavenumber 1 phase speed of 46 to 92 ms^{-1}, and vertical wavelengths of about 20 km. Hirota (1979), using Nimbus-5 SCR data, showed the period of the fast Kelvin waves to have a maximum in January and July and a minimum in March and September.

The fast Kelvin waves observed by Hirota, and other even faster waves, have been recorded by Salby et al (1984) using satellite limb soundings from the Nimbus 5 LIMS instrument. Figure 2.11 shows temperature power for wavenumber 1 at 3 pressure levels. An eastward moving wave with a period between 6.7 and 8.6 days (corresponding to phase speeds between 54 and 69 ms^{-1}) exists in the tropics at all 3 pressure levels, while a faster eastward disturbance with a period of 3.5 to 4.0 days (corresponding to phase speeds between 115 and 135 ms^{-1}) is observed at 0.2 mb. Salby et al also used plots of temperature power (not shown) to identify wavenumber 2 disturbances with periods of 6.0 to 7.5 days (corresponding to a phase speed of 31 to 39 ms^{-1}) and 3.8 to 4.3 days (corresponding to a phase speed of 55 to 62 ms^{-1}).

2.2.4. Gravity Waves and Tides

Tides, and gravity waves in particular, are believed to play an important role in determining the large-scale circulation and temperature structure in the middle atmosphere. The sources of gravity waves are not completely understood. However, they are thought to be forced in the lower atmosphere by frontal disturbances, orographic forcing, convective activity, geostrophic adjustment and shear instabilities (for a review of middle atmosphere gravity wave observations and theory see Fritts, 1984). Gravity waves typically have

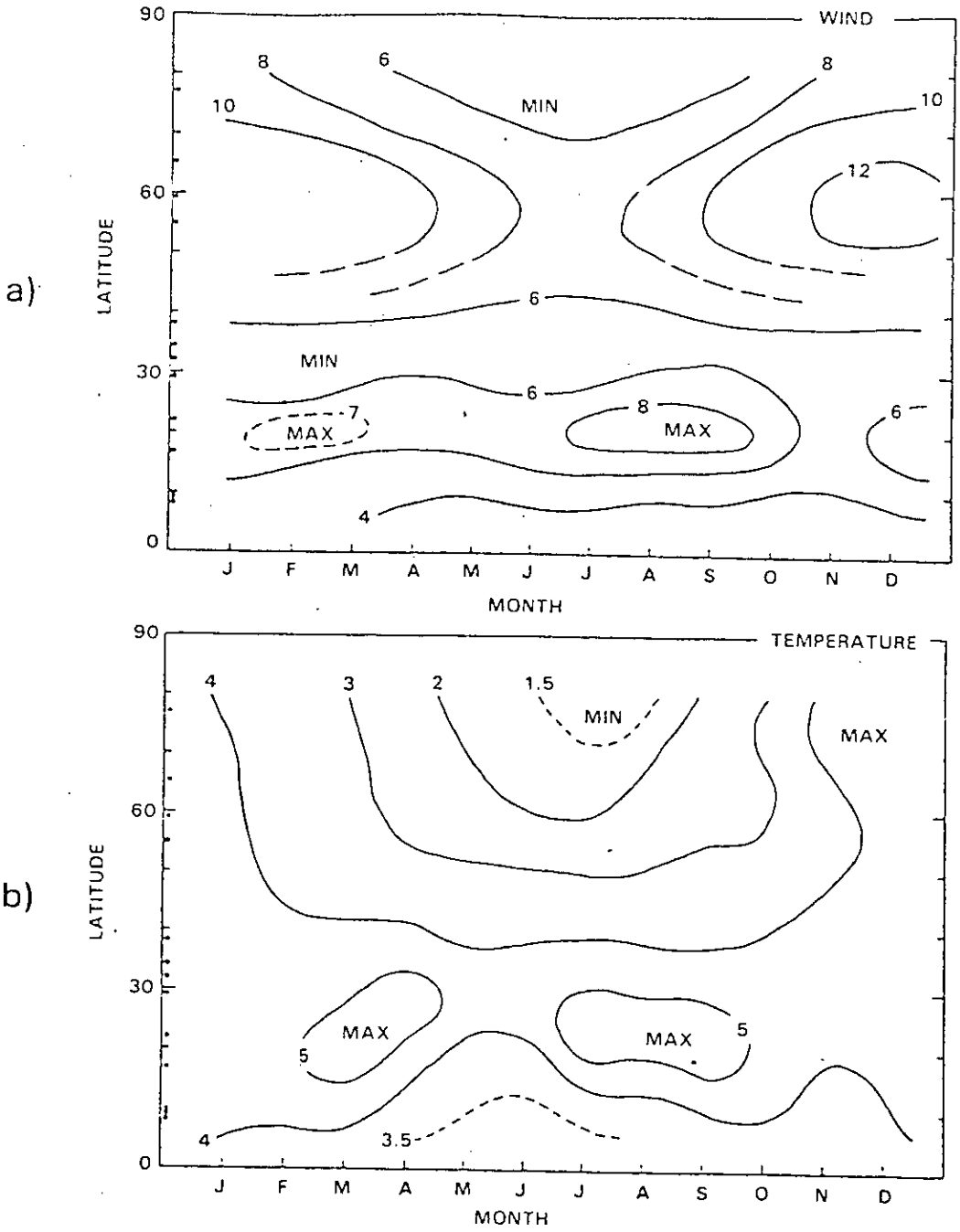


Figure 2.12 Latitude-time cross-sections of r.m.s fluctuations in 'wave intensity'. a) wind (units: $ms^{1}km^{-2}$); b) temperature (units: Kkm^{-2}) (from Hirota, 1984). The values are obtained by first applying a high pass filter to rocketsonde measurements; after filtering, only fluctuations with characteristic scales of about 10 km remain. A measure of 'wave intensity' is then obtained by integrating the second derivative of the filtered values over the height range of the observations.

horizontal wavelengths of 10 to 1000 km and vertical wavelengths of about 1.6 km near 60 km, rising to 3 km at 100 km. There are only a few indirect estimates of gravity wave phase velocities in the middle atmosphere, but what measurements there are give values in the range 20 to 100 ms^{-1} for mesospheric waves (Vincent and Reid, 1983). Periods are believed to be of a few hours or less. The waves have a comparatively small amplitude in the lower stratosphere, but because of their high frequency and rapid vertical propagation they are attenuated relatively little as they propagate upwards and thus have large amplitudes in the mesosphere, since the temperature amplitude of unattenuated waves is inversely proportional to the square root of density. Observations show the gravity wave spectrum in the middle atmosphere to have pronounced latitudinal, seasonal and temporal variability. For example, an indication of a semi-annual cycle in low latitudes (both in wind and in temperature) is shown in Figure 2.12 from Hirota (1984), who used data from Meteorological Rocket Network stations in the northern hemisphere.

Solar tides are thermally forced in the troposphere through infrared absorption by water vapour and in the stratosphere by ultraviolet absorption by ozone (for a review of tidal theory and observations see eg Lindzen, 1979; Forbes, 1984). Typically excitation will take place over local scales (about 1000 km), whilst the vertical scale of the tides is between 10 and 20 km. In the height range 25 - 45 km amplitudes are small and diurnal tides do not seem to play a significant role in the dynamics of the region. On the other hand, tides have larger amplitudes in the mesosphere, and have a more important role in the dynamics of that region.

2.3. Ability to Retrieve and Analyse Stratospheric Wave Motions Using Satellite Measurements

In this thesis methods of retrieving and analysing stratospheric temperatures from satellite spectral radiances are tested. The radiance measurements are made by radiometers which are mounted on the TIROS-N series of polar-orbiting satellites and which view the Earth at angles close to the local vertical. The vertical resolution of radiometer measurements is limited by the half-width of the channel weighting functions. For example, the Stratospheric Sounding Unit (SSU), which sounds exclusively in the stratosphere, has a vertical resolution of about 10 km according to statistical estimates (Clough et al, 1985). This resolution is usually sufficient to retrieve transient and stationary planetary waves. However, radiosonde and rocketsonde measurements suggest that dramatic changes in temperature can occur within quite a thin layer of atmosphere during sudden warmings. In addition, when waves are 'breaking', differential advection can cause, for example, warm layers to overlay cold layers, leading to large temperature shears which cannot be resolved by the SSU. Both wave breaking and sudden warming phenomena are frequently studied using isentropic maps of Ertel's potential vorticity, which is calculated using the vertical gradient of temperature. Hence the inability of satellite measurements to resolve large shears imposes a limit on the usefulness of isentropic potential vorticity maps. The vertical wavelengths of gravity waves and most equatorial waves are too small to be resolved by nadir sounding radiometers. However, limb sounding instruments, with their narrower weighting functions, have provided an opportunity to observe equatorial disturbances.

Since satellite observations are made asynoptically, their analysis presumes some form of time/space interpolation. The TIROS-N series satellite orbits the

Earth 14 times in a day, and thus the smallest scale wave that can be analysed in the zonal direction is of wavenumber 6. In general, quasi-stationary planetary waves are satisfactorily analysed, since by Charney-Drazin's linear theory, the smallest waves will be wavenumber 3. Transient waves with frequencies of up to 0.5 cpd can, in principle, be resolved. If one attempts to identify these waves by calculating power spectra, faster, unresolvable, waves can distort the spectra. Such distortions are caused by: 1) aliasing, where frequencies too high to be resolved appear falsely as low frequency components; and 2) leakage, in which frequencies present which do not exactly match with Fourier frequencies cause the energy associated with them to spill over into neighbouring bands. If the power spectra of the resolvable waves are weak then they will be masked by the distortions caused by the unresolvable waves. However, as has been described above, satellite measurements have been successfully used to identify a number of extratropical transient waves. In sudden warmings, where the flow is highly non-linear, the horizontal scales of motion often become too small to be adequately resolved. Moreover, the use of isentropic maps of Ertel's potential vorticity implies not only a high degree of spatial variance, but also a high degree of temporal variability through the advection of parcels by the flow field. It has been noted that these maps only offer a coarse grain view of reality, as the satellite measurements are unable to resolve the smallest scale features in either time, the horizontal or the vertical.

Observations from a single sun-synchronous orbiting satellite (such as TIROS-N) can be used to determine day/night differences in stratospheric temperature (eg Pick and Brownscombe, 1981). However, it is important to note that diurnal tides are not resolved by observations from a single satellite in such an orbit. This is because the orbit drifts westward at the same rate as the feature and thus views the same relative point with each latitude crossing (global observations of such phenomena require the use of measurements from

two or more satellites (eg Brownscombe et al, 1985)). Since the satellite views the same relative point of the tide *, this may lead to a bias in the measurement of zonal mean temperature. However, since stratospheric tides are small, this effect is probably also small in the stratosphere.

To summarise, nadir sounding instruments mounted on a polar-orbiting satellite are able to observe quasi-stationary planetary waves and transient extratropical waves. However, they are unable to resolve equatorial waves, gravity waves and tides, which all have a small vertical scale. For our purposes the inability to observe gravity waves and tides is not critical, as we are interested in the stratosphere; these phenomena are of more importance in the mesosphere. Vertical shears associated with sudden warmings are often too small to be resolved by radiometers, and in addition the temporal and horizontal variability present in sudden warmings is sometimes too small to be resolved by the satellite observational pattern. Hence in this thesis methods of retrieving and analysing satellite measurements are chiefly tested using fields which are affected by a sudden warming, as this phenomenon provides a stringent test of such methods.

* As an illustration of this point consider observations made at the equator, where the local time difference between ascending and descending node observations is 12 hrs. Since 12 hrs is equal to half the period of the diurnal tide, successive ascending and descending node observations view parts of the tide that are 180^0 out of phase, and hence the biasing effects of the tide cancel out. On the other hand, successive ascending and descending node observations of the semi-diurnal tide (which has a period of 12 hrs) view the same relative point of the tide, and this gives rise to a bias in the observed zonal mean temperature.

CHAPTER 3

PRINCIPLES OF THE REMOTE SOUNDING OF TEMPERATURE

Before performing tests of retrieval and analysis schemes it is important to describe the theory of radiative transfer on which temperature remote sounding is based. This theory appears in Section 3.1, together with a list of criteria used to determine which absorption bands are useful for temperature sounding, whilst Section 3.2 describes the satellite instruments used to observe radiation from these bands. In order to accurately retrieve temperature from satellite measurements it is important to accurately calculate the transmission profile (and hence weighting function) of each instrument. Details of such calculations appear in Section 3.3. Section 3.4 contains a description of current retrieval and analysis methods: the subsection on retrieval describes a number of commonly used methods (including multiple linear regression, a technique which is tested in Chapter 6), and discusses the error characteristics of retrievals. Such a discussion is important when evaluating results of tests of retrieval methods; the subsection on analysis contains an overview of a variety of analysis methods followed by a description of the two techniques to be tested in this thesis – time/space interpolation and sequential estimation of Fourier components.

3.1. Radiative Transfer Theory

At any frequency in the infrared or microwave regions where an atmospheric constituent absorbs radiation, it also emits radiation according to Kirchhoff's law. The electromagnetic radiation leaving the top of the atmosphere is a function of the distribution of the emitting gas and the distribution of temperature throughout the atmosphere. If we choose to examine radiation from gases such as carbon dioxide or oxygen, which are nearly uniformly mixed, then some details of the temperature structure may be deduced; this is

a principle exploited in several remote sounding satellites. The discussion in this chapter assumes that we observe radiation leaving the atmosphere in directions near to the local vertical (although other satellite instruments, 'limb sounders', observe the radiation leaving the atmosphere nearly tangentially). Thermal emission comes from (and thus has the characteristic temperature of) a region of the atmosphere about 10 – 15 km thick whose altitude depends on the absorption coefficient, and the higher the absorption coefficient the higher this emitting layer is in the atmosphere. By making measurements in several spectral regions a range of altitudes can be sounded.

Several criteria have to be satisfied for a particular absorption band to be employed usefully for temperature sounding.

1. As indicated above, the emitting constituent should be uniformly mixed in the atmosphere so that the radiance measurement gives information about the atmospheric temperature structure only.

2. The absorption should not be overlapped by absorption bands of other constituents.

3. Local thermodynamic equilibrium should apply, so that emission from the band is proportional to the Planck function. As altitude increases, local thermodynamic equilibrium becomes less of a good assumption: the assumption becomes poor at $>$ 100 km for the 5 mm O_2 band, and at $>$ 80 km for the 15 μm CO_2 band.

4. The wavelength should be longer than about 4 to 5 μm to ensure that scattered solar radiation is insignificant compared to thermal emission.

We now consider the radiative transfer theory on which the idea of temperature sounding is based (Kaplan, 1959; Houghton and Smith, 1970). One considers first a slice of an infinitely deep horizontally stratified atmosphere. The slice has a temperature T and path length du of absorber in the vertical direction. At frequency ν, with an absorption coefficient k_ν, Kirchhoff's law

states that the emitted radiance from this slice in the vertical direction will be k_ν du B_ν (T), where B_ν (T) is the Planck function at frequency ν and temperature T, and is given by

$$B_\nu (T) = c_1 \nu^3 / \{ \exp (c_2 \nu / T) - 1 \} \qquad (3.1)$$

where c_1 = 1.19096 x 10^{-5} mW m^{-2} cm^4 $ster^{-1}$ and c_2 = 1.43879 cm K. Lambert's law states that a proportion, τ_ν, will reach the top of the atmosphere, and this is expressed as

$$\tau_\nu = \exp (- \int k_\nu \, du) . \qquad (3.2)$$

This integral is over the region between the slice and the top of the atmosphere. Integrating over all such 'slices' leads to the total radiance R_ν at the top of the atmosphere. Thus, assuming negligible transmission from the Earth's surface,

$$R_\nu = \int B_\nu (T) \, k_\nu \, du \exp (- \int k_\nu \, du) . \qquad (3.3)$$

Alternatively, (3.3) can be written as

$$R_\nu = \int_0^1 B_\nu (T) \, d\tau_\nu . \qquad (3.4)$$

It is convenient to use an altitude-dependent coordinate. We use $\eta = - \ln (p/p_o)$, where p is pressure and p_o is a reference pressure. Thus

$$R_\nu = \int_0^\infty B_\nu (T) (d\tau_\nu / d\eta) \, d\eta , \qquad (3.5)$$

or, setting $K (\eta) = d\tau_n / d\eta$,

$$R_\nu = \int_0^\infty B_\nu (T)\, K (\eta)\, d\eta \qquad (3.6)$$

where K (η) is the weighting function. Equations (3.4), (3.5) and (3.6) are versions of the Radiative Transfer Equation.

3.2. Instruments used for Stratospheric Temperature Retrievals

In this thesis we use satellite radiances measured by the TIROS Operational Vertical Sounder (TOVS) which flies on the TIROS-N series of polar-orbiting satellites. TOVS consists of three instruments - the High Resolution Infrared Radiation Sounder (HIRS), the Microwave Sounding Unit (MSU) and the Stratospheric Sounding Unit (SSU). The HIRS and SSU channels used for stratospheric temperature retrievals measure infrared radiation from the 15 μm emission band of CO_2, whilst the MSU channels measure microwave radiation from the 5 mm band of O_2. These emission bands satisfy the four criteria required for temperature retrieval listed above.

Both HIRS and SSU instruments measure radiation from the 15 μm emission band of CO_2. Many emission lines on this band have widths of less than 0.1 cm^{-1}. The conventional spectroscopic techniques used in the HIRS instrument, for example, are unable to resolve details of the structure near the line centres and are therefore unable to select radiation from spectral regions where the absorption coefficient is very high. Thus HIRS channel weighting functions have peaks in the lower stratosphere. To perform temperature sounding of higher levels (eg the mid and upper stratosphere) it is necessary to resolve details of these individual lines and to use an instrument which is sensitive to emission only from a narrow range of frequencies close to the centre of one such spectral line. The SSU achieves this using the technique of pressure modulation.

The SSU is a version of the Pressure Modulator Radiometer (PMR) (Taylor et al, 1972; Miller et al, 1980; Pick and Brownscombe, 1981). Its optical system is shown in Figure 3.1. Filters ensure that the 15 μm band of CO_2 radiation is passed but that all other incoming radiation is excluded. The PMR admits radiation through a single cell of CO_2 whose pressure is increased and reduced by a piston. In effect the CO_2 acts as the spectrally selective element: as the pressure in the cell varies, the transmission of the cell and hence the radiation falling on the detector is modulated only at wavenumbers which lie within the absorption lines of the gas. Hence the pressure modulator is sensitive to energy at the required wavenumbers, but rejects the rest. By using the pressure modulator technique, the SSU provides weighting functions peaking at heights of up to 1.5 mb (about 45 km).

The MSU, which sounds in the microwave region of the electromagnetic spectrum, is the third instrument mounted on TOVS. An advantage of microwave sounders over radiometers which observe infrared emission in the CO_2 bands is that clouds are substantially transparent in the microwave region (although this is of little importance in the stratosphere). In addition the MSU has a very high spectral resolution, so its weighting functions are essentially monochromatic. On the other hand the field of view of any microwave instrument cannot be as narrow as that of an infrared radiometer (unless antenna size is increased) because of the diffraction limit at the longer wavelength. In addition, for the same accuracy of temperature measurement the accuracy of the radiance measurement has to be higher (for example a 1 K temperature difference leads to a change of about 0.33 % in radiance at 5 mm (microwave region) but a change of about 1 % at 15 μm (infrared region)).

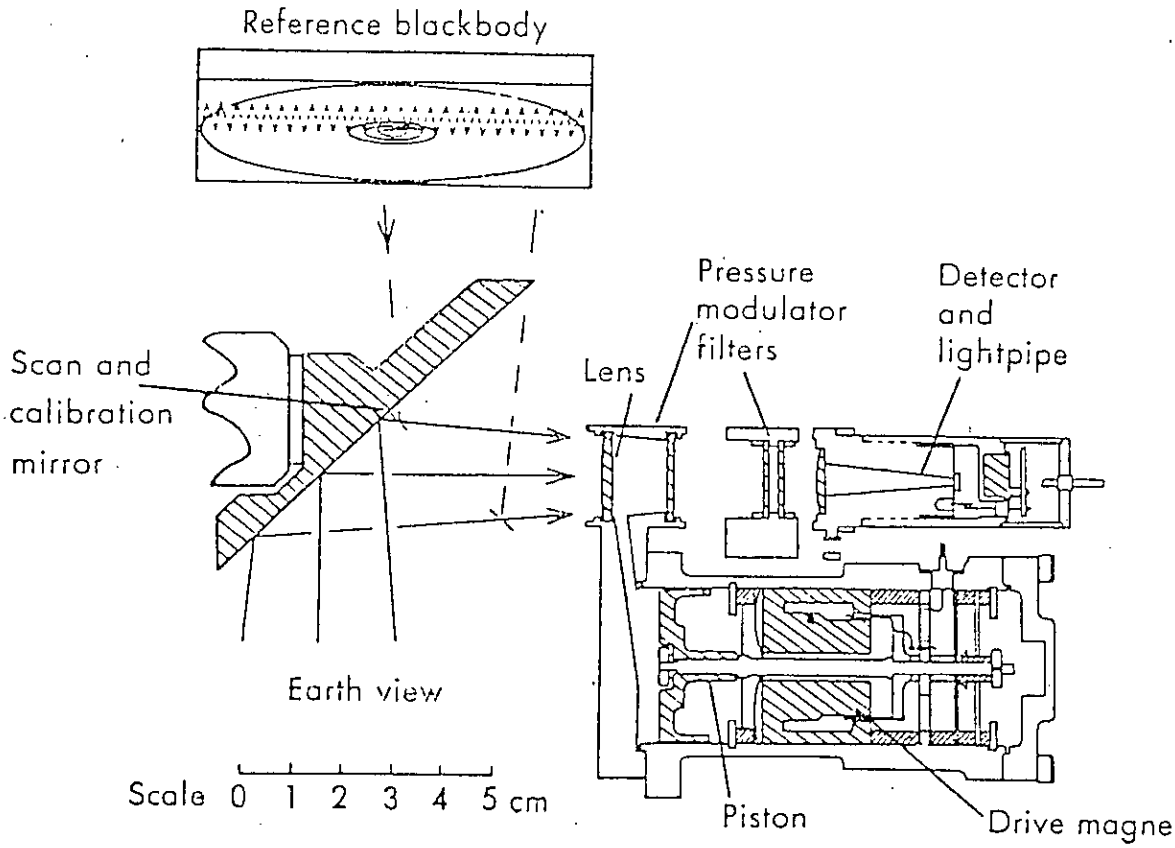


Figure 3.1 Pressure modulator and SSU optical system (from Pick & Brownscombe, 1981).

3.3. Calculation of Instrument Weighting Functions

In order to retrieve Planck function (and hence temperature) from satellite radiance measurements, it is necessary to know the weighting functions from each channel. It is important that these weighting functions are calculated accurately; the calculation should not introduce errors which contribute appreciably to the errors in the retrieval process. Calculations of transmission profiles, and hence weighting functions, require knowledge of the optical characteristics of the radiometer and detailed spectral information regarding absorption line positions, intensities, widths and shapes.

Th first step is to calculate transmission profiles for each channel using detailed spectral line data. For example, transmission profiles in the 15μm CO_2 region are calculated using the line-by-line method (Drayson, 1966), which sums the absorption coefficients in very narrow spectral intervals over each contributing line and then integrates over the atmospheric path. Because of possible uncertainties in line strengths, shapes and positions, the numerical method, and in instrumental response, it is essential to verify these calculated transmittances against measurements. This is especially important in the case of the SSU, since the filling pressure of the pressure modulator cell changes while the SSU is on the ground awaiting launch. These changes are caused by CO_2 outgassing from the cell, and by air leaking into the cell. Before the satellite launch, therefore, line-by-line calculations were compared with laboratory measurements made by the SSU of the transmission of a path through CO_2 at constant pressure (Pick et al, 1976). The result of this comparison is the specification of an 'effective mean pressure', which is the value of mean cell pressure needed for the line-by-line calculations to agree with experimental measurements at a transmission value of 0.5. The experimental measurements also show a small systematic shape bias in the transmission against pressure profile, compared to the line-by-line calculations.

Hence, before being used in the retrieval scheme, the transmission profile must be calculated at the effective mean pressure using the line-by-line methods, and then the systematic bias revealed by the experimental measurements must be removed.

Laboratory measurements are a very useful way of verifying instrument weighting functions. However, they have the disadvantage that the variable pressure path in the atmosphere cannot be simulated. In principle, a closer correspondence between measurement and atmospheric weighting functions can be obtained by making measurements of the transmission of solar radiation through the atmosphere from a balloon platform. This method was used by Batey and Abel (see Houghton, 1972) to test the Nimbus 4 Selective Chopper Radiometer and by Pick and Barwell (1978) to test the SSU. The latter tests demonstrated that the calculation of atmospheric transmission profiles for the SSU is within the experimental accuracy of the measurement.

Since transmission profiles vary with temperature and pressure we need to calculate a separate transmision profile for each retrieval. However, accurate line-by-line spectroscopic techniques are computationally too expensive for such an application. Instead, spectral line data are used to calculate transmission profiles for a small number of representative and extreme atmospheres, and these pre-calculated profiles are then interpolated to any arbitrary temperature profile using a fast numerical model. The fast model used in this thesis was developed by McMillin and Fleming (1976). This model is appropriate to observations made at the local vertical and to gases with a constant mixing ratio. However, the approach has been extended to account for slant path observations (Fleming and McMillin, 1977), and gases with variable mixing ratios (McMillin et al, 1979). Comparison between transmission profiles calculated by the line-by-line method and by McMillin and Fleming's fast model

(McMillin and Fleming, 1976) show the maximum absolute error to be 0.0031, and the maximum r.m.s error (over all pressure levels) to be 0.0011. Such errors are smaller than probable errors in the line-by-line technique caused by uncertainty in basic spectroscopic parameters, and it is concluded that the accuracy of the fast model is adequate for most operational sounding applications.

3.4. Current Retrieval and Analysis Methods

The inference of temperature from satellite measurements can be thought of as having two parts, though it is possible to devise procedures in which they are combined. They are 1) 'retrieval' in which a single temperature profile is calculated from a more-or-less instantaneous set of observations, and 2) 'analysis' in which the state of the atmosphere at a given instant is deduced on a regularly spaced grid of points from the retrieved profiles which are asynoptic and distributed according to the shifting satellite orbit. Below we present a review of several well-used methods.

3.4.1. Retrieval Methods

3.4.1.1. Introduction

In this section we discuss retrieval theory (for a more comprehensive discussion see, for example Rodgers (1976b)). The inferral of Planck function (and hence temperature) from radiance is a mathematically underconstrained problem since we are trying to determine a continuous variable from a finite number of measurements. The first step of the solution is to represent the continuous variables (Planck function and weighting function) in the radiative transfer equation (3.6) discretely at a large number of levels (typically 30 to 50). The problem is now to calculate the unknown profile at these levels from a smaller number of radiance measurements (typically 4 to 8). The radiative

transfer equation is rewritten in vector form as

$$\mathbf{y} = \mathbf{K}\,\mathbf{x} \tag{3.7}$$

where **y** is a vector of channel spectral radiances, **K** is a discretised form of the weighting function, and **x** is a vector representing Planck function at a number of pressure levels. In our discussion of retrieval methods below, we also assume that the retrieval of Planck function from radiance is a linear problem (ie **K** does not depend on **x**). Since the number of radiance measurements is smaller than the number of levels at which we wish to calculate the unknown profile, the problem is still underconstrained, and it is possible to calculate an infinite number of solutions from one set of radiance measurements. The radiance measurements must therefore be supplemented by enough constraints to make the problem well posed. One way of doing this is by specifying a 'first guess' estimate of the unknown profile, and its error covariance. This estimate is usually based on climatological or forecast profiles.

The forecast or climatological estimate of the unknown profile can be regarded as another measurement of the profile – a 'virtual measurement' – and its error covariance as the uncertainty of this measurement. If there are two measurements of any vector **x**, namely $\mathbf{x}_1$ and $\mathbf{x}_2$, they can be combined by weighting each with the inverse of their respective measurement error covariances $\mathbf{S}_1$ and $\mathbf{S}_2$. The combined estimate $\hat{\mathbf{x}}$ is thus

$$\hat{\mathbf{x}} = (\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1})^{-1} (\mathbf{S}_1^{-1} \mathbf{x}_1 + \mathbf{S}_2^{-1} \mathbf{x}_2) \tag{3.8}$$

and the error covariance of this estimate is

$$\hat{\mathbf{S}} = (\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1})^{-1} . \tag{3.9}$$

The full derivation of equations (3.8) and (3.9) appears in Appendix I. In the

case of the retrieval problem, the radiance measurement, $\mathbf{y} = \mathbf{K}\,\mathbf{x}$, is combined with the virtual measurement which is denoted by $\mathbf{x}_o$, and has a error covariance S_x. We wish to re-express equations (3.8) and (3.9) in terms of these measurements. We assume that $\mathbf{D}\,\mathbf{y}$ is an estimate of $\mathbf{x}$, and that $\mathbf{D}$ is any exact solution such that $\mathbf{K}\,\mathbf{D} = \mathbf{I}$, where $\mathbf{I}$ is the unit matrix. Equation (3.8) thus becomes

$$\hat{\mathbf{x}} = (S_x^{-1} + S_{Dy}^{-1})^{-1} (S_x^{-1} \mathbf{x}_o + S_{Dy}^{-1} \mathbf{D}\,\mathbf{y}) \tag{3.10}$$

where S_{Dy} is the covariance of $\mathbf{D}\,\mathbf{y}$ and is

$$S_{Dy} = \mathbf{D}\, S_\varepsilon\, \mathbf{D}^T$$

where S_ε is the error covariance matrix of the radiance measurement, often referred to as 'instrumental noise'. Instrumental noise from one channel will often be uncorrelated with noise from other channels. The inverse of S_{Dy} is

$$S_{Dy}^{-1} = \mathbf{D}^{T\,-1}\, S_\varepsilon^{-1}\, \mathbf{D}^{-1} \quad .$$

Substituting into equation (3.10) gives

$$\hat{\mathbf{x}} = (S_x^{-1} + \mathbf{D}^{T\,-1}\, S_\varepsilon^{-1}\, \mathbf{D}^{-1})^{-1} (S_x^{-1}\ \mathbf{x}_o + \mathbf{D}^{T\,-1}\, S_\varepsilon^{-1}\, \mathbf{y}) \tag{3.11}$$

Since $\mathbf{K}\,\mathbf{D} = \mathbf{I}$, equation (3.11) can be rewritten as

$$\hat{\mathbf{x}} = (S_x^{-1} + \mathbf{K}^T\, S_\varepsilon^{-1}\, \mathbf{K})^{-1} (S_x^{-1}\, \mathbf{x}_o + \mathbf{K}^T\, S_\varepsilon^{-1}\, \mathbf{y}) \tag{3.12}$$

and equation (3.9) is rewritten as

$$\hat{S} = (S_x^{-1} + \mathbf{K}^T\, S_\varepsilon^{-1}\, \mathbf{K})^{-1} \quad . \tag{3.13}$$

Equations (3.12) and (3.13) are not useful for practical computation because

they require the inversion of large matrices. By making use of a matrix identity, however, they can be written in a more convenient form, viz

$$\hat{x} = x_o + S_x K^T (K S_x K^T + S_\varepsilon)^{-1} (y - K x_o) \qquad (3.14)$$

and

$$\hat{S} = S_x - S_x K^T (K S_x K^T + S_\varepsilon)^{-1} K S_x \ . \qquad (3.15)$$

Full details of the algebraic steps required to rearrange equations (3.12) and (3.13) into the form of equations (3.14) and (3.15) appear in Appendix I.

Temperature retrieval may be carried out in two ways: the minimum information method and regression. Minimum information methods retrieve temperature by inverting the radiative transfer equation, with x_o explicitly being the optimal 'first guess' estimate of the unknown profile. In Section 3.4.1.2 we derive equations (3.14) and (3.15) using two methods: 1) Maximum Likelihood chooses from the infinite number of possible solutions which are consistent with the observations the one which is most probable; 2) Minimum Variance finds some linear combination of the observations such that the expected value of the error variance of the estimate is minimised. This more general approach of inverting the radiative transfer equation therefore gives more insight into the nature of the solution. Regression relates Planck function (and hence temperature) to radiance using a set of regression coefficients which are calculated from a statistically representative sample of co-located sonde and satellite measurements. Here x_o is the mean of this statistical sample. Regression is described in Section 3.4.1.3, and an analysis of the errors of both minimum information and regression methods is presented in Section 3.4.1.4.

3.4.1.2. Minimum Information Methods

The optimum solution for $\hat{\mathbf{x}}$ is obtained by a variety of seemingly different methods, and two commonly used methods are described below.

a) Maximum Likelihood Solution

The maximum likelihood solution is found by choosing from the infinite number of possible solutions which are consistent with the observations the one which is most probable. This is done by maximising the conditional probability density function $P (x|y)$ of the solution x given the observation y. Bayes theorem states

$$P (x|y) = P (y|x) P (x) / P (y)$$

where $P (x)$ and $P (y)$ are the probability density functions of x and y, respectively, and $P (y|x)$ is the conditional probability density function of y given x. If there were no instrumental noise $P (y|x)$ would be a delta function centred at $y = K x$, the theoretical value of the observation. However, the presence of noise turns it into a Gaussian distribution centred at the same place. $P (x)$ can be estimated from rocketsonde and radiosonde observations, whilst $P (y)$ is not required, as it is constant when we maximise with respect to x. We must now assume some algebraic form for $P (x)$, and the most convenient is a Gaussian distribution. Examination of the available data reveals that if all latitudes and all seasons are considered together the statistics are by no means Gaussian. However, smaller regions of space and time give distributions which are closer to Gaussian.

Assuming, therefore, that $P (x)$ has a Gaussian distribution, then we write

$$P (x) \alpha \exp \{ - 1/2 (x - x_o)^T S_x^{-1} (x - x_o) \} \qquad (3.16)$$

where S_x is the error covariance of the mean profile x_o. Similarly, $P (y|x)$ is

written as

$$P (y|x) \alpha \exp \{ - 1/2 (y - K x)^T S_\varepsilon^{-1} (y - K x) \} \quad (3.17)$$

where S_ε is the instrumental noise error covariance. Equations (3.16) and (3.17) can be rewritten as

$$- 2 \ln \{ P (x) \} = (x - x_o)^T S_x^{-1} (x - x_o)$$

and

$$-2 \ln \{ P (y|x) \} = (y - K x)^T S_\varepsilon^{-1} (y - K x)$$

To find $\hat{x}$, the most likely value of x, we maximise P (x|y) with respect to x, or we minimise minus its logarithm, viz

$$d/dx \{ (\hat{x} - x_o)^T S_x^{-1} (\hat{x} - x_o) + (y - K \hat{x})^T S_\varepsilon^{-1} (y - K \hat{x}) \} = 0$$

This leads to

$$(\hat{x}^T - x_o^T) S_x^{-1} - (y^T - \hat{x}^T K^T) S_\varepsilon^{-1} K = 0$$

Rearranging, using the matrix identity $(A B)^T = B^T A^T$ gives

$$((S_x^{-1} + K^T S_\varepsilon^{-1} K)^T \hat{x})^T = (S_x^{-1\ T} x_o)^T + (K^T S_\varepsilon^{-1\ T} y)^T$$

leading to

$$(S_x^{-1} + K^T S_\varepsilon^{-1} K)^T \hat{x} = S_x^{-1\ T} x_o + K^T S_\varepsilon^{-1\ T} y$$

If A is a symmetric matrix, then $A^T = A$. Since the error covariance matrices are symmetric, we write

$$\hat{x} = (S_x^{-1} + K^T S_\varepsilon^{-1} K)^{-1} (S_x^{-1} x_o + K^T S_\varepsilon^{-1} y) \quad (3.18)$$

which is the same as equation (3.12), and the error covariance of this solution is thus given by equation (3.13).

b) Minimum Variance Solution

The principle behind the minimum variance methods is to find a linear predictor, such that the expected value of the variance of the error in the estimate is minimised. First, one must calculate *a priori* statistics from a sample of temperature profiles and radiance measurements. Next, the means of these observations are subtracted from the radiative transfer equation (3.7), and the equation is re-expressed as

$$\hat{x} - x_o = W (y - y_o) \tag{3.19}$$

where $\hat{x}$ is the estimated Planck profile, x_o is the mean Planck profile, y_o is the vector of radiances simulated from x_o, given by

$$y_o = K x_o$$

and W is a linear predictor, which is calculated by minimising the variance of the estimate error, ie we minimise

$$r = E [(x - x_o - W (y - y_o))^T (x - x_o - W (y - y_o)] .$$

We wish to find W such that dr/ d W = 0. Expanding terms and differentiating gives

$$E [-2 (\hat{x} - x_o) (y - y_o)^T + 2 W (y - y_o) (y - y_o)^T] = 0$$

Since $y - y_o = K (x - x_o) + \epsilon$, where ϵ is the measurement error, with covariance S_ϵ, then clearly $E [(\hat{x} - x_o) (y - y_o)^T] = S_x K^T$ and

$E [(y - y_o) (y - y_o)^T] = K S_x K^T + S_\epsilon$, and hence

$$W = S_x K^T (K S_x K^T + S_\varepsilon)^{-1} \qquad (3.20)$$

Substituting in equation (3.19) gives

$$\hat{x} = x_o + S_x K^T (K S_x K^T + S_\varepsilon)^{-1} (y - K x_o) \qquad (3.21)$$

which is the same as equation (3.14). By making use of a matrix identity (see Appendix I), equation (3.18) can be rearranged to give equation (3.21). Hence, provided Gaussian statistics are assumed, the maximum likelihood solution is the same as the minimum variance solution.

3.4.1.3. Regression

Planck function, and hence temperature, can also be retrieved from radiances by linear regression. Regression coefficients relating Planck function to radiance are deduced from regression analyses of large samples of radiance and temperature measurements. In the troposphere these comprise satellite-observed radiance and colocated radiosonde observations; in the stratosphere, due to the scarcity of *in situ* observations, the sample often has to consist of rocketsonde observations and radiances which have been simulated from these rocketsonde measurements. It is important that the sample is similar to the atmospheric conditions we are trying to achieve. The estimated deviation of profile Planck function, $\Delta \hat{x}_i$, from the mean of the sample is written as a linear combination of the deviation of the radiances from the mean radiance, namely

$$\Delta \hat{x}_i = a_{0,i} + \sum_{j=1}^{N} (y_j - \bar{y}_j) a_{i,j} \qquad (i = 1,...,M) \qquad (3.22)$$

where N is the number of channels and M is the number of levels at which we wish to retrieve Planck function. The retrieved Planck function, $\hat{x}_i$, for level i is given by

$$\hat{x}_i = \bar{x}_i + \Delta \hat{x}_i \qquad (i = 1,\ldots,M)\,. \tag{3.23}$$

Before using the regression model, the assumptions of a linear relation between Planck function and radiance, and of the number of channels of data to be used, must be examined. A detailed description and examination of the regression model we test in this thesis appears in Chapter 6. The brief, general discussion of regression in this section appears only so that we can relate regression, and the other retrieval methods described in this chapter, to the discussion of retrieval error characteristics in Section 3.4.1.4.

The regression technique has the advantage of being rapid and simple to use, and the regression coefficients can easily be updated to account for instrumental drifts. In addition, if the coefficients are calculated using radiance measurements and colocated sonde observations, then no knowledge of the channel weighting functions is required (when the radiances are simulated from rocketsonde measurements errors in the calculation of the weighting function will lead to errors in the regression coefficients). A major disadvantage of regression is that the technique relies on a large collection of radiance and sonde data, which if not representative of atmospheric conditions, can lead to large errors in the solution.

3.4.1.4. Error Characteristics of Retrievals

The discussion in this section is useful in helping to draw conclusions from tests of a retrieval scheme, such as those that are described in Chapter 6. Estimates of Planck function deduced from radiance measurements are sensitive to the 'first guess' of the atmospheric state which is used in the inversion process. Below we derive the error structure of individual retrievals (see, for example, Eyre, 1987), and thus exhibit how much of the total retrieval error is

due to errors in the 'first guess', and how much is due to observational errors. In the following discussion we ignore errors due to uncertainties in the calculation of the weighting functions (often referred to as the 'forward problem'). This is a reasonable assumption for our purposes, since in this thesis the retrieval method is examined in a simulation experiment, but in operational retrievals the effects of such errors are important and should not be ignored; forward model errors in TOVS channels are unlikely to be less than 0.2 K in equivalent blackbody temperature (Eyre, 1989, personal communication)). Linear inversions from satellite radiances can be expressed in the form of equation (3.19), viz

$$\hat{\mathbf{x}} - \mathbf{x}_o = \mathbf{W} (\mathbf{y} - \mathbf{y}_o) \tag{3.24}$$

where $\hat{\mathbf{x}}$ is the vector of retrieved atmospheric parameters, $\mathbf{x}_o$ is the 'first guess' value of the vector, $\mathbf{y}$ is the vector of channel radiance measurements, $\mathbf{y}_o$ is the corresponding vector appropriate to the first guess, and $\mathbf{W}$ is the 'inverse matrix'. In 'physical' inversion schemes, such as minimum variance or maximum likelihood methods, the first-guess vector $\mathbf{x}_o$ is usually the starting point. It is usually obtained from climatological values or from a numerical forecast model. In regression schemes $\mathbf{x}_o$ is implicit – it is the mean of data from which the regression coefficients are calculated.

The inverse matrix, $\mathbf{W}$, can be obtained by a variety of methods. One such method is the 'minimum variance' solution, which is described in Section 3.4.1.2. There an attempt is made to minimise the variance between $\hat{\mathbf{x}}$ and the true value of $\mathbf{x}$ over a large number of 'representative' cases. Another example of this type of solution is linear regression. The optimal solution to both minimum variance and linear regression methods is obtained by least squares analysis of a sample of rocketsonde and simulated radiance data. Thus, despite superficial appearances to the contrary, both linear regression and minimum variance

inversion of the radiative transfer equation are mathematically equivalent if they use the same first guess and make the same assumptions about the first-guess and measurement errors and the linearity of the problem.

Let us now look at the error characteristics of equation (3.24). The linear approximation to the forward radiative transfer equation is given by

$$y - y_o = K (x_T - x_o) + \varepsilon_m \tag{3.25}$$

where x_T is the vector of the true atmospheric parameters, and ε_m is the vector of measurement errors, which are assumed to be random, Gaussian and unbiased. Substituting equation (3.25) into equation (3.24) gives

$$\hat{x} - x_o = W K (x_T - x_o) + W \varepsilon_m$$

or

$$\hat{x} - x_o = R (x_T - x_o) + W \varepsilon_m \tag{3.26}$$

where

$$R = W K$$

Equation (3.26) may be rewritten in terms of the retrieval, first-guess and measurement errors as

$$\hat{x} - x_T = (I - R) (x_o - x_T) + W \varepsilon_m \tag{3.27}$$

where I is an identity matrix. The first term on the right hand side represents the mapping of the first-guess errors into retrieval errors, while the second maps the measurement errors into retrieval errors. In the case of the minimum variance solution W is given by equation (3.20), and hence R is expressed as

$$R = S_x K^T (K S_x K^T + S_\varepsilon)^{-1} K$$

If the first-guess error covariance is large enough to make $| K S_x K^T | >> S_\varepsilon$ then the value of R will be close to the identity matrix. From equation (3.27), this means that the first guess error is small, but there will be a high sensitivity to noise in the measurements. On the other hand, if the first guess covariance is small, the sensitivity to measurement noise is less (i.e smaller values of W), but the retrieval may have a larger bias due to a larger first-guess error. Reference to such points is made in the discussion of the error characteristics of the retrieval scheme tested in Chapter 6.

3.4.2. Analysis

3.4.2.1. Introduction

To be of meteorological use, we wish to represent the state of the atmosphere at a given instant on a regular grid of points. However satellite measurements are made asynoptically at locations determined by a shifting satellite orbit. Figure 3.2 illustrates the problem of 'analysing' observations on a grid. The Figure shows satellite 'superobservations' made by TOVS on the NOAA-7 satellite on 18/1/87 between -20° and 20° latitude, -160° and -120° longitude and 1020 and 1340 GMT (a description of what we mean by 'superobservations' appears in Section 5.3). The grid has a spacing of 5° in both latitude and longitude and we wish to interpolate the observations to the space/time gridpoints.

Many numerical analysis schemes were developed for use in numerical weather prediction models. Early schemes generally consisted of simple procedures for interpolation of observed data onto a regular network of grid points. These included the time/space interpolation scheme (described below), and the fitting of a polynomial surface to the observed data in the close vicinity of a gridpoint. Later model analysis schemes used a short-range numerical forecast ('first-guess') to improve the analysis. An example of such a scheme is

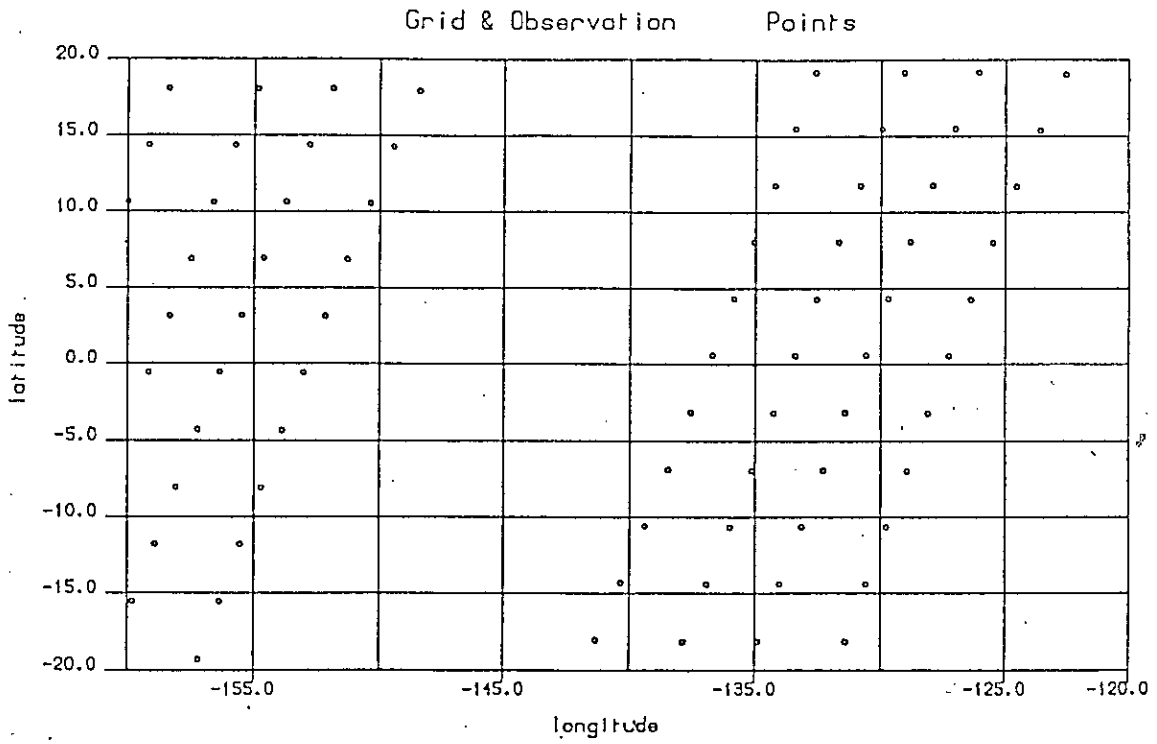


Figure 3.2 TOVS superobservation points on 18/1/87 between -20° and 20° latitude, -160° and -120° longitude, and 1020 and 1340 GMT. Also plotted is a grid with a spacing of 5°.

the successive correction method (Bergthorsson and Doos, 1955; Cressman, 1959), where the analysed field is expressed as a linear combination of the preliminary field at the gridpoint and the weighted observed deviations from the preliminary field. These weights are computed explicitly and the preliminary field is iteratively corrected to the observed values during several analysis scans. Successive correction has now been superceded by statistical interpolation (e.g Lorenc, 1981), which combines first guess and observed quantities in a way consistent with the estimated accuracy of each.

Such schemes have been developed specifically for numerical weather prediction models, where a short-range forecast is available. However, in studies of the stratosphere such a forecast does not usually exist. In this thesis we test two schemes commonly used to analyse stratospheric satellite data : 1) 'time/space interpolation' in which each observation is given a time and distance weight which decreases the further the observation point is from the gridpoint, and 2) 'sequential estimation', where Fourier components of the analysed field are deduced using a version of the Kalman Filter. These are described below.

3.4.2.2. Time/Space Interpolation

In the time/space interpolation scheme, each observation is given a time and distance weight. The weights decrease the further an observation is away from the gridpoint and analysis time. The description below assumes the weights to be linear; however other weights (e.g cosine, negative exponential) can be used. Only observations which lie within a specified time or distance of the analysis time or gridpoint (called 'search radii') are employed.

The time weight for the kth observation point within the search time interval is given by

$$w_t(k) = \begin{cases} (r_t - |t - t_o(k)|) / r_t & , |t - t_o(k)| \leq r_t \\ 0 & , |t - t_o(k)| > r_t \end{cases} \quad (3.28)$$

where r_t is the search radius, t is the analysis time and t_o (k) is the time of the kth observation within the search radius. The distance weight for the kth observation within the distance search circle is given by

$$w_d(k) = \begin{cases} (r_d - d(k)) / r_d & , d(k) \leq r_d \\ 0 & , d(k) > r_d \end{cases} \quad (3.29)$$

where r_d is the distance search radius and d (k) is the physical distance between the gridpoint and the observation point. Using a standard geometric argument, d (k) is expressed as

$$d(k) = a((\cos\phi \cos\lambda - \cos\phi_k \cos\lambda_k)^2 + (\cos\phi \sin\lambda - \cos\phi_k \sin\lambda_k)^2 + (\sin\phi - \sin\phi_k)^2)^{1/2}$$

where (ϕ, λ) is the latitude and longitude of the gridpoint and (ϕ_k, λ_k) is the latitude and longitude of the kth observation point within the search radii. It is not obvious that linear weights produce the best analyses, and hence in Chapter 4 results of tests of the time/space interpolation scheme made with linear weights are compared with results of tests made using cosine and negative exponential weights.

The combined weight w (k) is taken to be the product of the time and distance weights, viz

$$w'(k) = w_t(k)\, w_d(k) \quad . \tag{3.30}$$

The weights are then normalised to ensure that their sum is 1, ie we form

$$w'(k) = w(k) / \sum_{j=1}^{n} w(j) \quad , \tag{3.31}$$

where n is the number of observations within both time and distance radii.

The interpolated value, $\hat{T}$, at time t and gridpoint (ϕ,λ) is expressed as a linear sum of the product of the weight, $w'(k)$, and the kth observation within the search radii, $T_o(k)$, i.e

$$\hat{T} = \sum_{k=1}^{n} w'(k)\, T_o(k) \tag{3.32}$$

Time/space interpolation is used operationally to analyse TOVS measurements, and the method is easy to understand and use. A slight drawback is that since the method produces essentially a weighted average of a set of observations maximum values of the field are always underestimated and minimum values are always overestimated. However, such errors are very small if appropriate time and distance search radii are chosen. The effects of this choice of search radii on the quality of the analysis are investigated in Chapters 4 and 7.

3.4.2.3. Sequential Estimation of Fourier Field Components

Two dimensional analyses of satellite measurements can be performed using sequential estimation (Rodgers, 1976c), which is a form of the Kalman Filter (Kalman, 1960; Kalman and Bucy, 1961). This technique has been used to analyse measurements made by the LIMS instrument (see for example Gille and Russell, 1984). We assume that the observed quantity at any latitude can be

represented by a Fourier series in longitude whose coefficients vary with time, thus

$$Y(\lambda,t) = \bar{Y}(t) + \sum_{n=1}^{M} \left(a_n(t) \cos(n\lambda) + b_n(t) \sin(n\lambda) \right) + N_t \tag{3.33}$$

where $Y(\lambda,t)$ is the observation at longitude λ and time t, $\bar{Y}(t)$ is the zonal mean and $a_n(t)$ and $b_n(t)$ are Fourier coefficients. N_t is measurement error. M is the number of wavenumbers present. When the sequential estimation method is used, M is chosen to be 6, since as the TIROS-N satellite makes 14 orbits a day no more than 13 coefficients, or 6 wavenumbers, can be resolved using one day's data.

Equation (3.33) may be rewritten in vector form as

$$\mathbf{Y}_{\lambda t} = \mathbf{K}_t \mathbf{x}_t + \mathbf{N}_t \tag{3.34}$$

where $\mathbf{Y}_{\lambda t}$ is a vector of radiance observations at latitude λ and time t, $\mathbf{x}_t$ is a column vector of Fourier coefficients and $\mathbf{K}_t$ is a vector of cosines and sines. Each latitude is operated on independently. Hence this equation is similar in form to the radiative transfer equation used to retrieve Planck function, and thus temperature, from radiances. The Kalman Filter is a set of equations which allows an estimator to be updated once a new observation becomes available. The first step consists of forming the optimal predictor of the Fourier coefficients at the next observation time, given all the information currently available. We must make some estimate of the time evolution of the Fourier coefficients, and we therefore exploit the temporal coherence of the field by choosing the predictor, or 'first guess', of **x** to be the optimal estimate from the previous observation time. This assumption is called a random walk. The 'first guess' estimate at time τ_t is

$$x_{ot} = \hat{x}_{t-1} + \eta_t \qquad (3.35)$$
$$S_{ot} = \hat{S}_{t-1} + (\tau_t - \tau_{t-1}) \Delta S$$

where x_{ot} is the first guess estimate of the Fourier coefficients, η_t is the innovation at time t (i.e it represents information in x at time t which is not present at time time t−1); S_{ot} is its error covariance, $\hat{x}_{t-1}$ is the optimal estimate of the Fourier coefficients at time t−1 (ie at the previous observation time), and $\hat{S}_{t-1}$ is its error covariance, τ_t is the time of measurement t and Δ S is a measure of the increase in the uncertainty per unit time (note that prior to performing the sequential estimation, we must specify x_{ot} and S_{ot} at t = 0, and Δ S). The $(\tau_t - \tau_{t-1}) \Delta S$ term implies that the uncertainty in x_{ot} increases linearly as the time interval between observations increases; we choose to write the equation for S_{ot} in this form because the time interval between the observations is variable. This 'first guess' estimate is then combined with measurement t, $Y_{\lambda t}$, viz

$$\hat{x_t} = x_{ot} + S_{ot} K^T_t (K_t S_{ot} K^T_t + S_\varepsilon)^{-1} (Y_t - K_t x_{ot}) \qquad (3.36)$$

$$\hat{S_t} = S_{ot} - S_{ot} K^T_t (K_t S_{ot} K^T_t + S_\varepsilon)^{-1} K_t S_{ot} \qquad (3.37)$$

where $\hat{x}_t$ is the optimal estimate of the Fourier coefficients, and $\hat{S}_t$ is its error covariance. Note that equations (3.36) and (3.37) are similar to the optimal estimation equations (3.14) and (3.15) for temperature retrievals.

Equation (3.36) estimates Fourier components at time τ_t using only measurements at or before τ_t. In most research applications the data will be available both before and after the analysis time, and thus it is beneficial to combine estimates of the state at t made both forwards and backwards in time. Let us assume there are T observations and we require an analysis at time τ_t.

We wish to combine the forward 'updated' estimate' (performed using observations 1 to t) with the backward 'first guess' estimate (performed using observations T to t+1; this is to avoid using observation t twice). The forward and backward estimates are combined in a proper statistical manner using equation (3.8), and hence the combined estimate, $\hat{x}_t^c$ at time τ_t is given by

$$\hat{x}_t^c = (\hat{S}_t^{-1} + S_{ot}^{-1})^{-1} (\hat{S}_t^{-1} \hat{x}_t + S_{ot}^{-1} x_{ot}) \tag{3.38}$$

where x_{ot} and S_{ot} are the first guess backward estimate and its error covariance, and $\hat{x}_t$ and $\hat{S}_t$ are the forward estimate and its error covariance.

Sequential estimation allows us to estimate Fourier coefficients around a latitude circle at any observation time. However, observation and analysis times do not usually coincide, and so the estimate of the Fourier coefficients, $\hat{x}_{Ta}$, at the analysis time T_a, is made using a version of the random walk equation (3.35), viz

$$\hat{x}_{Ta} = \hat{x}_{To}^c + \eta_{Ta} \tag{3.39}$$

where $\hat{x}_{To}^c$ is the optimal estimate of x at time T_o, η_{Ta} is the innovation in x at time Ta, and T_o is the observation time nearest to the analysis time.

Since the Kalman Filter enables the estimate of the Fourier coefficients to be updated once a new observation becomes available, sequential estimation should not over-smooth field maxima and minima in the way that time/space interpolation does. However, the quality of the estimate depends on how appropriate the model of the time evolution of the Fourier coefficients is. A random walk model is probably reasonable when the time interval between the observations is small (i.e a few minutes or less), but is likely to be poor at larger time intervals. Tests of the sequential estimation method are presented in Chapter 8, together with suggestions for future research.

3.5. Summary

In this chapter we have described:

- the radiative transfer theory on which temperature sounding is based.
- the instruments used to measure upwelling radiance from appropriate absorption bands.
- methods of calculating channel weighting functions, and the importance of this for accurate temperature retrievals.
- techniques to retrieve temperature from radiances, and to interpolate satellite observations to a synoptic grid.

The aim of this chapter is not only to clearly describe the principles of temperature sounding, but also to aid the understanding and interpretation of results of tests of retrieval and analysis methods which appear in subsequent chapters. In the remainder of this thesis, we test a number of retrieval and analysis schemes which are commonly used to infer stratospheric temperature from satellite measurements. The multiple linear regression technique (described in Section 3.4.1.3) is tested in Chapter 6, and the subsequent discussion of the results of these tests utilises important points from the description of retrieval error characteristics, which appears in Section 3.4.1.4. Two analysis methods are tested in this thesis: time/space interpolation is tested in Chapters 4 and 7, whilst the sequential estimation of Fourier components is tested in Chapter 8. The description of both these methods, which appears in Section 3.4.2, helps to identify each scheme's possible strengths and weaknesses, and this is useful when evaluating the results of the tests in Chapters 4, 7 and 8.

CHAPTER 4

PRELIMINARY TESTS OF THE TIME/SPACE INTERPOLATION METHOD USING ANALYTICAL RADIANCE FIELDS

In this chapter we make preliminary tests of the time/space interpolation scheme described in Chapter 3. The sizes of the time and distance radii selected for the scheme are important. If a large radius is chosen, then the analysis may be oversmoothed; the magnitude of this oversmoothing depends on the temporal and spatial variability of the field. In Chapter 7 we test the interpolation scheme by analysing retrieved temperature profiles from 18/1/87 and then comparing the analysed fields with corresponding Met. Office stratosphere/mesosphere model fields from that day. However, the task of simulating radiances from model temperatures and then retrieving and interpolating them requires considerable computer time. Accordingly the effect of varying the search radii used in the interpolation scheme is initially tested in this chapter using an analytically calculated radiance field. The tests are computationally inexpensive, and thus can be repeated using a large variety of time and distance radii.

The radiance field is expressed as a sum of spherical harmonics; radiances are easily calculated at grid points and 'observations' are calculated from this field at satellite 'observation' points determined using a model of a TIROS-N satellite orbit. These 'observations' are then interpolated in time and space using the time/space interpolation scheme and compared with the analytical radiance field.

Details of the calculation of the model radiance are given in Section 4.1, and the method of calculating the satellite orbit is described in Section 4.2. Results are presented in Section 4.3.

4.1. The Model Radiance Field

4.1.1. Theory

An atmospheric field can be expressed on a sphere using

$$R(\phi, \lambda, t) = \sum_{j=0}^{J} \sum_{i=0}^{I} \alpha_{j,i,t} P_{j+i}^{i}(\phi) \cos i\lambda + \sum_{j=0}^{J} \sum_{i=1}^{I} \beta_{j,i,t} P_{j+i}^{i}(\phi) \sin i\lambda \qquad (4.1)$$

where $P_{j+i}^{i}(\phi)$

is an associated Legendre polynomial, i and j represent the zonal and meridional wavenumbers, whilst $\alpha_{j,i,t}$ and $\beta_{j,i,t}$ are spherical coefficients; ϕ is latitude, λ is longitude and t is time. The model fields we simulate here are chosen to travel at a fixed speed and with a constant amplitude. Thus, it is convenient to re-express the spherical harmonics in equation (4.1) as a single combined harmonic, viz

$$\alpha_{j,i,t} P_{j+i}^{i}(\phi) \cos i\lambda + \beta_{j,i,t} P_{j+i}^{i}(\phi) \sin i\lambda$$

is re-written as

$$A_{ji} P_{j+i}^{i}(\phi) \cos(i\lambda - \varepsilon_{jit})$$

where

$$\alpha_{j,i,t} = A_{ji} \cos \varepsilon_{jit}$$

and

$$\beta_{j,i,t} = A_{ji} \sin \varepsilon_{jit}$$

and hence equation (4.1) is re-written as

$$R(\phi, \lambda, t) = \sum_{j=0}^{J} \sum_{i=0}^{I} A_{ji} P_{j+i}^{i}(\phi) \cos(i\lambda - \varepsilon_{jit}) \qquad (4.2)$$

If equation (4.2) represents waves of wavenumber i rotating as a solid body with solid rotation rate ω_i, then

$$\varepsilon_{jit} = \omega_i t + \varepsilon_{ji0} \quad ,$$

where t is time and ε_{ji0} is the phase at t = 0. Moreover, for such a wave A_{ji} is a constant. It is clearly straightforward to simulate a high-amplitude 5-day wave (for example) by choosing appropriate values of ω_i, A_{ji} and ε_{ji0}. Details of the values chosen here appears in Section 4.1.2.

4.1.2. Numerical Details

We perform tests on three fields: Field 1 resembles the stratosphere, as observed by an SSU channel, during a Northern hemisphere winter in that there is a small amplitude wavenumber 1 wave in the Southern Hemisphere and tropics and there are higher amplitude wavenumber 1 and 2 waves in the Northern hemisphere middle and high latitudes. Field 2 is like Field 1, except that a wavenumber 3 wave is also simulated, and thus allows us to examine how well the scheme can estimate smaller-scale waves. If the model field changes sufficiently during the time period of the observations, then the interpolation scheme can over-smooth the model field features. Accordingly, to provide a stringent test of the scheme we have chosen high wave speeds for both Fields 1 and 2. The simulated wavenumber 1 wave, for example, completes a latitude circle in 5 days; such a wave is commonly observed in the stratosphere (see eg Rodgers, 1976a). The stratospheric climatology is, however, more typically dominated by stationary waves (see Section 2.2.1), and hence further tests are made using Field 3, which is like Field 2 except that the modelled waves are stationary.

In Fields 1 and 2 we choose wave speeds equivalent to that of a

wavenumber 1 5-day wave, and hence we set $\omega_{i1} = 1.45 \times 10^{-5}$ rad s^{-1}. The values of A_{ji} and ε_{ji0} used to calculate the Field 1 and 2 model radiances from (4.2) are shown in Tables 4.1 and 4.2, respectively. Model radiance fields are calculated at t = 0, 24 and 48 hrs.

Field 1

	i			
j	0	1	2	3
0	51.90/0.0	1.18/0.0	0.66/0.0	1.442?/0.0
1	7.05/0.0	2.42/0.0	0.53/0.0	0.0071/0.0
2	7.47/0.0	0.57/0.0	0.25/0.0	0.0045/0.0
3	4.12/3.14	0.036/0.0	0.066/0.0	0.0011/0.0

Table 4.1 Wave amplitudes (in radiance units (r.u)) and initial phase used to simulate field 1, written in form A_{ji}/ε_{ji0} (r.u = mW m^{-2} cm $ster^{-1}$). Initially fields were calculated on a rectangular grid. This field was then Fourier analysed and spherical coefficients, α_{jit} and β_{jit}, were calculated. Then A_{ji} and ε_{ji0} were calculated from these spherical coefficients.

Figure 4.1a shows Field 1 at t = 24 hrs.

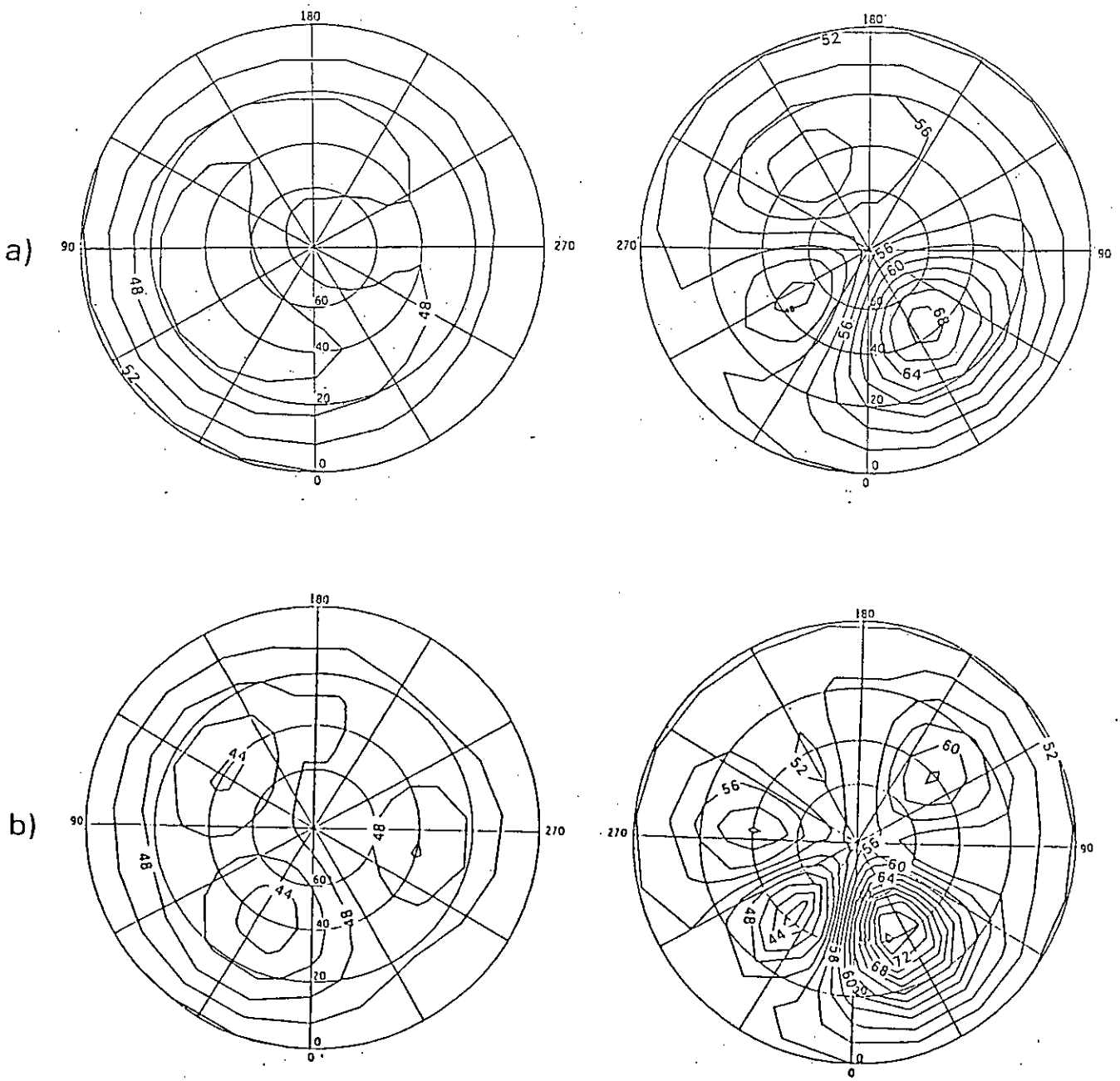


Figure 4.1 Model radiance field. The northern hemisphere field is plotted on the right and the southern hemisphere on the left. The contour spacing is 2 r.u. a) Field 1 at t = 24 hrs; b) Field 2 at t = 24 hrs;

c)

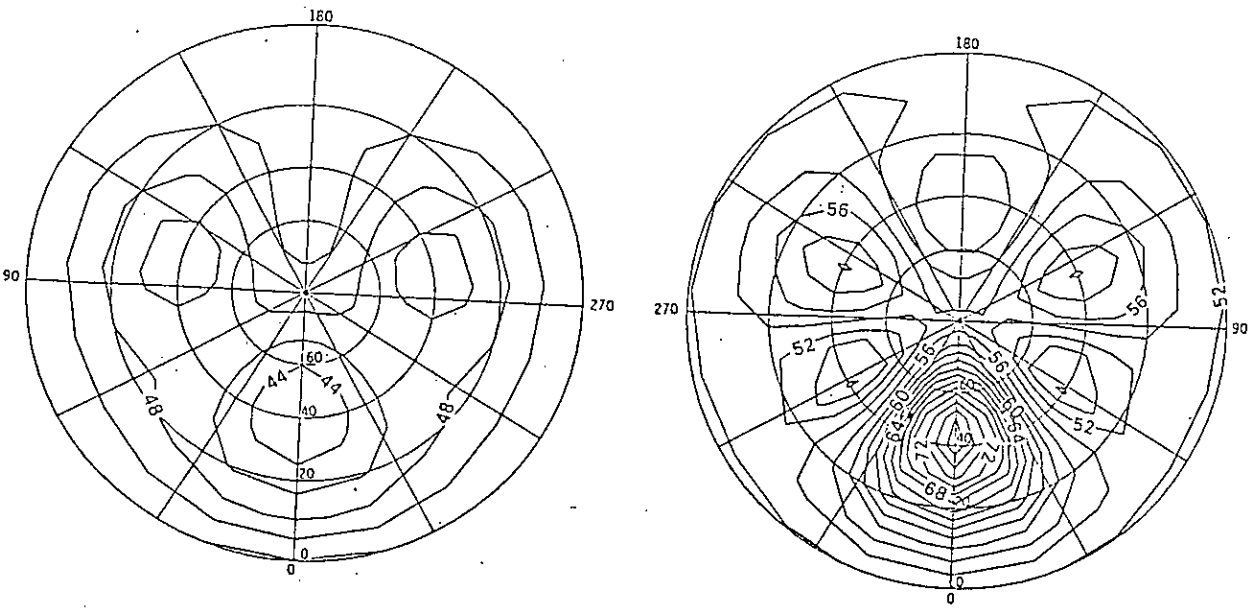


Figure 4.1(cont.) c) Field 3 at t = 24 hrs.

Field 2

j \ i	0	1	2	3	4
0	51.88/0.0	1.26/0.0	0.59/0.0	0.13/0.0	0.0026/0.0
1	6.98/0.0	2.56/0.0	0.43/0.0	0.16/0.0	0.0033/0.0
2	0.015/3.14	0.61/0.0	0.22/0.0	0.041/0.0	0.0/0.0
3	4.06/3.14	0.039/0.0	-0.055/0.0	0.019/0.0	0.0/0.0
4	0.029/0.0	0.15/-3.14	0.0034/3.14	0.0023/0.0	0.0/0.0

Table 4.2 Wave amplitudes (in radiance units (r.u)) and initial phase used to simulate Field 2, appearing in the pattern A_{ji}/ϵ_{ji0} (r.u = mW^{-2}. cm $ster^{-1}$). Initially fields were calculated on a rectangular grid. This field was then Fourier analysed and spherical coefficients, α_{jit} and β_{jit}, were calculated. Then A_{ji} and ϵ_{ji0} were calculated from these spherical coefficients.

Field 2 at t = 24 hrs is shown in Figure 4.1b. Field 3 is like Field 2 except all wave speeds are zero (ie $\omega_{it} = 0$), and the field at t = 24 hrs is shown in Figure 4.1c.

4.2. Calculation of the 'Observed' Radiance Field

'Observed' radiances are calculated at the observational latitude and longitude using equation (4.2). To simulate instrumental noise a randomly generated number is added to this calculated radiance. The random number is taken from a Normal distribution of mean 0.0 r.u and standard deviation 0.3 r.u. The choice of these values is based on validation tests of the SSU instruments (Pick and Brownscombe, 1979). The satellite observation points are calculated using a model of the TIROS-N satellite orbit, which is described below. For further details of satellite orbital mechanics see eg Duck and King (1983).

The modelled orbit is similar to an actual TIROS-N series satellite orbit. The orbit has an inclination to the Earth, i, of 80° and a height, h, of 850 km. The node longitude is 0° and the equator crossing time is 0230 hrs. The scanning motion of the satellite instrument is also modelled: observations are simulated at angles of -30°, -10°, 10° and 30° and at time intervals of 16s, for a period of 48 hrs.

The latitude, ϕ, of a satellite in a circular orbit is given by

$$\phi = \sin^{-1} (\sin i \sin \theta) \qquad (4.3)$$

where $\theta = \omega t$, ω being the rate of rotation of the satellite about the Earth. Longitude, λ, relative to the node (i.e the intersection of the equatorial plane and the plane of the subsatellite track) is given by

$$\lambda = \tan^{-1} (\cos i \tan \theta) + \Lambda_n t . \qquad (4.4)$$

where t is the time since the satellite last crossed the nodal point, and Λ_n is the nodal rotation rate, given by $\Lambda - \Omega$. Λ is the Earth's rotation rate and Ω, the nodal precession rate, is given by

$$\Omega = - 3/2 \, J \, r_e^2 \, (GM)^{1/2} \, (r_e + h)^{-7/2} \cos i$$

where $J = 0.00108263$, r_e is the Earth's radius, G is Newton's gravitational constant and M is the mass of the Earth.

Equations (4.3) and (4.4) describe the subsatellite track, but we wish to calculate the location at which a scanning radiometer observes. The off-nadir longitude, λ_D, is

$$\lambda_D = \tan^{-1} (\cos i \tan \theta) + \Lambda_n t + \sin^{-1} (\sin \psi \sin (\beta + \gamma) / \cos \phi_D) \qquad (4.5)$$

and the off-nadir latitude is

$$\phi_D = \sin^{-1} (\cos \phi \cos \psi + \cos \phi \sin \psi \sin (\beta + \gamma)) \qquad (4.6)$$

where ϕ is the latitude of the subsatellite point, β is the orbit azimuthal angle given by

$$\beta = \tan^{-1} (1 / (\tan i \cos \theta)) ,$$

and

$$\psi = \sin^{-1} (((r_e + h) / r_e) \sin \delta) - \delta ,$$

γ is the scan azimuthal angle (90° for TOVS instruments), and δ is the off-nadir scan angle.

4.3. Results

First tests of the interpolation scheme were made on Field 1 and results appear in Section 4.3.1. The ability of the interpolation scheme to estimate smaller waves is examined by repeating these tests on Field 2, while corresponding tests on Field 3 examine the effect of stationary waves on the analysis. Results appear in Sections 4.3.2 and 4.3.3 respectively. All these tests assume that the time and distance weight assigned to each observation linearly decreases the further the observation is from the gridpoint. The effect of using different types of weights is examined in Section 4.3.4; results are shown of analyses of Field 2 made using cosine and negative exponential weights. A summary of the results appears in Section 4.3.5, whilst Section 4.3.6 contains recommendations for further tests of the time/space interpolation scheme.

'Observed' radiances are interpolated to an analysis time of 24 hrs, on a grid of resolution 10° latitude by 20° longitude, and the analysed field is compared to the model field at t = 24 hrs. R.m.s errors are calculated at t = 24 hrs using the expression

$$RMS = \left(\left(1 / N \right) \sum_{i=1}^{N} \left(R_{Ti} - R_{Ai} \right)^2 \right)^{1/2} \qquad (4.7)$$

where N is the total number of gridpoints, R_{Ti} is the model radiance at t = 24 hrs and gridpoint i, and R_{Ai} is the corresponding analysed radiance. Equation (4.7) tells us how much R_{Ai} differs from R_{Ti}, but on its own does not tell us how well the model field structure is estimated. Hence it is also necessary to examine qualitatively the estimates by examining plots of the analysed fields. Tests are performed on both Fields 1 and 2 using a variety of distance radii between 500 and 5000 km, and time radii between 6 and 24 hrs. Analyses made with a time radius of 6 hrs and a distance radius of 1200 km or less have a number of gridpoints which have no observations within the corresponding search radii. This also happens with analyses which use larger time radii and a distance radii of 500 km. Results of these tests are not discussed below.

4.3.1. Field 1

R.m.s analysis errors for the tests made on Field 1 are shown in Table 4.3.

	r_t / hrs			
r_d / km	6	12	18	24
5000	1.80	1.66	1.69	1.76
3000	0.97	0.84	0.88	0.99
2000	0.65	0.50	0.54	0.67
1500	0.57	0.44	0.45	0.58
1200	...	0.44	0.40	0.53
1000	...	0.45	0.39	0.51
750	...	0.59	0.41	0.50

Table 4.3. R.m.s analysis errors in radiance units (r.u) for Field 1 (r.u = mWm^{-2} cm $ster^{-1}$). The analysis was made using linear weights.

Reducing the distance radius clearly decreases the r.m.s error, provided the

distance radius is greater than 1200 km. Detailed consideration of the original and interpolated fields reveals that reduction of the size of the distance radius leads to an improvement in the estimate of the model field's wavenumber 1 and 2 features. An examination of analyses made with a time radius of 12 hrs shows, in particular, that

1. The analysis made with a time radius of 12 hrs and a distance radius of 5000 km (Figure 4.2a) has an r.m.s error of 1.66 r.u. Comparison between this field and the model field (Figure 4.1a) shows that most of the model wavenumber 1 structure has been reproduced, but that the size of the analysed wavenumber 2 structure is smaller than in the model field. Figure 4.2b reveals that, in general, the difference between the analysed and model fields does not exceed 6 r.u.

2. Changing the distance radius from 5000 to 3000 km reduces the r.m.s error by about 50 % and the resultant analysis (Figure 4.2c) estimates both the wavenumber 1 and wavenumber 2 features better. Moreover, Figure 4.2d shows that the difference between model and analysed fields does not exceed 4 r.u.

3. Changing the distance radius from 3000 to 2000 km again reduces the r.m.s error by nearly 50 % and both wavenumber 1 and 2 features of the model field are better estimated. Figure 4.2e shows the analysed field and Figure 4.2f shows that the difference between model and analysed fields is usually less than 2 r.u. However, changing the distance radius from 2000 to 1500 km improves the estimate little.

The r.m.s error decreases by a smaller amount when the distance radius is changed from 2000 to 1500 km, and this also happens with analyses made with time radii of 6, 18 and 24 hrs. This is probably because the smallest distance radius required to estimate the model field features lies between 1500 and 2000 km. Reducing the distance radius below 1500 km causes different changes in the estimated field. Analyses made with a time radius of 12 hrs and a distance radius of 1200 km or less have r.m.s errors which are the same or slightly higher than the error associated with the analysis made with a distance radius of 1500 km. All wavenumber 1 and 2 features are satisfactorily estimated but

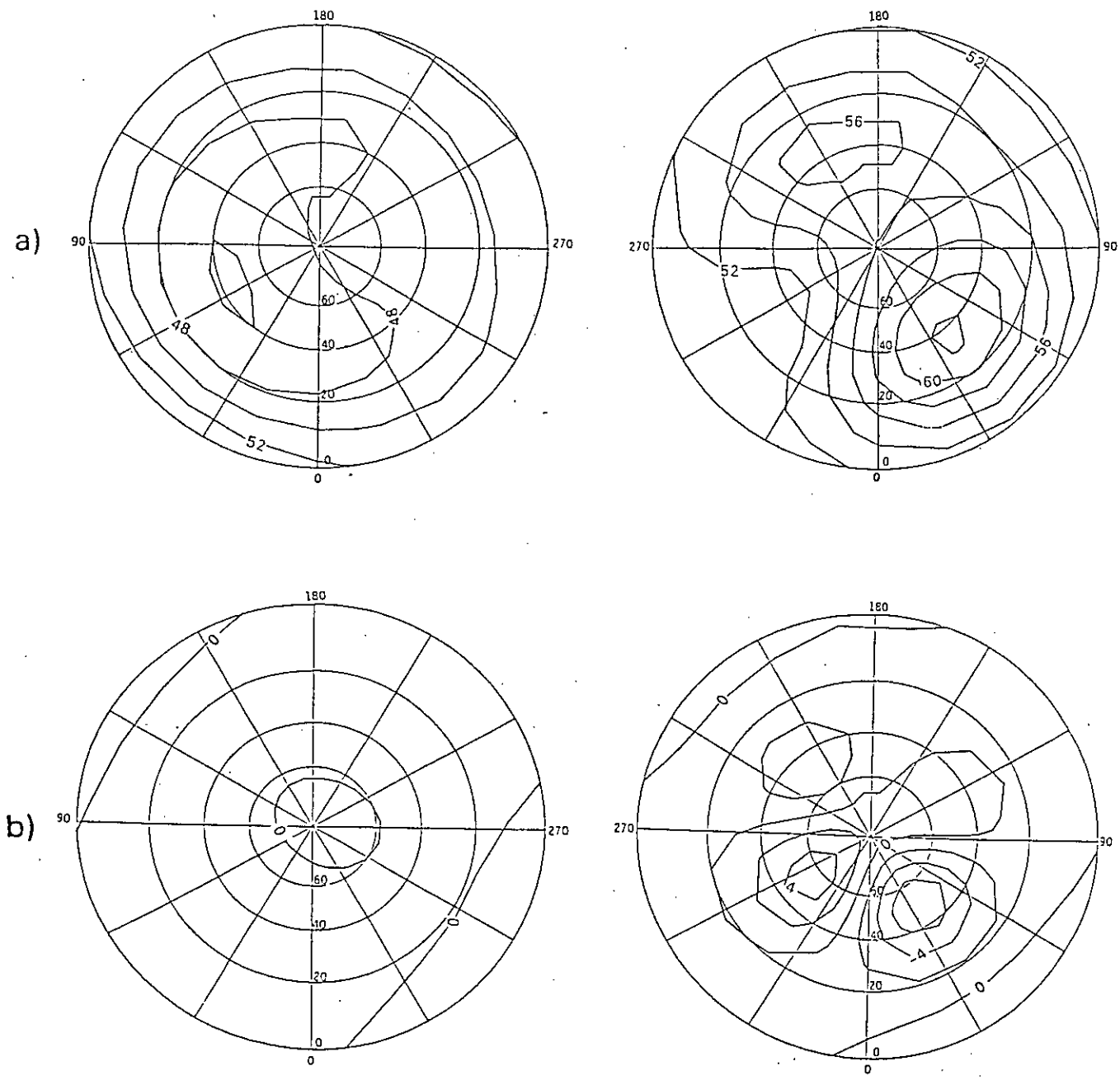


Figure 4.2 Field 1 analysed at t = 24 hrs using linear weights. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 2 r.u. a) analysed values obtained using a time radius of 12 hrs and a distance radius of 5000 km; b) difference between field in a) and model field;

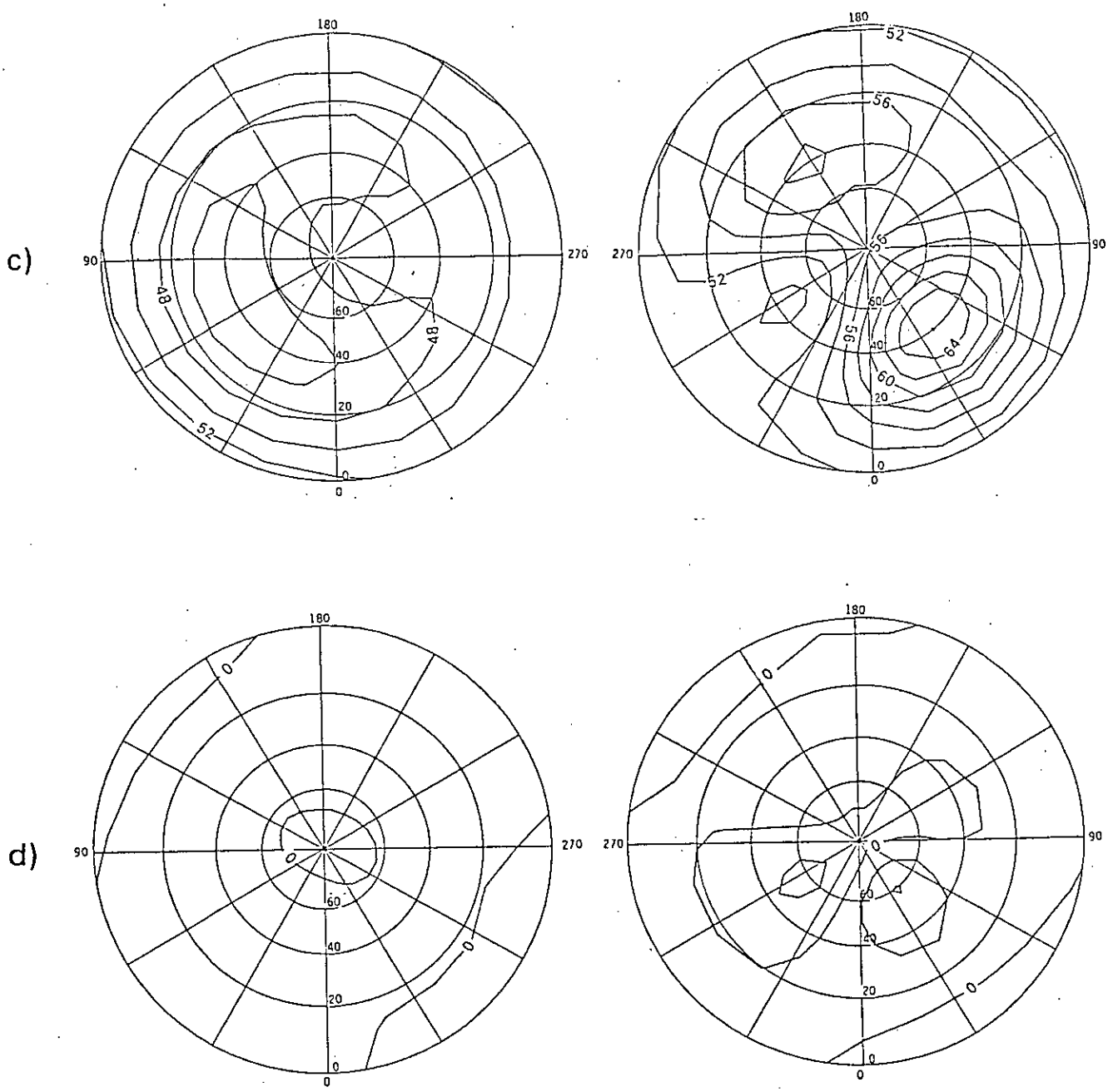


Figure 4.2 (cont.) c) as a), except the distance radius used is 3000 km; d) difference between field in c) and model field;

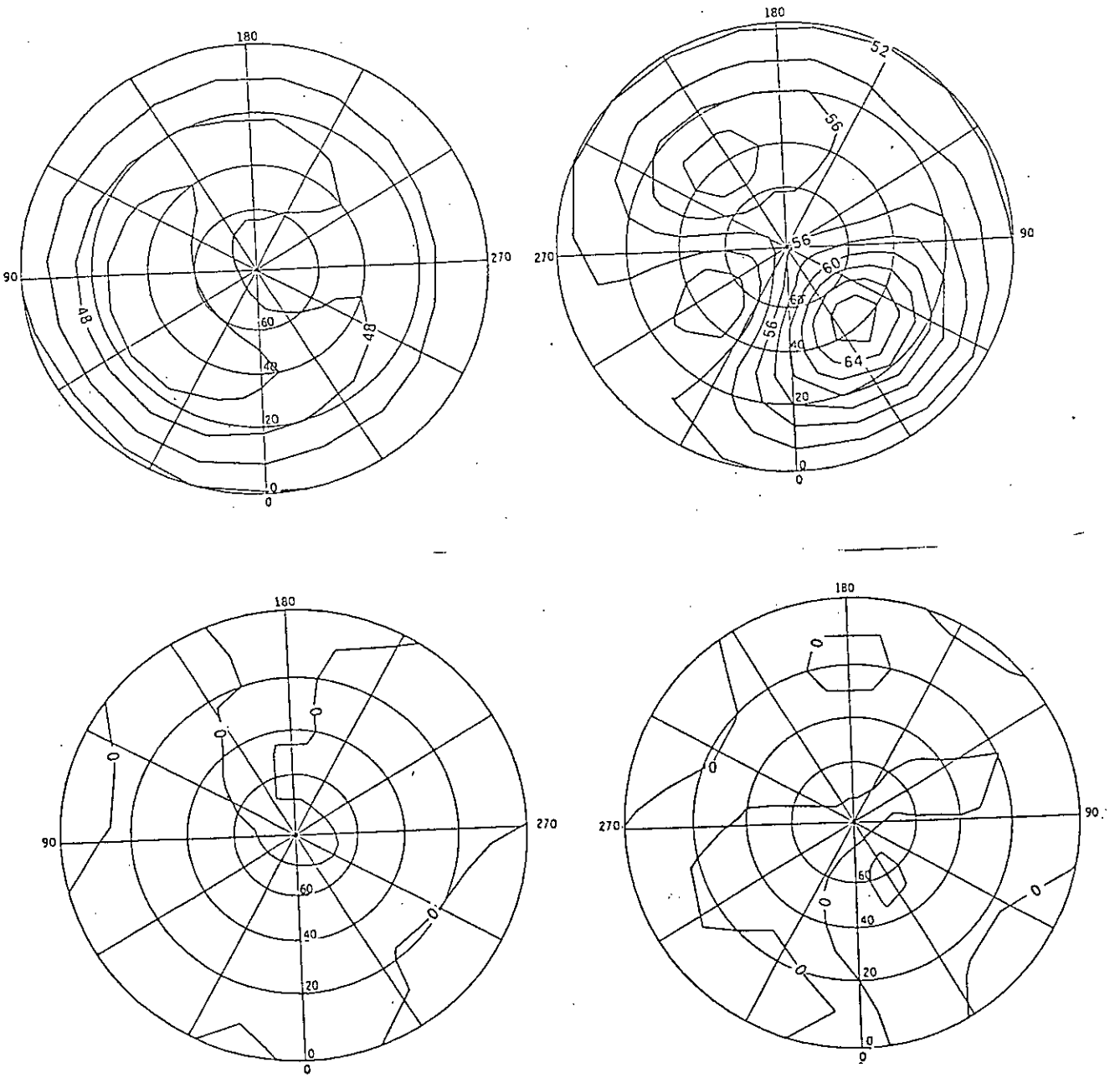


Figure 4.2 (cont.) e) as a), except the distance radius used is 2000 km; f) difference between field in e) and model field;

g)

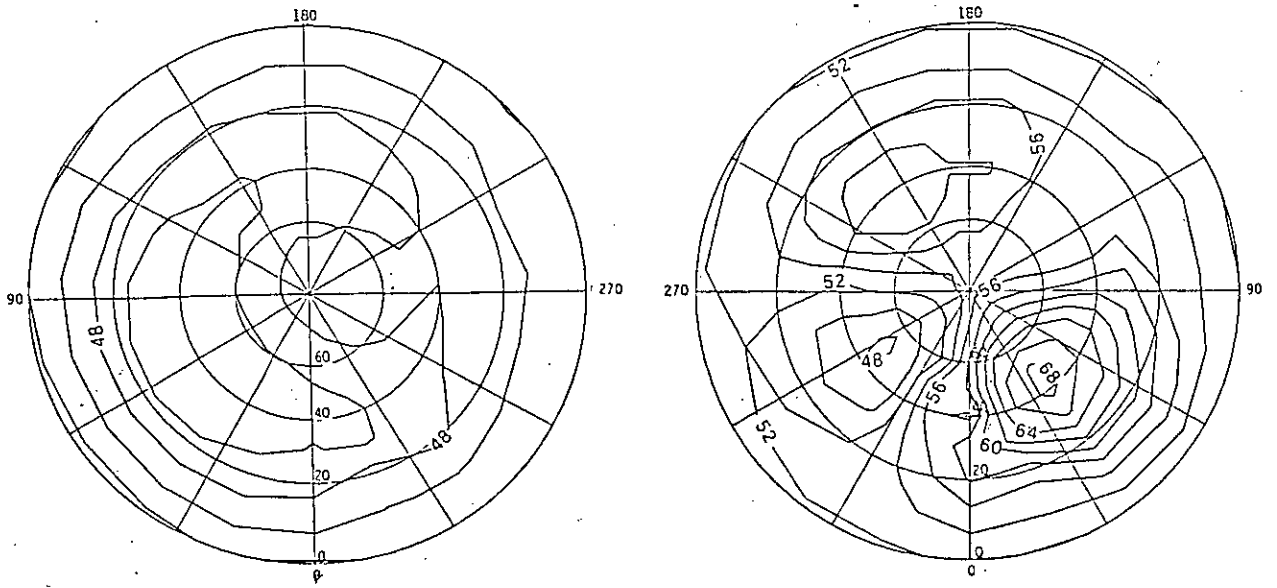


Figure 4.2 (cont.) g) as a), except the distance radius used is 750 km.

the analysed field is less smooth than the model field because at a number of gridpoints there is only one observation within the corresponding search radii. An example of this is the analysis made with a distance radius of 750 km (Figure 4.2g). A similar effect also happens with analyses made with a time radius of 18 hrs, but here the r.m.s error increases when the distance radius is changed from 1000 to 750 km. This is due to the larger number of observations available to these analyses. Analyses made using a time radius of 24 hrs have an even larger number of observations available; hence results of such analyses reveal that, within the range of distance radii tested, the r.m.s error always decreases when the distance radius decreases.

Despite the high model wave speeds, changing the time radius has little effect on the r.m.s error. Tests made using a 2000 km distance radius, for example, show that changing the time radius from 12 to 18 hrs causes a slight increase in the r.m.s error, and that the increase is also small when the time radius is further raised to 24 hrs. Theoretically, the fastest wave that can be estimated by measurements from a polar-orbiting satellite has a period of 2 days. However, the fastest wave in Field 1 has a period of 5 days, and hence is satisfactorily estimated by the scheme. In addition, the large horizontal scale of the field (ie no smaller than wavenumber 2) means that the estimated field features are over-smoothed very little, even when a large time radius (eg 18 hrs, 24 hrs) is used. The reduction in the r.m.s error when the time radius is changed from 6 to 12 hrs may be explained by the lower number of observations available in the 6 hr analysis.

4.3.2. Field 2

The analysis scheme is tested using Field 2 in a similar manner to the tests made with Field 1. R.m.s analysis errors for a variety of time and distance radii are shown in Table 4.4.

	r_t/hrs			
	6	12	18	24
r_d/km				
5000	2.67	2.47	2.51	2.60
3000	1.49	1.29	1.37	1.47
2000	1.01	0.76	0.83	0.94
1500	0.93	0.70	0.70	0.78
1200	...	0.74	0.66	0.71
1000	...	0.78	0.65	0.68
750		0.99	0.66	0.66

Table 4.4 R.m.s analysis errors in radiance units (r.u) for Field 2 (r.u = mWm^{-2} cm $ster^{-1}$). The analysis was made using linear weights.

The inclusion of a wavenumber 3 wave in model Field 2 results in higher r.m.s errors than for Field 1. However, the chief conclusion drawn from these tests is similar to that drawn from the tests made on Field 1, namely that reducing the distance radius clearly reduces the r.m.s error, down to a limit where the number of observations within the search radii is so small that the estimated field becomes less smooth than the model field. Concentrating on analyses made with a time radius of 12 hrs, we note in particular that:

1. Comparison between the analysis made with a distance radius of 5000 km (Figure 4.3a) and the model field (Figure 4.1b) reveals that the model field's wavenumber 1 structure is well reproduced. However, only some of the field's wavenumber 2 structure is estimated, and the scheme fails to reproduce the wavenumber 3 behaviour in the northern hemisphere. This is also demonstrated in Figure 4.3b, which shows the difference between analysed and model fields. The largest differences (over 12 r.u) occur in the northern hemisphere in regions of large amplitude model wavenumber 3 behaviour.

2. Changing the distance radius to 3000 km reduces the r.m.s error by about 50 %. The analysis (Figure 4.3c) estimates the model wavenumber 1 and 2 structure well, and the wavenumber 3 structure is also better estimated. However,

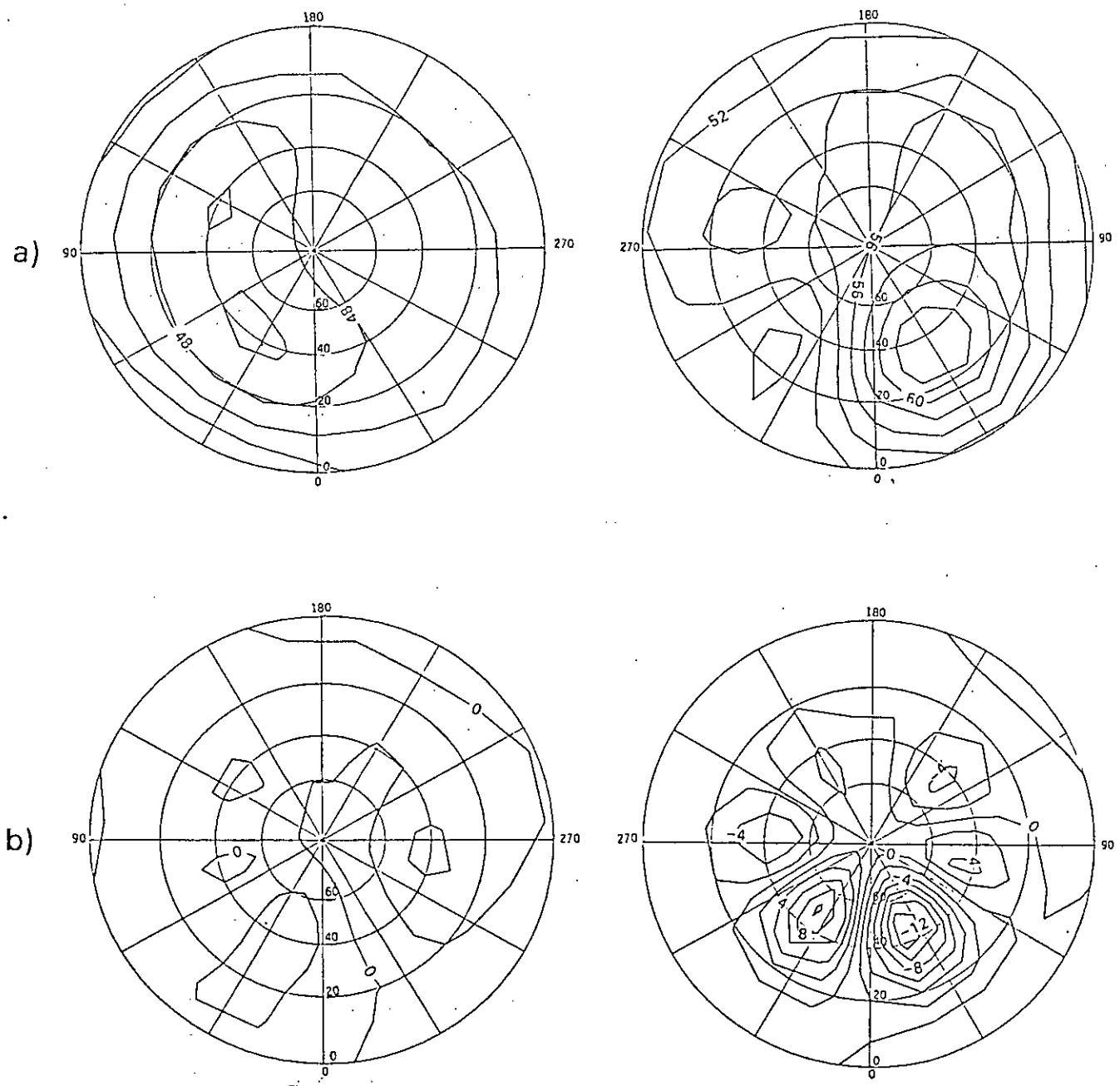
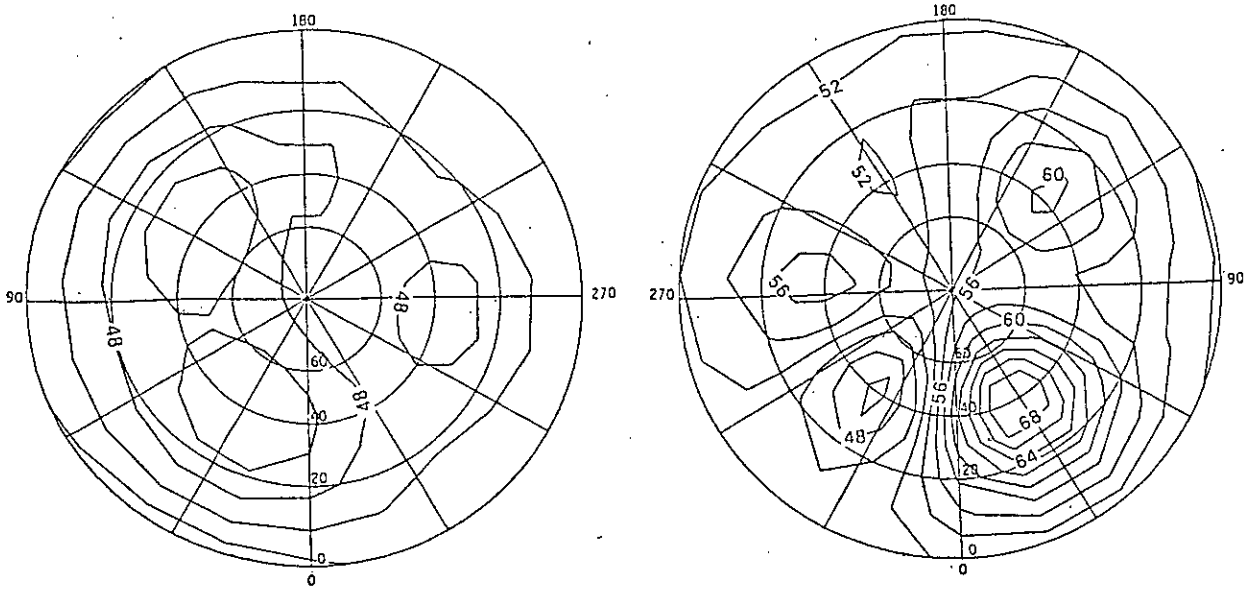


Figure 4.3 Field 2 analysed at t = 24 hrs using linear weights. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 2 r.u. a) analysed values obtained using a time radius of 12 hrs and a distance radius of 5000 km; b) difference between field in a) and model field;

c)



d)

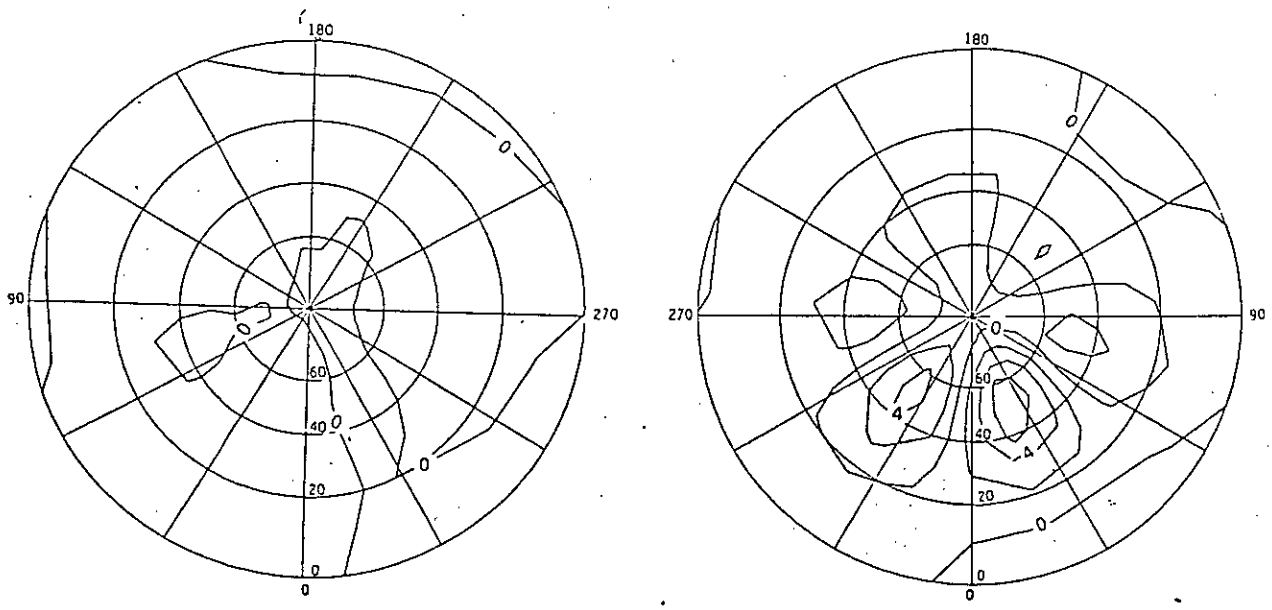
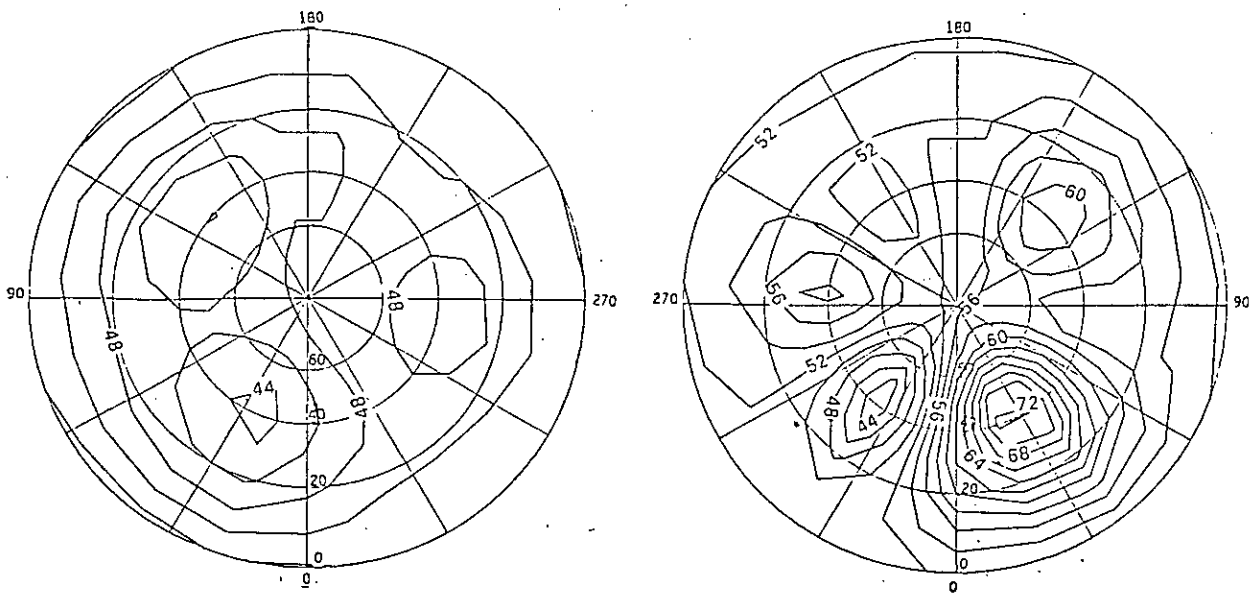


Figure 4.3 (cont.) c) as a), except the distance radius used is 3000 km; d) difference between field in c) and model field;

e)



f)

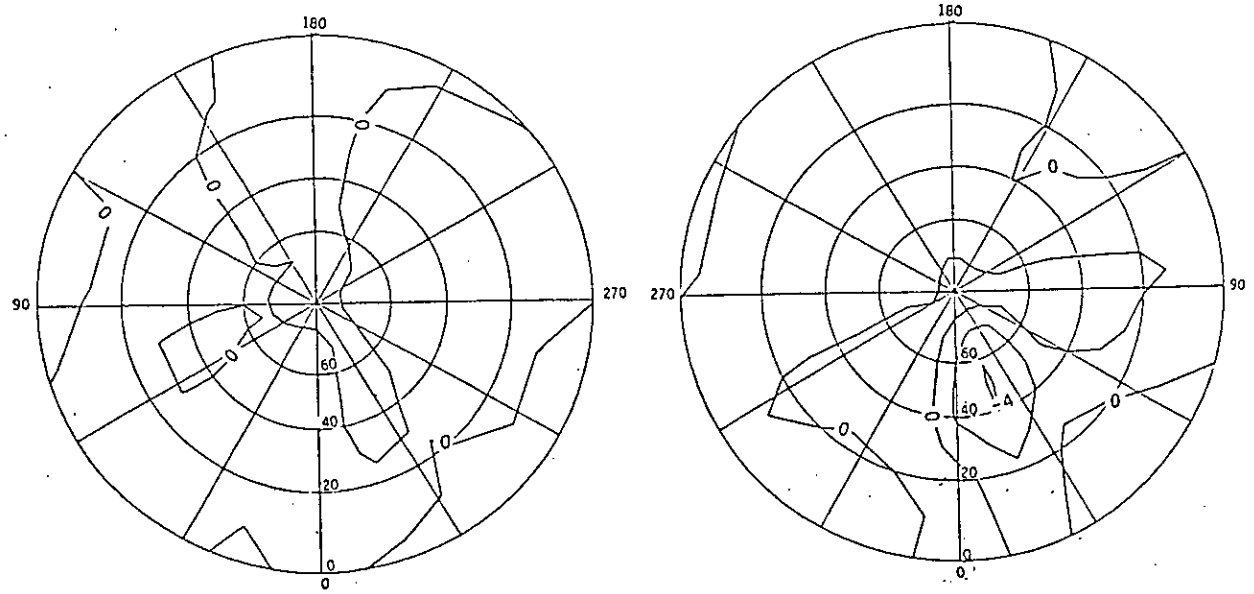


Figure 4.3 (cont.) e) as a), except the distance radius used is 2000 km; f) difference between field in e) and model field;

g)

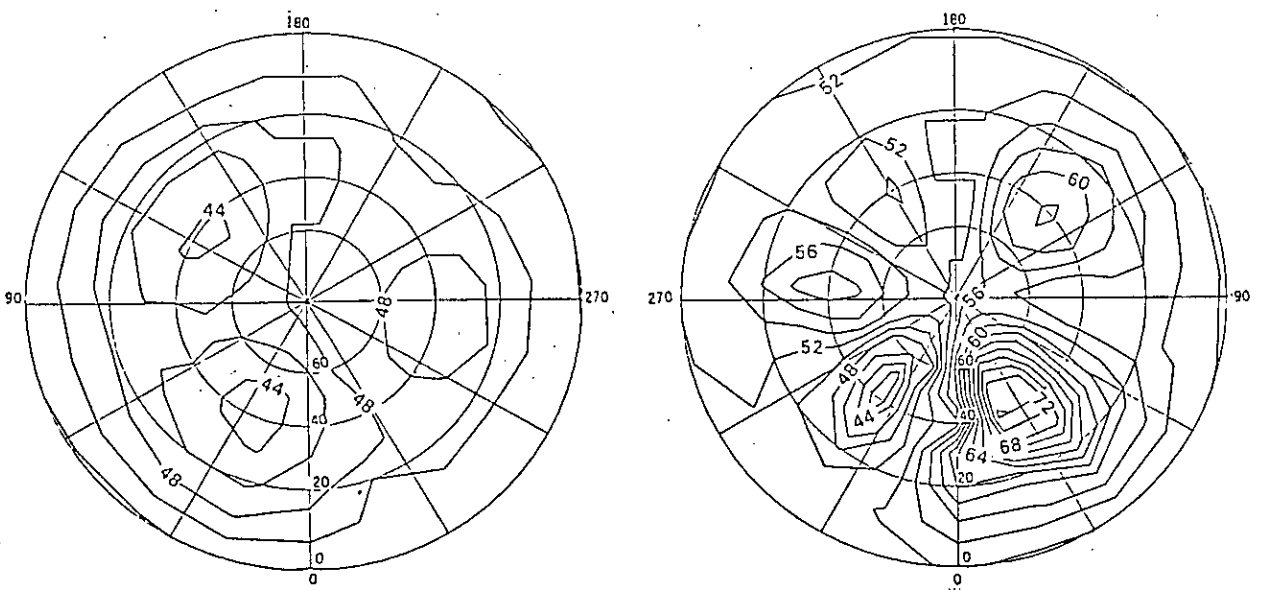


Figure 4.3 (cont.) g) as a), except the distance radius used is 750 km.

a plot of the difference between model and analysed fields (Figure 4.3d) shows that differences exceed 6 r.u in parts of the northern hemisphere.

3. Changing the distance radius to 2000 km (Figure 4.3e) produces a better estimate of the wavenumber 3 structure than the analysis made with a distance radius of 3000 km, and the difference between model and analysed fields (Figure 4.3f) is usually no larger than 4 r.u. Reducing the distance radius to 1500 km improves the estimate (not shown) even further, and differences between the fields (not shown) are reduced to less than 2 r.u.

As with Field 1, analyses made with a time radius of 12 hrs and distance radii of less than 1500 km have slightly larger errors than the analysis made with a distance radius of 1500 km. These analysed fields are less smooth than the model field because at a number of gridpoints there is only one observation within the corresponding search radii. An example is Figure 4.3g, which shows the analysis made with a distance radius of 750 km. In a similar manner, the analysis made with a time radius of 18 hrs and a distance radius of 750 km has a larger r.m.s error than the corresponding analysis made with a distance radius of 1000 km. Such a lack of smoothness (which also occurs in corresponding analyses of Field 1) implies the presence of high-wavenumber waves. The size of such waves can be demonstrated by Fourier analysing the relevant estimated field. Table 4.5 shows the Fourier components (to wavenumber 8) of: A) the field in Figure 4.3g; B) the field estimated using a time radius of 12 hrs and a distance radius of 2000 km; this field appears in Figure 4.3e and does not appear to contain waves of a smaller scale than those that are in the model field; C) the model field at t = 24 hrs (Figure 4.1b).

Wavenumber	Field A		Field B		Field C	
	a	b	a	b	a	b
1	1.31	5.41	1.89	5.22	1.97	5.42
2	0.11	3.90	0.82	3.90	1.65	5.39
3	1.68	7.55	2.19	6.79	2.95	8.48
4	−0.37	2.19	0.44	2.18	0.41	1.60
5	−1.10	−0.48	−0.11	−0.43	0.0	0.0
6	−0.95	0.25	6.96×10^{-3}	0.28	0.0	0.0
7	−1.20	−0.94	-3.68×10^{-3}	−0.57	0.0	0.0
8	−1.45	−0.43	−0.12	1.39×10^{-2}	0.0	0.0

Table 4.5 Fourier coefficients of three fields at t = 24 hrs: A) analysed with a time radius of 12 hrs and distance radius of 750 km; B) analysed with a time radius of 12 hrs and a distance radius of 2000 km; C) model field. The Fourier analysis was performed around the 40°N latitude circle. In the Table cosine Fourier coefficients are denoted by a, and sine coefficients by b. Units are radiance units (r.u) (r.u = mWm^{-2} cm $ster^{-1}$).

Table 4.5 shows that both Fields A and B adequately estimate the model field wavenumber 1 to 3 structure: with a few exceptions (notably the wavenumber 2 cosine coefficient of Field A) Field A and B coefficients are between 66 and 100 % of model values. However, both Field A and Field B contain more shorter-scale waves: wavenumber 4 coefficients are about the same size or larger than model values. In addition, whereas model coefficients for wavenumbers greater than 4 are zero, corresponding coefficients for Fields A and B are non-zero, and in general the coefficients are larger for Field A than for Field B. This is unsurprising, since Figure 4.3g clearly shows that Field A contains structure of a smaller scale than present in the model field, whilst Figure 4.3e reveals that Field B has negligible structure which is of a smaller scale than that present in the model field. Table 4.5 reveals that for wavenumbers of 5 and more, the cosine Fourier coefficients of Field A are at

least 10 times as large as corresponding coefficients of Field B; sine Fourier coefficients of wavenumber 5 and 6 are about the same size for both fields, but the wavenumber 7 and 8 sine coefficients of Field A are about 2 and 30 times larger, respectively, than the corresponding coefficients of Field B. These results underline how a small distance radius can lead to the estimation of waves of scales smaller than those that are actually present in the 'real' field. This can cause further problems if retrieved temperatures are used to calculate other variables; an example is thermal wind, which is calculated using horizontal derivatives of the temperature field. If such small scale waves are present in the analysis, the finite differencing method used to calculate the temperature derivative can produce even larger errors in the thermal wind field than those that appear in the analysed temperature field. It is therefore important to exercise care both when choosing an appropriate distance radius for the analysis, and when using analysed fields to calculate dynamical diagnostics.

Despite the high model wave speeds, changing the time radius causes much smaller changes in r.m.s error than those caused by changes in distance radius. The model waves are sufficiently slow, and are of sufficiently large a scale, for the analysis not to over-smooth the model field features noticeably, even when large time radii are used. Hence, when using the time/space interpolation scheme to analyse real stratospheric fields the choice of time radius is less important than the choice of distance radius.

4.3.3. Field 3

The analysis scheme is tested with Field 3 in a similar way to the tests made on Field 2, and comparisons of the tests are made to indicate the effect the stationary waves in Field 3 have on the results. R.m.s analysis errors appear in Table 4.6.

r_d/km	r_t/hrs 6	12	18	24
5000	2.65	2.51	2.50	2.51
3000	1.43	1.27	1.26	1.27
2000	0.82	0.67	0.66	0.66
1500	0.64	0.50	0.46	0.45
1200	...	0.46	0.38	0.36
1000	...	0.43	0.34	0.32
750	...	0.37	0.28	0.26

Table 4.6 R.m.s analysis errors in radiance units (r.u) for Field 3 (r.u=mW m^{-2} cm $ster^{-1}$). The analysis was made using linear weights.

In general, the quality of the analysis changes with changing distance radius in a similar manner to tests made on Field 2. For example: a) comparison of the model field (Figure 4.1c) and the analysis made with a time radius of 12 hrs and a distance radius of 5000 km (Figure 4.4a) shows that the model wavenumber 1 features are reproduced well, and wavenumber 2 features quite well, but that the analysis fails to reproduce the model wavenumber 3 features; b) an analysis made with a similar time radius, but a distance radius of 2000 km (Figure 4.4b) satisfactorily reproduces all wavenumbers. Results of analyses a) and b) are similar to corresponding analyses made on Field 2. However, there are some differences in the results caused by the stationarity of Field 3, viz

1. Field 3 r.m.s error values are generally smaller than corresponding values for Field 2: the smallest error is 0.26 r.u (for an analysis using a time radius of 24 hrs and a distance radius of 750 km), while the smallest r.m.s error for tests on Field 2 was 0.65 r.u (for an analysis made using a time radius of 18 hrs and a distance radius of 1000 km).

2. Errors of analyses made with a fixed distance radius but different time radii do not in general increase as time radius

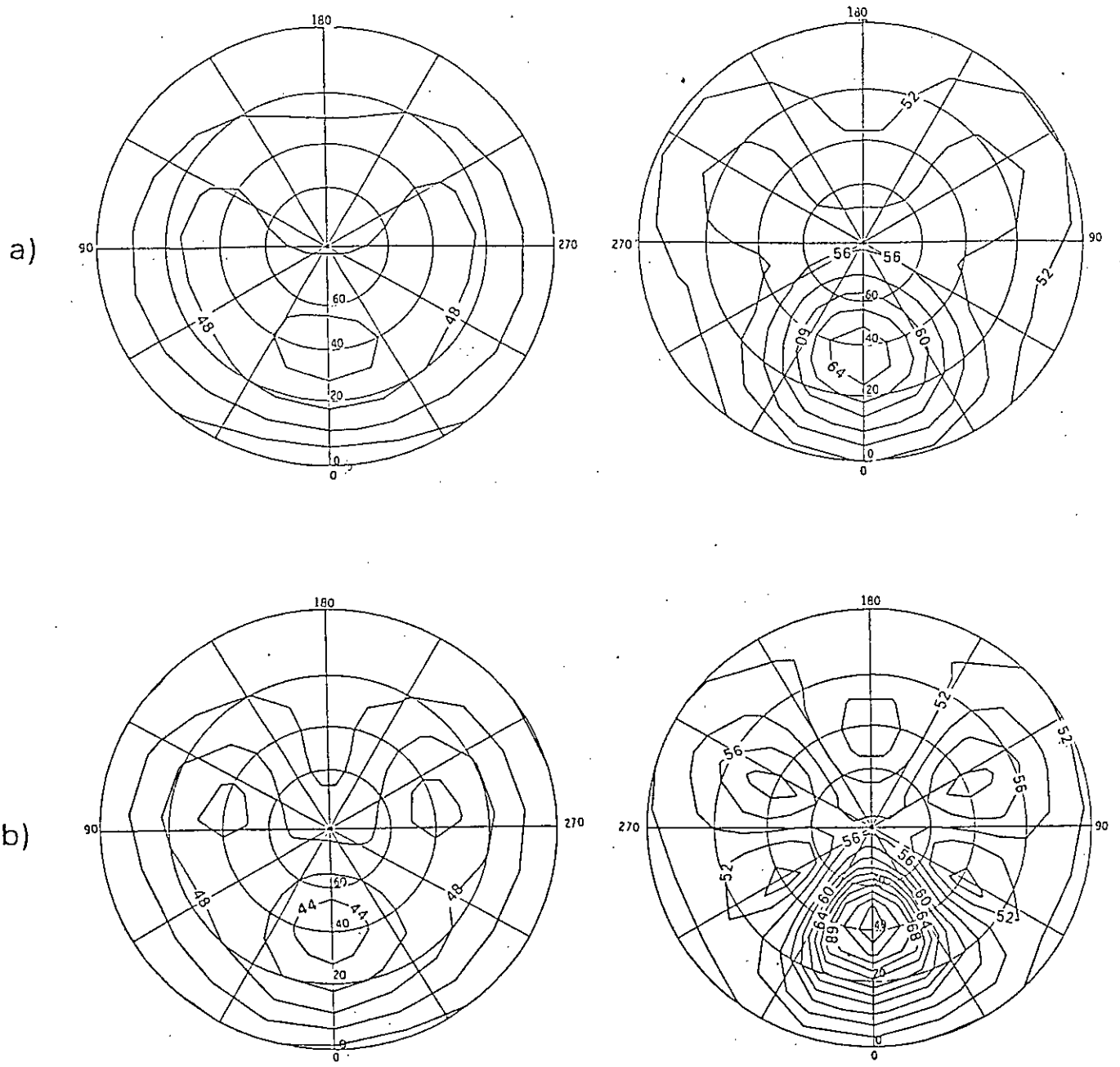


Figure 4.4 Field 3 analysed at t = 24 hrs using linear weights. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 2 r.u. a) analysed values obtained using a time radius of 12 hrs and a distance radius of 5000 km; b) as a), except the distance radius used is 2000 km;

c)

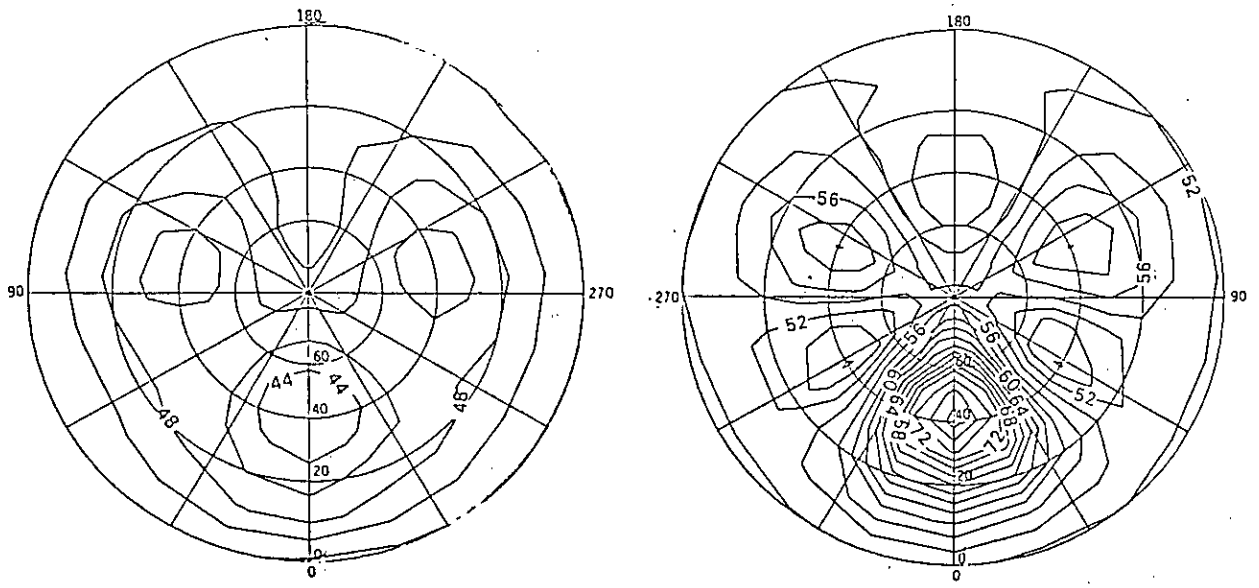


Figure 4.4 (cont.) c) as a), except the distance radius used is 750 km.

increases. In fact, when distance radii of less than 2000 km are used the error tends to decrease slightly, due to the larger number of observations available when large time radii are used.

3. With Field 2, analyses made with small distance radii often had higher r.m.s errors than analyses made with slightly larger distance radii because at a number of gridpoints there was only one observation within the corresponding search radii. Tests made on Field 3 reveal no such increase in error at small distance radii – because the field is stationary the fact that only one observation may be inside the search radii is less critical. For example, Figure 4.4c shows an analysis made with a time radius of 12 hrs and distance radius of 750 km, and one can see that it is much smoother than the corresponding analysis made using Field 2 (Figure 4.3g). This is to be expected: for a stationary field, as the distance radius tends to zero and the time radius tends to infinity the analysis error should be explained solely by instrumental errors.

The results of tests made with Fields 1, 2 and 3 reveal that the choice of distance radius is more important than the choice of time radius. Increasing the time radius changes the r.m.s analysis error only slightly. On the other hand, reducing the distance radius reduces the r.m.s error by a larger amount, provided the distance radius is 1500 km or greater. When a distance radius of less than 1500 km is used, the r.m.s errors of Field 3 analyses made with a time radius of 12 hrs continue to fall, but the r.m.s errors of corresponding Field 1 and 2 analyses rise slightly because these estimates contain waves of a smaller scale than those present in the 'real' field. A similar effect is noticed in analyses made using a time radius of 18 hrs when the distance radius is reduced below 1000 km. The small-scale waves are present because at small distance radii the number of observations within such radii is too small to resolve the travelling waves of Fields 1 and 2 adequately. A similar problem does not exist with corresponding Field 3 analyses because that field is stationary. However, as discussed in Chapter 2, in the real stratosphere we may more typically observe a combination of travelling and stationary waves. Hence, when analysing real

satellite data, one should choose a distance radius that is small enough to resolve the major field features, but be careful that the radius is not so small as to introduce erroneous small scale waves to the estimate.

4.3.4. Analyses Using Different Weights

The results above were obtained using an analysis scheme which gives each observation a weight which linearly decreases the further away the observation is from the analysis time or gridpoint. It is not clear that this type of weight produces the best results. Thus further tests on Field 2 are made using two different kinds of weight: a cosine weight and a negative exponential weight. Such functions have been used in schemes to assimilate observations into a numerical weather prediction model (eg Schlatter et al, 1976; Seaman, 1988).

The cosine time weight is written as

$$w_t(i) = \begin{cases} \cos(\pi \,|\, t - t_o(i) \,|\, / 2\, r_t), & |\, t - t_o(i) \,|\, \leq r_t \\ 0, & |\, t - t_o(i) \,|\, > r_t \end{cases} \quad (4.8)$$

where t is the analysis time, $t_o(i)$ is the time of observation i and r_t is the time search radius. The cosine distance weight, $w_d(i)$, is similar to equation (4.8) except r_t is replaced by the distance search radius, r_d, and $|\, t - t_o(i) \,|$ is replaced by the distance, $d(i)$, between observation and grid points.

The negative exponential time weight is given by

$$w_t(i) = e^{-W \,|\, t - t_o(i) \,|\, / r_t} \quad (4.9)$$

W is a coefficient which ensures that $w_t(i) \sim= 0$ at $|\, t - t_o(i) \,| = r_t$. In the tests described below two values of W are used, namely 5 and 10. The negative exponential distance weight, $w_d(i)$, is similar to equation (4.9), except that $|\, t - t_o(i) \,|$ is replaced by $d(i)$ and r_t is replaced by r_d.

Linear, cosine and negative exponential weights are plotted in Figure 4.5. It is clear that the cosine weight gives more weight to observations near to the analysis point than does the linear weight. The negative exponential weight gives a negligible weight to observations more than about half (W = 10 weight) or three quarters (W = 5 weight) of a search radius away from the analysis point.

4.3.4.1. Cosine Weights – Results

R.m.s errors of analyses made using both linear and cosine weights are shown in Table 4.7.

	r_t / hrs			
	6	12	18	24
r_d / km				
5000	2.73 (2.67)	2.59 (2.47)	2.63 (2.51)	2.69 (2.60)
3000	1.53 (1.49)	1.38 (1.29)	1.47 (1.37)	1.57 (1.47)
2000	1.03 (1.01)	0.82 (0.76)	0.90 (0.83)	1.02 (0.94)
1500	0.94 (0.93)	0.74 (0.70)	0.74 (0.70)	0.84 (0.78)
1200	...	0.76 (0.74)	0.68 (0.66)	0.76 (0.71)
1000	...	0.82 (0.78)	0.66 (0.65)	0.72 (0.68)
750	...	1.02 (0.99)	0.66 (0.66)	0.69 (0.66)

Table 4.7 R.m.s analysis error in radiance units (r.u) for Field 2 (r.u = mW m^2 cm $ster^{-1}$). Figures not in brackets are errors for analyses made with cosine weights, whilst figures in brackets are errors for analyses made with linear weights.

R.m.s errors are very slightly higher when using cosine weights than when using linear weights. Further inspection of Table 4.7 shows that, for a constant time radius, the difference between cosine-weight and linear-weight r.m.s

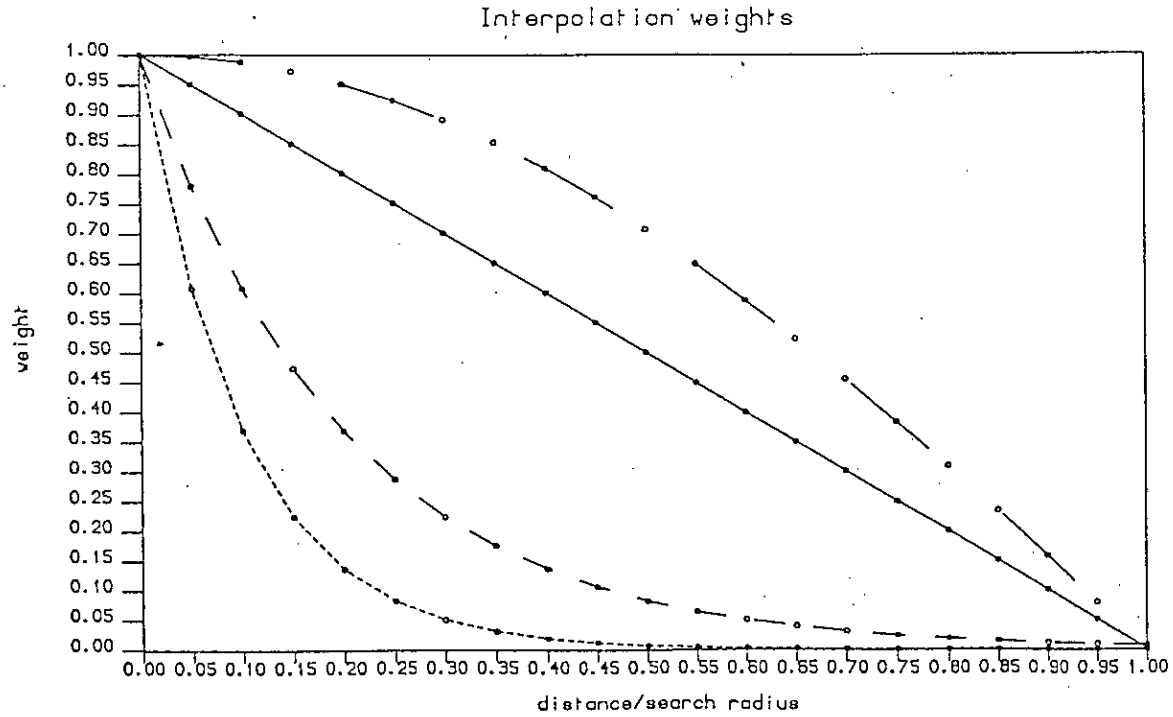


Figure 4.5 Interpolation weights used, expressed between 0 and 1 search radius from the gridpoint: linear (solid); cosine (long dashes); negative exponential (W=5) (short dashes); negative exponential (W=10) (dotted).

errors is greater for larger distance radii (ie 5000 or 3000 km) than for smaller distance radii. Figure 4.5 shows that all observations between 0.0 and 0.2 search radii from a gridpoint are given cosine weights close to 1.0. On the other hand, linear weights in the same region range from 1.0 to 0.8. This means that cosine-weight analyses are smoothed more than linear-weight analyses, and hence have larger r.m.s errors. This effect is more noticeable at larger distance radii, where there are a greater number of observations close to the gridpoint. However, the smoothing is very small, and fields analysed with cosine weights (not shown) are very similar to corresponding fields derived using linear weights. Thus the use of cosine weights instead of linear weights makes only a slight difference to the quality of the analysis.

4.3.4.2. Negative Exponential Weights – Results

Analysis tests using negative exponential weights are also made using Field 2. R.m.s errors are shown in Table 4.8 along with corresponding errors from analyses made with linear weights.

r_d / km	r_t / hrs 6	12	18	24
5000	2.67/2.36/1.98	2.47/1.79/1.13	2.51/1.71/0.93	2.60/1.72/0.86
3000	1.49/1.29/1.13	1.29/0.93/0.77	1.37/0.89/0.71	1.47/0.93/0.68
2000	1.01/0.93/0.97	0.76/0.68/0.77	0.83/0.65/0.74	0.94/0.66/0.70
1500	0.93/0.86/0.91	0.70/0.69/0.78	0.70/0.64/0.76	0.78/0.63/0.72
1200	...	0.74/0.72/0.80	0.66/0.66/0.77	0.71/0.63/0.73
1000	...	0.78/0.71/0.74	0.65/0.65/0.71	0.68/0.62/0.72
750	...	0.99/0.93/0.94	0.66/0.70/0.82	0.66/0.67/0.86

Table 4.8 R.m.s analysis error in radiance units (r.u). for Field 2 (r.u = mW m^{-2} cm $ster^{-1}$). Errors are shown for analyses performed with both linear and negative exponential weights, in the pattern (linear/exponential (W=5)/exponential (W=10)).

These results follow a similar pattern to the results obtained using linear weights, namely that in general reducing the distance radius reduces r.m.s error. In particular

1. When a distance radius of 5000 km is used the errors from the W = 10 negative exponential weighted analysis are smaller than those from the W = 5 negative exponential analysis. The W = 5 analysis in turn has smaller errors than the analysis performed with linear weights. Figure 4.6a shows the analysis made with W = 10 negative exponential

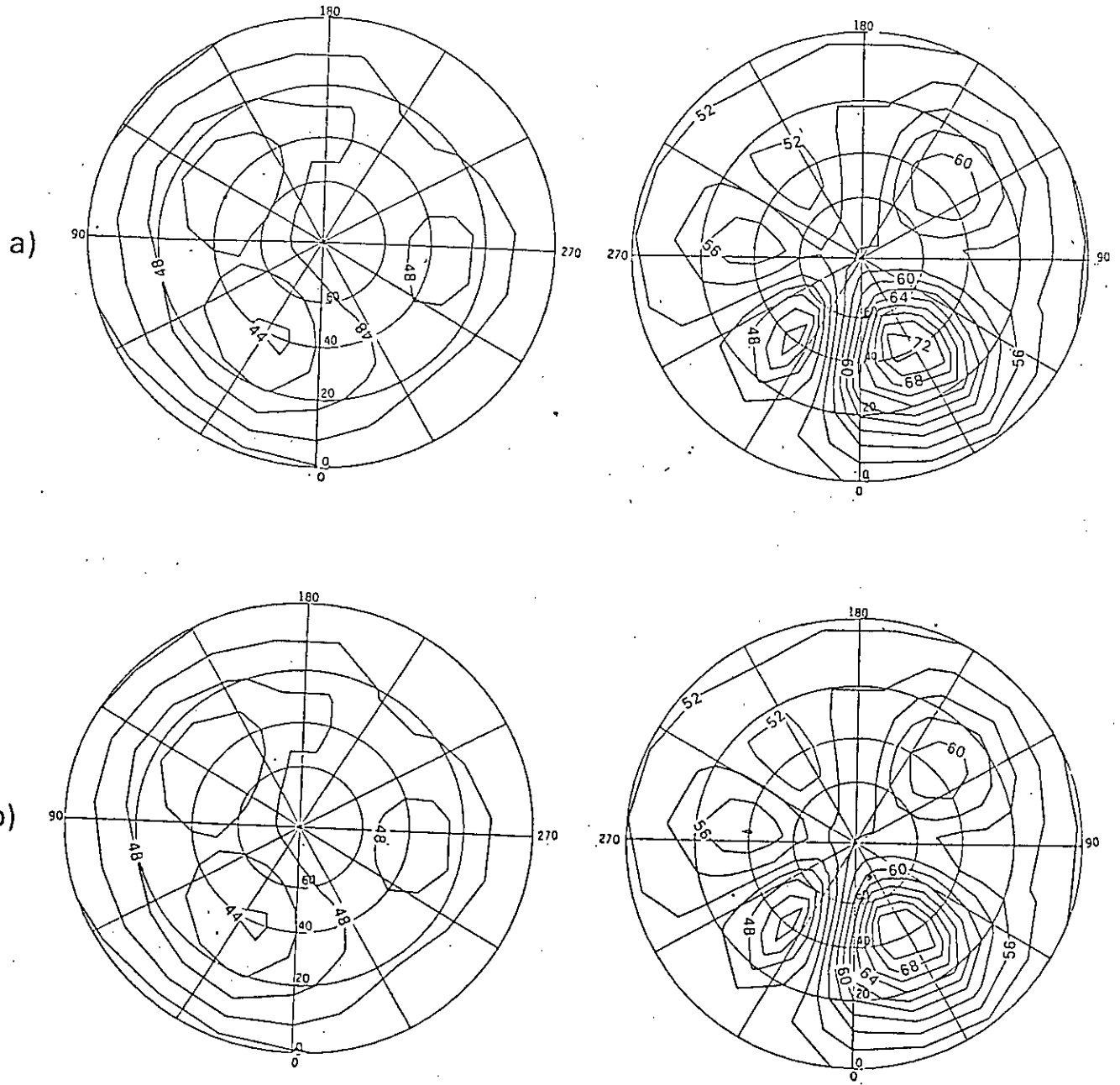


Figure 4.6 Field 2 analysed at t = 24 hrs using negative exponential weights. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 2 r.u. a) analysed values obtained using W=10 weights, a time radius of 12 hrs and a distance radius of 5000 km; b) as a), except W=5 weights and a distance radius of 3000 km were used;

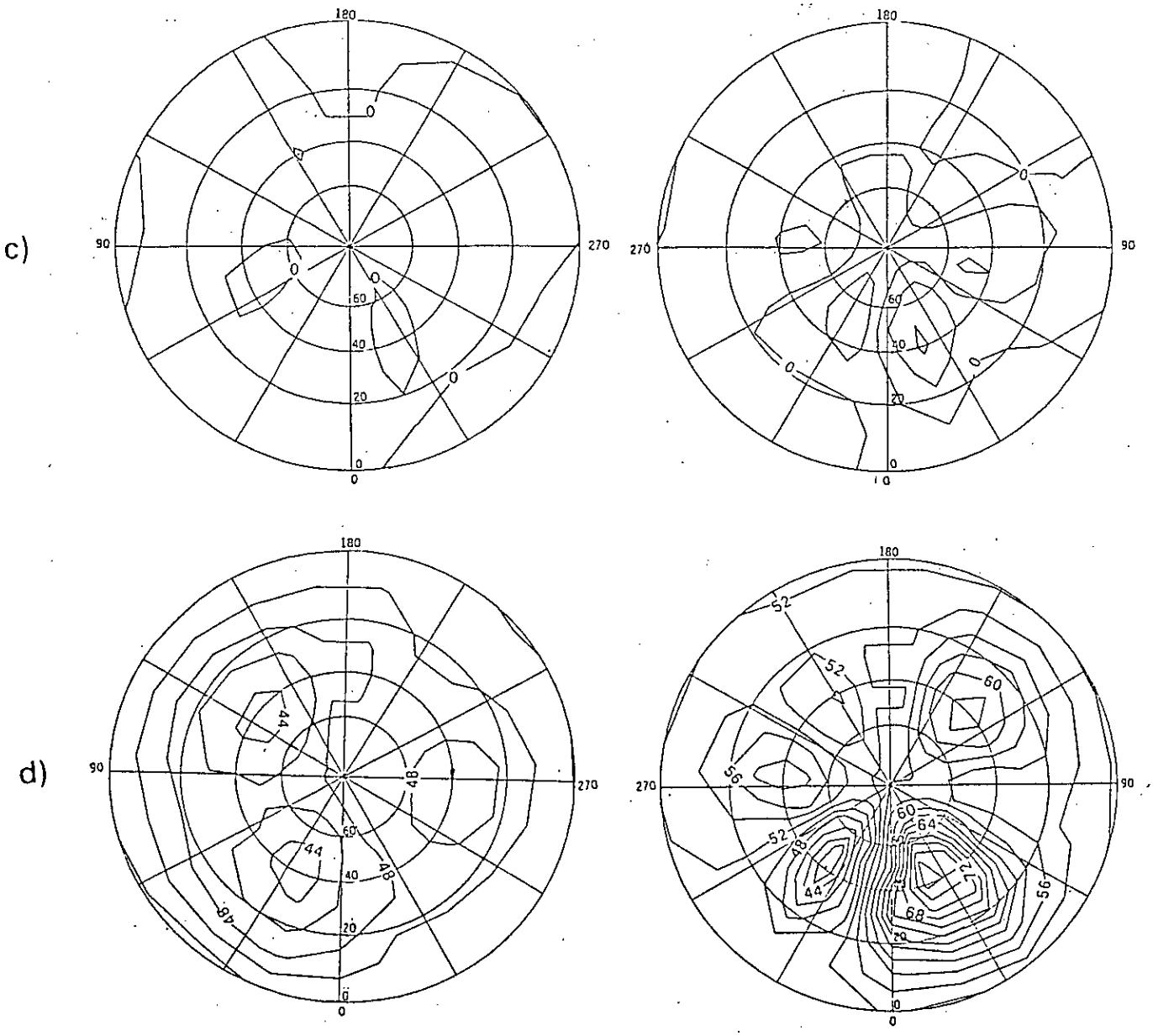


Figure 4.6 (cont.) c) difference between the field in b) and the model field; d) as a), except the distance radius used is 2000 km.

weights using a time radius of 12 hrs and a distance radius of 5000 km. Comparison of this field with the model field (Figure 4.1b) shows that the model wavenumber 1, 2 and 3 structure is satisfactorily estimated. This is in contrast to the linear-weighted analysis (Figure 4.3a) which fails to estimate any wavenumber 3 structure.

2. R.m.s errors of both analyses performed with negative exponential weights and using a distance radius of 3000 km are smaller than the error of the corresponding analysis made with linear weights, and again the W=10 analysis error is even smaller than the corresponding W=5 error. Both estimate the model field features slightly better than the linear analysis: Figure 4.6b shows the W = 5 negative exponential estimate and Figure 4.6c reveals that the difference between this estimate and the model field usually does not exceed 4 r.u (compared to 6 r.u for the corresponding linear analysis (Figure 4.3d)). Differences between analysed and model fields for the W=10 analysis (not shown) are even smaller.

Analyses made using distance radii of less than 3000 km reveal that r.m.s errors of all three weighting methods are generally of similar size. Fields analysed with negative exponential weights are less smooth than the model field, an example being Figure 4.6d, which shows the field analysed with W=10 negative exponential weights, using a time radius of 12 hrs and distance radius of 2000 km. This field is similar to the linear weighted analysis made using a time radius of 12 hrs and a distance radius of 1000 km (not shown), and the r.m.s errors are also similar.

4.3.4.3. Discussion of Results Made Using Negative Exponential Weights

Analyses made with negative exponential weights and using distance radii larger than 2000 km produce a better estimate of model fields 1 and 2 than do corresponding analyses made with linear weights. However when analyses are made with distance radii of 2000 km or less all three r.m.s errors are about the same size. This is because the exponentially weighted analyses are less smooth than the model field. A similar problem affects linear analyses at slightly smaller distance radii (about 1200 km). A possible explanation of this is given

by examining Figure 4.5. Compared to the linear weight, both negative exponential weights give very little weight to observations more than about 0.5 to 0.75 of a search radius away from the analysis point. Thus, for a given value of r_d or r_t, these weights have smaller *effective* radii than the linear weight has. As an example, consider an analysis made with W = 10 negative exponential weights and using a distance radius of 2000 km. Suppose that at a chosen grid point there are only two observations within the search radius: (A) 200 km (0.1 of a search radius) from the grid point, (B) 1300 km (0.65 of a search radius) from the grid point. The weight given to observation B is miniscule. After normalisation, the weight given to observation A will be close to 1 and the weight given to observation B will be close to 0. Suppose we then perform another analysis, this time with linear weights and using a search radius of 1200 km. At the chosen gridpoint, observation B is outside the search radius and hence, after normalisation, observation A will have a weight of one. The weightings given to the two observations in the linear scheme are thus similar to those given by the negative exponential scheme. This helps explain why minimum r.m.s errors are about the same size for analyses made using all three types of weight.

4.3.5. Summary of Results

Despite the fact that the analytical fields only crudely mimic the real atmosphere, one can conclude that, within the range of search radii tested :

1. R.m.s errors of analyses made using linear weights decrease when the distance radius is reduced, down to some limit. This limit is reached when the number of observations within the search radii is so small that the estimated field is less smooth than the model field.

2. Despite the high wave speeds of Fields 1 and 2, the size of the time radius appears to have little impact on the r.m.s error.

3. When analyses are made using small distance radii, there may be some gridpoints where there is only one

observation within the corresponding search radii. In a moving field (eg Fields 1 and 2) such analyses are less smooth than the model field and r.m.s errors are slightly higher than those of analyses made with slightly larger distance radii. Corresponding analyses made using a stationary field (eg Field 3) do not have this problem – the estimate is smooth and r.m.s errors are smaller than those of analyses made using slightly larger distance radii.

4. Analyses made using cosine or negative exponential weights are no better than those made with linear weights, although the way in which the r.m.s errors of negative exponential weighted analyses change with changing distance radius is slightly different.

4.3.6. Recommendations for Further Tests of the Time/Space Interpolation Scheme

The tests of the time/space interpolation scheme described here are made using idealised radiance fields. The next step is to test the scheme using more realistic stratospheric fields. Accordingly, in Chapter 7 we test the scheme using a field from the Met Office stratosphere/mesosphere model which is affected by a sudden warming.

The results of tests made on idealised fields indicate that changing the time radius causes only small changes to the quality of the estimate of the field. Thus the first tests made using the Met Office model field should examine this by comparing two analyses made with the same distance radius, but different time radii. The distance radius chosen for such tests is 2000 km, and the time radii are 6 hrs and 12 hrs. Operational Met Office analyses are usually made using 24 hrs of data. Since these tests will help users of such analyses to evaluate the confidence which can be placed in them, the tests using the model field are also performed using 24 hrs of data, and hence 12 hrs is the largest time search radius that can be used.

Results of tests presented in this chapter also demonstrate that reducing

the distance radius reduces the analysis error, provided the distance radius is greater than 1500 km. Concentrating on tests made with a time radius of 12 hrs we note that when a distance radius of less than 1500 km is used the r.m.s error continues to fall (for Field 3), or rises slightly (for Fields 1 and 2). This effect is due to the small number of observations within the distance radius, and is governed by the motion of the field; Fields 1 and 2 contain travelling waves only, whilst Field 3 is stationary. However, when the time/space interpolation scheme is used to estimate more realistic stratospheric fields we do not know how the quality of the analysis will change when the distance radius is reduced to less than 1500 km because, as described in Chapter 2, the field may typically consist of a combination of travelling and stationary waves. In addition, many real stratospheric fields have length scales smaller than those of Fields 1, 2 and 3. Accordingly, the effect on the analysis of changing the distance radius has to be examined further, and hence further tests are performed on the Met Office model field using a constant time radius and two distance radii. The time radius used is 12 hrs, and the distance radii used are 1000 km and 2000 km. In addition, a further test is made using the radii used operationally by the Met Office, namely a time radius of 12 hrs and a distance radius of 500 km. Results of these tests appear in Chapter 7.

CHAPTER 5

TESTS OF RETRIEVAL/ANALYSIS SCHEMES USING MET. OFFICE STRATOSPHERE/MESOSPHERE MODEL FIELDS : METHOD OF SIMULATION AND DATA USED

5.1. Purpose of the Tests

The aim of the tests to be described in the next three chapters is to evaluate the performance of two retrieval/analysis schemes for obtaining stratospheric temperatures from the TOVS instrument on the TIROS-N series of polar-orbiting satellites. The retrieval/analysis scheme tested in Chapters 6 and 7 is similar to that used by the UK Meteorological Office, so these tests will help users of the analyses to evaluate the confidence which can be placed in them. The temperatures retrieved by the method used in Chapter 6 are also analysed using the sequential estimation analysis scheme, and results of these tests appear in Chapter 8. Comparison of results of the tests in Chapters 7 and 8 are made, and the relative strengths and weaknesses of each analysis scheme are identified.

A difficulty in testing any retrieval/analysis scheme using real observations is the absence of adequate 'ground truth' observations. Rocketsondes observe temperature in the upper stratosphere but rocket flights are infrequent and badly spaced. Radiosondes observe in the lower stratosphere and give better global coverage, but even so there are few observations made over oceans or in the southern hemisphere. Many workers (eg Nash and Brownscombe, 1983; Pick and Brownscombe, 1979, 1981; Schmidlin, 1984; Barnett et al, 1975) have compared retrieved temperature profiles with coincident rocket and radiosonde ascents. However, the sondes themselves have instrumental errors, and it is often unclear whether differences between sonde and satellite profiles are due to systematic errors of one of the methods or to differences in spatial and time

resolution (see eg McMillin et al , 1983 (and references therein), for a fuller discussion). Comparisons of satellite and sonde-derived fields have been made by eg Pick and Brownscombe (1979, 1981) Grose and Rodgers (1986), Rodgers (1984). Because of the very poor spatial coverage of rocketsonde observations, fields have been derived using mainly radiosonde data. Hence it is not practical to derive such fields at levels above the maximum ascent level of the radiosondes (about 10 mb). However, even at stratospheric levels below 10 mb, the poor radiosonde coverage in the southern hemisphere and over the oceans means that sonde-derived fields may be of poor quality in these regions, and field values in these regions are sometimes subjectively estimated by the human analyst. In addition, Rodgers (1984) notes that stratospheric radiosonde errors can be substantial above 100 mb. Over 15 types of radiosonde are in operational use, and since systematic errors vary from sonde-type to sonde-type, stratospheric analyses can become nearly impossible unless the temperature fields are adjusted to ensure observational compatibility. Comparison with ground truth for fields analysed from satellite data is also difficult. Possible problems with analyses have been reported by Al-Ajmi et al (1985) and by Clough et al (1985). The former paper gives evidence that there may be temporal variations too rapid for proper resolution, and the latter provides evidence that the vertical temperature structure is not always adequately resolved. Accordingly, because of these difficulties with ground truth observations, we have chosen to test the retrieval and analysis schemes in a simulation experiment which uses an atmosphere calculated in a numerical model. Simulated observations are then calculated by computing the radiances which would be observed from this model atmosphere from a TOVS-like instrument, including the effects of instrumental noise. Details of the numerical model are given in Section 5.2, whilst Section 5.3 describes the simulation of observed radiances from the model field.

Temperatures are retrieved from the simulated observations using a regression model similar to that used by the UK Meteorological Office (Pick and Brownscombe, 1981), the main difference being that their scheme gives thicknesses of fairly thick layers of atmosphere, whereas ours gives the temperature profile at 31 pressure levels from 0.2 to 570 mb. Chapter 6 describes the regression model and the method of obtaining the regression coefficients, together with the results of tests of the retrieval scheme. The retrieved temperatures are interpolated to a grid by two methods: 1) a time/space analysis method which is used operationally by the UK Meteorological Office; 2) the sequential estimation of Fourier components, a technique which has been used to analyse LIMS observations. Both techniques are described in Chapter 3. Results of tests of the time/space interpolation scheme appear in Chapter 7, whilst results of tests of the sequential estimation scheme appear in Chapter 8.

5.2. The Stratosphere/Mesosphere Model

The retrieval and analysis schemes are tested below in a simulation experiment which uses an atmosphere calculated by a numerical model. The model gives a reasonable representation of the 'true' atmosphere, and has been used in a number of studies of the stratosphere (eg O'Neill and Pope, 1988; Fairlie and O'Neill, 1987; Shine, 1987). However, it is not capable of reproducing certain phenomena, such as tides, which exist in the real stratosphere. The model used is the UK Meteorological Office stratosphere/mesosphere multi-level model (Fisher, 1987) based on the primitive equations. These equations are solved to fourth-order accuracy in the horizontal, and to second-order accuracy in both the vertical and in time, using energy conserving 'box' type finite differences and leapfrog integration. The model utilises a regular grid in spherical coordinates with gridpoints at intervals of 5° in latitude and longitude and 33 levels between 100 and 0.001 mb, which are

equally spaced in log pressure, and are approximately 2 km apart. To avoid having to represent the troposphere, a lower boundary condition is imposed near the tropopause, namely the geopotential height of the 100 mb surface specified from analysed observations.

Our simulation uses one day's output at 1 hour intervals from a run with lower boundary heights corresponding to 18/1/1987. On that day a sudden warming (described in Chapter 2) was present in the modelled northern hemisphere. Such warmings cause large and rapid changes to the temperature structure of the stratosphere – there may be temperature rises of the order of 50 K over a few days. As discussed in Section 2.3, this phenomenon provides the most stringent circumstances for testing both the retrieval and analysis schemes.

5.3. Simulation of Radiances

To calculate a simulated radiance we use a temperature profile which extends from the surface to above the level of the topmost weighting function (2 mb). The Met Office model calculates temperatures between 100 mb and 0.001 mb, and hence at pressure levels below 100 mb we must use temperatures from another source. Here, we use Met Office Central Forecasting Office (CFO) fields; the fields have been calculated chiefly from radiosonde observations, and provide us with temperatures from the surface to 100 mb. The TIROS-N satellite views the atmosphere by scanning from one side of the vertical to the other at 8 scan angles. Adjacent observations are then averaged in blocks of 4 so that the 16 observations of 2 successive scans are combined to give 4 'superobservations' at effective angles of $-30°$, $-10°$, $10°$ and $30°$ from the vertical. Figure 5.1 shows the superobservations made by a TIROS-N series satellite (NOAA-7) in a 24 hr period. There is almost global data coverage, though some areas in the tropics are free of observations. Model or CFO

temperatures are linearly interpolated in space and time to these observation points, and the interpolated temperatures are used to calculate radiances that the satellite would 'observe'. In these tests no data loss due to calibration sequences is assumed.

Observations are simulated for 8 TOVS channels using the radiative transfer equation (3.6), which expresses the spectral radiance as the vertical integral of the product of Planck function and weighting function. The largest contribution to the radiance comes from pressure levels close to the peak of the weighting function. Indeed, channel radiance may be considered to give a measure of the temperature of a layer 10–15 km thick situated about the peak of the weighting function. The weighting functions for the 8 channels used in the retrieval scheme are shown in Figure 5.2. Table 5.1 shows the pressure level at each channel's weighting function peak and its central wavenumber ν.

CHANNEL	WEIGHTING FN. PEAK/mb	CENTRAL WAVENUMBER/cm^{-1}
HIRS 1	30	669
HIRS 2	60	679
HIRS 3	100	690
MSU 23	280	1.832
MSU 24	100	1.932
SSU 25	15	667
SSU 26	5	667
SSU 27	2	667

Table 5.1 TOVS weighting functions for a vertical view. The acronyms refer to the instruments mounted on TOVS. HIRS denotes the High Resolution Infrared Radiation Sounder, MSU the Microwave Sounding Unit, and SSU the Stratospheric Sounding Unit.

The weighting functions depend upon the angle of view. However, to

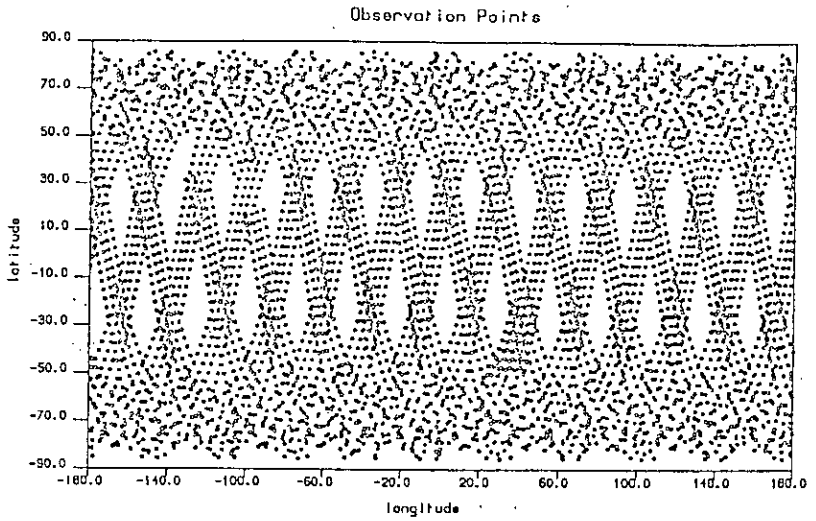


Figure 5.1 Superobservation points of the TOVS instrument on NOAA-7 for 18/1/87. Observations are made at the 4 scan angles -30°, -10°, 10° and 30° from the local vertical.

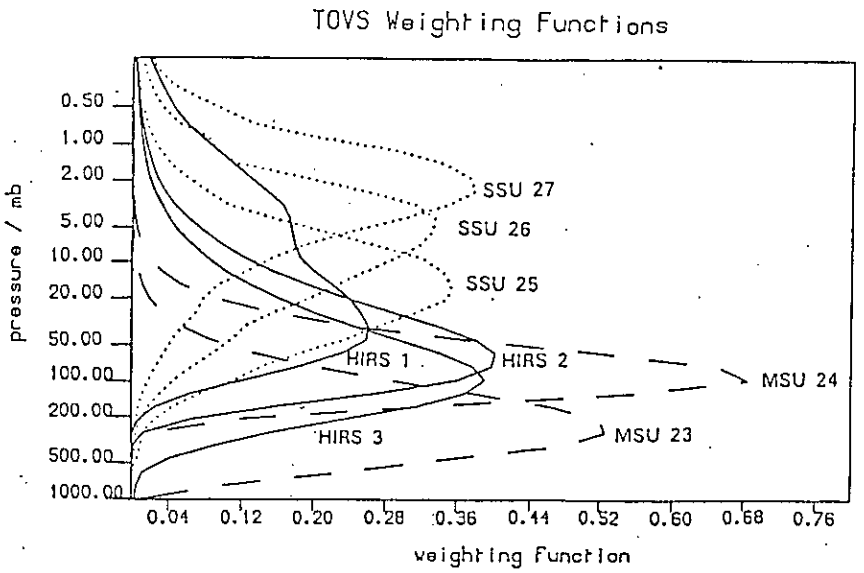


Figure 5.2 The 8 TOVS channel weighting functions expressed between 0.2 and 1000 mb, for a vertical view.

simplify the calculations we have simulated all radiances using the weighting function appropriate to the vertical view. Most operational retrieval schemes also use this simplification: a statistical correction scheme is applied to radiance observations made at an angle to the vertical in order to calculate the radiances which would have been observed at the vertical. However, this need for a 'nadir correction' introduces a slight error to operational retrievals which is not present in our simulation experiment (see eg Koehler, 1989; Le Marshall and Schreiner, 1985).

The transmission profile, and hence weighting function, for each channel is calculated using a numerical technique developed by McMillin and Fleming (1976): first, transmission profiles are calculated for a small number of representative and extreme atmospheres using the line-by-line method (Drayson, 1966), and then these pre-computed profiles are interpolated to any arbitrary profile. 'Observed' radiances can thus be calculated by evaluating equation (3.6) and then adding randomly generated numbers to simulate the radiometric noise of the instrument. These numbers are taken from a Normal distribution which has a mean of 0.0 r.u and a standard deviation of 0.3 r.u (HIRS and SSU channels), or a mean of 0.0 K and a standard deviation of 0.15 K in equivalent blackbody temperature (MSU channels). The choice of these values is based on validation studies of these instruments (Pick and Brownscombe, 1979; Eyre, personal communication, 1989).

CHAPTER 6

TESTS OF THE REGRESSION RETRIEVAL SCHEME

In this chapter, a regression retrieval scheme is tested in a simulation experiment which uses an atmosphere calculated in a numerical model. The scheme is described in Section 6.1, whilst the data used to calculate the regression coefficients is discussed in Section 6.2, along with tests of assumptions made in the formulation of the regression model. Results appear in Section 6.3, and these results are discussed in Section 6.4.

6.1. The Regression Model

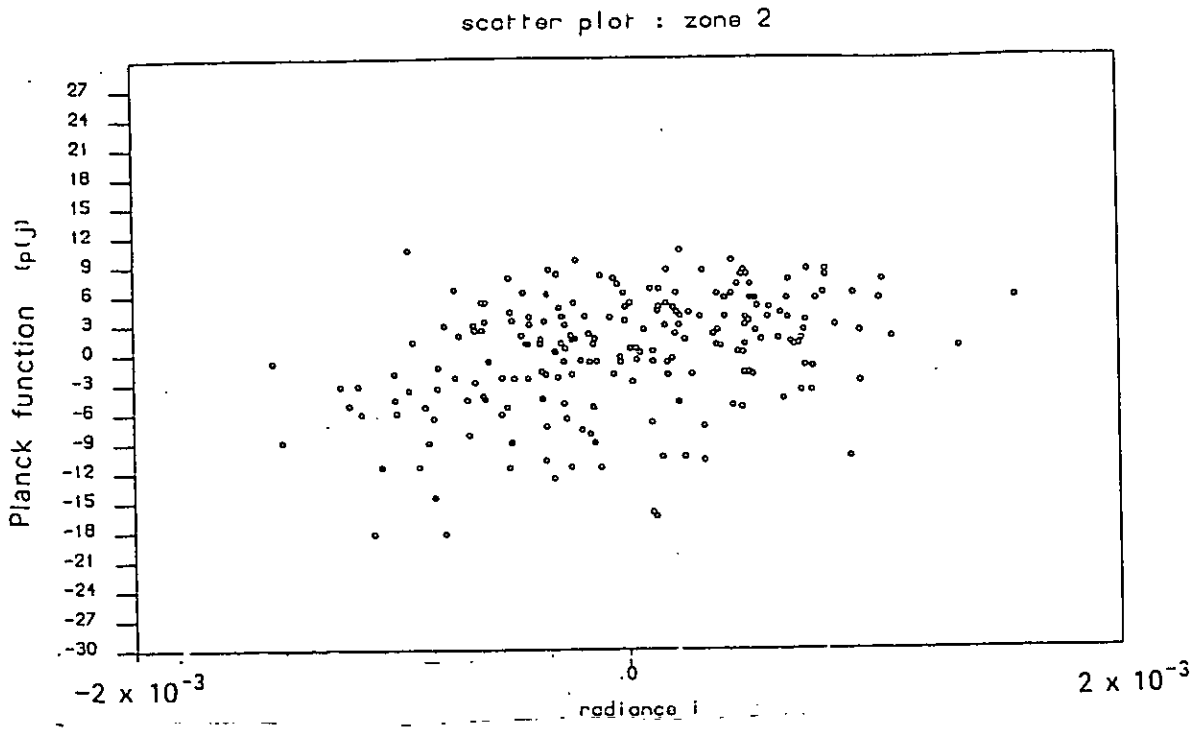
Planck function, and hence temperature, is retrieved from radiance measurements by regression. In the regression model we must express the Planck function at a reference wavenumber U. Since 6 of the 8 channels used in the regression model are in the infrared region (see Table 5.1), it is convenient to choose a reference wavenumber which is also in the infrared; we choose U to be 668 cm^{-1}. It is an assumption of the regression model that radiance is linearly related to the Planck function at 668 cm^{-1}. Whilst this is approximately true in the infrared, it is not the case in the microwave region. An illustration is Figure 6.1a, which shows a scatter plot of deviations of MSU channel 25 'observed' radiances from the mean (ie $(R_j - \bar{R}_j)$ against deviations of Planck function at 85 mb (calculated at 668 cm^{-1}) from the mean (ie $(B_i - \bar{B}_i)$). There is no clear linear relation between $(R_j - \bar{R}_j)$ and $(B_i - \bar{B}_i)$. We need to re-express the observed radiance, R_j, in such a way that the relation between radiances at all wavenumbers and Planck function (expressed at 668 cm^{-1}) is close to linear. This is done by calculating a standardised radiance: the standardised radiance, χ_j, of channel j is easily calculated from R_j, the measured radiance of channel j via

$$\chi_j = c_1 U^3 / [\exp (c_2 U / r) - 1] \qquad (6.1)$$

where $r = c_2 \nu / \ln (1 + c_1 \nu^3 / R_j)$,

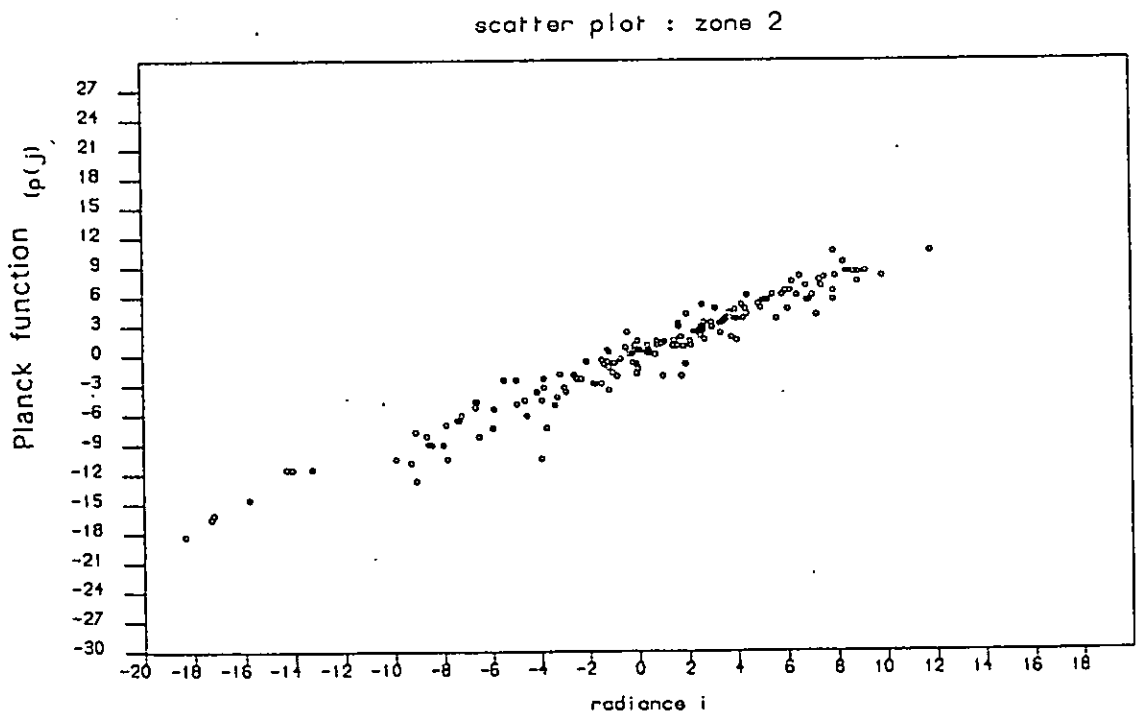
ν is the channel wavenumber, $c_1 = 1.19096 \times 10^{-5}$ mW m^{-2} cm^4 $ster^{-1}$ and $c_2 = 1.43879$ cm K. Figure 6.1b is like Figure 6.1a, except that deviations of

a)



Mean measured radiance, $\bar{R}_j$ = 6.68 x 10^{-3} r.u

b)



Mean standardised radiance, $\bar{X}_j$ = 43.45 r.u

Figure 6.1 Scatter plot of deviations of Planck function at 85 mb (evaluated at 668 cm^{-1}) from the mean against: a) deviations of MSU channel 25 'observed' radiances from the mean; b) deviations of MSU channel 25 standardised radiances from the mean. Data is for zone 2.

standardised radiance from the mean (ie $(\chi_j - \bar{\chi}_j)$) is plotted instead of deviations of measured radiance from the mean (ie $(R_j - \bar{R}_j)$). There is evidence here of a linear relation between standardised radiance and Planck function.

The estimated deviation of profile Planck function, $\hat{\Delta B_i}$, from the mean is written as a linear combination of the deviation of the radiances from the mean radiance, namely

$$\hat{\Delta B_i} = a_{0,i} + \sum_{j=1}^{N} (\chi_j - \bar{\chi}_j) a_{j,i} \qquad (i= 1,...,31) . \tag{6.2}$$

The $a_{o,i}$ and $a_{j,i}$ are predetermined in advance by ordinary least squares as described in Section 6.2. The mean Planck function and the mean standardised radiance, $\bar{\chi}_j$, are calculated from a set of rocketsonde temperature measurements. N is the total number of channels and χ_j is the observed standardised radiance. Retrievals are performed at 31 pressure levels equally spaced in log pressure between 0.2 and 570 mb.

The retrieved Planck function, $\hat{B}_i$, for level i is then given by

$$\hat{B_i} = \bar{B}_i + \hat{\Delta B_i} \qquad (i = 1,...,31) \tag{6.3}$$

where $\bar{B}_i$ is the mean Planck function for pressure level i. It is straightforward to calculate temperature from the retrieved Planck function.

6.2. Calculation of Regression Coefficients

The regression analysis is based on a dataset of 1200 temperature profiles. Each profile is calculated using a combination of a radiosonde and a rocketsonde measurement: the profile uses radiosonde temperatures from the surface to the lower stratosphere (which is the vertical range of the instrument), and temperatures from a quasi-coincident rocketsonde at higher

levels. Standardised radiances are calculated from these measurements using equations (3.6) and (6.1). The data are divided into 7 zones (Table 6.1) according to the latitude and season of each rocketsonde measurement, and regression coefficients are calculated for each of the zones. This is done to restrict the range of atmospheric conditions over which the regression analyses are applied.

ZONE	LATITUDE / SEASON	NO. OF ASCENTS IN EACH ZONE
1	70° – 90°, winter	64
2	50° – 70°, "	212
3	30° – 50°, "	124
4	30°N –30°S, all seasons	400
5	30° – 50°, summer	125
6	50° – 70°, "	204
7	70° – 90°, "	71

Table 6.1. Latitude/season zones for which regression coefficients are calculated. 'Winter' is the six months between October and March (for the northern hemisphere) or between April and September (for the southern hemisphere). 'Summer' is the six months between April and September (for the northern hemisphere) or between October and March (for the southern hemisphere).

6.2.1. Relation of The Sonde Dataset to the Numerical Model

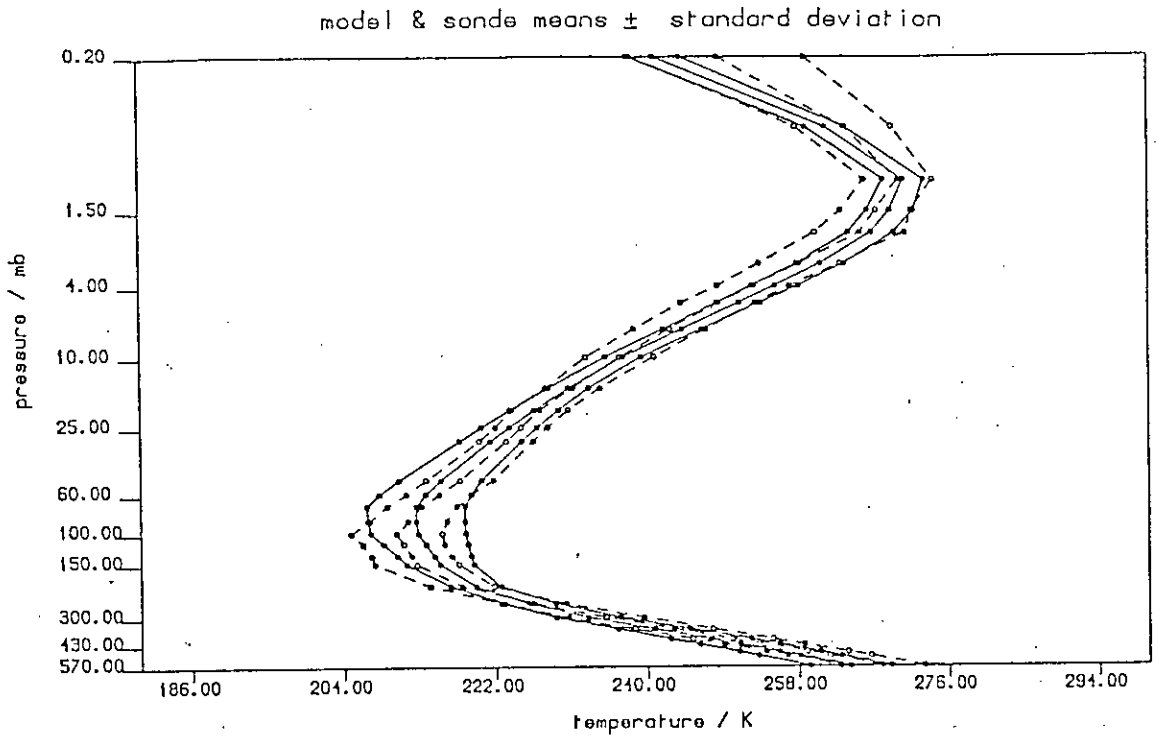
The dataset used to calculate the regression coefficients contains rocketsonde measurements of the *real* atmosphere. However, in this thesis the retrieval and analysis schemes are tested in a simulation experiment which uses an atmosphere calculated by a numerical model. It is important that the model adequately reproduces the real atmosphere observed by the rocketsondes. This is examined by comparing the means and standard

deviations of the two datasets.

We consider first the southern hemisphere, zone 5. The model field in this zone is similar to the climatology of the sonde data. Figure 6.2a shows means and standard deviations of sonde and model datasets in zone 5. The dashed lines show the mean of the sonde temperatures together with departures of one standard deviation. The solid lines show the corresponding statistics for the *model* dataset in this zone. Both sonde and model standard deviations are low, indicating the lack of variation in the summer stratosphere. Moreover, at most pressure levels the means and standard deviations of both datasets are similar, implying that the model field in zone 5 is similar to that observed by rocketsondes. In zones 6 (Figure 6.2b) and 7 (Figure not shown) the model mean also lies within one sonde standard deviation of the sonde mean, and both model and sonde standard deviations are small. By contrast, the model mean in zone 4 (Figure 6.2c) is more than one sonde standard deviation away from the sonde mean between 1.5 and 2 mb and15 and 75 mb. This suggests that the model field is somewhat unrealistic in the tropics.

Whilst sonde and model statistics are similar for zones 5 to 7, the same is not true for zones 1 and 2. On 18/1/1987 the model reproduces a sudden warming in these latter zones. In Table 6.1 'winter' is defined as the 6 month period between October and March (for the northern hemisphere) or between April and September (for the southern hemisphere). Within such a large time period only a small proportion (if any) of rocketsondes will observe a sudden warming, thus it is likely that sonde data in zones 1, 2 and 3 will not be representative of sudden warming conditions. Figure 6.2d is like Figure 6.2a, except that data for zone 1 are shown. As one might expect in a winter stratosphere, both sonde and model datasets have high standard deviations. The two mean profiles differ greatly at a number of pressure levels. Between 3

a)



b)

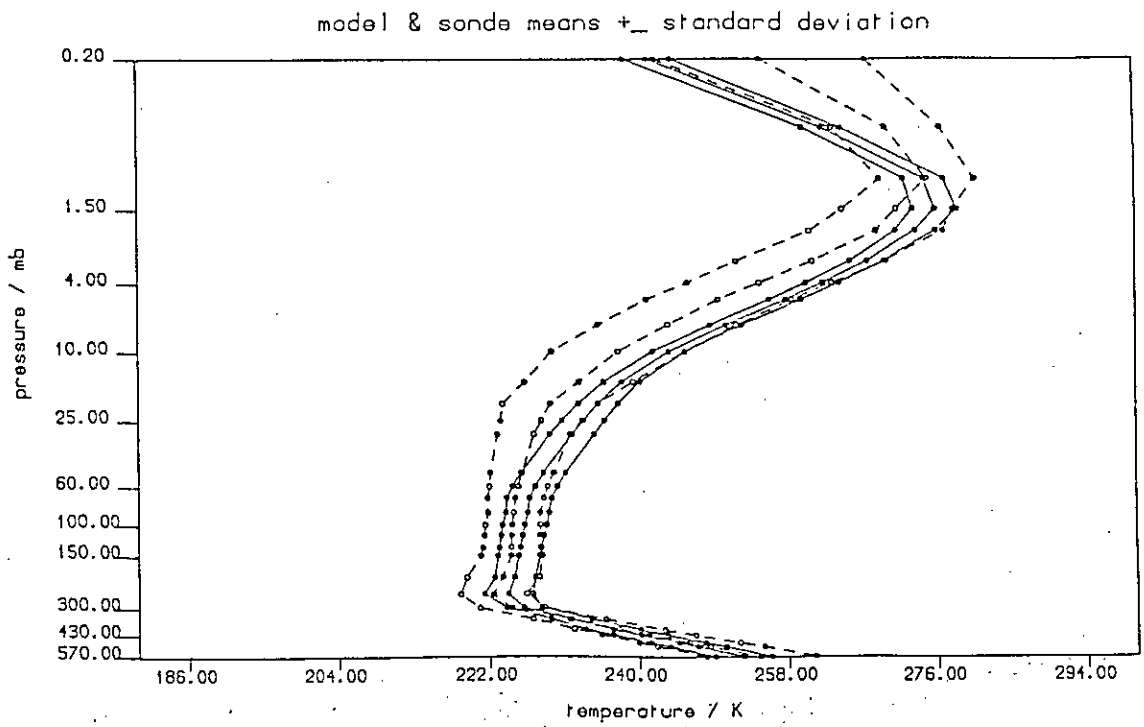
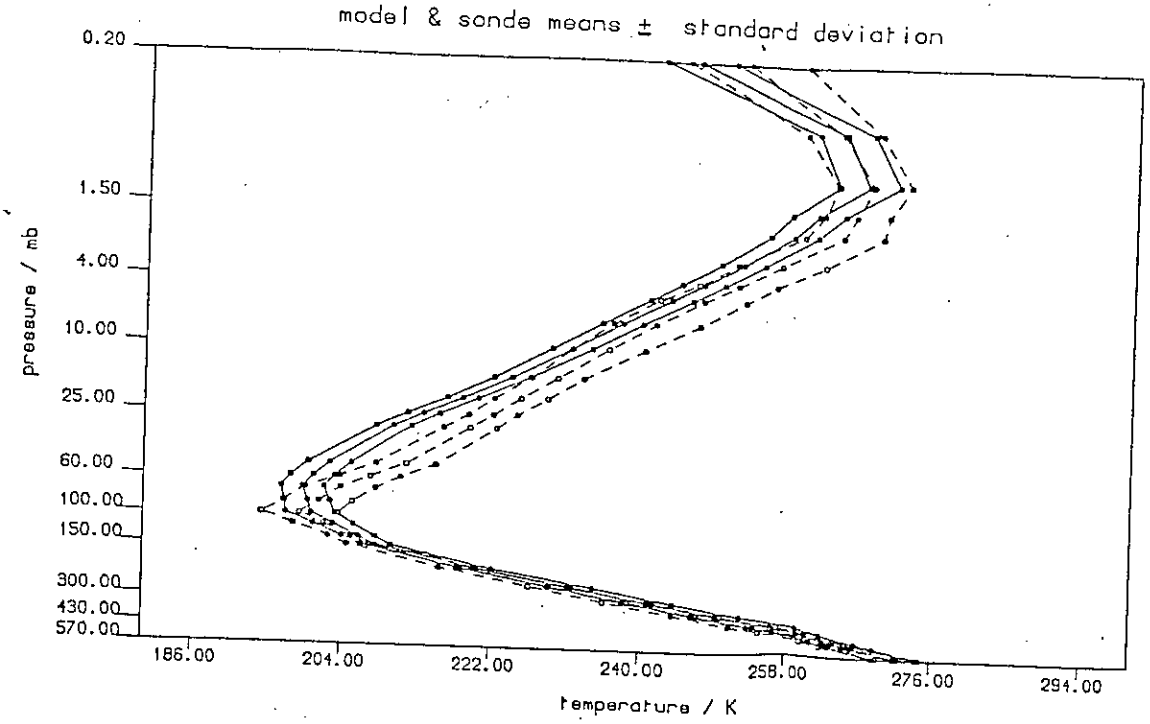


Figure 6.2 Mean temperature profile (K), and this profile plus and minus the sonde standard deviation for the sonde dataset (dashed lines) and model (solid lines). a) zone 5; b) zone 6;

c)



d)

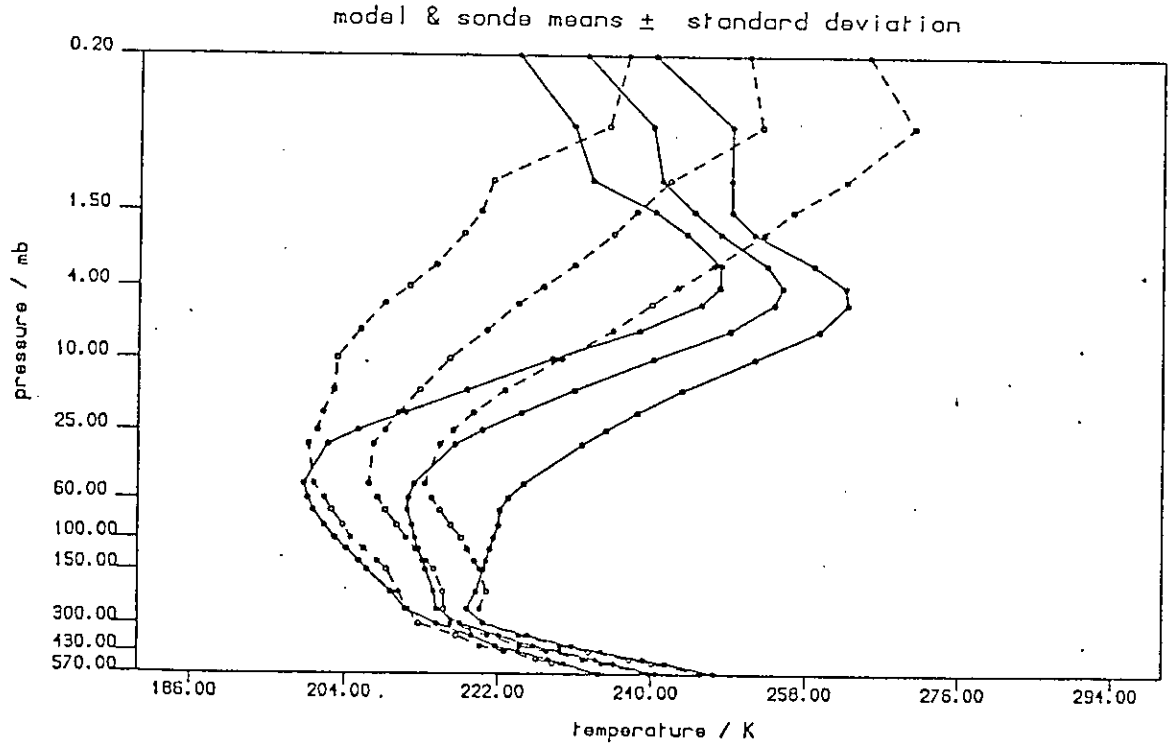


Figure 6.2 (cont.) c) zone 4; d) zone 1.

and 50 mb the model mean temperature is greater than one sonde standard deviation away from the sonde mean temperature. Thus model conditions in zone 1 differ greatly from those observed by rocketsonde. This is also true in zone 2 (not shown), while in zone 3 (not shown) differences are also large, although in general the model mean differs from the sonde mean by less than one standard deviation. It is important to test the retrieval scheme in a sudden warming precisely because of these extreme differences between sudden warming conditions and mean sonde values. If the scheme performs well in a sudden warming, it is reasonable to conclude that it will perform well in most conditions.

These results have important implications for the method of retrieval by regression, since the sonde dataset used to calculate the regression coefficients is used widely (for example, in the NOAA regression retrieval scheme (Smith and Woolf, 1976; Smith et al, 1979); many major research groups use NOAA-retrieved stratospheric temperatures). The discussion above suggests that the sonde dataset appears to be unrepresentative of sudden warming conditions. However, it is likely that the limitations of the sonde data are even more widespread, because the measurements are made at a limited number of sonde stations (24), and most of these are in the northern hemisphere (out of the 1200 ascents in the dataset, only 118 are made in the southern hemisphere). The regression zones described in Table 6.1 are based on the assumption that stratospheric climatology depends on latitude and season, but not hemisphere. However, recent work (eg Andrews, 1989) indicates that there are noticeable differences between the climatologies of the northern and southern hemispheres of the middle atmosphere. Accordingly, future stratospheric retrieval research might profitably utilise methods which have been used to improve tropospheric retrievals. For example, 'stratified climatology' (eg Uddstrom and Wark, 1985) uses the radiances to 'point' to a

class of atmospheres to which the profile probably belongs. The regression coefficients are then obtained from the statistics of the class. A related technique employs a library search (eg Chedin et al, 1985). Another possibility is using datasets which are taken from a forecast model rather than climatology.

6.2.2. Testing of the Regression Model

It is an assumption of the regression model that radiances from all 8 channels should be used. Clearly it is of interest to investigate if a certain amount of effort could be saved by using less than 8 channels in the regression. Accordingly, we have produced scatter plots of radiances in pairs of channels to discover the degree of correlation between them. Several of the channels show a high degree of association; an example is Figure 6.3, which shows a plot of HIRS-3 radiance against MSU-24 radiance in zone 7. Regression coefficients were calculated using systematically fewer and fewer channels. The resulting minimal regression models were then used to estimate sonde temperatures. The residual r.m.s error (i.e the r.m.s error of the difference between true and estimated sonde temperature) suggested that the quality of the retrieval decreases when data from one or more channels are omitted. An example is Figure 6.4, which shows the residual r.m.s error of four zone 3 minimal regression models. The models use: a) all eight TOVS channels; b) seven channels (SSU channel 27 is omitted); c) five channels (all three SSU channels are omitted); d) one channel (HIRS channel 1). The residual error of the 8-channel model is smaller than the error of the other models. The error of the 7-channel model is the same or slightly larger than the 8-channel model error at most pressure levels, but is over 1 K larger near 1.5 mb, which is close the level of the (missing) SSU channel 27 weighting function peak. The error profile of the 5-channel model matches that of the 8-channel model from 570 to 50 mb, but is noticeably larger at pressures less than 50 mb. This is to be

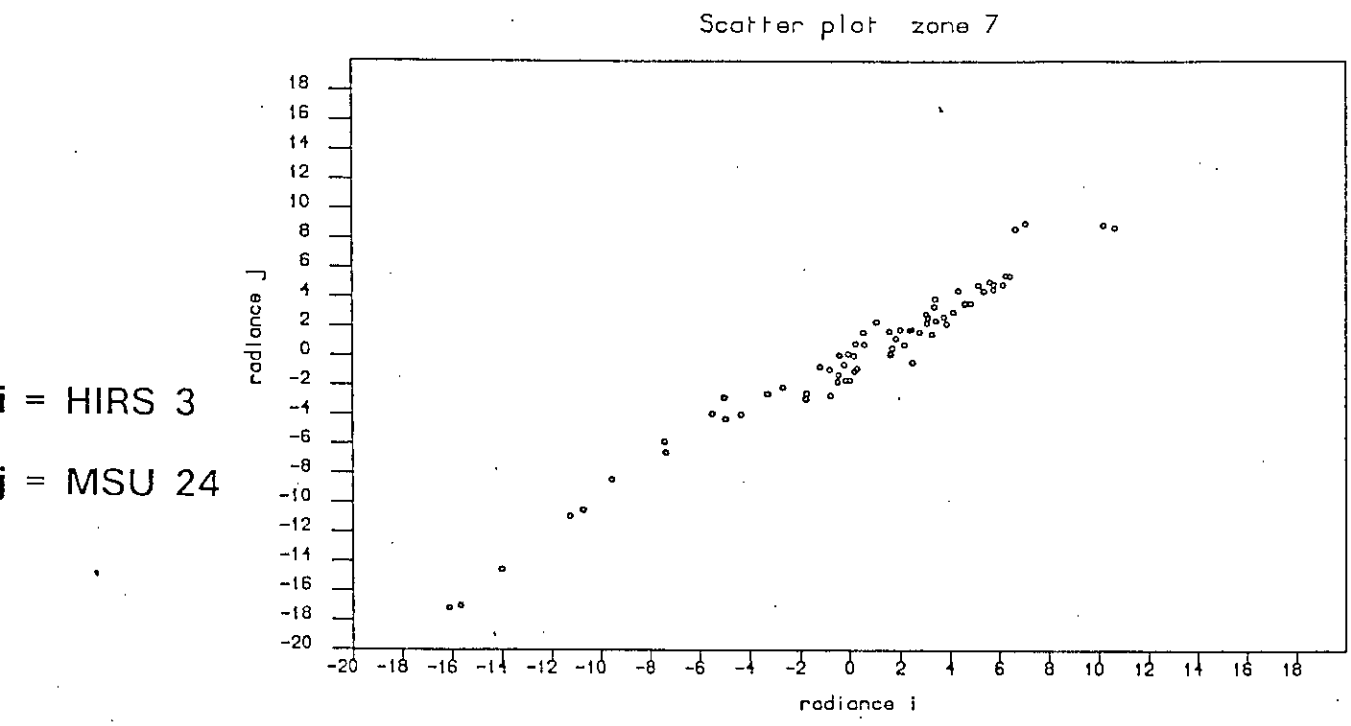


Figure 6.3 Scatter plot of deviations of standardised HIRS-3 radiances from the mean against corresponding values of MSU-25 radiance. Data is for zone 7.

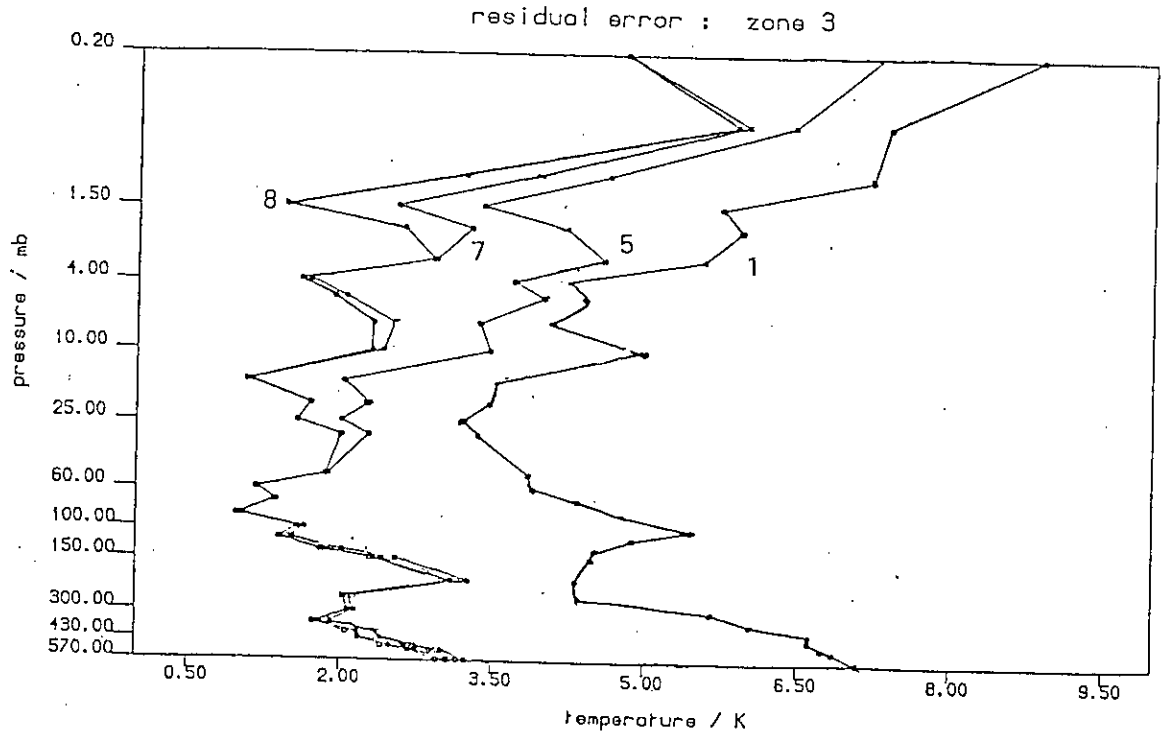


Figure 6.4 Residual r.m.s error (in K) of four zone 3 minimal regression models: i) 8-channel model which uses all eight TOVS channels; ii) 7-channel model which excludes SSU channel 27; iii) 5-channel model which excludes all three SSU channels; iv) 1-channel model which uses only HIRS channel 1.

expected as the 5-channel model does not use the three SSU channels, whose weighting functions peak at 15, 5 and 2 mb. The error of the 1-channel model is the largest of the four error profiles and exceeds the 8-channel error by at least 1.5 K at all levels. Such results are equivalent to showing that each channel makes a significant contribution to information content (Peckham, 1974)

The regression model also assumes that Planck function is linearly related to radiance. To test the validity of this assumption scatter plots of Planck function at level k against radiance for channel j for a number of k,j pairs were produced. These plots provide no strong evidence that the relationship between Planck function and radiance is anything other than linear. An example of such a plot is Figure 6.1b, which shows the standardised radiance for MSU channel 24 plotted against the Planck function at 85 mb. We conclude from these scatter plots that an appropriate model is linear regression of all 8 channel radiances against Planck function.

6.3. Retrieval Results

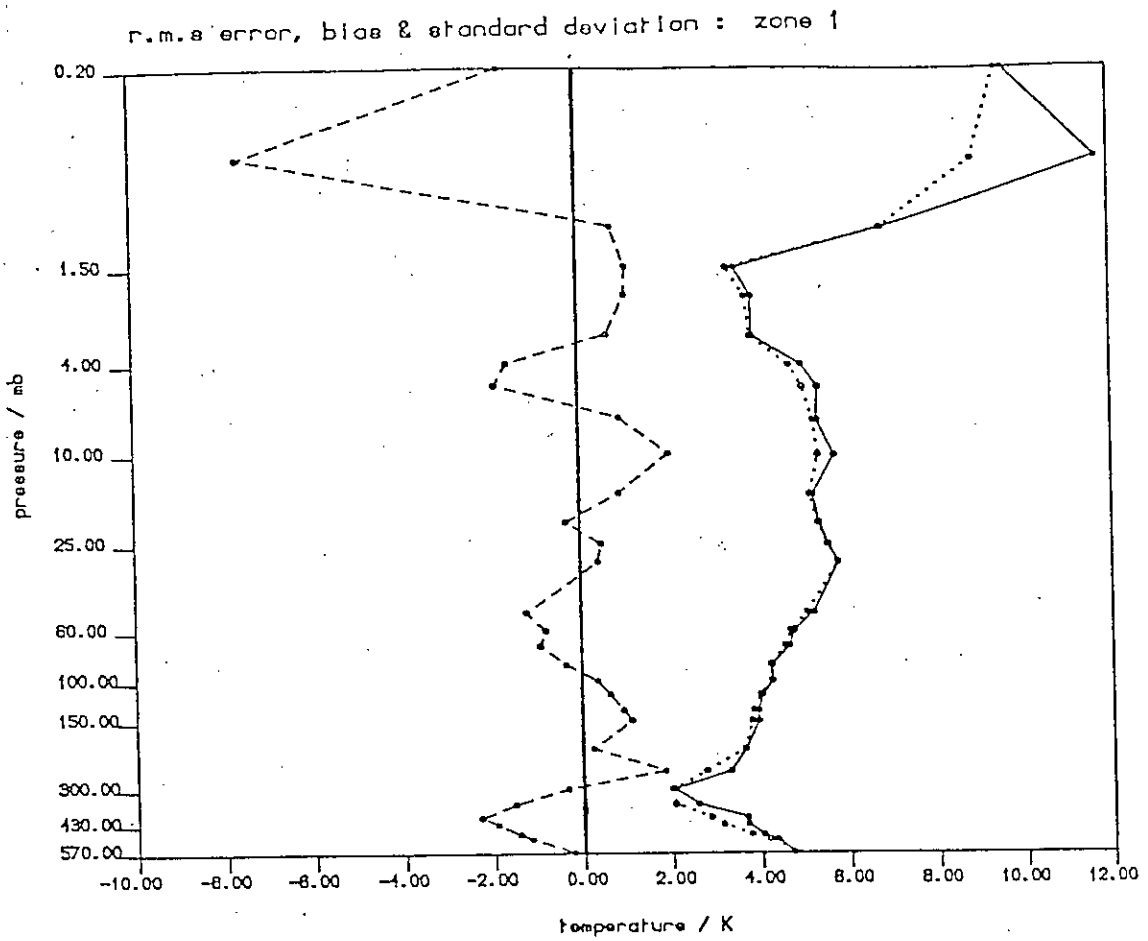
We now turn to investigate how well the regression scheme works in practice. our overall aim is to test the behaviour of the combined retrieval/analysis scheme, but it is important to be able to study the retrieval and analysis parts separately. To this end, we perform a test in this Chapter which does not involve time/space analysis. The tests are made using model profiles at 31 pressure levels. The model profiles are interpolated to the satellite observational points shown in Figure 5.1, and these are taken to be the 'true profiles'. 'Observed' radiances corresponding to these profiles are calculated using equation (3.6) and these are retrieved using the regression model described in Section 6.1 to give 'retrieved profiles'. We now compare 'true' and 'retrieved' profiles by examining the r.m.s retrieval error, the retrieval

bias, and the standard deviation of the difference between retrieved and model temperatures.

For all regression zones the r.m.s retrieval error at pressures greater than than that of the peak of the bottommost weighting function (280 mb) is generally higher than at 300 mb, and the r.m.s error at pressures less than that of the topmost weighting function peak (2 mb) is generally larger than the r.m.s error at 1.5 mb. This is of course to be expected, as away from the region containing the weighting function peaks the satellite measurements provide little information about the temperature. Other workers have also noticed this behaviour. Rodgers (1984) and Grose and Rodgers (1986) compared fields derived from both satellite and sonde data, and concluded with the recommendation that operational retrievals should not be made above the level of the topmost weighting function peak.

As anticipated, errors in the zones containing a sudden warming (zones 1 and 2) are high. Difficulties arise because of the small vertical temperature structure present in a sudden warming (discussed further in Section 6.3.1) and because the sonde measurements in these zones were made in conditions different to those in the sudden warming (see, for example, Figure 6.2d). Figure 6.5a shows the r.m.s error (solid), bias (dashed) and standard deviation (dotted) profiles for zone 1. For pressures higher than 1.5 mb the error is between 2 and 6 K. The standard deviation has nearly the same values as the r.m.s errors at most levels and hence the bias is generally much smaller, rarely exceeding 2 K. The corresponding profiles for zone 2 (Figure 6.5b) are broadly similar to those of zone 1. In zone 3 (not shown) the standard deviation is close to the r.m.s errors but their values are smaller than in zones 1 and 2. In zone 4 the r.m.s error profile (Figure 6.5c) has a zig-zag behaviour. This is due to biases in the retrievals, since peaks in the error profile occur at the same pressure levels

a)



b)

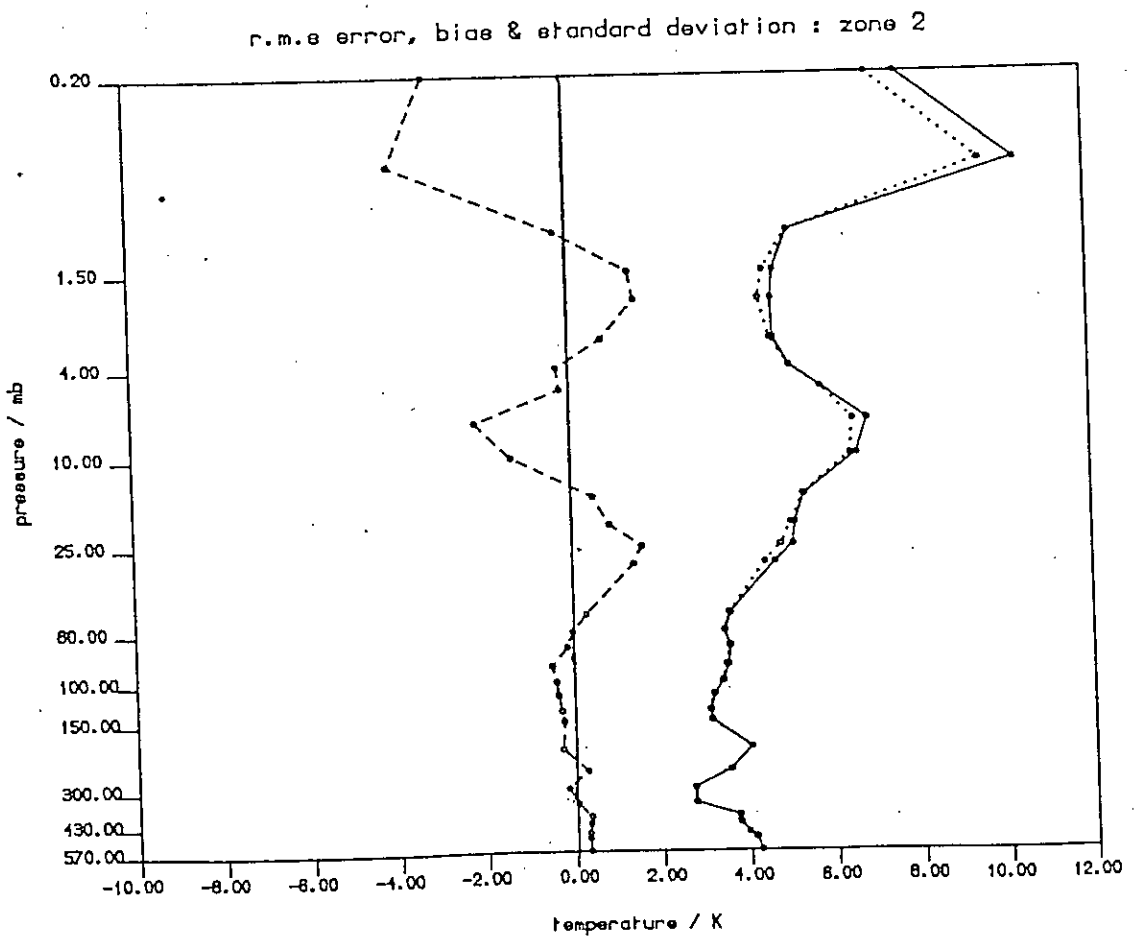
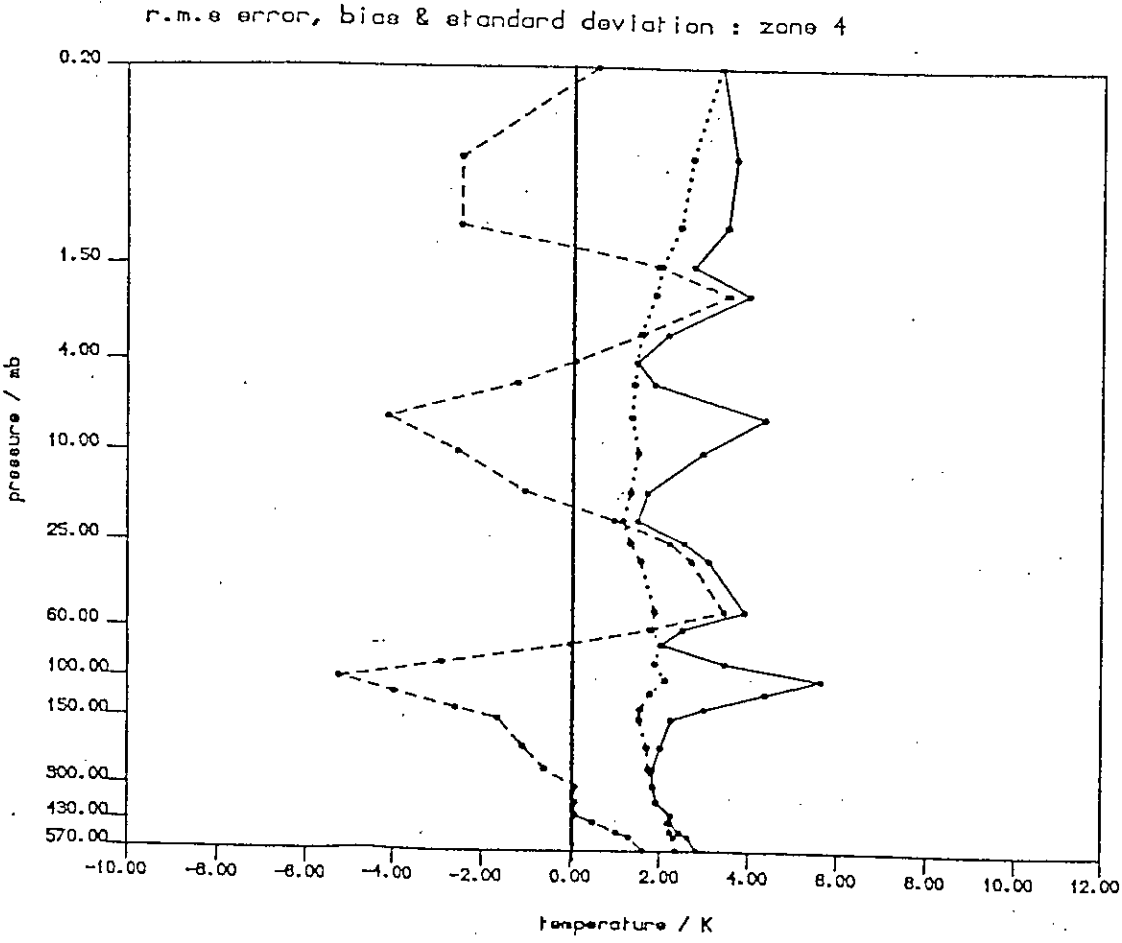


Figure 6.5 R.m.s temperature retrieval error (K) (solid), bias (dashed) and standard deviation (dotted). a) zone 1; b) zone 2;



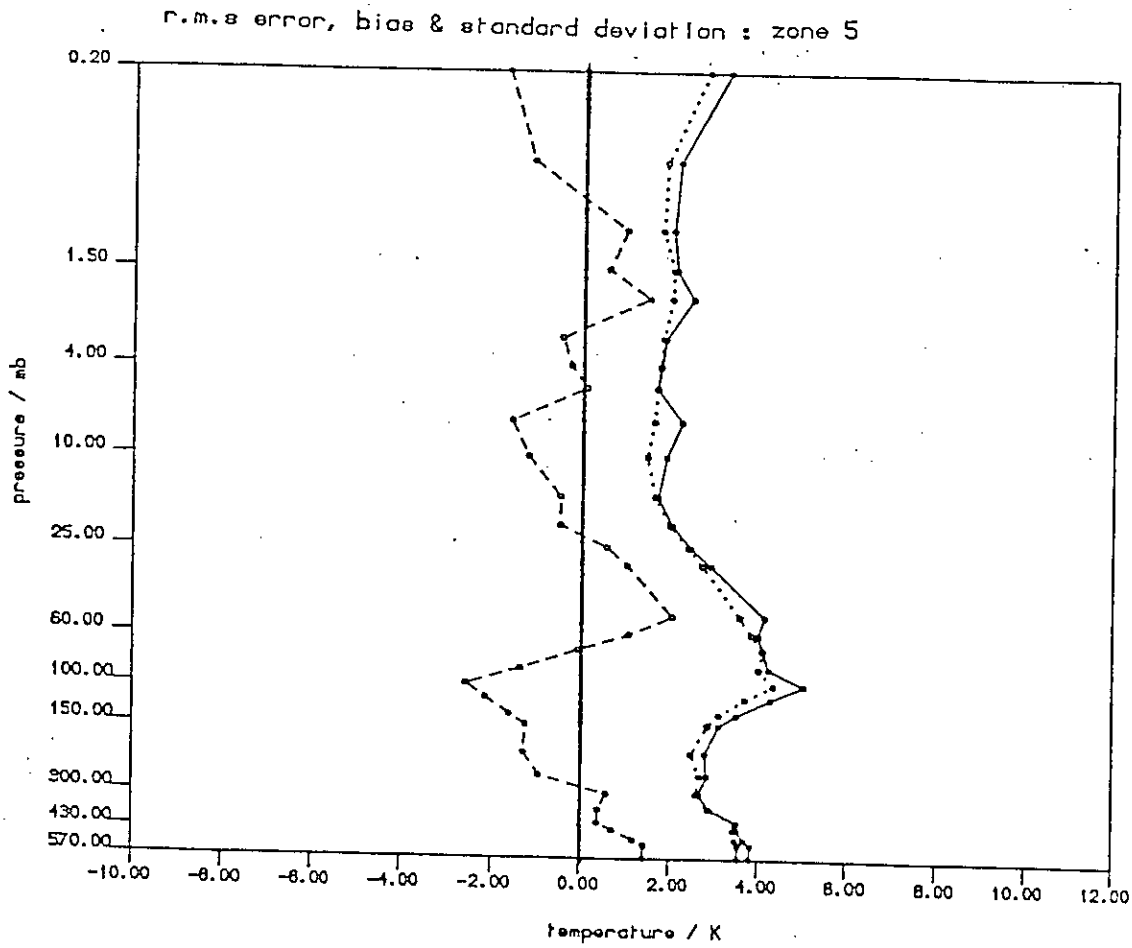


Figure 6.5 (cont.) c) zone 4; d) zone 5;

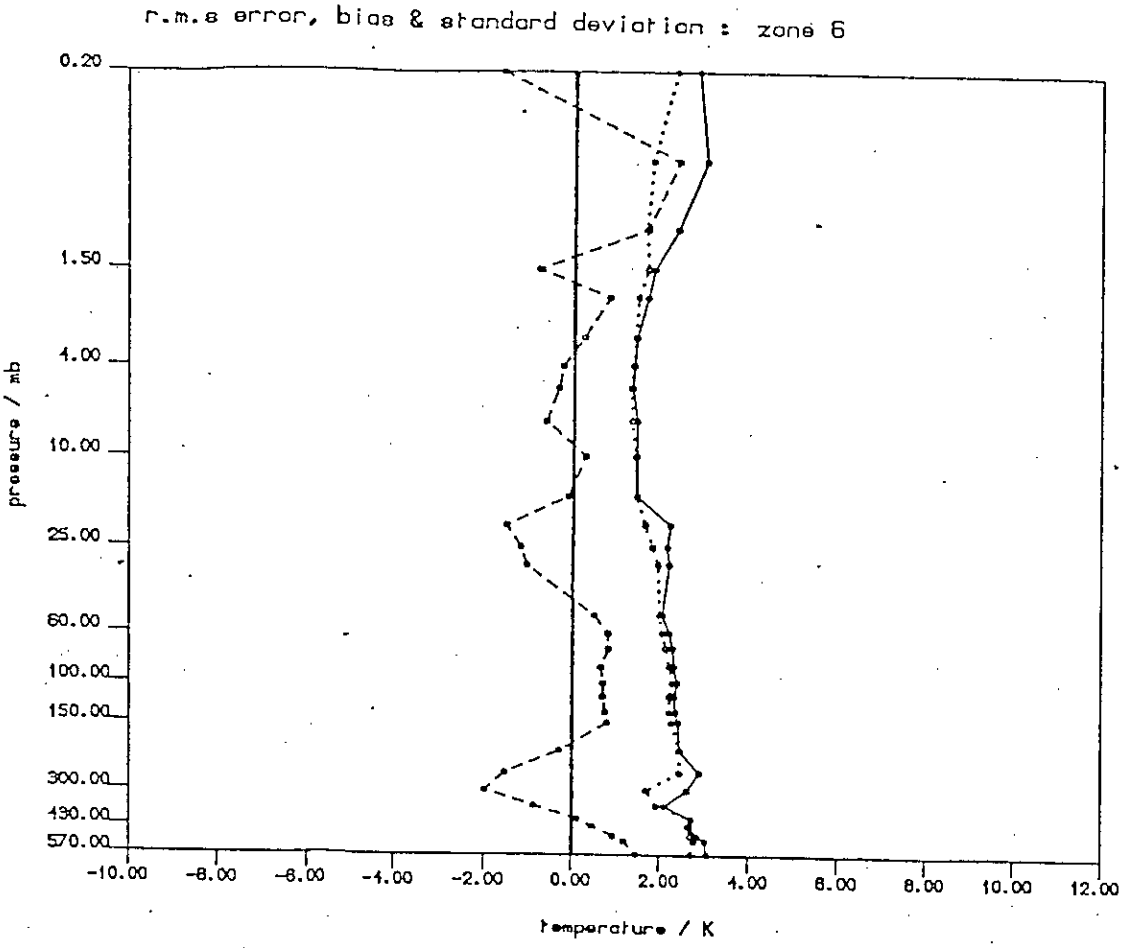


Figure 6.5 (cont.) e) zone 6.

as high values in the bias profile. For example, at 7 mb the r.m.s error is 4.33 K and the bias is −4.13 K and at 100 mb the r.m.s error is 5.61 K and the bias is −5.20 K. This shape of the bias profile is closely related to the difference between the means for model and sondes (Figure 6.6a) and seems therefore to be related to an unrealistic structure in the modelled tropics. The low standard deviation of the zone 4 sonde measurements used to calculate the regression coefficients (Figure 6.6b) constrains retrievals to the mean of the sonde temperatures. R.m.s errors for zone 5 (Figure 6.5d) are lower than for zones 1 to 3 and the profile has a zig-zag pattern, which is also due to bias in the retrievals. However, these zig-zags are much smaller than for the corresponding profiles in zone 4, the largest bias value being −2.55 K (compared to −5.20 K in zone 4). This is probably because the sonde temperatures in zone 5 are more representative of zone 5 model conditions on 18/1/87. Figure 6.6a shows that the difference between model and sonde means for zone 5 is smaller than the corresponding difference for zone 4. In regression zones 6 to 7, where there was little wave activity, r.m.s retrieval errors are low. The zone 6 r.m.s error profile (Figure 6.5e), for example, does not exceed 3 K. The standard deviation of the difference between model and retrieved temperatures is close to the r.m.s error and biases are much smaller than in zone 4. This is because, although Figure 6.6a shows that differences between model and sonde means for zones 6 and 7 are of similar size to the corresponding difference for zone 4, the sonde standard deviations in zones 6 and 7 are larger than the corresponding sonde standard deviation in zone 4 (Figure 6.6b), and hence the retrievals are less constrained towards the sonde mean profile.

It is of interest to note that for tropical and summer latitudes, the standard deviations are in general low. That is of significance for the computation of thermal winds, which, being proportional to horizontal temperature gradients, are unaffected by bias except possibly at transitions between zones.

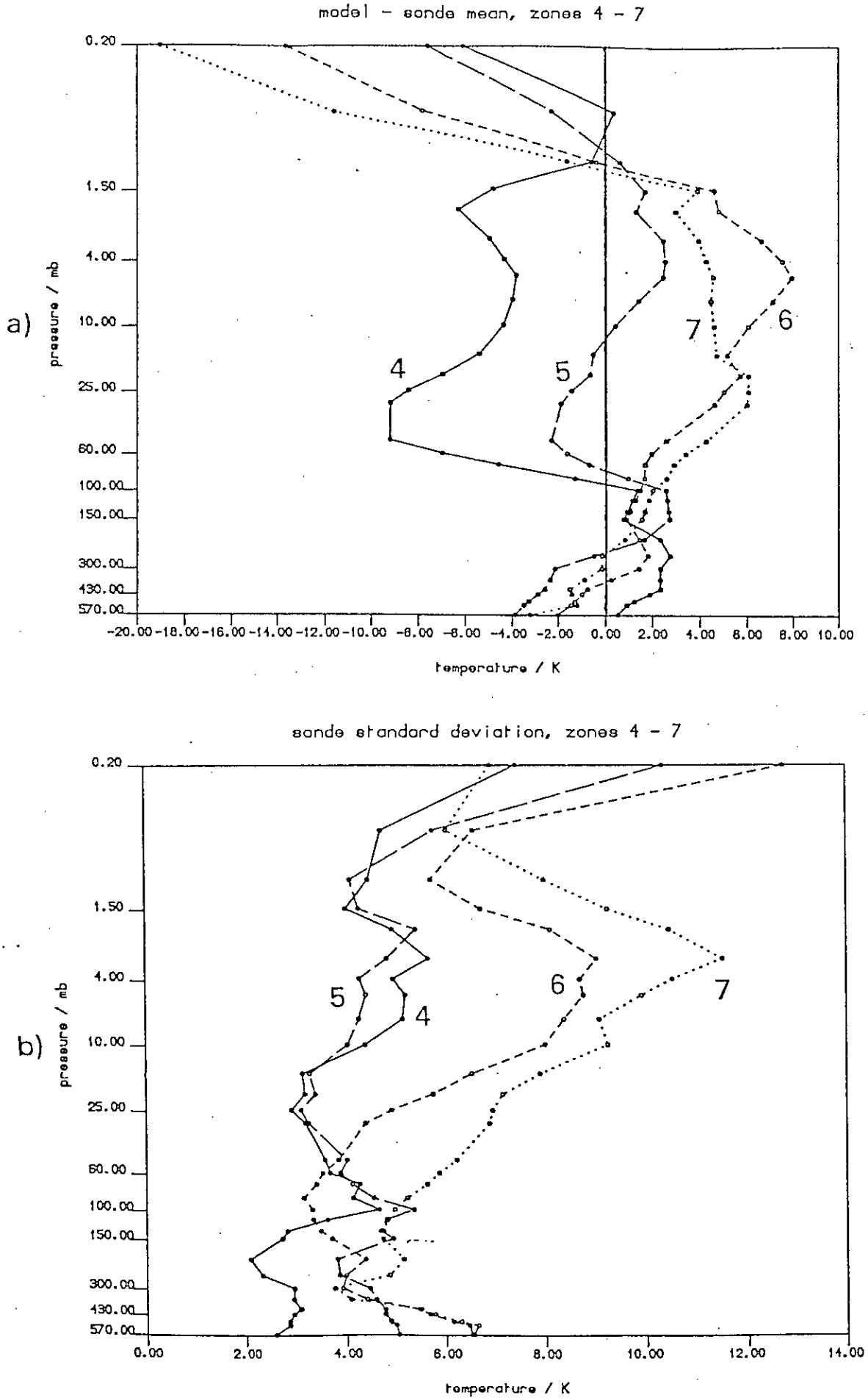


Figure 6.6 a) Difference between model and sonde mean profiles for zones 4 (solid), 5 (long dashes), 6 (short dashes) and 7 (dotted); b) sonde standard deviations for zones 4 (solid), 5 (long dashes), 6 (short dashes) and 7 (dotted).

The results of these tests show that differences between sonde and model datasets influence the size of the retrieval error. The largest differences between the datasets (and hence the largest r.m.s retrieval errors) occur in zones affected by the sudden warming. Clearly operational retrieval errors will also be large in situations where the dataset used to calculate the regression coefficients differs greatly from atmospheric conditions. Ways in which such differences may be reduced are discussed in Section 6.4. Another conclusion from these tests is that high retrieval errors in the zones affected by the sudden warming are partly due to vertical structure which is of too small a scale to be resolved. This is discussed further in Section 6.3.1.

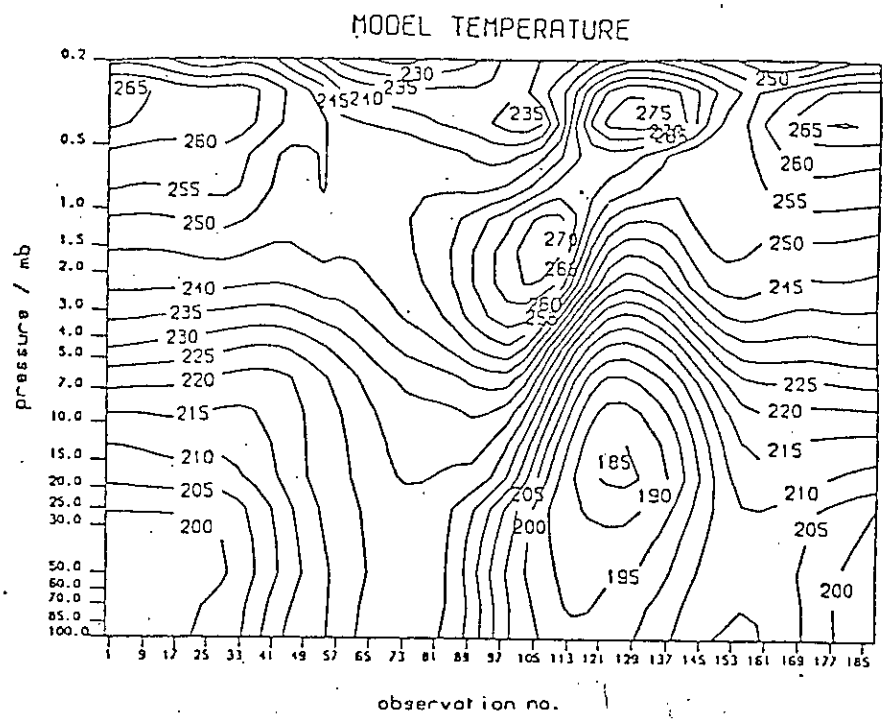
6.3.1. Cross-Sections

The high retrieval errors in zones 1 and 2 are partly attributable to the large vertical temperature gradients present in a sudden warming. This is illustrated by plotting a cross-section of the model temperature field along half a satellite orbit's observation points and comparing it with a corresponding cross-section of retrieved temperatures.

The half-orbit we considered starts at the equator and passes through the region affected by the sudden warming before returning to the equator about 50 minutes later. Observations are made at 189 points. Broadly speaking, observations 1 to 81 and 153 to 189 are made outwith the sudden warming region (in zones 3 and 4) and observations 82 to 152 are made within the sudden warming (in zones 1, 2 and 3).

Figure 6.7a shows a cross-section of the model temperature field between 100 and 0.2 mb and for observations 1 and 189. Within the region of the sudden warming there is a deep low centred at 15 mb and observation number 125. Above this there is a region of high temperature, with peaks at 1.5 mb (observation 100), 0.4 mb (observation 125), and 0.4 mb (observation 177).

a)



b)

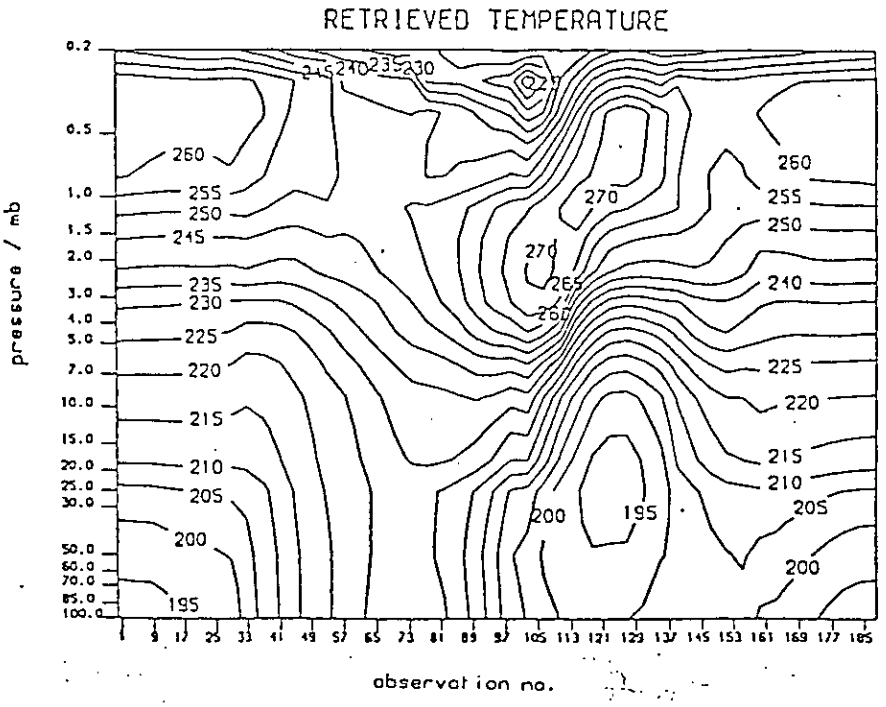


Figure 6.7 'Cross-section' of temperature (°K) at 189 observation points along a satellite orbit. a) model; b) retrieved values;

c)

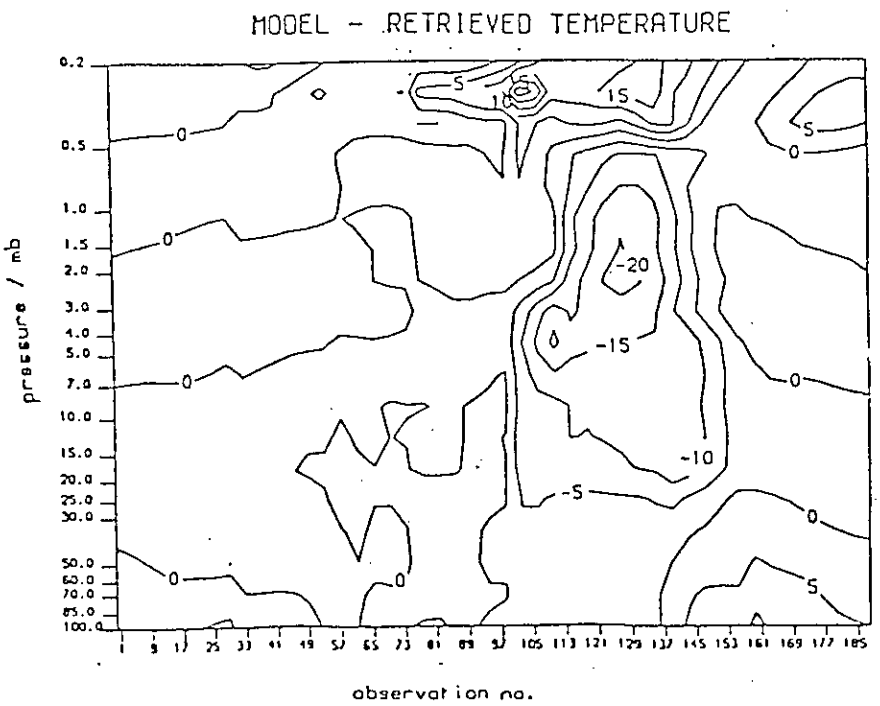


Figure 6.7 (cont.) c) difference between model and retrieved values.

Below the the first two of these highs there is a strong temperature gradient of approximately 4 K / km.

Figure 6.7b shows the corresponding cross-section of the retrieved temperatures. Both outwith and within the sudden warming retrieved values are less than model values at pressures below 0.5 mb. A good example of this is the low at observation 97. Figure 6.7c shows that the difference between model and retrieved temperatures at observation 97 exceeds 10 K. This suggests that the retrieval is biased at pressures less than 0.5 mb, and this is confirmed by Figures 6.5a and 6.5b, which show a large bias in zone 1 and 2 retrievals near 0.5 mb. This bias occurs because these pressure levels are far away from weighting function peaks.

At pressures higher than 0.5 mb the model field is well retrieved outside the sudden warming region (observations 1 to 81 and 153 to 189) - differences between model and retrieved fields are generally less than 5 K, which is small compared with the temperature variation over the northern hemisphere (eg 45 K at 5 mb). However within the sudden warming the field is not well retrieved. The model low centred at 15 mb, observation 125 has been retrieved in the right place, but its retrieved depth is around 10 K greater than its model magnitude. The model maximum at 1.5 mb (observation 100) is well reproduced, but the nearby maximum at peak at 0.4 mb (observation 125) is underestimated by 15 K. The retrieved temperature gradient between the highs and the major low is not as strong as in the model field, especially between 3 and 1 mb and observations 109 to 141 - differences between model and retrieved fields are greater than 20 K in places. These results are consistent with those of Rodgers (1984). He examined a vertical cross-section of a region, affected by a sudden warming, which has a similar structure to the example shown here. He compared a cross-section derived from satellite observations

made by the SAMS (Stratospheric and Mesospheric Sounder) instrument with a cross-section subjectively analysed from sonde data, and found that the vertical gradient in the SAMS-derived field was less steep than the corresponding gradient in the sonde-derived field. Such results emphasise the difficulty in retrieving such vertical structure.

The cross-sections demonstrate that some of the vertical structure in a sudden warming is too small to be retrieved. The first term on the right hand side of equation (3.27) can be regarded as the 'null-space error' of an individual retrieval (Rodgers, 1987), as it corresponds to those portions of profile space that cannot be measured by the observing system. The temperature field within the sudden warming contains vertical structure smaller than the vertical resolution of the TOVS instruments (which is about 10 to 15 km). Hence the 'null-space error' makes a large contribution to the total retrieval error. If, on the other hand, we decrease the vertical resolution of what we are attempting to retrieve by estimating instead the mean temperature of 15 km-thick layers of atmosphere, the 'null-space error' will be smaller (provided the first guess is optimal). The total retrieval error will consequently also be smaller, and will depend more on the size of the instrumental noise (ie on the $\mathbf{W}\ \varepsilon_m$ term in equation (3.27)). There is hence a trade-off between resolution and retrieval error.

6.4. Discussion

The results indicate that temperature retrieval errors were highest within the region of the stratospheric sudden warming (ie zones 1 and 2, and part of zone 3). This is because the vertical temperature structure in the sudden warming was too small to be observed by the satellite instrument, and because the sonde data used to calculate the regression coefficients were inevitably not representative of sudden warming conditions. The former errors, due to

small-scale vertical structure, are inherent in the method of observation rather than the method of temperature retrieval. In contrast, the latter errors are caused by dissimilarities between the sonde dataset and sudden warming conditions. These errors can be described as 'first guess' errors, since the mean of the sonde dataset can be thought of as a first guess estimate of the retrieved temperature. Temperature was adequately retrieved in zones 4 to 7, but the r.m.s retrieval error profiles for these zones also highlight the way the first guess can influence retrievals. For example, the r.m.s error profile for zone 4 (Figure 6.5c) has a zig-zag pattern which is caused by large biases in the retrievals. These biases occur because the retrieval is constrained to the mean of the sonde dataset by the small covariance of the sonde data, and are consequently largest when the difference between model and sonde means is largest. This explanation is consistent with the discussion of error retrievals in Section 3.4.1.4. Ideally one would wish to reduce such errors by using a first guess more appropriate to the atmospheric conditions we are trying to estimate. Before discussing different approaches it is important to consider the sonde dataset we are presently using. It consists of 1200 observations, most of which have been made in the northern hemisphere. These observations are divided into 7 zones according to latitude and season and regression coefficients are calculated for each zone. Hence, the implicit assumption is that there is no difference between the climatologies of the northern and southern hemispheres in the middle atmosphere. However, recent work (eg Andrews, 1989) suggests that there is. A new approach may be to calculate regression coefficients in zones which are determined by latitude, season and hemisphere. However, the sparsity of rocketsonde data in the southern hemisphere may make this difficult.

Another possible way of reducing the first guess error is to use 'stratified climatology'. This is used in tropospheric retrievals (eg Uddstrom and Wark,

1985), and uses the radiance to 'point' to a class of atmospheres to which a profile probably belongs. The regression coefficients are then obtained from the statistics of the class. A related technique is that of the library search (eg Chedin et al, 1985). Observed radiances are compared with a dataset of observations which describe a variety of atmospheric conditions. Normalised least-squares differences between observed and dataset profiles are calculated, and the dataset profile which is closest to the observations is used as the first guess solution to the retrieval. It would be possible to apply these methods to stratospheric retrievals, although the sparsity of stratospheric sonde data may impose a limitation on the usefulness of these techniques. For example, there may not be enough rocketsonde observations of a sudden warming to be able to define an atmosphere of that class. Another possible approach is to utilise numerical model forecasts of the stratosphere. Future research may involve the development of a retrieval scheme which uses for the *a priori* information the output at the previous analysis time from a numerical weather prediction model (this has already been attempted in the troposphere eg Susskind et al (1984)). At present such models do not produce operational forecasts for the stratosphere, but it is expected that a number of such models will be extended up to the stratosphere in the near future.

CHAPTER 7

TESTS OF THE TIME/SPACE INTERPOLATION METHOD

In Chapter 4 the time/space interpolation method (described in Section 3.4.2.2) was tested using analytical fields. These tests reveal that in general reducing the distance radius reduces the r.m.s analysis error, whilst changing the time radius has little effect on the analysis. The tests also reveal that cosine and negative exponential weights produce essentially the same results as linear weights, and hence all tests in this chapter will be performed using linear weights. The analytical fields used in these tests, however, are an over-simplification of observed stratospheric structure. For a more complete examination of the time/space interpolation scheme, it is therefore necessary to perform further tests using more realistic fields, and in particular: 1) to examine whether the abovementioned conclusions from the tests made with analytical fields are still valid; and 2) to obtain an impression of the kind of misrepresentation of 'true' fields which will occur when the scheme is used operationally. Accordingly, the scheme is further tested in a simulation experiment which uses an atmosphere calculated in a numerical model. The model atmosphere is a more realistic representation of the observed stratosphere than are the fields used in the preliminary tests described in Chapter 4. To provide a stringent test of the analysis scheme, we use a model field which is affected by a sudden warming. The simulation experiment, the design of which is described more fully in Chapter 5, actually tests a retrieval/analysis scheme. The results of the retrieval part of the scheme appear in Chapter 6, whilst here we concentrate on the time/space interpolation part of the scheme.

The tests are described in more detail in Section 7.1. A description of the model field used here, and a discussion of what results we might expect the time/space interpolation scheme to produce, appear in Section 7.1.1. An evaluation of estimates made with a variety of search radii is effected in Sections 7.1.2 and 7.1.3 using r.m.s errors and a comparison of maps of the

analysed fields, whilst a discussion of the results appears in Section 7.2.

7.1. Interpolation of Retrieved Temperatures

Retrieved temperatures from 18/1/87 are interpolated using four combinations of search radii. Preliminary tests of the interpolation scheme (Chapter 4) indicated that changing the time radius causes only slight changes to the quality of the estimated field; this is further examined here by comparing two estimates made with the same distance radius (2000 km), but with different time radii, viz 6 hrs and 12 hrs. The tests in Chapter 4 also revealed that reducing the distance radius generally produced a better estimate of the field, but that the estimate became slightly poorer at distance radii less than 1500 km if the field consisted totally of travelling waves (ie Fields 1 and 2): this is further examined by comparing the analysis made with a time radius of 12 hrs and a distance radius of 2000 km to another made with the same time radius but with a distance radius of 1000 km. The fourth analysis uses the search radii employed in the United Kingdom Meteorological Office's operational stratospheric analysis scheme, namely 12 hrs and 500 km, and enables us to determine whether the Met Office's choice of search radii is more suitable than ours. When analyses are performed with distance radii of 2000 and 1000 km every gridpoint has at least one observation within its corresponding search radii. However, when a distance radius of 500 km is used there are a number of gridpoints which have no observations within their search radii. These missing values are filled by linearly interpolating the nearest 'good' observations from gridpoints east and west of those with no observations, and then smoothing the whole field with a 3-point smoother with weights of the form (0.25,0.5,0.25). The temperatures are interpolated to the model grid and to the analysis time of 1200 GMT. To keep the task of interpretation within reasonable bounds, analyses are made only at selected pressure levels. These include some close to the SSU weighting function peaks (1.5, 5.0 and 15.0 mb), two in between

these peaks (3.0 and 10.0 mb), and two outwith the range of these weighting function peaks (0.2 and 25.0 mb). These levels are chosen to examine whether analyses at pressure levels away from from weighting function peaks are poorer than analyses at levels close to weighting function peaks.

7.1.1. Examination of the Model Field Used in the Tests

Before performing tests of the time/space interpolation scheme, it is necessary to examine in more detail the model field used, and to consider how this field differs from the fields used in the preliminary test in Chapter 4. Since the Met. Office model has been used in many dynamical studies of the middle atmosphere (eg O'Neill and Pope, 1988; Shine, 1987; Fairlie and O'Neill, 1987), we expect it to represent the 'real' stratosphere better than the idealised fields. In addition, by examining the model field we can identify what model field features provide a stringent test of the time/space interpolation scheme, and discuss what sort of results we expect the scheme to produce.

We examine plots of the model temperature at 5 mb at 6 hour intervals on 18/1/87. Figure 7.1a shows the model field at 0000 GMT. There is mainly zonal flow in the southern hemisphere. The interpolation scheme should estimate this zonal pattern quite easily. Hence our discussion will instead concentrate on the northern hemisphere, which is affected by a sudden warming. There is a low centred near 55°N and 20°E which is separated from a high centred near 75°N and 90°W by a strong temperature gradient. This structure is of a smaller spatial scale than any of the fields used in the preliminary tests in Chapter 4, and suggests that reducing the size of the distance radius from 2000 to 1000 km should noticeably improve the quality of the estimated field. The northern hemisphere field at 0600 GMT (Figure 7.1b) has a larger temperature gradient between the major high and low. This larger gradient occurs because, although magnitudes of the high and low have changed little, the high has moved nearer

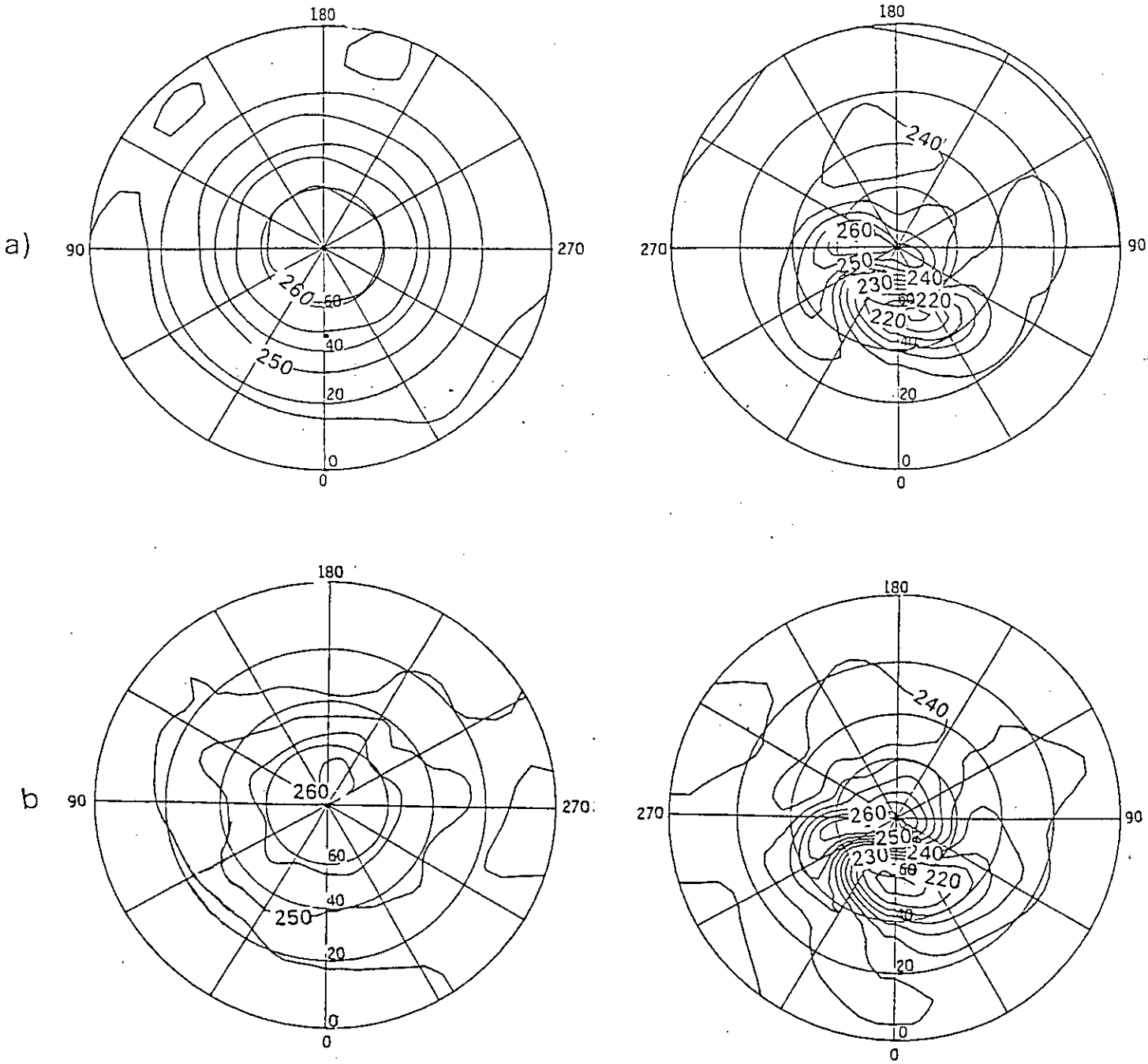


Figure 7.1 Temperature on 18/1/87 at 5 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. a) model field at 0000 GMT; b) as a), except field at 0600 GMT is plotted;

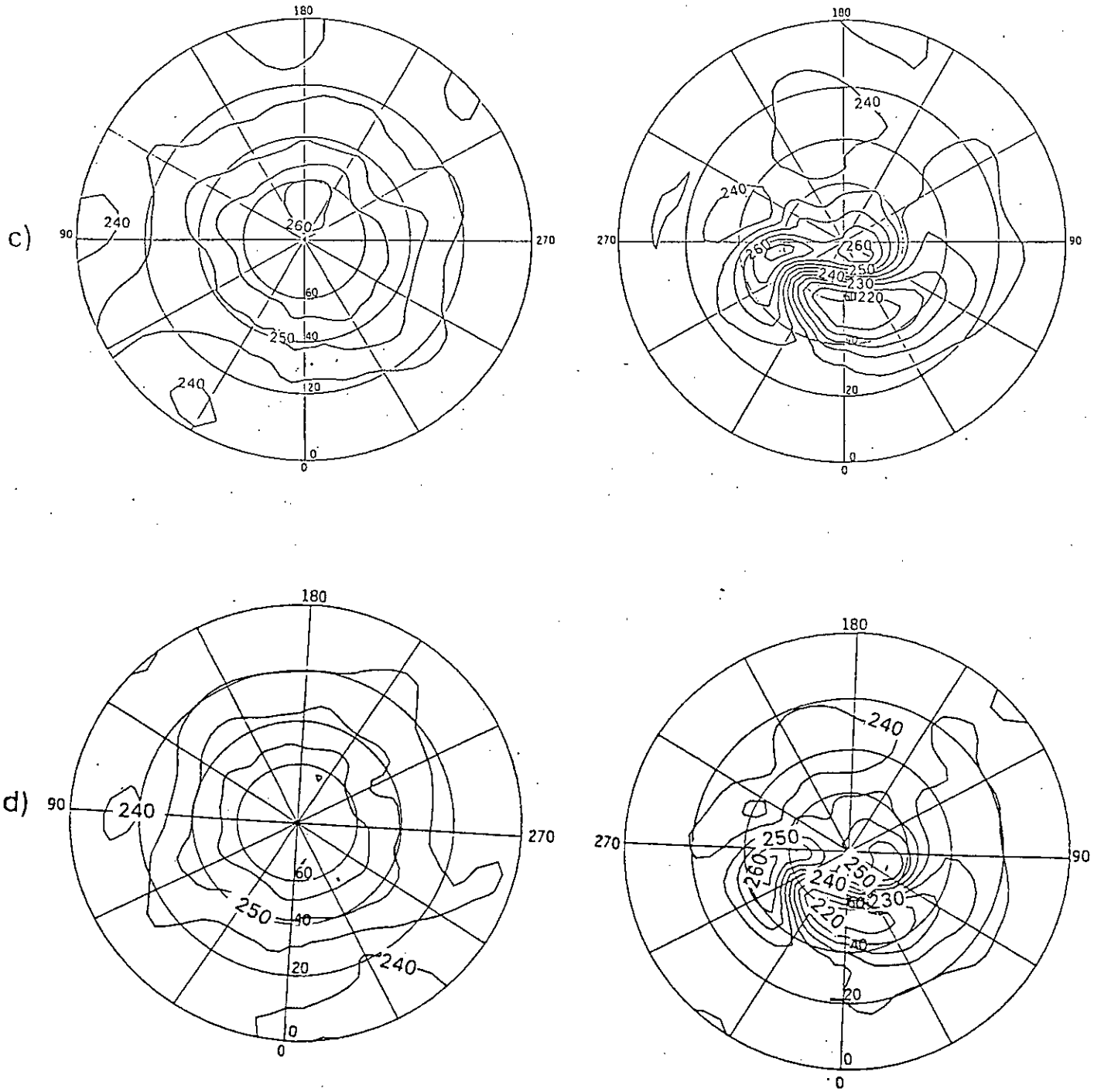


Figure 7.1 (cont.) c) as a), except field at 1200 GMT is plotted; d) as a), except field at 1800 GMT is plotted;

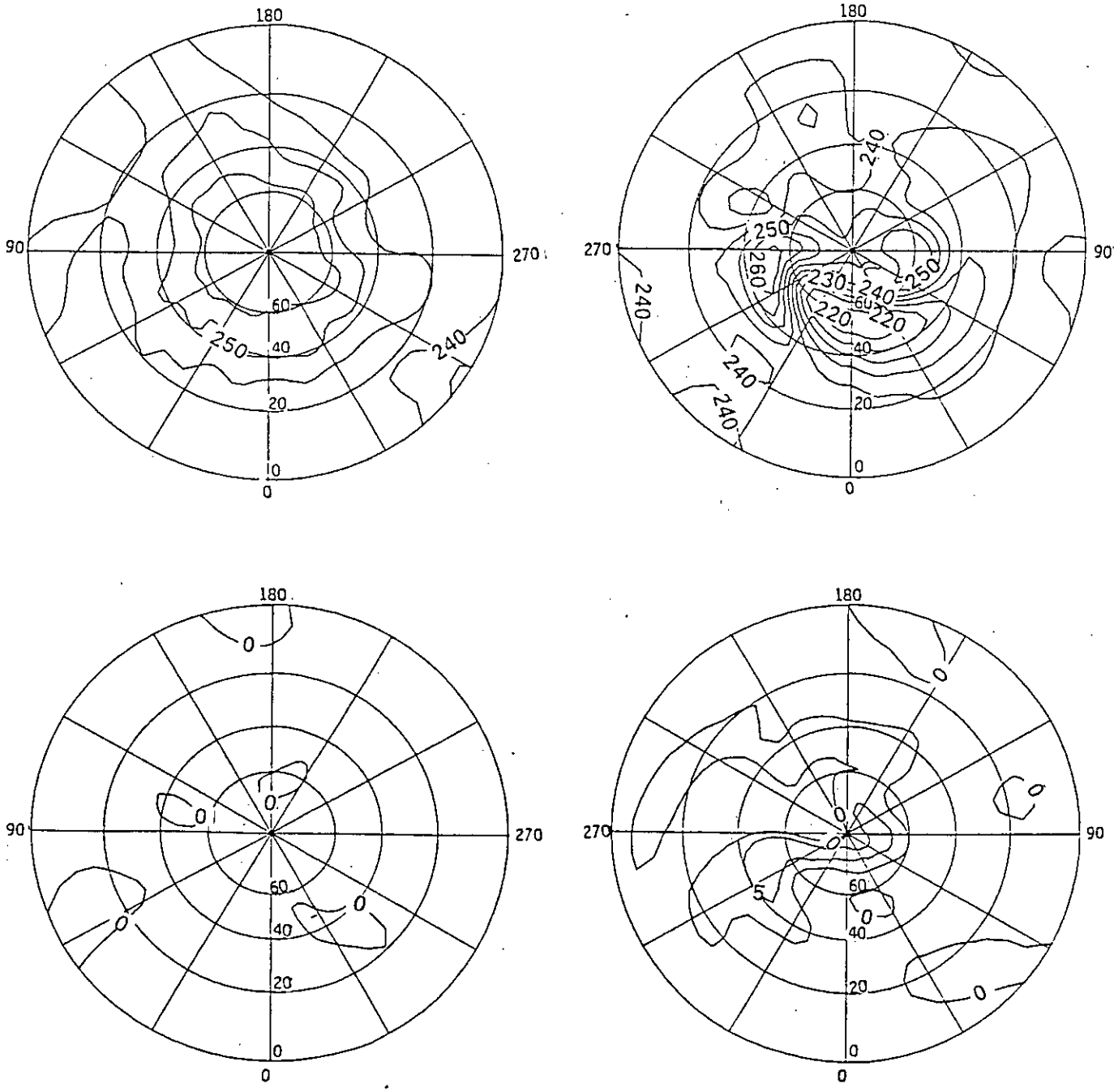


Figure 7.1 (cont.) e) as a), except field at 2400 GMT is plotted; f) difference between fields in c) and a);

g)

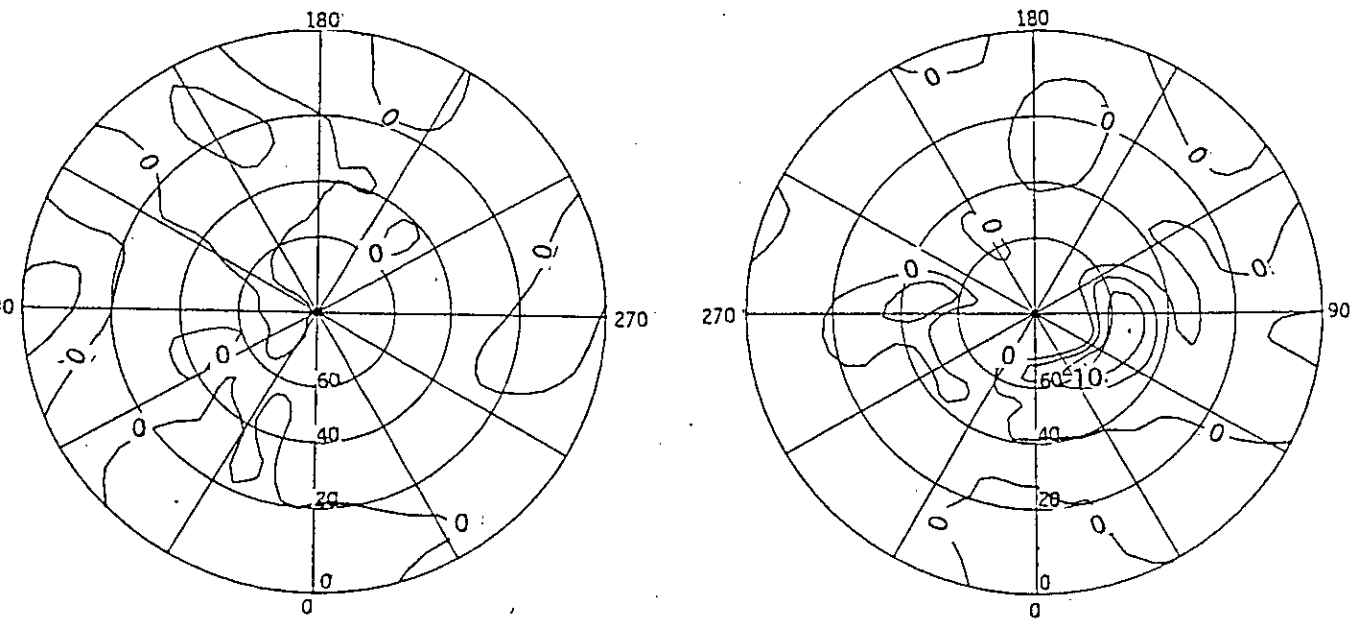


Figure 7.1 (cont.) g) difference between fields in e) and c).

the low. At 1200 GMT (Figure 7.1c) the location of the high and low has changed little, but the high has become bifurcated. This 'splitting' of the major high is even more noticeable at 1800 GMT (Figure 7.1d) and 2400 GMT (Figure 7.1e). It is clear from all these plots that, unlike the fields used in Chapter 4, the model highs and lows change their location little with time. Instead, changes are due to relatively small scale distortion and twisting of field features. These points suggest that increasing the time radius should not degrade the analysis, but that the estimate of the magnitude of the major high and low in the northern hemisphere, and of the large temperature gradient between them, should improve when the distance radius is reduced from 2000 to 1000 km. However, since the scale of these features appears to be about 1000 km or more, it is unclear whether the estimate will improve further when the distance radius is reduced from 1000 to 500 km.

It has been noted that the northern hemisphere model high gradually 'splits'. This bifurcation does not exist at 0000 GMT or 0600 GMT, whilst the bifurcation at 1800 GMT and 2400 GMT is stronger than that at 1200 GMT. Because of this, and because the feature is located at high latitudes, where observations in a 12 or 24 hr period are nearly symmetrically spaced in time about the analysis time (ie 1200 GMT) (see for example Figure 5.1), one should expect the bifurcation in the high at 1200 GMT to be satisfactorily estimated. However, other changes to the field, which are more local in time, may prove harder to estimate. Figure 7.1f shows the difference between the model fields at 1200 GMT and 0000 GMT. There is a tongue between 40°N, 60°W and 80°N, 150°E where the difference between the fields exceeds 5 K. At other locations the difference is less than 5 K. Contrast this to Figure 7.1g, which shows the difference between model fields at 2400 GMT and 1200 GMT. The differences here are greater, but more localised than the differences shown in Figure 7.1f. For example, between 60°N, 30°E and 50°N, 90°E differences exceed 10 K. Since there is little similarity

between Figures 7.1f and 7.1g, this means that the field changes highlighted do not persist for the whole 24 hrs of the analysis and hence may be poorly estimated when a time radius of 12 hrs is used.

7.1.2. R.m.s Errors

The r.m.s error of the combined retrieval and analysis process is calculated, under the assumption that the model temperature at 1200 GMT is the "true" temperature. The biases of the estimated temperatures are also calculated. As discussed in Chapter 4, r.m.s errors do not always explain fully how well (or how badly) the field has been estimated, and so to aid interpretation of the results, we frequently also make qualitative comparisons of maps of the analysed fields. Both r.m.s errors and biases were calculated for all latitudes, and also for the 7 latitude/season retrieval zones. R.m.s errors for the analyses made using time radii of 6 and 12 hrs and a distance radius of 2000 km are shown in Table 7.1a. Errors for analyses made with a time radius of 12 hrs and distance radii of 2000 km and 1000 km are shown in Table 7.1b; errors for analyses made with a time radius of 12 hrs and distance radii of 1000 and 500 km are shown in Table 7.1c.

PRESSURE / mb

	0.2	1.5	3.0	5.0	10.0	15.0	25.0
ZONE	---	---	---	---	---	---	---
1	5.88/6.50	4.28/4.14	3.36/2.98	5.96/5.39	6.33/5.18	6.78/5.30	7.82/5.72
2	8.54/8.61	4.73/4.63	4.88/4.32	5.49/4.78	4.94/4.13	5.13/4.76	4.37/4.37
3	5.27/5.71	4.31/4.06	2.72/2.60	3.16/3.10	2.94/2.74	2.09/2.01	1.98/1.99
4	4.99/4.96	4.01/3.78	3.21/2.95	1.96/1.87	3.01/3.00	1.74/1.74	2.63/2.63
5	3.47/3.58	2.08/1.93	1.44/1.44	1.35/1.22	1.31/1.08	0.97/0.81	1.27/1.30
6	3.74/4.05	3.26/3.58	1.98/2.28	1.49/1.66	0.80/0.81	0.65/0.65	1.22/1.16
7	3.50/2.83	1.60/2.14	0.89/0.89	0.67/0.66	0.59/0.57	0.50/0.48	2.12/2.08
Global	5.25/5.37	3.73/3.64	2.98/2.76	3.19/2.92	3.38/2.99	3.11/2.69	3.54/3.05

Table 7.1a R.m.s errors for the combined retrieval and analysis in degrees K. Errors for the analysis made using a time radius of 6 hrs and a distance radius of 2000 km are shown to the left of the slash; errors for the analysis made with a time radius of 12 hrs and a distance radius of 2000 km are shown to the right of the slash.

Inspection of Table 7.1a reveals that the r.m.s errors at 0.2 mb are generally higher than at other levels. There are two possible explanations for this. Firstly, 0.2 mb is far away from a weighting function peak, so one would expect the retrieval at this level to be poor. Rodgers (1984) and Grose and Rodgers (1986) have noted the poor quality of satellite-derived fields compared to sonde fields above the level of the topmost weighting function peak. Secondly, the 0.2 mb model field (Figure 7.2a), unlike the fields at other levels, has considerable small-scale structure (typically between 200 and 300 km) that the interpolation scheme might have difficulty resolving. The table also shows that the global r.m.s error of the analysis made at 0.2 mb with a time radius of 12 hrs is slightly larger than the error of the corresponding analysis made with a 6 hr time radius. Comparison of plots of analyses made with time radii of 6 hrs (Figure 7.2b) and 12 hrs (Figure 7.2c), however, shows there is little qualitative

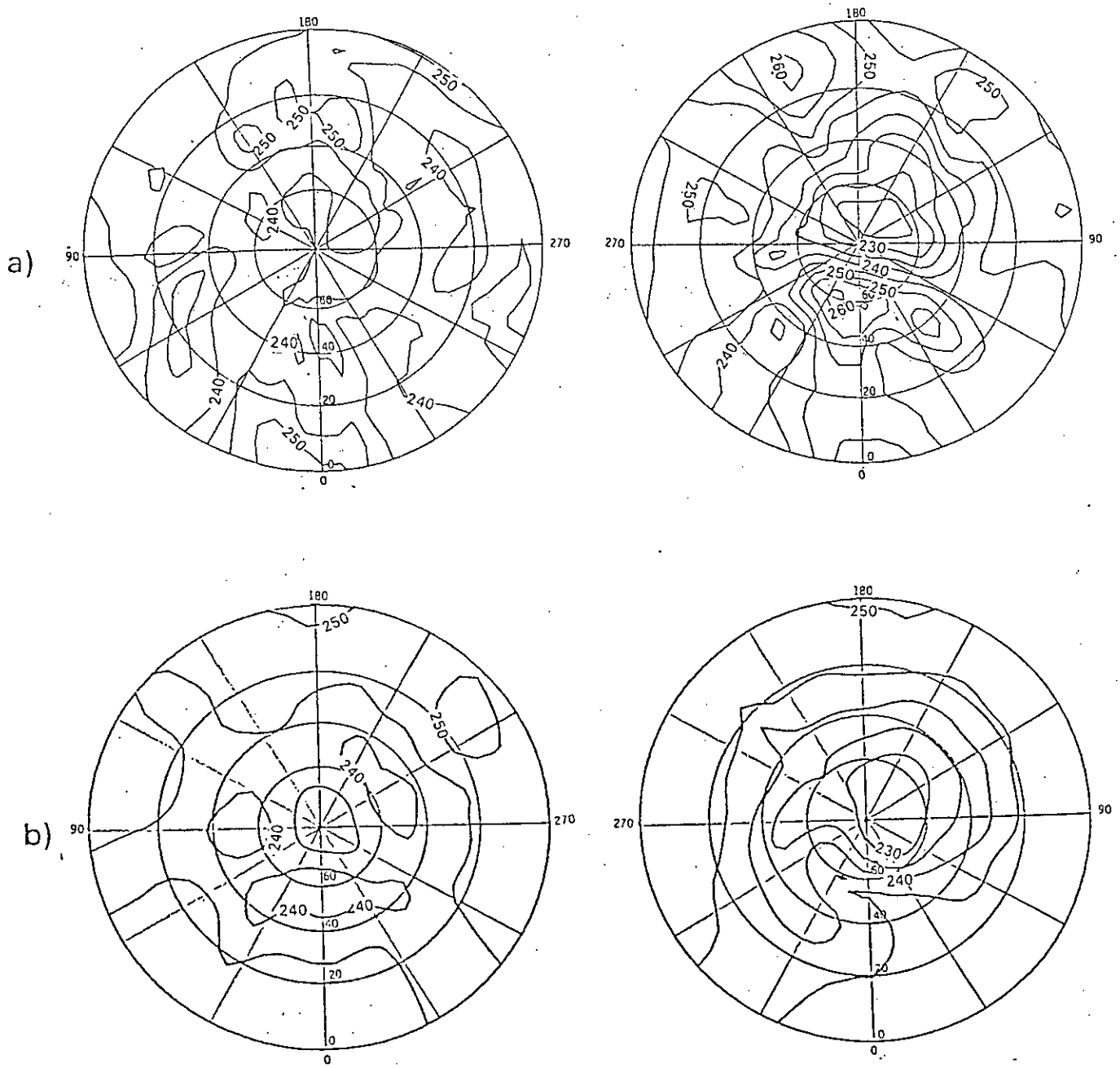


Figure 7.2 Temperature field for 1200 GMT on 18/1/87 at 0.2 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. a) model; b) analysed values obtained using a time radius of 6 hrs and a distance radius of 2000 km;

c)

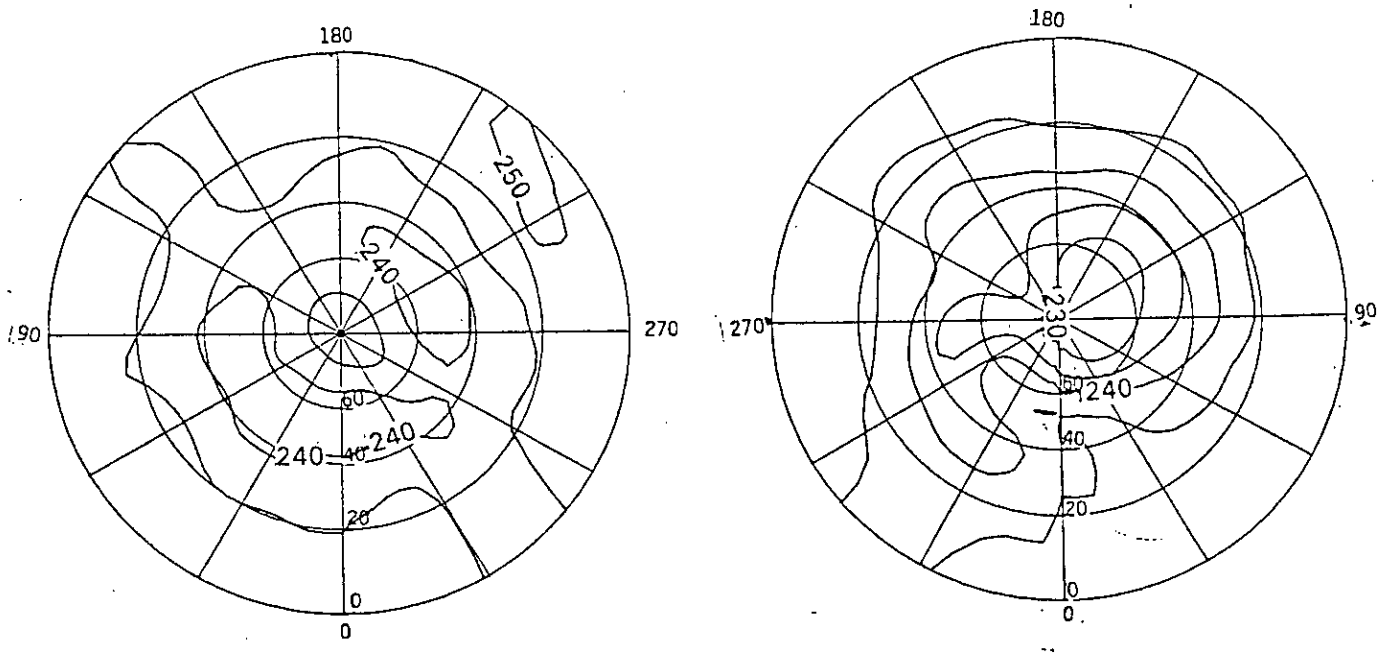


Figure 7.2 (cont.) c) as b), except the time radius used is 12 hrs.

difference between the analyses. The model field at 1.5 mb (Figure 7.3a) also has some small-scale structure, though not as much as at 0.2 mb. The analyses made with a time radius of 6hrs (Figure 7.3b), and 12 hrs (Figure 7.3c), both fail to estimate much of the model small-scale structure. The SSU channel 27 weighting function peak is close to 1.5 mb, and hence the limitations in the quality of the analysis at 1.5 mb are not due to poor retrievals but instead must be because some of the model field structure is too small to be resolved by the analysis scheme. It is possible that such small-scale structure is due to inaccuracies in the model, but nevertheless it is clear that if such scales do exist in the atmosphere an operational time/space interpolation scheme will not be able to resolve them.

The model fields at 3.0, 5.0, 10.0, 15.0 and 25.0 mb are quite similar to each other. Therefore the discussion in Section 7.1.1 of the model field at 5 mb is applicable to all fields between 3 and 25 mb. Typically, the flow in the southern hemisphere is mainly zonal, whilst there is strong wavenumber 1 activity in the northern hemisphere middle and high latitudes. There is little structure of a smaller scale than this. Table 7.1a shows that changing the time radius from 6 to 12 hrs in general causes a slight drop in r.m.s error. However, comparison of analysed fields (see Section 7.1.3) reveals that qualitatively the analysis changes little when the time radius is changed. The explanation for this is as follows: increasing the time radius decreases the r.m.s error because, as more observations are available, the accuracy of the estimate is increased. In addition, as discussed in Section 7.1.1, increasing the time radius should not greatly increase the analysis error, since the field's major highs and lows change little with time. However, since the distance radius is unchanged, the ability of the analysis scheme to resolve the smaller spatial features of the field is unchanged, and hence qualitatively the analysis appears to be unchanged.

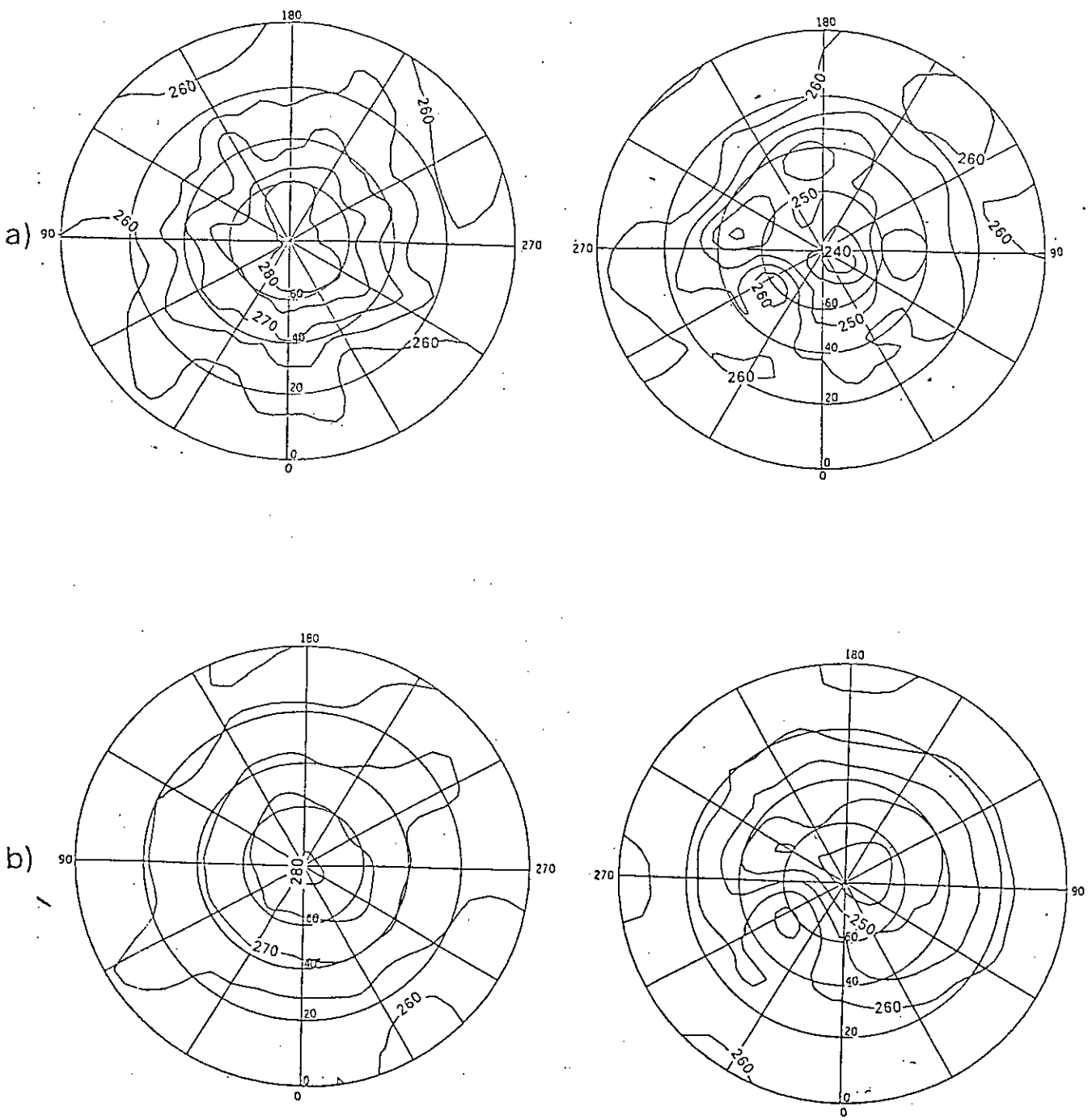


Figure 7.3 Temperature field for 1200 GMT on 18/1/87 at 1.5 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. a) model; b) analysed values obtained using a time radius of 6hrs and a distance radius of 2000 km;

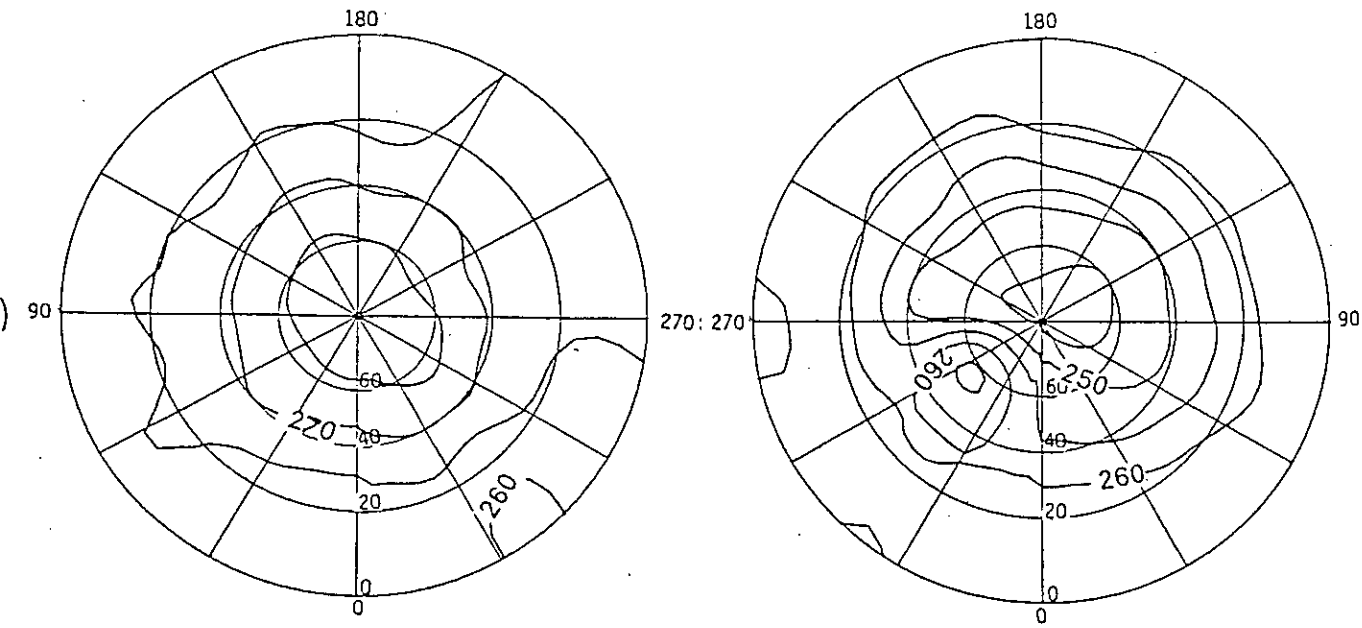


Figure 7.3 (cont.) c) as b), except the time radius used is 12 hrs.

PRESSURE / mb

	0.2	1.5	3.0	5.0	10.0	15.0	25.0
ZONE	---	---	---	---	---	---	---
1	6.50/7.59	4.14/4.14	2.98/4.07	5.39/4.37	5.18/2.73	5.30/2.46	5.72/3.36
2	8.61/7.63	4.63/4.71	4.32/2.87	4.78/2.38	4.13/2.35	4.76/2.76	4.37/3.43
3	5.71/6.13	4.06/3.88	2.60/2.24	3.10/2.46	2.74/2.21	2.01/1.24	1.99/1.89
4	4.96/5.16	3.78/3.95	2.95/3.05	1.87/1.99	3.00/3.03	1.74/1.78	2.63/2.52
5	3.58/3.98	1.93/1.97	1.44/1.58	1.22/1.23	1.08/1.26	0.81/0.83	1.30/1.17
6	4.05/4.75	3.58/3.82	2.28/2.10	1.66/1.58	0.81/0.91	0.65/0.67	1.16/1.33
7	2.83/3.56	2.14/1.98	0.89/0.97	0.66/0.59	0.57/0.64	0.48/0.63	2.08/1.78
Global	5.37/5.62	3.64/3.71	2.76/2.67	2.92/2.28	2.99/2.32	2.69/1.71	3.05/2.41

Table 7.1b. R.m.s errors for the combined retrieval and analysis in degrees K. Errors for the analysis made using a time radius of 12 hrs and a distance radius of 2000 km are shown to the left of the slash, errors for the analysis made using a time radius of 12 hrs and a distance radius of 1000 km are shown to the right of the slash.

PRESSURE / mb

	0.2	1.5	3.0	5.0	10.0	15.0	25.0
ZONE	---	---	---	---	---	---	---
1	7.59/8.17	4.14/4.26	4.07/4.74	4.37/4.66	2.73/2.76	2.46/1.82	3.36/2.65
2	7.63/7.32	4.71/4.80	2.87/2.73	2.38/2.02	2.35/2.48	2.76/2.39	3.43/3.30
3	6.13/6.45	3.88/3.83	2.24/2.29	2.46/2.37	2.21/2.17	1.24/1.14	1.89/1.96
4	5.16/5.20	3.95/3.95	3.05/3.02	1.99/2.08	3.03/3.14	1.78/1.90	2.52/2.46
5	3.98/4.30	1.97/2.00	1.58/1.67	1.23/1.24	1.26/1.28	0.83/0.82	1.17/1.17
6	4.75/5.06	3.82/3.94	2.10/2.08	1.58/1.58	0.91/0.98	0.67/0.72	1.33/1.43
7	3.56/3.81	1.98/1.98	0.97/1.21	0.59/0.73	0.64/0.73	0.63/0.73	1.78/1.68
Global	5.62/5.79	3.71/3.74	2.67/2.81	2.28/2.33	2.32/2.39	1.71/1.59	2.41/2.26

Table 7.1c. R.m.s errors for the combined retrieval and analysis in degrees K. Errors for the analysis made using a time radius of 12 hrs and a distance radius of 1000 km are shown to the left of the slash, errors for the analysis made using a time radius of 12 hrs and a distance radius of 500 km are shown to the right of the slash.

Inspection of Table 7.1b shows that the r.m.s errors at 0.2 mb are higher than for any other pressure level. Fields analysed using a 2000 km distance radius (Figure 7.2c), and a 1000 km distance radius (Figure 7.4), both fail to estimate much of the model field structure. One might have expected the analysis made with the smaller distance radius to have estimated the small-scale model structure better, but r.m.s errors at 0.2 mb are in general higher using a 1000 km distance radius than when using a 2000 km radius. This suggests that the high error is due to poor retrievals.

The model field at 1.5 mb (Figure 7.3a) also has some small-scale structure, though not as much as at 0.2 mb. The analysis made with a time radius of 12 hrs and a distance radius of 2000 km (Figure 7.3c) fails to estimate this small-scale structure, although the large-scale structure is reasonably estimated. Figure 7.5 shows that when the distance radius is reduced to 1000

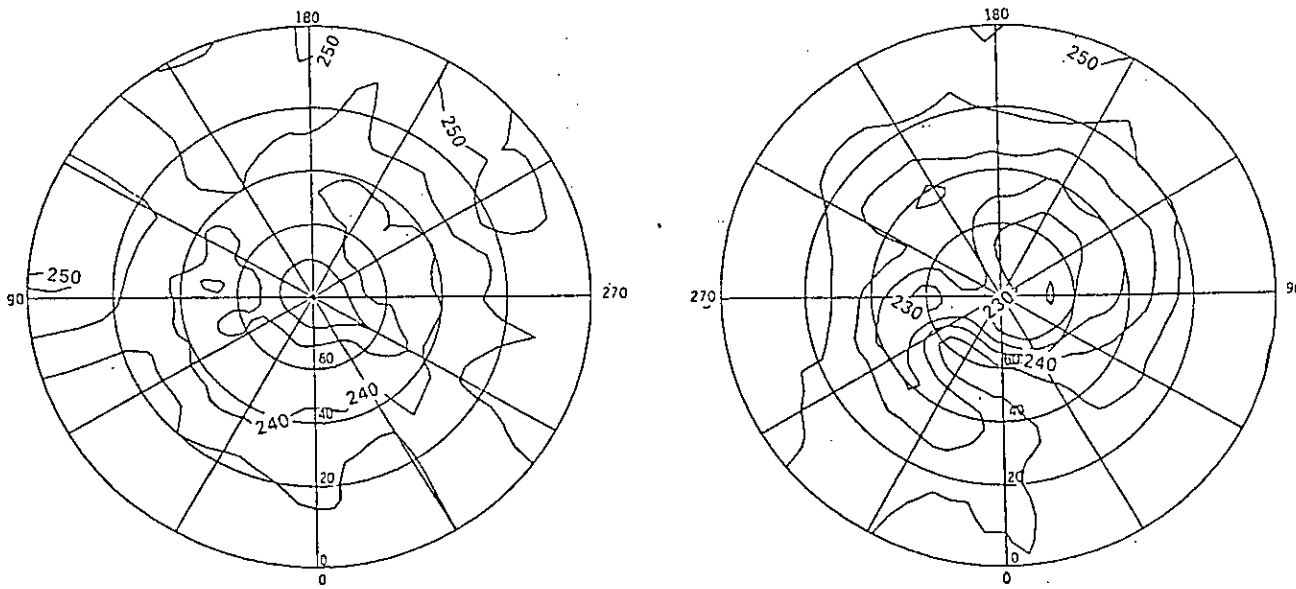


Figure 7.4 Temperature field for 1200 GMT on 18/1/87 at 0.2 mb analysed using a time radius of 12 hrs and a distance radius of 1000 km. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K.

km the interpolation scheme still fails to estimate most of the model small-scale structure. Moreover, Table 7.1b shows that the r.m.s error changes little when the distance radius is changed from 2000 to 1000 km. Hence we again conclude that the analysis scheme fails to resolve the small-scale structure present at pressures lower than 1.5 mb, and users of analysed fields should be aware of the possible errors present in analyses of fields which contain such small-scale structure.

The model fields at 3.0, 5.0, 10.0, 15.0 and 25.0 mb are quite similar to each other, and the way their r.m.s errors change when the distance radius changes from 2000 to 1000 km is also similar. With a distance radius of 2000 km, global r.m.s errors range between 2.69 and 3.05 K. Reducing the distance radius to 1000 km reduces the global r.m.s error to between 1.71 and 2.67 K. A qualitative comparison of maps of the two analyses (see Section 7.1.3) shows that both satisfactorily estimate the sudden warming, but, in agreement with the discussion in Section 7.1.1, the analysis made using the smaller distance radius (1000 km) estimates the magnitude of the highs and lows in the sudden warming region better. The quality of the analyses in regions not affected by the sudden warming is similar for both analyses. To underline these points, an inspection of the errors in each zone reveals that when the distance radius is reduced from 2000 km to 1000 km, the largest fall in r.m.s error generally occurs in the region of the sudden warming, but away from the sudden warming r.m.s errors in general change little when the distance radius is changed. This means that we require a small distance radius when analysing the high spatial variability within a sudden warming, but that a larger distance radius is adequate to analyse the field elsewhere. In general terms, since the scheme satisfactorily estimates the sudden warming here, one would expect it to produce a reasonable estimate of most dramatic dynamical events observed in the stratosphere. However, it would have more difficulty estimating

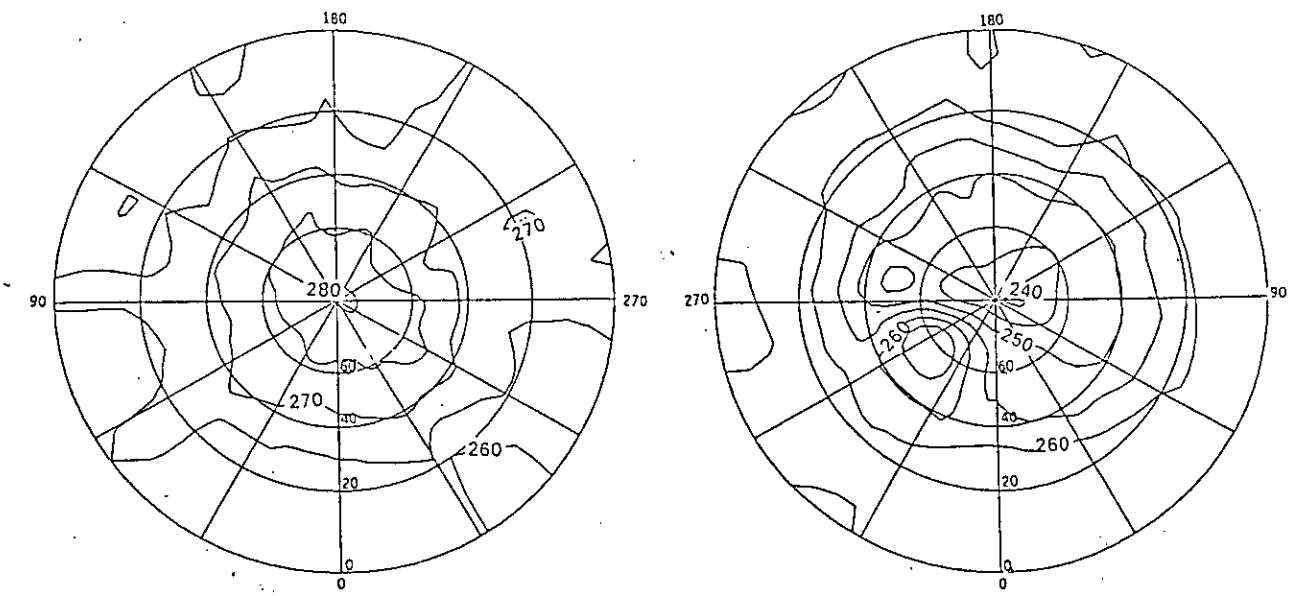


Figure 7.5 Temperature field for 1200 GMT on 18/1/87 at 1.5 mb analysed using a time radius of 12 hrs and a distance radius of 1000 km. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K.

small-scale structure such as that in the model fields at 0.2 and 1.5 mb.

Table 7.1c reveals that changing the distance radius from 1000 to 500 km and interpolating between gridpoints for those with no nearby observations produces little change in the r.m.s error, and maps of fields analysed with a 500 km distance radius (see Section 7.1.3) are little different from corresponding fields analysed using a 1000 km distance radius. This suggests that the use of a 500 km distance radius is unnecessary, as comparably good analyses can be obtained using a distance radius of 1000 km without problems caused by missing observations.

7.1.3. Comparison of Maps at 5 mb

To study the performance of the interpolation scheme more closely, we compare model and analysed fields at 5 mb. The results at 5 mb provide a good example of the improvement in the analysis caused by reducing the distance radius from 2000 to 1000 km. They also demonstrate the lack of qualitative change in the analysis when the time radius is changed from 6 to 12 hrs. Figure 7.1c shows the 5mb model field at 1200 GMT. There is strong wavenumber 1 activity in the middle and upper latitudes of the Northern hemisphere. There is a major high situated between 40°N and 90°N and 90°E and 90°W with two peaks at 70°N, 70°W and at 80°N, 30°E. There is also a large low situated between 50°N and 70°N, and 30°E and 30°W, having a minimum value of less than 220 K. The high and the low are separated by a region of strong temperature gradients. Outwith the middle and upper latitudes in the Northern hemisphere the flow is generally zonal.

A comparison of two analyses of the 5 mb field reveals that changing the time radius produces little qualitative change to the analysis. Figure 7.6a shows the 5 mb field analysed with a time radius of 6 hrs and a distance radius of 2000 km. The shape and location of model field features have been reproduced

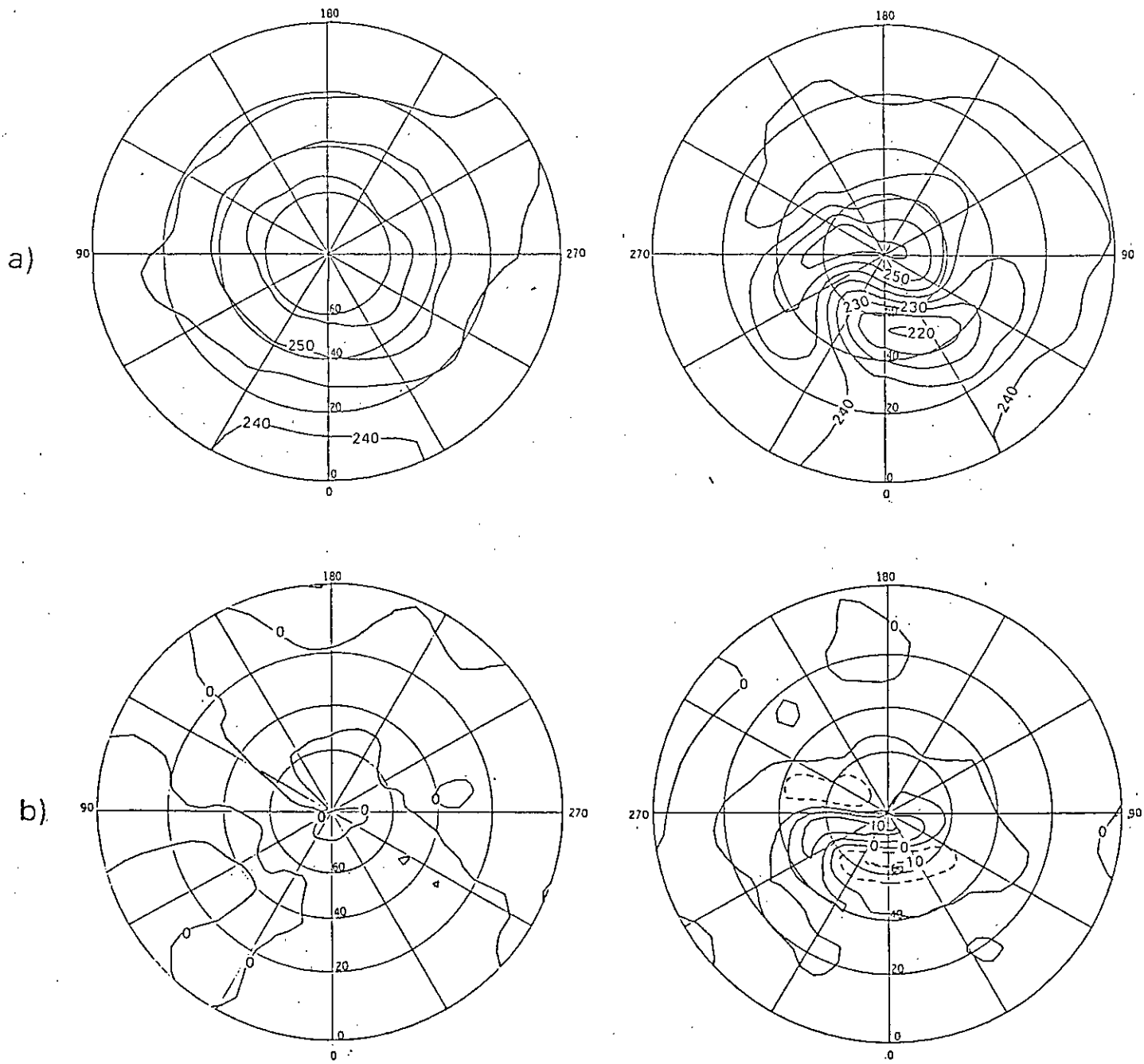


Figure 7.6 Temperature field for 1200 GMT on 18/1/87 at 5 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. Dashed contours represent negative values. a) analysed values obtained using a time radius of 6 hrs and a distance radius of 2000 km; b) difference between model field (Figure 7.1c) and field in a);

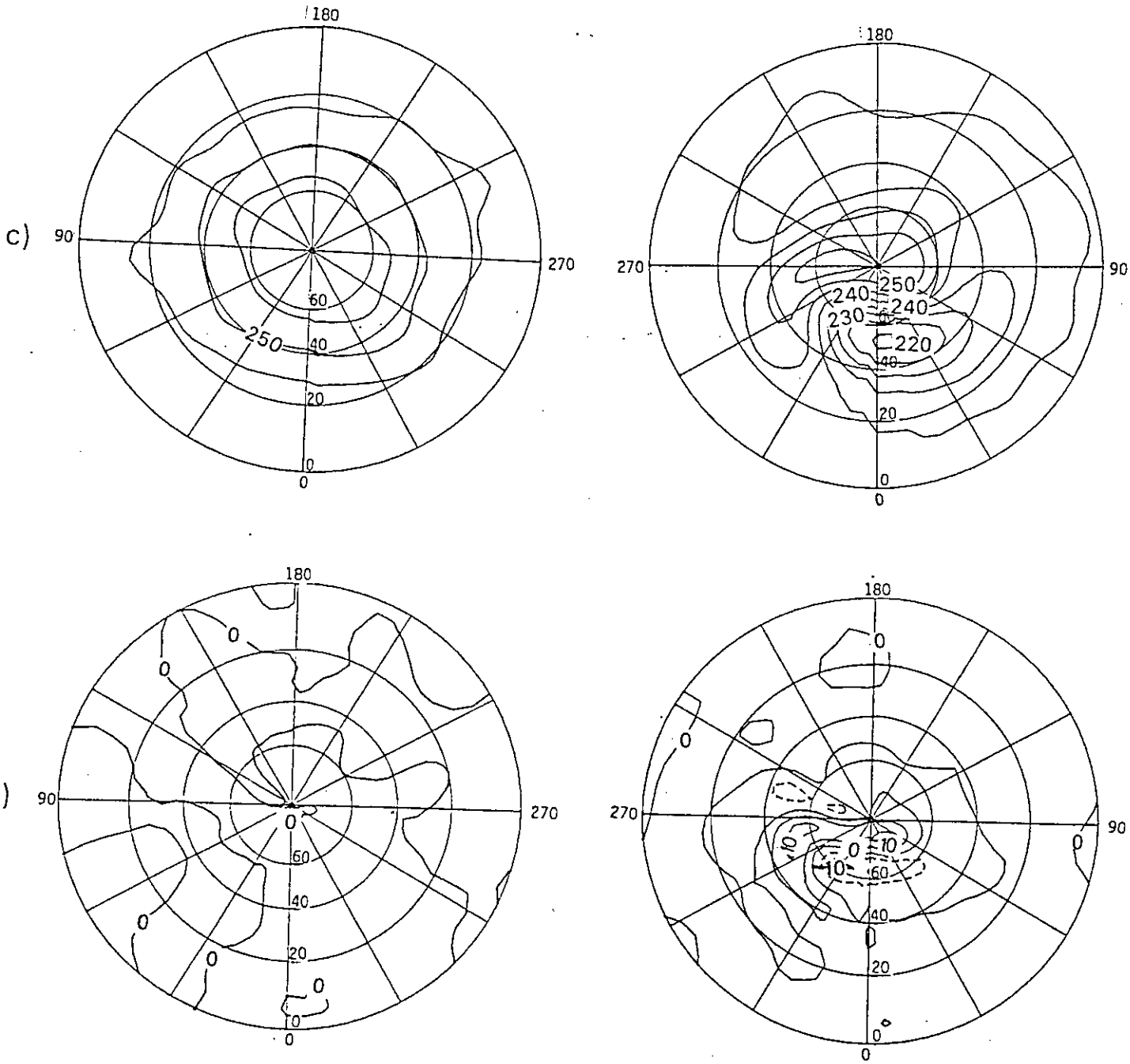
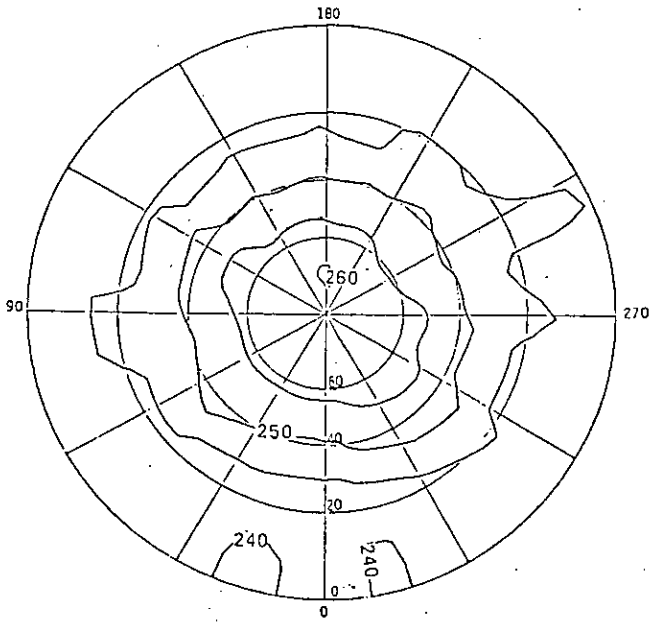
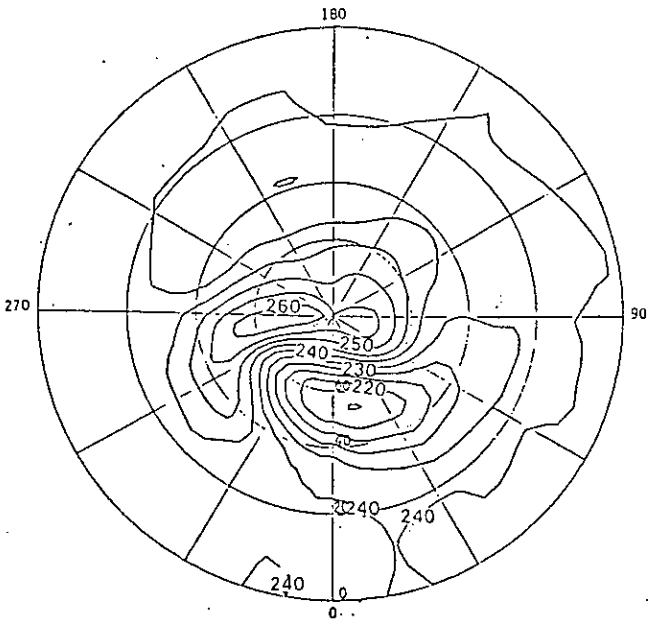


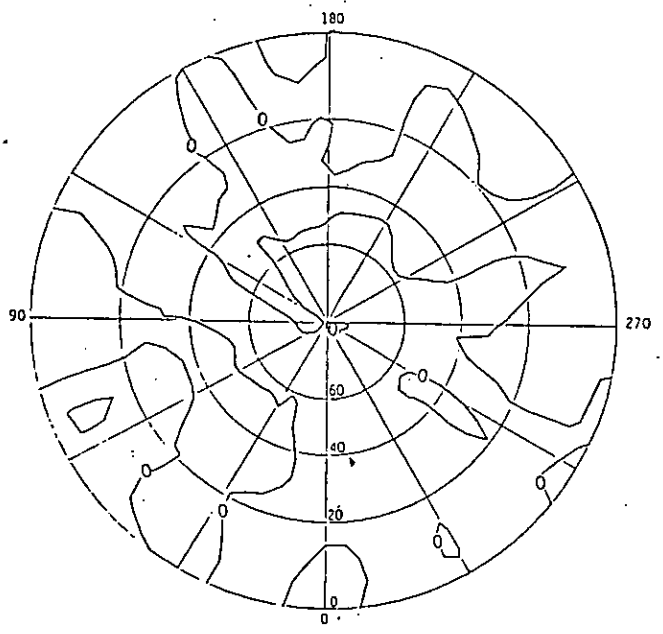
Figure 7.6 (cont.) c) as a), except the time radius is 12 hrs and the distance radius is 2000 km; d) difference between model field (Figure 7.1c) and field in c);

e)





f)



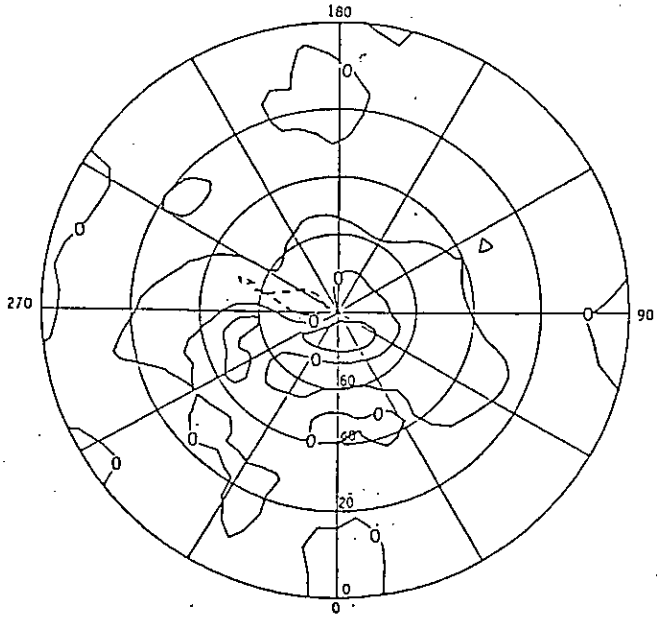
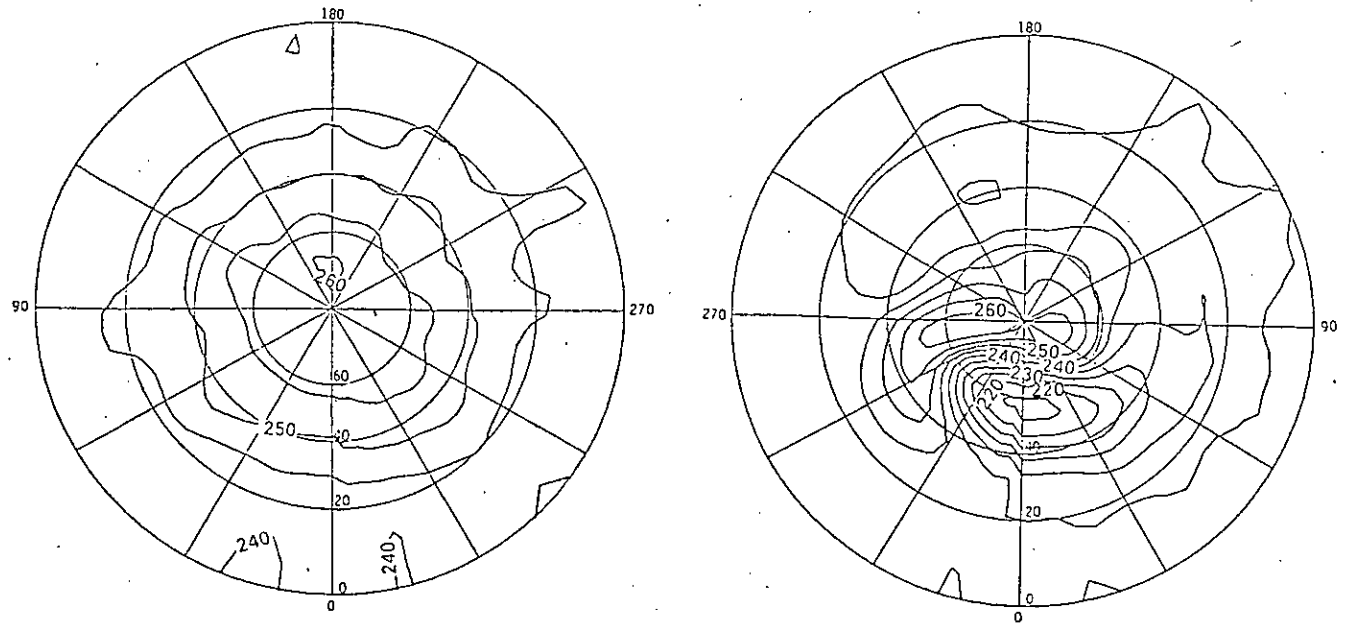


Figure 7.6 (cont.) e) as a), except the time radius is 12 hrs and the distance radius is 1000 km; f) difference between model field (Figure 7.1c) and field in e);

g)



h)

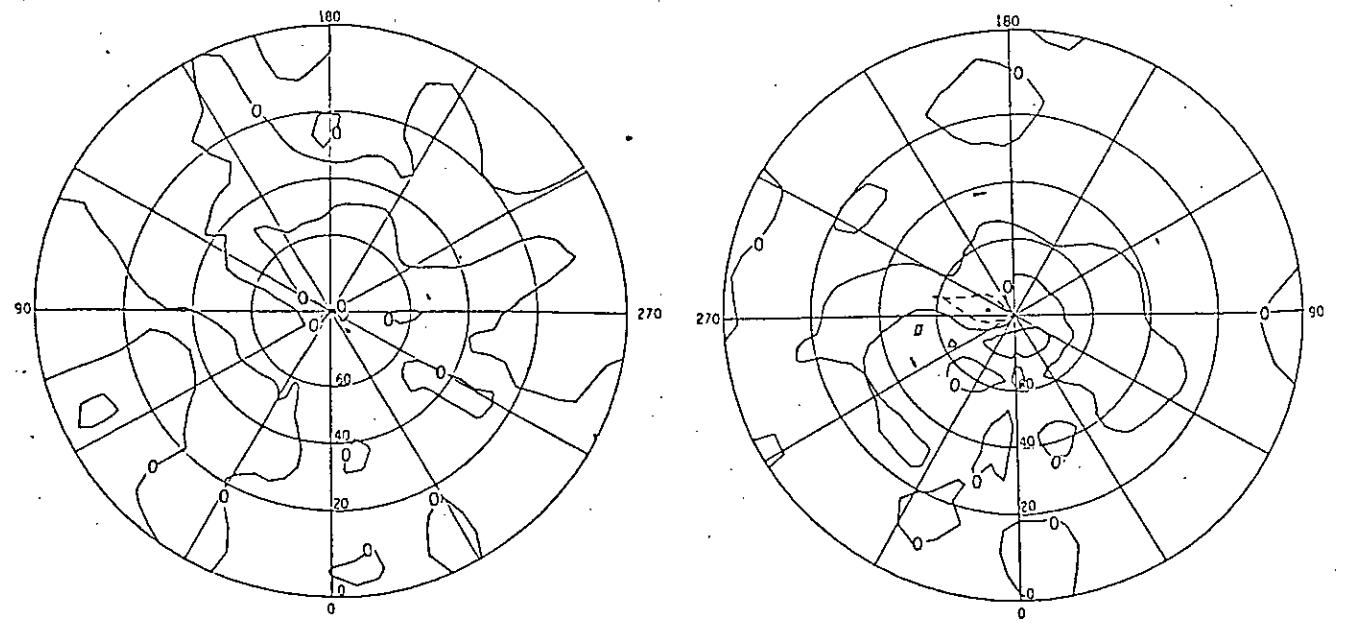


Figure 7.6 (cont.) g) as a), except the time radius is 12 hrs and the distance radius is 500 km; h) difference between model field (Figure 7.1c) and field in g).

satisfactorily, however Figure 7.6b (which shows the difference between model and analysed fields) reveals that the major high in the Northern hemisphere has been underestimated by over 10 K. In addition, the major low has been overestimated by 10 K, and in consequence the large gradient between the major high and low has been poorly estimated. Table 7.1a shows that changing the time radius from 6 to 12 hrs decreases the global r.m.s error slightly. However, Figure 7.6c, which shows the field analysed with a time radius of 12 hrs and a distance radius of 2000 km, reveals that qualitatively the estimate is little different from that made using a similar distance radius and a time radius of 6 hrs (Figure 7.6a). In particular, the magnitudes of the major high and low in the northern hemisphere have not been well estimated. Figure 7.6d, which shows the difference between the model field and the analysed field, reveals, for example, that the major high in the northern hemisphere has been underestimated by over 10 K. This field is similar to the corresponding field derived from the analysis made with a time radius of 6 hrs (Figure 7.6b). These results are in accord with results of tests in Chapter 4, which show that the quality of the estimated field changes little as the time radius changes.

Comparison of Figures 7.6c and 7.6d with corresponding fields analysed using a smaller distance radius (1000 km) confirms that use of a smaller distance radius improves the quality of the analysis. Figure 7.6e shows the field analysed with a time radius of 12 hrs and a distance radius of 1000 km, and shows that the gradient between the major high and low in the northern hemisphere has been much better estimated than when a distance radius of 2000 km is used (ie Figures 7.6a and 7.6c), and is only slightly weaker than the gradient in the model field (Figure 7.1c). In addition, the estimated magnitudes of the major high and low are closer to model values. Figure 7.6f shows that the underestimate of the major high has been reduced to between 5 and 10 K and that the major low has been estimated to within less than 5 K of the model

value. These results are consistent with those of Pick and Brownscombe (1979, 1981) who compared satellite-derived geopotential height fields at 10 and 20 mb with corresponding fields which were subjectively analysed from sonde data. They found that both fields were qualitatively similar, but that the satellite data caused rapid gradient changes to be slightly smoothed out. In addition,the amplitude of highs and lows in the satellite-derived field was slightly lower.

Comparison of maps of the analysis made with a 1000 km distance radius (Figure 7.6e) and of the analysis made with a distance radius of 500 km (Figure 7.6g) reveals them to be very similar. In addition, maps of the difference between model and analysed fields (Figure 7.6f for the 1000 km analysis; Figure 7.6h for the 500 km analysis) are also similar. This suggests that there is no advantage in using the Met. Office operational distance radius of 500 km, which requires gap-filling when orbits fall far from gridpoints, over a distance radius of 1000 km, which requires gap-filling only where observations are missing due to calibration sequences or drop-outs.

7.2. Conclusions

Preliminary tests of the time/space interpolation scheme were made in Chapter 4. In this chapter we have described further tests of the scheme using a more realistic field. The findings of these tests can thus be more confidently applied to the evaluation of operational analyses of satellite measurements.

In general, the conclusions drawn from the preliminary tests also hold for the tests described here. In particular, changing the time radius from 6 to 12 hrs causes little qualitative change to the analysis, whilst changing the distance radius from 2000 to 1000 km improves the analysis. A corresponding change to the distance radii used to estimate Fields 1 and 2 in Chapter 4 led to a slight decrease in the quality of the analysis, but this apparent discrepancy between the tests can be explained by

1. Fields 1 and 2 consist entirely of travelling waves, which means that when a small distance radius is used, the number of observations within the search radii is too small to resolve the moving field. However, the model field used in the tests in this chapter changes less in time than do Fields 1 and 2, so the small number of observations does not lead to a decrease in the quality of the analysis. This result is in accord with results of preliminary tests of Field 3, which is a stationary field.

2. The model field used here has smaller spatial scales than the fields used in Chapter 4, so that changing the distance radius from 2000 to 1000 km is more likely to cause an increase in the quality of the estimate.

To provide a stringent test of the scheme, these tests were performed using a field affected by a sudden warming. Results show that the interpolation scheme appears broadly able to resolve dramatic dynamical events such as this. Whilst the performance of the interpolation scheme is degraded by increasing the distance radius beyond 1000 km, there appears to be little advantage in the current Met. Office operational distance radius of 500 km, which requires gap-filling where orbits fall far from gridpoints, over the 1000 km radius, which requires gap-filling only where observations are missing due to calibration sequences or drop-outs.

The small scale features of the model fields at 0.2 and 1.5 mb are poorly estimated. The high errors at 0.2 mb may be due to poor retrievals, since that pressure level is far away from a weighting function peak. However, 1.5 mb is close to the SSU channel 27 weighting function peak, and hence the errors here may be caused by the scheme's inability to resolve the small scale features (ie about 200 to 300 km) of these model fields. On the other hand, such errors may arise because the vertical resolution of the satellite instrument is too low to measure such small scales. These small scale features may be partially due to inaccuracies in the model; however, it is clear that if these scales do exist in the real atmosphere, an operational time/space interpolation scheme will be unable to reproduce them.

CHAPTER 8

TESTS OF THE SEQUENTIAL ESTIMATION ANALYSIS METHOD

In this chapter satellite observations are analysed by sequentially estimating Fourier field coefficients at grid latitudes, a method which is described in more detail in Chapter 3. This application of sequential estimation was originally proposed by Rodgers (1976c), but, although it has been used to analyse observations from LIMS (see eg Gille and Russell, 1984), it has not been used to analyse TOVS measurements. Hence in this chapter we test the sequential estimation method using two fields: 1) an analytically calculated radiance field (described in Chapter 4); and 2) a field calculated by the Met. Office stratosphere/mesosphere model (described in Chapter 5). We shall then compare the results of these tests with corresponding results of tests of the time/space interpolation scheme (which appear in Chapters 4 and 7, respectively), and discuss the strengths and weaknesses of each scheme. The analytically calculated radiance field contains fast-moving waves which provide a stringent test of the sequential estimation scheme. However, because of the simplicity of the field, the results of such tests cannot reasonably be used to evaluate the ability of the scheme to estimate real stratospheric fields. Hence, the sequential estimation scheme is further tested using a Met Office model field, which is more like the real stratosphere than is the analytical field. We use a model field from a day affected by a sudden warming because the large temperature gradients and small-scale structure present in its Northern Hemisphere on that day (see Section 7.1.1) provide a severe test of the sequential estimation scheme. It is reasonable to conclude that if the sequential estimation scheme can satisfactorily estimate the model field, then it will be able to satisfactorily estimate most 'true' stratospheric fields when it is used operationally. Details and results of tests using both analytical and model fields appear in Section 8.1 and Section 8.2, respectively, whilst conclusions and

suggestions for future research are in Section 8.3.

8.1. Tests Using Analytically Calculated Fields – Details and Results

The sequential estimation scheme is first tested using an idealised radiance field – Field 2 (Figure 8.1). This field resembles the stratosphere as observed by an SSU channel during a northern hemisphere winter in that there are small amplitude waves in the southern hemisphere and tropics and higher amplitude wavenumber 1, 2 and 3 waves in the northern hemisphere middle and high latitudes. The model field is fast-moving, and so provides a stringent test of the analysis scheme. Other details of the tests, such as the model used to simulate the TIROS-N satellite orbit, and the method of simulating observations, are also as described in Chapter 4. 48 hours of observations are used to produce an analysis at time t = 24 hrs on a grid with a resolution of 10° in latitude and 20° in longitude. At each observation point an estimate of the Fourier fields is made using equation (3.36), which combines the 'first guess' estimate (equation (3.35)) with the observation. Two sequential estimates are made: 1) a forward estimate of the Fourier field coefficients using observations from t = 0 hrs to the observation time closest to t = 24 hrs; 2) a backward estimate of the Fourier coefficients using observations from t = 48 hrs to the observation time closest to t = 24 hrs. The 'updated' forward estimate at the observation time closest to t = 24 hrs is then combined with the corresponding *a priori* backward estimate. As mentioned in Section 3.4.2.3, this is done to avoid using the observation nearest t = 24 hrs twice. These estimates are combined using equation (3.38), and the combined estimate is then smoothed to the analysis time using equation (3.39). The smoothed estimates of the Fourier coefficients are then used to calculate the analysed field in physical space at the space/time gridpoints.

Prior to performing the sequential estimation, observations must be

b)

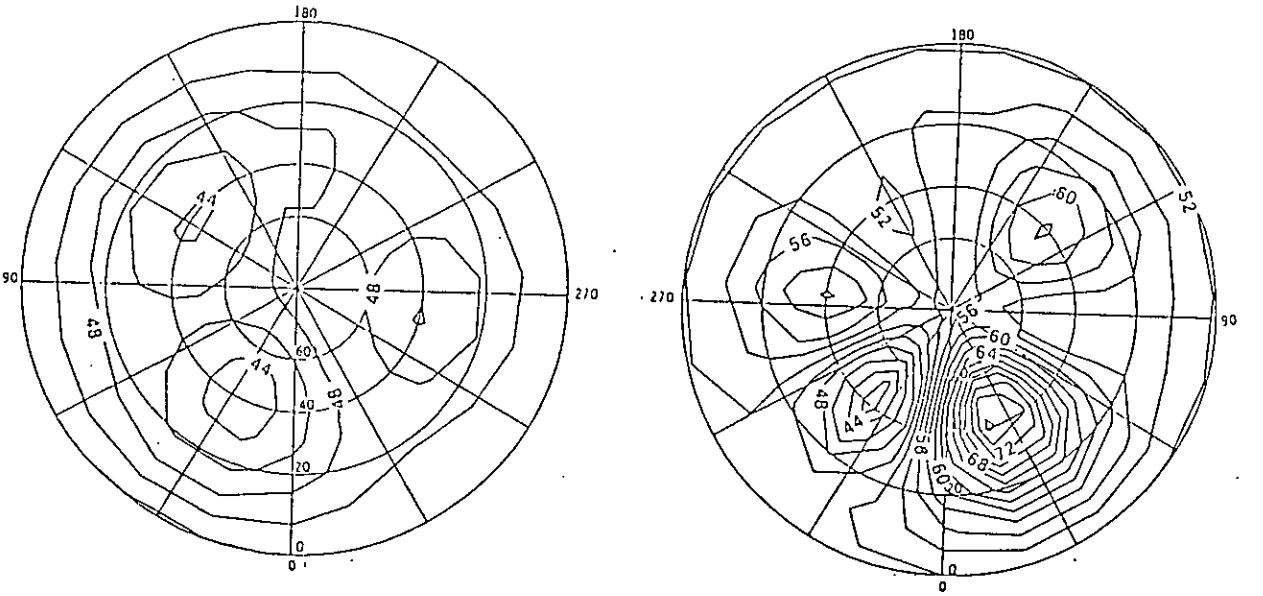


Figure 8.1 Model radiance field 2 at t = 24 hrs. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 2 r.u.

interpolated to grid latitudes, and initial 'first guess' values of the Fourier coefficients and its error covariance must be specified. Details of these preliminaries appear in Subsection 8.1.1, whilst results are in Subsection 8.1.2.

8.1.1. Preliminaries – Interpolation and Calculation of Initial 'First Guess' Estimates

Since we perform independent estimates of Fourier field coefficients around grid latitude circles, we must first express the observations at these grid latitudes. The scanning pattern of the radiometer means that interpolation of observations along the orbital path is difficult because, in the ascending (descending) node part of an orbit, the observational latitude does not increase (decrease) monotonically. Hence, we initially consider the observations made at each of the four TOVS scan angles as separate time series. Observations in each of these series are linearly interpolated along the orbital path to grid latitudes, and then the four sets of interpolated time series are merged to produce one, large time series. At any grid latitude, this time series consists of groups of about 10 interpolated observations each separated by a larger time interval. The time interval between interpolated observations within these groups is similar to the interval between successive TOVS observations (about 16s). On the other hand, the larger time intervals are more variable: the sum of two consecutive 'larger' time intervals equals the orbital period (about 100 minutes), but, because ascending and descending nodes are not treated separately, individual 'larger' time intervals vary with latitude. For example, at certain high latitudes the 'large' time intervals will alternate between 10 minutes and 90 minutes, whilst at the equator, the 'large' time interval is always 50 mins (if the field to be analysed contains tidal effects, a better approach may be to treat the ascending and descending nodes separately; however, there are no tides in the model fields used in the tests described here).

Because the time interval between most of the observations is small, we exploit the temporal coherence of the field by assuming that the Fourier coefficients evolve in time according to a random walk (equation (3.35)). In these tests we set the innovation, η_t, in equation (3.35) to zero. Further research may investigate the effect of using different values for η_t. Before starting the sequential estimation, we need to calculate what we will refer to as 'initial values', viz: the 'first guess' of the Fourier coefficients, $\mathbf{x}_{ot}$, and its error covariance, $\mathbf{S}_{ot}$, at time t = 0 hrs, and also Δ **S**, which is a measure of the increase in the uncertainty of the 'first guess' per unit time. Since the observations are derived from an analytical radiance field, it is straightforward to calculate Fourier field coefficients at the observation points. From equation (4.1)

$$a_{it} = \sum_{j=0}^{J} \alpha_{jit} P_{j+i}^{i}(\phi)$$

$$b_{it} = \sum_{j=0}^{J} \beta_{jit} P_{j+i}^{i}(\phi)$$

where a_{it} and b_{it} are the cosine and sine Fourier components, respectively, for zonal wavenumber i and time t, j is the meridional wavenumber, α_{jit} and β_{jit} are spherical coefficients, $P_{j+i}^{i}(\phi)$ is an associated Legendre polynomial, and ϕ is latitude. The means and covariances of the Fourier components are then calculated at each grid latitude, and these values are used to represent $\mathbf{x}_{ot}$ and $\mathbf{S}_{ot}$ at t = 0 hrs. Subsequent values for $\mathbf{x}_{ot}$ and $\mathbf{S}_{ot}$ are, of course, calculated using equation (3.35). Values for Δ **S** are represented by the covariance of a sample of Fourier components whose spacing in time is similar to the larger time interval between interpolated observations, namely of the order of 50 mins.

8.1.2. Results

Initially, the sequential estimation method is tested using 'initial values' calculated by the method described in Section 8.1.1 (Run A in Table 8.1). The 'initial values' are calculated using the observations we are trying to analyse (this is because we only have a short time series of data available; the sequential estimation technique is usually applied to long time series), but in reality we would not know the Fourier coefficients and their variability to such precision. However, operationally these values could be calculated from the previous day's analysis (although one must be careful, especially during development of the scheme, since a poorly-estimated previous day's analysis could lead to poorer analyses on subsequent days). We assume that the difference between these values and the actual field values of the day on which the analysis is made would not be more than 20 %: Labitzke and Goretzki (1982) examined time series of wavenumber 1 and 2 components of the observed temperature field in 16 northern hemisphere winters from 1965/6 to 1980/1. Results show that daily changes during sudden warmings are usually about 20 %, and that daily changes are about 10 % or less at other times (note, however, that in most cases the day-to-day variability is much less than 10 or 20 %. For example Figure 8 in Boville and Randel (1986) shows that the variability of observed temperature in the tropical and summer stratosphere is about a fifth of the variability in regions most strongly affected by sudden warmings.) Hence we examine the effect of using 'initial values' which are different from those calculated from observations (i.e the values used in Run A) by 10 % (Runs B and C) and by 20 % (Runs D and E) (these changes are applied to both zonal mean and wave components). R.m.s errors are calculated by comparing estimated and model fields, and are shown in Table 8.1. Further tests change each of the three 'initial values' individually by a large amount (50 %) in order that we might assess the effect that each has on the analysis. R.m.s errors for these tests appear in Table 8.2.

Initially, the sequential estimation method is tested using 'initial values' calculated by the method described in Section 8.1.1 (Run A in Table 8.1). The 'initial values' are calculated using the observations we are trying to analyse, but in reality we would not know the zonal mean Fourier coefficients and their variability to such precision. However, operationally these values could be calculated from the previous day's analysis (although one must be careful, especially during development of the scheme, since a poorly-estimated previous day's analysis could lead to poorer analyses on subsequent days). We assume that the difference between these values and the actual field values of the day on which the analysis is made would not be more than 20 %: Labitzke and Goretzki (1982) examined time series of wavenumber 1 and 2 components of the observed temperature field in 16 northern hemisphere winters from 1955/6 to 1980/1. Results show that daily changes during sudden warmings are usually about 20 %, and that daily changes are about 10 % or less at other times. In addition, Figure 8 in Boville and Randel (1986) shows that the variability of observed temperature in the tropical and summer stratosphere is about a fifth of the variability in regions most strongly affected by sudden warmings. Hence we examine the effect of using 'initial values' which are different from those calculated from observations (i.e the values used in Run A) by 10 % (Runs B and C) and by 20 % (Runs D and E). R.m.s errors are calculated by comparing estimated and model fields, and are shown in Table 8.1. Further tests change each of the three 'initial values' individually by a large amount (50 %) in order that we might assess the effect that each has on the analysis. R.m.s errors for these tests appear in Table 8.2.

Run	'Initial Values'	r.m.s error/r.u
...	.....	.....
A	calculated from observations	1.68
B	0.9 times values for Run A	1.81
C	1.1 times values for Run A	1.77
D	0.8 times values for Run A	2.14
E	1.2 times values for Run A	2.05

Table 8.1 R.m.s analysis errors in radiance units (r.u) (r.u = mW m^{-2} cm $ster^{-1}$). The analysis was performed by sequential estimation, using a variety of 'initial values'.

Figure 8.2a shows the field estimated using Run A values. A comparison with the model field (Figure 8.1) shows that the shape, location and magnitude of the model features have been well estimated everywhere south of about 50°N. Figure 8.2b, which shows the difference between analysed and model fields, reveals that the magnitudes of the major highs and lows in the northern hemisphere middle latitudes have been estimated to within about 2 r.u of the model values. This is as good as, or better than, estimates of the model field made by the time/space interpolation method. North of 50°N, the analysis 'closes off' the major low (centred near 30°N and 40°W) and major high (centred near 40°N and 20°E), instead of extending them to the North Pole (as in the model field (Figure 8.1)), and Figure 8.2b shows that the difference between model and analysed fields exceeds 6 r.u in this region.

The global r.m.s error for this analysis is 1.68 r.u, while the smallest corresponding r.m.s error using the time/space interpolation method is 0.65 r.u (see Table 4.4). However, as discussed in Section 4.3, one should not over-rely on r.m.s error as a means of assessing the quality of an estimate, especially when the r.m.s error is calculated for the whole globe. For example, comparison

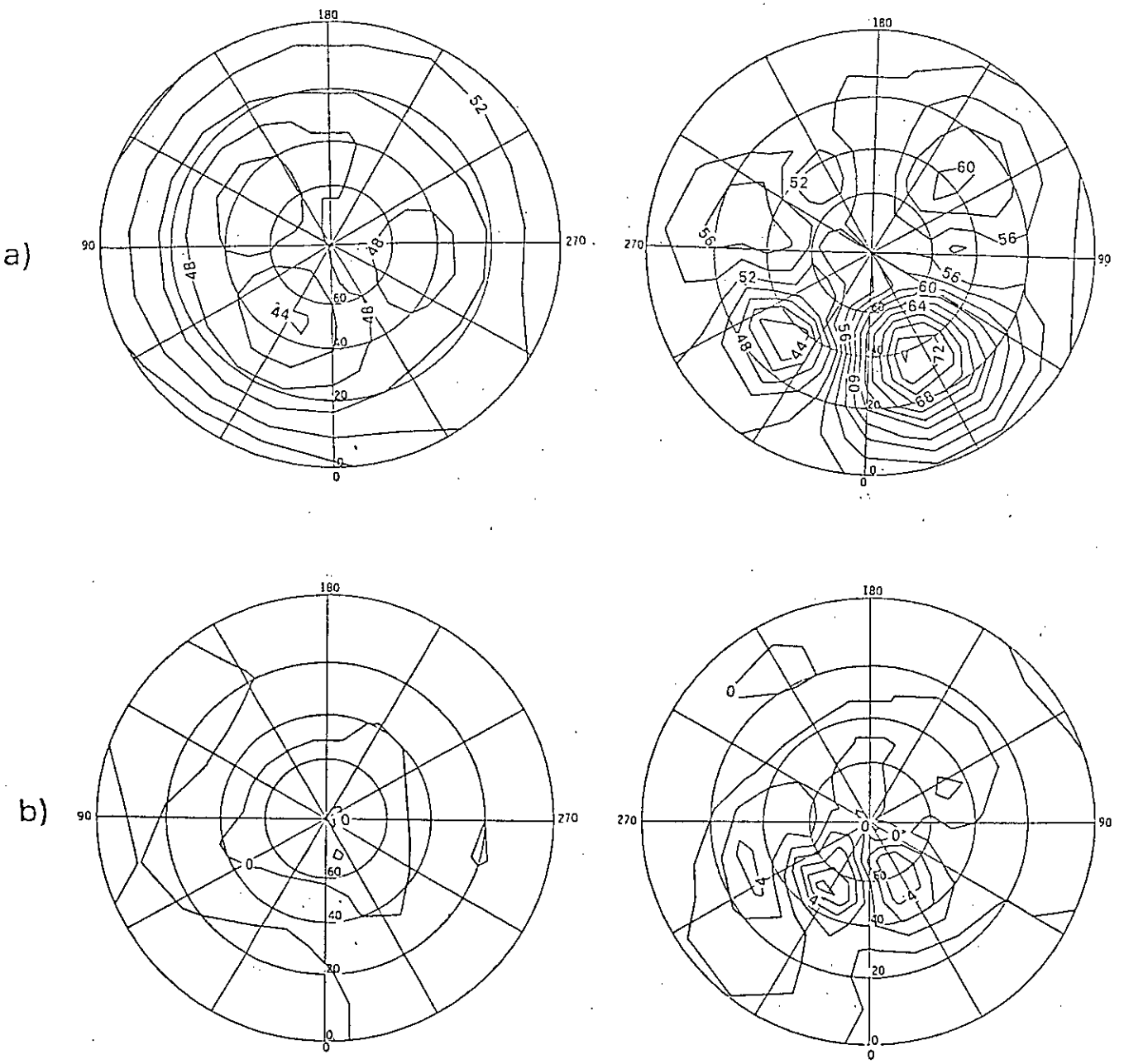
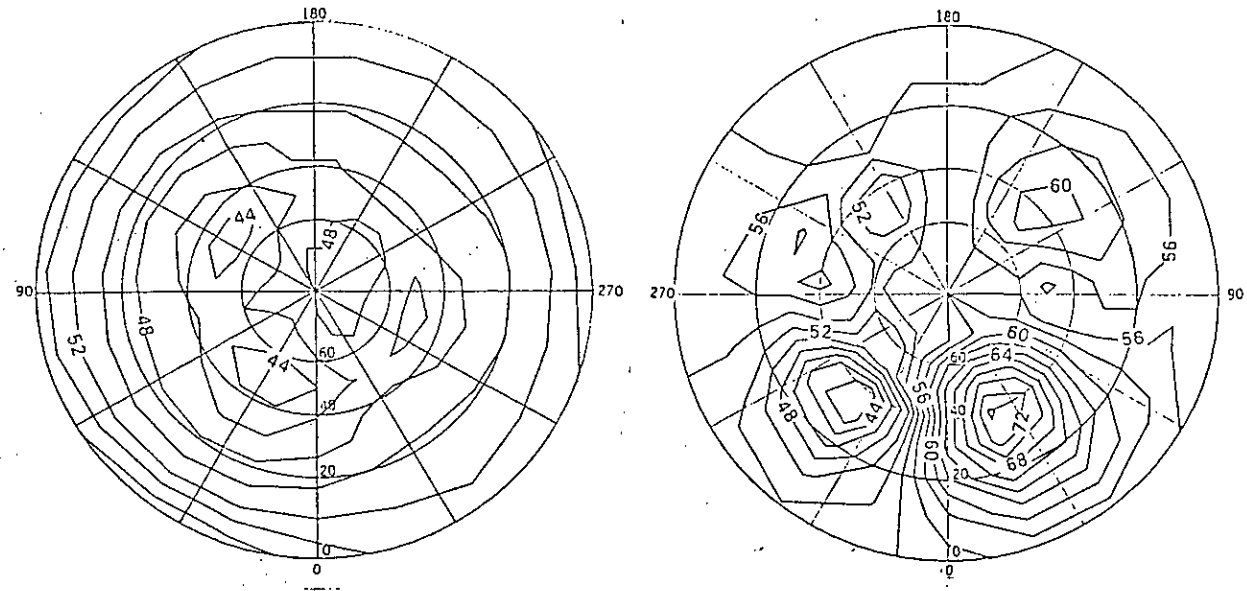


Figure 8.2 Field 2 analysed at t = 24 hrs using the sequential estimation of Fourier field coefficients. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing in 2 r.u. a) analysed field obtained using initial values calculated from the observations. b) difference between a) and the model field (Figure 8.1).

c)



d)

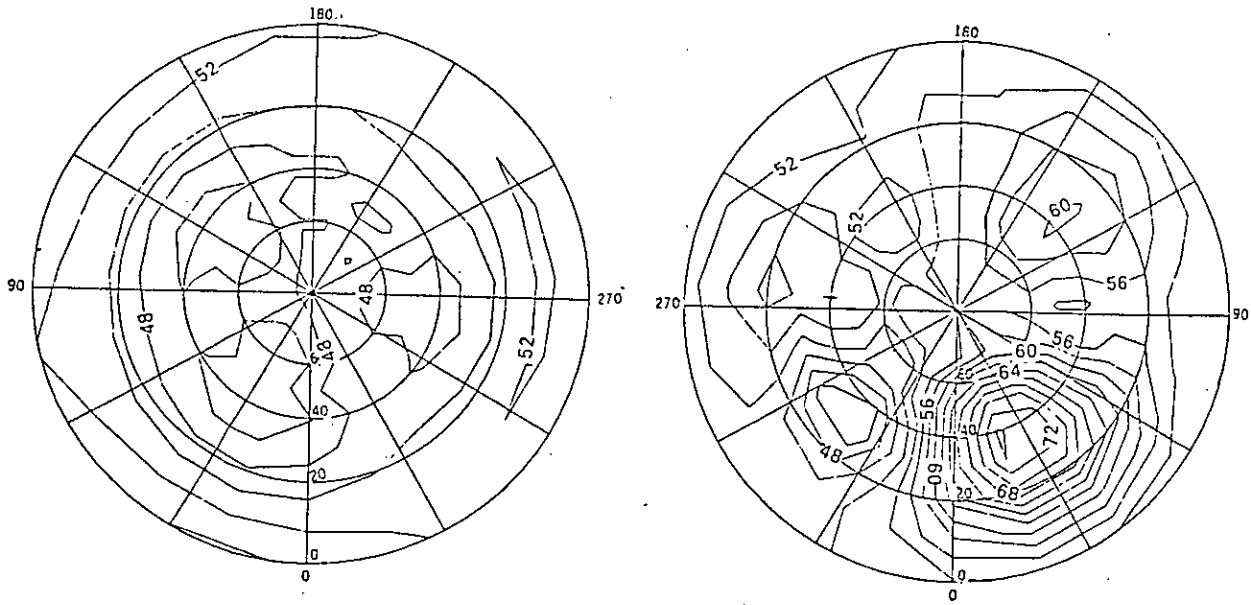
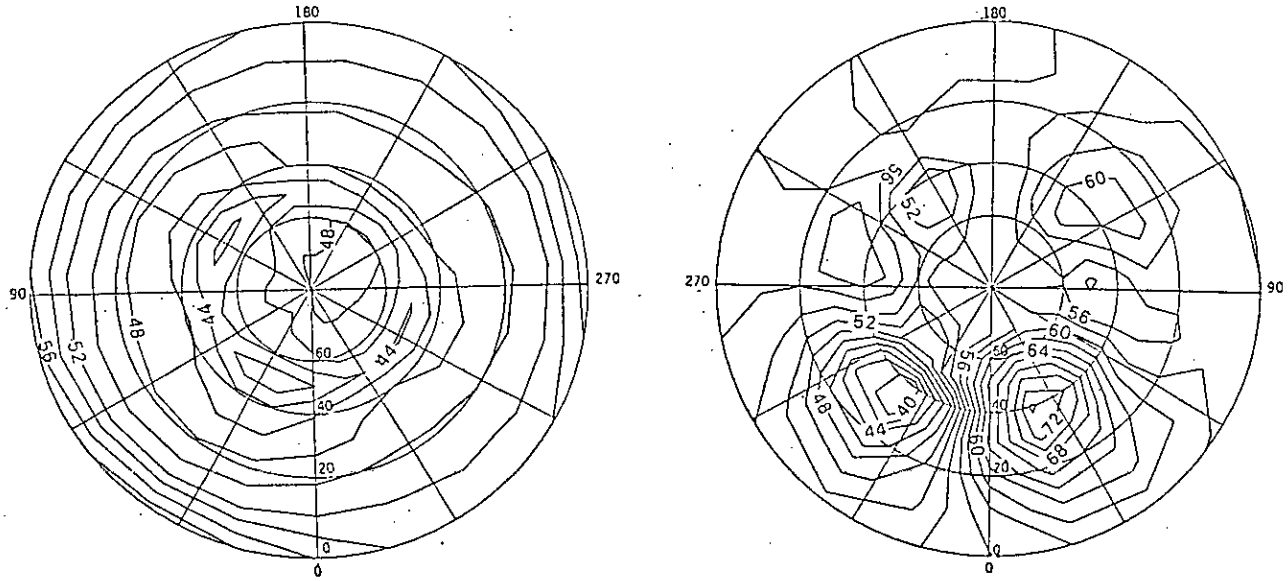


Figure 8.2 (cont.) c) as a), except initial values are 0.9 times those calculated from observations. d) as a), except initial values are 1.1 times those calculated from observations.

e)



f)

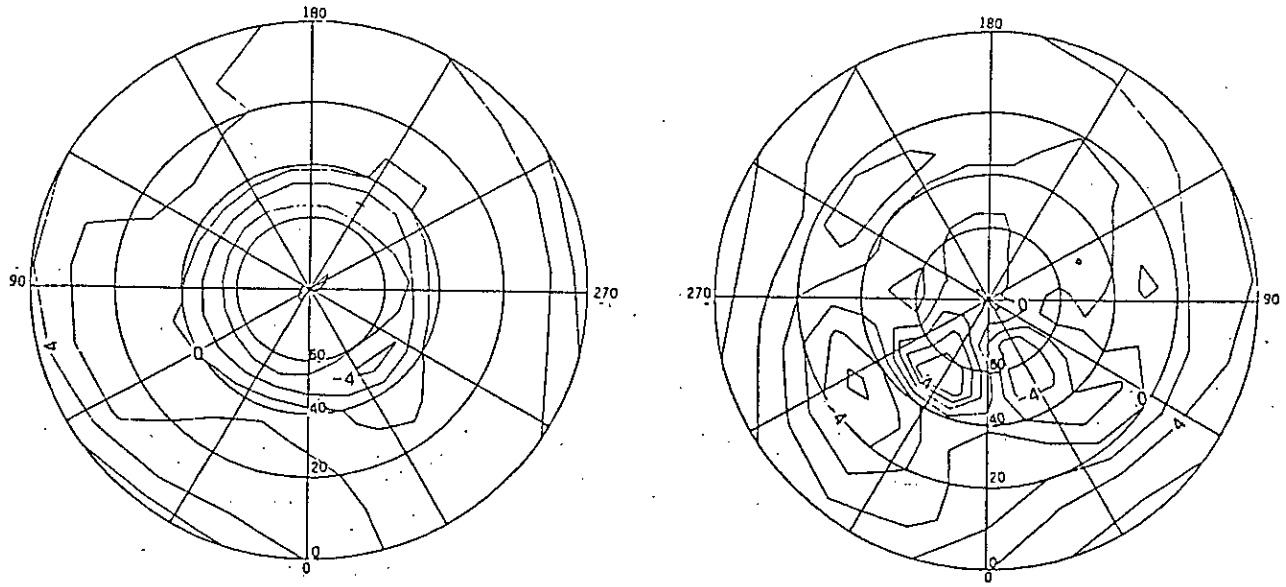
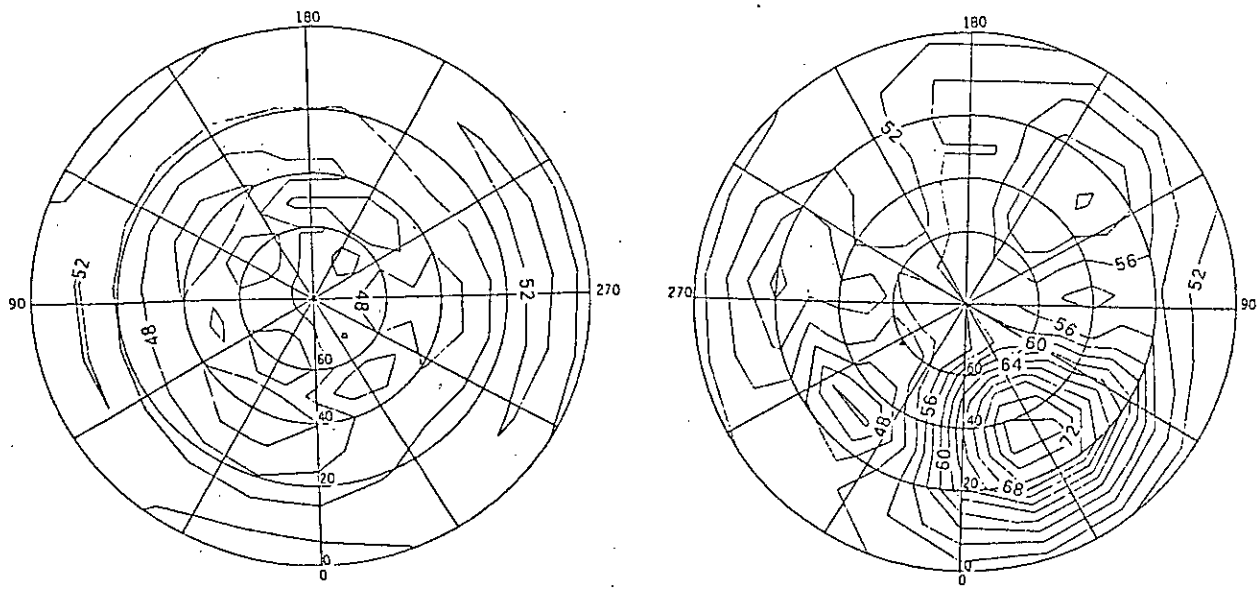


Figure 8.2 (cont.) e) as a), except initial values are 0.8 times those calculated from observations. f) difference between e) and the model field (Figure 8.1).

g)



h)

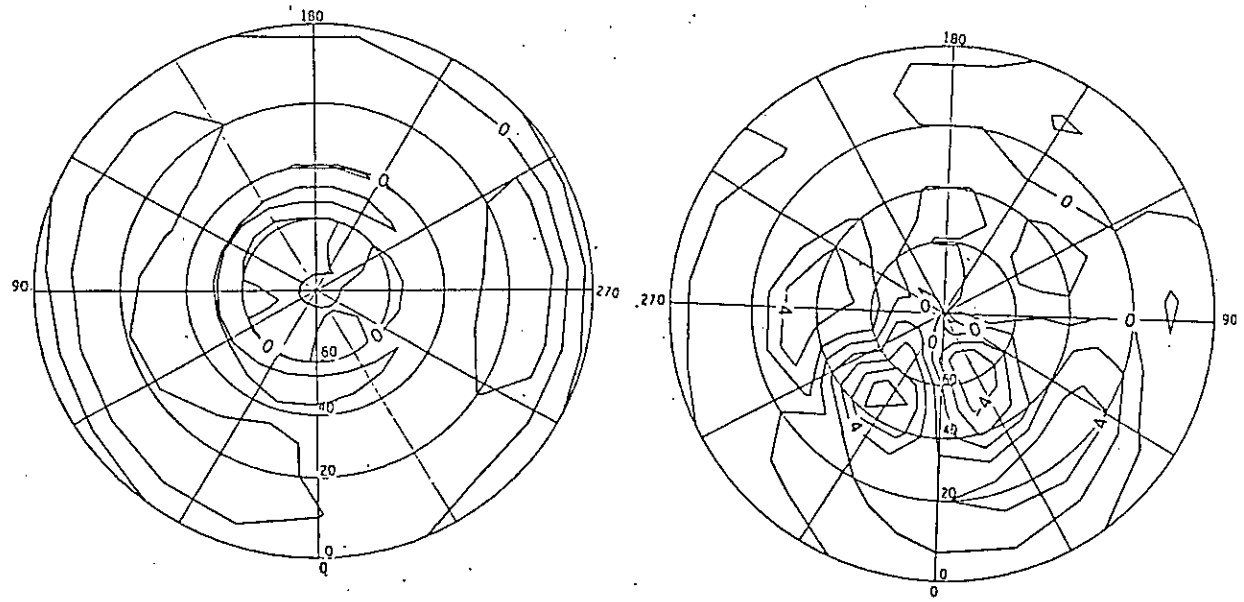


Figure 8.2 (cont.) g) as a), except initial values are 1.2 times those calculated from observations. h) difference between g) and the model field (Figure 8.1).

of the sequentially estimated field (Figure 8.2b) with a time/space interpolated field produced using a time radius of 12 hrs and a distance radius of 2000 km (Figure 4.3c) shows that the shape, location and magnitude of most model field features have in general been reproduced equally well by both techniques. However, the sequentially estimated field fails to reproduce the major model highs and lows north of 50°N, and the shape of the model low near 30°N, 45°W is also poorly estimated. The poor estimates in these regions may occur because the random walk model used to estimate the time evolution of the Fourier coefficients is inappropriate in these regions. However, in general the ability of the sequential estimation method to successfully estimate the shape, location and magnitude of fast-moving large amplitude waves is encouraging.

When Run A 'initial values' are multiplied by 0.9 and 1.1, the resultant analyses are only slightly different from the Run A estimate. The analysed field for Run B (Figure 8.2c) is slightly different than that for Run A in the southern hemisphere: two lows enclosed by the 46 r.u contour appear between 0°E , 150°W and 30°S, 70°S in the Run A estimate, but have been smoothed together in the Run B estimate. The northern hemisphere fields are quantitatively similar, however. The analysed field for Run C (Figure 8.2d) is also very similar to the Run A analysis, but is less smooth, especially in the southern hemisphere. When Run A 'initial values' are multiplied by 0.8 and 1.2, differences between the resultant analyses and the Run A analysis are not unexpectedly larger than corresponding differences between Run B and C analyses and the Run A analysis. Figure 8.2e shows the analysed field for Run D. The two small lows in the southern hemisphere model field have been underestimated, whilst the shape of the model high at 90°W and 40°S has been poorly reproduced. In addition, analysed values exceed model values by about 4 to 6 r.u near the equator, as evidenced by Figure 8.2f, which shows the difference between analysed and model fields. In the northern hemisphere the major high near

40°N, 20°E is as well estimated as when Run A statistics are used, but the major low near 30°N, 40°W has been underestimated by about 4 r.u. Comparison of Figures 8.2b and 8.2f shows that the difference between the model field and the analysed field in the southern hemisphere increase when initial values are multiplied by 0.8. However, differences between model and analysed fields in the northern hemisphere are about the same in both figures. Figure 8.2g shows the analysed field for Run E. The shape, magnitude and location of the major southern hemisphere model features have been reproduced, but additional small-scale features exist near the poles, and, like Run C, the analysed field is generally not as smooth as the Run A analysed field. In the northern hemisphere the major high near 40°N, 20°E is well estimated, but the neighbouring low near 30°N, 40°W has been overestimated by about 4 r.u. In addition, the high near 90°W, 30°N has been underestimated by about 4 r.u, as evidenced by Figure 8.2h, which shows the difference between analysed and model fields.

In summary, the main points from these tests are

1. The sequential estimate made using values calculated from the radiance field (ie Run A) successfully reproduces the shape, location and magnitude of major model features: the r.m.s error is over twice the size of the smallest corresponding error of time/space interpolated fields chiefly because of a relatively poor estimate of the model field poleward of 50°N, rather than because of any serious failure of the estimation scheme.

2. Sequential estimates made with 'initial values' 10 % different from those calculated from the radiance field (ie Runs B and C) successfully reproduce model field features, but the estimated fields are not as smooth as the corresponding Run A field.

3. Estimates made with initial values 20 % different from those calculated from the model (ie Runs D and E) contain considerable erroneous small-scale structure, but most major model field structure has been adequately reproduced.

The above results show that changing the 'initial values' can affect the analysed product. However, up to this point, the term 'initial values' has embraced three quantities, viz:

- x_{ot}, the 'first guess' Fourier coefficient at time t = 0
- S_{ot}, the error covariance of the 'first guess' at t = 0
- Δ S, the increase in the uncertainty of the 'first guess' per unit time

The results presented above do not indicate the contributions to the analysis error made by each of these quantities. This is examined below by altering each by a large amount (50 %), and leaving the others unaltered. Results of these tests are summarised in Table 8.2

Run	'Initial Values'	r.m.s error/r.u
F	Multiply x_{ot} by 1.5, other values as Run A	3.43
G	Multiply S_{ot} by 1.5, other values as Run A	1.68
H	Multiply Δ S by 1.5, other values as Run A	1.68
I	Multiply x_{ot} by 0.5, other values as Run A	3.53
J	Multiply S_{ot} by 0.5, other values as Run A	1.68
K	Multiply Δ S by 0.5, other values as Run A	1.68

Table 8.2 R.m.s analysis errors in radiance units (r.u) (r.u = mW m^{-2}cm $ster^{-1}$). Analyses are performed by sequential estimation, using a variety of 'initial values'.

It is clear from Table 8.2 that a large change in the value of x_{ot} at t = 0 has

a large effect on the quality of the analysed field. Changes to the specified values of S_{ot} and ΔS have little effect on the r.m.s error and the qualitative nature of the analysed fields (not shown), is similar to the field analysed using Run A statistics. This is probably because the values of S_{ot} and ΔS used are quite small, and since a random walk model is used, the evolving error covariance will become much larger than these values, especially when the time interval between interpolated observations is large (i.e of the order of half an orbital period). The size of the error covariance does not therefore seem to be strongly dependent on the initial values of S_{ot} and ΔS. However, changes to x_{ot} cause large changes to the estimated field. The analysed field for Run F is shown in Figure 8.3a. Multiplying x_{ot} by 1.5 leads to the field being generally overestimated. In the southern hemisphere the changes to x_{ot} have most impact near the equator and between 40°S and 60°S, where differences between model and analysed fields (Figure 8.3b) exceed 8 r.u. The analysis in the western part of the northern hemisphere fails to estimate any model wave structure at all, while in the eastern part the major high near 40°N, 20°E has been zonally stretched, its peak is estimated about 20° east of its model location, and is about 8 r.u higher than its model value. Whilst multiplying x_{ot} by 1.5 leads to an overestimation of the radiance field, multiplying x_{ot} by 0.5 not surprisingly causes most field features to be underestimated. Figure 8.3c shows the analysed field for Run I. Greatest differences between analysed and model fields in the southern hemisphere (Figure 8.3d), again occur near the equator and between 40° and 60°S. In the northern hemisphere the major low near 30°N and 40°W has been underestimated by over 10 r.u, and the gradient on the poleward side of the low is much stronger than in the model field. The major high near 40°N and 20°E has been estimated quite well, both in magnitude and location, but it is confined to mid-latitudes, whereas the model low exists in the whole hemisphere.

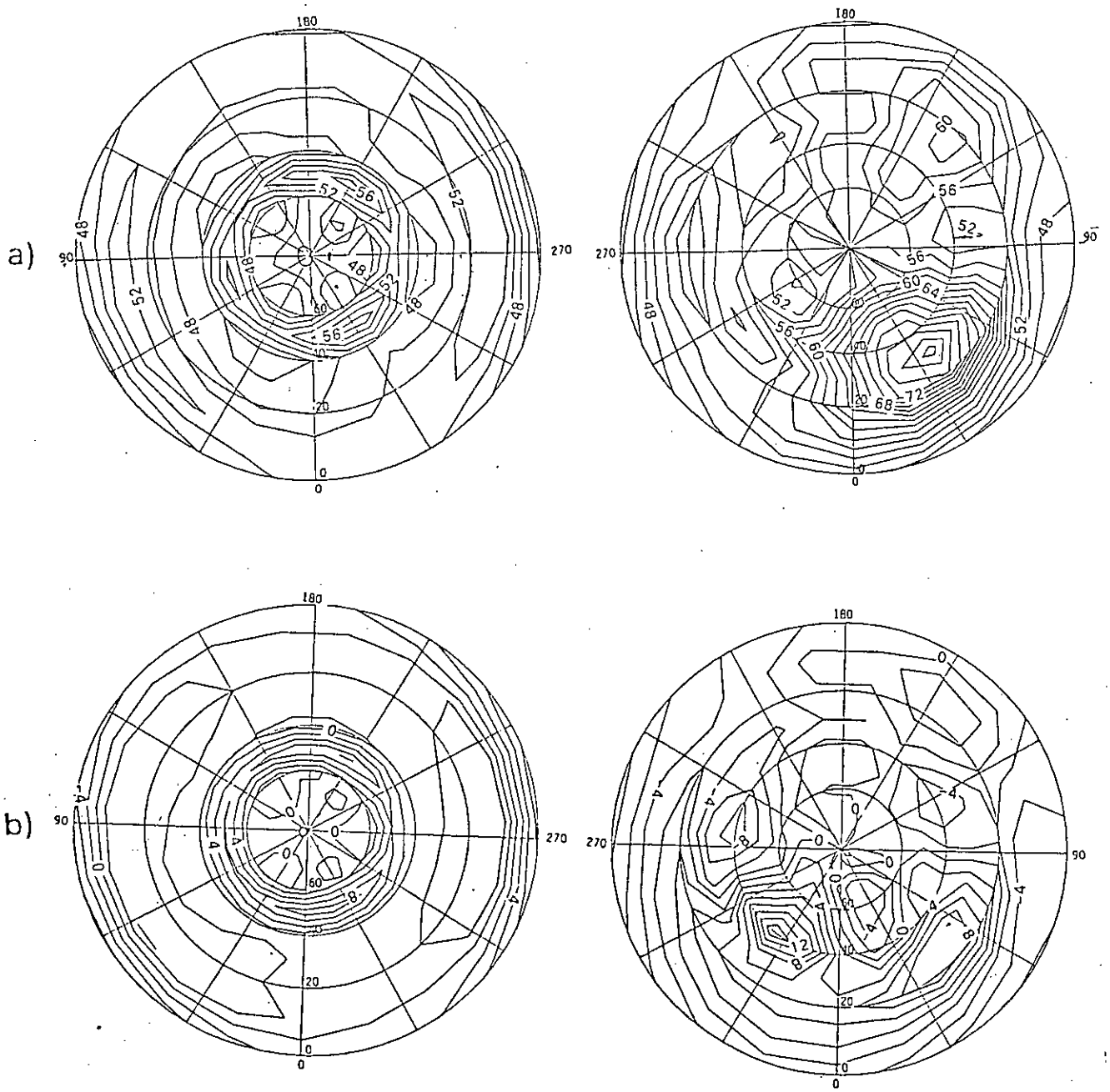
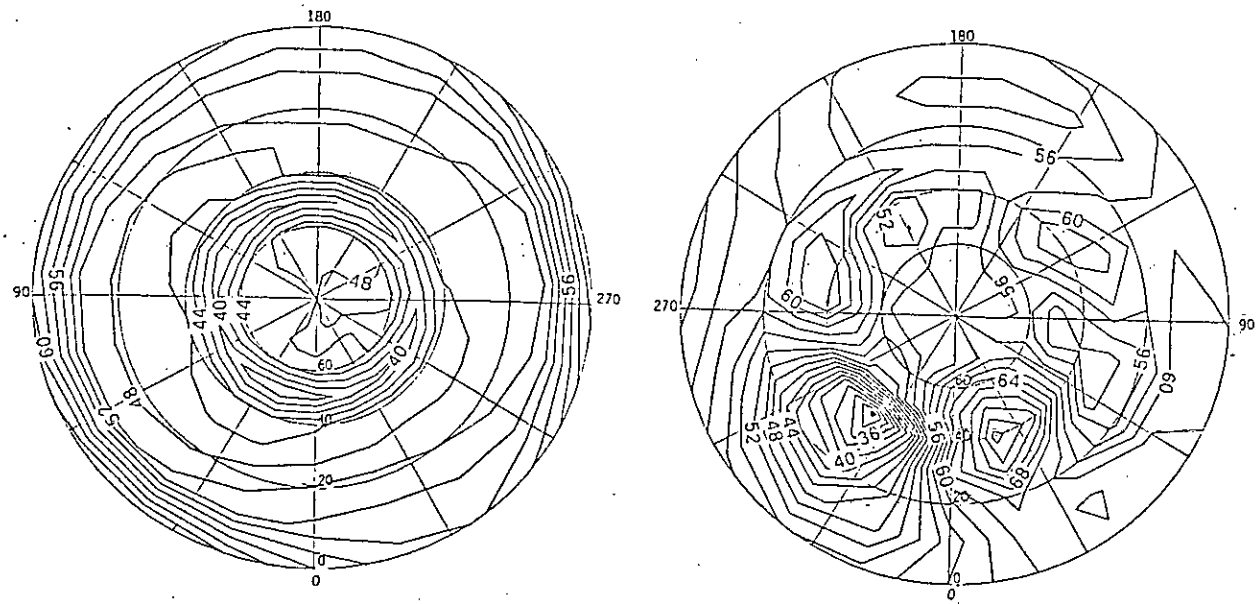


Figure 8.3 Field 2 analysed at t = 24 hrs using the sequential estimation of Fourier field coefficients. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing is 2 r.u. a) analysed field obtained using an initial value of x_{ot} which is 1.5 times the value calculated from observations. Δ S and the initial value of S_{ot} are calculated from observations. b) difference between a) and the model field (Figure 8.1).

c)



d)

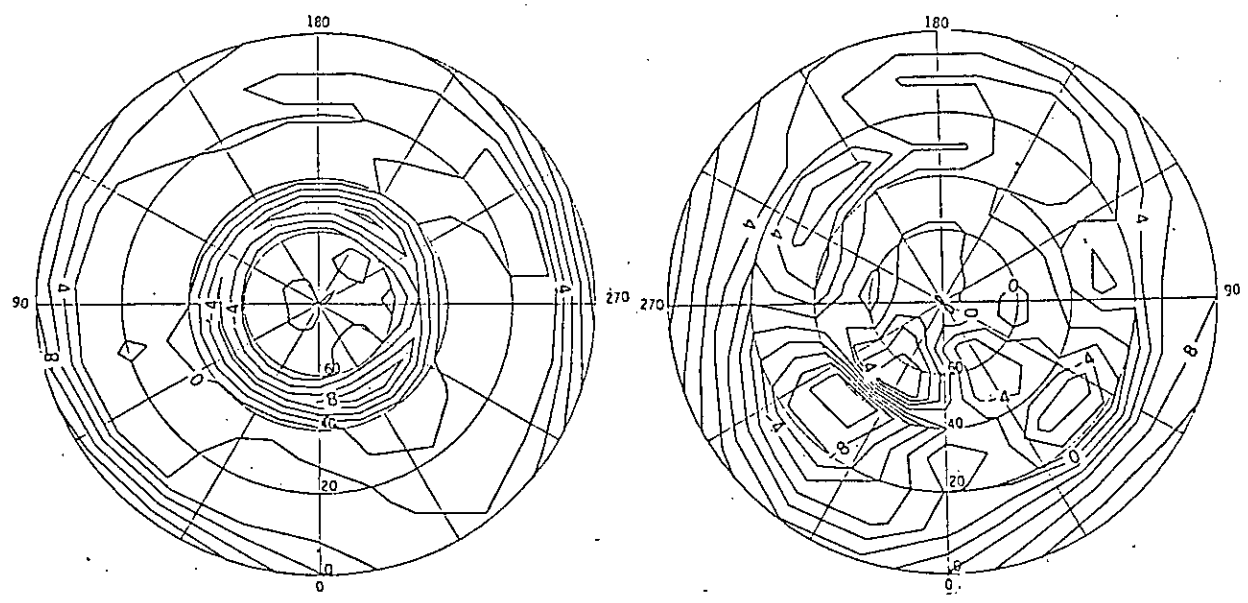


Figure 8.3 (cont.) c) as a), except the initial value of x_{ot} is 0.5 times the value calculated from observations. d) difference between c) and the model field (Figure 8.1).

These tests show that the sequential estimation method can produce a reasonable estimate of an analytical radiance field. The highest quality analysis was obtained using 'initial values' calculated from the field itself, but the quality of the analysis decreased only slightly when estimates were made with 'initial values' 10 or 20 % different from those calculated from the analytical field. Further tests show that the quality of the estimate depends strongly on the chosen value of x_{ot} at t = 0 hrs, but that the choice of S_{ot} at t = 0 hrs and of Δ S has little influence on the quality of the estimate. Although the analytical radiance field contains waves which are observed in the stratosphere, it does not adequately represent the true stratosphere. Hence, the tests described in this section are repeated in Section 8.2 using a more realistic field calculated by a numerical model.

8.2. Tests Using a Met. Office Model Field

Results from Section 8.1 show that the sequential estimation scheme satisfactorily estimates an idealised radiance field (ie the shape and location of model field features are accurately reproduced, and estimated values are within less than 6 r.u of 'true' values (compared to a variation over the model northern hemisphere of 24 r.u)). The next step is to examine how well the scheme can estimate a more realistic stratospheric field. To provide a strict test of the scheme, and to enable comparisons with results of corresponding tests of the time/space interpolation scheme (see Chapter 7), we use the same Met. Office model field from 18/1/87. On this day the model northern hemisphere is affected by a sudden warming. Results in Chapter 7 suggest that the characteristics of estimates made using the time/space interpolation scheme are of two types: 1) the estimates at 3, 5, 10, 15 and 25 mb resolve most of the model field features, and have broadly similar r.m.s error characteristics; 2) at 0.2 and 1.5 mb the time/space interpolation scheme has difficulty in resolving the small-scale model features. Accordingly, tests of the sequential estimation

scheme are made at one pressure level from each of these groups, namely 5 mb and 1.5 mb.

Prior to performing the tests, the retrieved temperatures from 18/1/87 must be interpolated to grid latitudes. This is done using the method described in Section 8.1.1, except that, whereas the radiance 'observations' used in the analytical field tests are interpolated to 19 grid latitudes (spaced at 10° intervals) between 90°S and 90°N, here the retrieved temperatures are interpolated to 36 latitudes (spaced at 5° intervals) between 87.5°S and 87.5°N. At any latitude, the time series of interpolated observations consists of groups of about 5 observations spaced about 16 s apart, each separated by a larger time interval which is approximately equal to half an orbital period (ie 50 min). Before starting the sequential estimation we also need to calculate 'initial values', namely $\mathbf{x}_{ot}$, the 'first guess' estimate of the Fourier coefficients at t = 0 hrs, $\mathbf{S}_{ot}$, the first guess error covariance at t = 0 hrs, and Δ S, which is a measure of the increase of the uncertainty of the 'first guess' per unit time. The 'initial values' are calculated using Met. Office model fields, expressed at hourly intervals, between 0000 GMT on 18/1/87 and 0000 GMT on 19/1/87. These fields are Fourier analysed, and the means and covariances of the Fourier coefficients are calculated at each grid latitude. These values are used to represent $\mathbf{x}_{ot}$ and $\mathbf{S}_{ot}$ at t = 0 hrs, respectively, whilst Δ S is represented by the covariance of the change in Fourier coefficients in 1 hr.

8.2.1. Results

These 'initial values' are calculated using the model field we wish to estimate, but in reality we would not know these values to such precision. Thus, in a similar fashion to the tests made using the analytical radiance field, further tests are made here using 'initial values' which are 10 and 20 % different from those calculated from the model field. R.m.s errors are calculated

for all latitudes, and also for each of the 7 latitude/season retrieval zones (described in Table 6.1), and appear in Table 8.3a and 8.3b. A third series of tests multiplies each of the three 'initial values' by a large amount (50 %) to examine the importance of each 'initial value' to the quality of the estimate. R.m.s errors for these tests appear in Table 8.4a and 8.4b.

ZONE	RUN IDENTIFIER				
	AA	BB	CC	DD	EE
1	4.32	5.71	7.00	9.38	11.28
2	4.73	8.19	10.90	16.30	18.51
3	3.72	10.21	8.26	18.63	16.14
4	3.71	10.88	6.89	19.21	14.53
5	2.22	9.91	8.46	18.99	17.06
6	3.13	5.51	7.02	10.68	12.10
7	1.52	3.61	5.21	7.50	9.60
Global	3.55	8.82	7.64	16.12	14.57

Table 8.3a R.m.s errors at 1.5 mb for the combined retrieval and analysis in degrees K. The analysis has been performed using the sequential estimation of Fourier coefficients. Details of each run are as follows: Run AA - 'initial values' calculated from the model field; Run BB - 'initial values' are 1.1 times those used in Run AA; Run CC - 'initial values' are 0.9 times those used in Run AA; Run DD - 'initial values are 1.2 times those used in Run AA; Run EE - 'initial values' are 0.8 times those used in Run AA. Errors in Runs BB to EE are large due to the extremely large errors in the 'initial values'.

Table 8.3a Kalis errors at 1.5 mb for the combined retrieval and analysis in
departures to the analysis has been performed using the sequential estimation of
Fourier coefficients. Details of each run are as follows: Run AA - 'initial values'
calculated from the model field; Run BB - 'initial values' are 1.1 times those
used in Run AA; Run CC - 'initial values' are 0.9 times those used in Run AA;
Run DD - 'initial values' are 1.2 times those used in Run AA; Run EE - 'initial
values' are 0.8 times those used in Run AA.

ZONE	RUN IDENTIFIER				
	AA	BB	CC	DD	EE
1	6.93	10.17	8.75	15.58	13.78
2	4.85	13.76	12.79	25.87	24.52
3	3.99	8.77	11.18	17.43	20.34
4	1.86	6.01	6.34	11.79	12.17
5	1.70	6.77	4.70	12.24	10.41
6	1.30	4.48	4.31	8.47	8.62
7	0.58	4.90	5.45	9.66	11.10
Global	3.37	7.94	7.86	14.71	14.74

Table 8.3b Like Table 8.3a, except r.m.s errors at 5 mb are shown.

Figure 8.4a shows the field at 1.5 mb estimated using Run AA values. A comparison with the model field (Figure 7.3a) shows that the magnitude, shape and location the large-scale features of the field are reasonably estimated, and Figure 8.4b shows that differences between model and estimated fields are generally less than 10 K. Comparison of Figure 8.4a with fields at 1.5 mb estimated by the time/space interpolation scheme (Figures 7.3b, 7.3c and 7.5) shows them to be qualitatively similar. In addition, their global r.m.s errors are similar, being 3.55 K for Run AA, and between 3.64 and 3.74 K for estimates made using time/space interpolation (see Tables 7.1a to 7.1c). However, the smaller-scale (ie 200 to 300 km) features of the model field are poorly estimated. These smaller-scale features are also not satisfactorily estimated by the time/space interpolation scheme, and hence this suggests that these features are of too small a scale to be resolved by the satellite observing pattern.

Figure 8.5a shows the field at 5 mb estimated using Run AA values. Comparison with the model field (Figure 7.1c) reveals that, in general, the

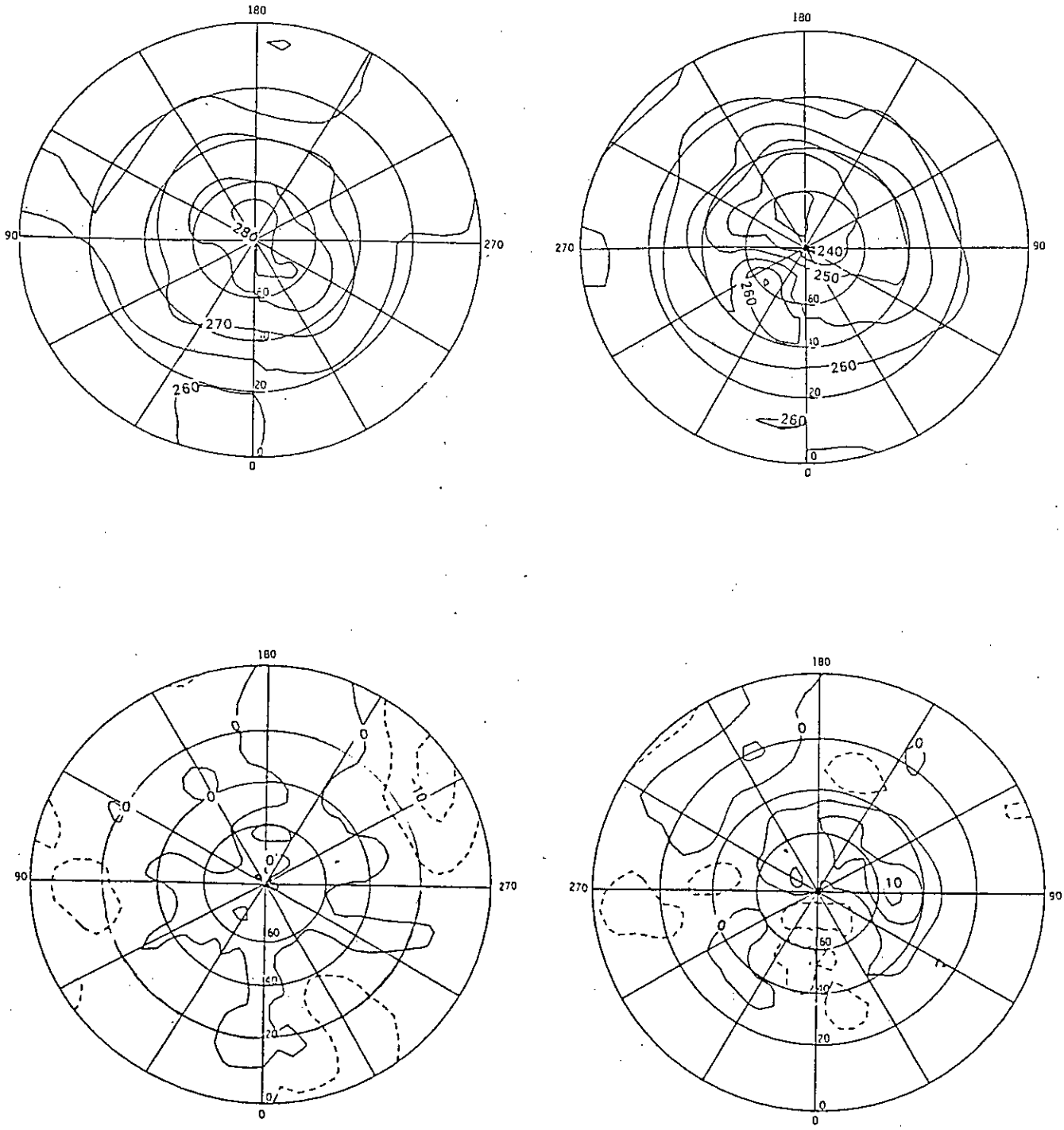


Figure 8.4 The Met Office model field on 18/1/87 at 1.5 mb analysed at t = 12 hrs using the sequential estimation of Fourier field coefficients. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing is 5 K. Negative values are dashed. a) analysed field obtained using initial values calculated from the model field. b) difference between model field (Figure 7.3a) and field in a).

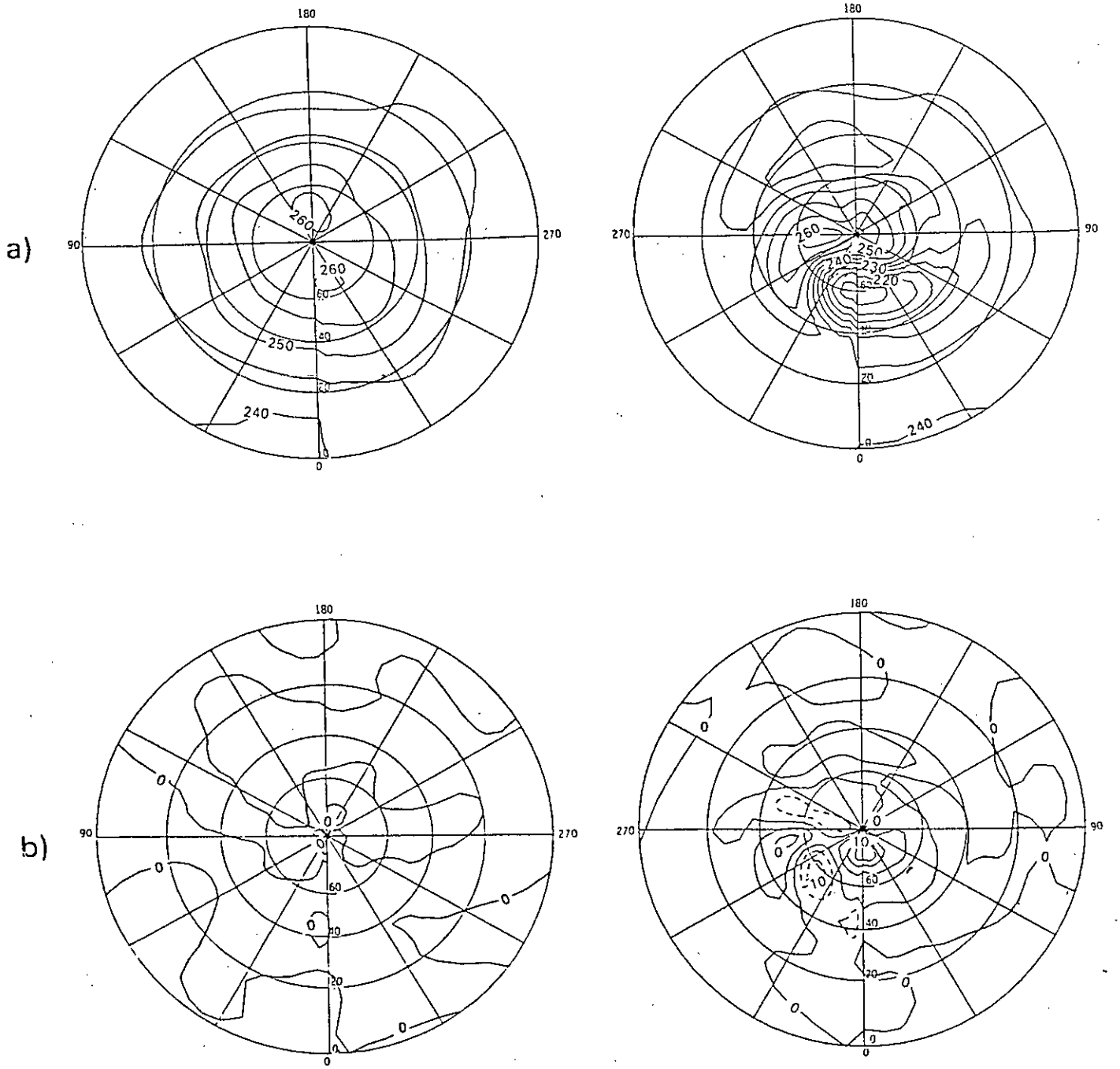


Figure 8.5 The Met Office model field on 18/1/87 at 5 mb analysed at t = 12 hrs using the sequential estimation of Fourier field coefficients. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing is 5 K. Negative values are dashed. a) analysed field obtained using initial values calculated from the model field. b) difference between model field (Figure 7.1c) and field in a).

location, magnitude and shape of the major model features have been reasonably estimated. Figure 8.5b shows that largest differences between model and estimated fields occur in the region of the strong temperature gradient between the major high and major low in the northern hemisphere. These differences occur because of inaccuracies in the shape and alignment of the analysed temperature gradient. The smallest r.m.s error of a corresponding estimate made using the time/space interpolation scheme is 2.28 K. This estimate, made with a time radius of 12 hrs and a distance radius of 1000 km, is shown in Figure 7.6e; the strong gradient between the major model high and low in the northern hemisphere has been accurately reproduced. It is thus encouraging that the sequentially estimated field (Figure 8.5a) is also able to reproduce the strength of this gradient. The r.m.s error of this estimate is higher than the error of the time/space interpolated estimate because of inaccuracies in the shape and alignment of the analysed temperature gradient. This suggests that the sequential estimation scheme is potentially as well equipped as the time/space interpolation scheme to estimate these strong gradients. However, further research, investigating chiefly the assumptions made about the time evolution of the Fourier coefficients, is essential to discover whether the inaccuracies present in the northern hemisphere estimated field can be reduced.

As discussed in Chapter 7, errors of estimates made using the time/space interpolation scheme increase as the distance radius is increased because large distance radii oversmooth 'true' field features. In certain situations the time/space interpolation scheme may have to use a large distance radius (for example, because the quantity to be estimated has a large diurnal variation) and in these cases it may be advantageous to analyse the field using the sequential estimation scheme instead.

Tables 8.3a and 8.3b show that changing Run AA 'initial values' by 10 % (Runs BB and CC) more than doubles the global r.m.s error, whilst changing 'initial values' by 20 % (Runs DD and EE) increases the global r.m.s error approximately fourfold (examination of Table 8.1 shows that corresponding changes to 'initial values' causes errors of estimates of the analytical radiance field to rise by only 10 and 30 %, respectively). Qualitatively, such estimated fields at 1.5 mb bear little resemblance to the model field. For example, Figure 8.6a shows the Run BB field. In the southern hemisphere the zonal flow of the model field has been contaminated by unrealistic small-scale structure, whilst in the northern hemisphere, values near the pole have been adequately reproduced, but the hemisphere model maximum near 60°N, 50°W has been overestimated by about 10 K. Similarly, the estimated field for Run CC (Figure 8.6b) contains unrealistic small-scale structure, but here most field values are underestimated (by up to 25 K), as revealed by Figure 8.6c, which shows the difference between model and estimated fields. Estimates for Runs DD and EE (not shown) are even more dissimilar to the model field than are Run BB and CC estimates. Differences between estimated and model fields show Run DD to be overestimated by up to 35 K (Figure 8.6d) and Run EE to be underestimated by up to 45 K (Figure 8.6e). These results suggest that the estimates here are even more strongly influenced by the choice of the initial value of $\mathbf{x}_{0t}$ than are estimates of the analytical field. Possible reasons for this are discussed in Section 8.3.

Examination of the Run BB estimate at 5 mb (Figure 8.7a) shows that qualitatively a lot of the model field structure is reproduced. The flow in the southern hemisphere is mainly zonal, whilst in the northern hemisphere a major low and a major high, separated by a strong temperature gradient, appear in approximately the same position as corresponding features in the model field. However, the bifurcation of the model high is much greater than in the model

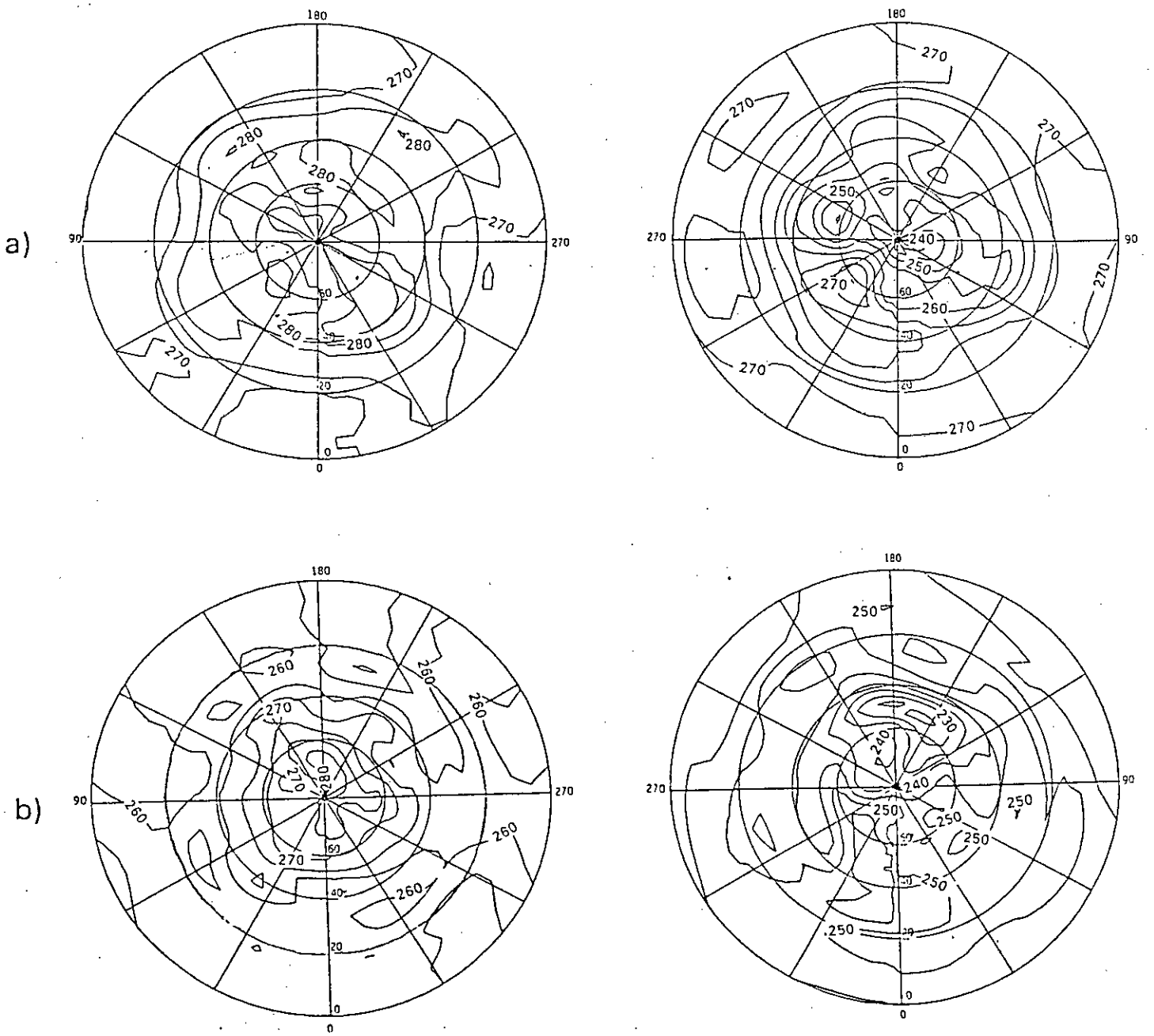
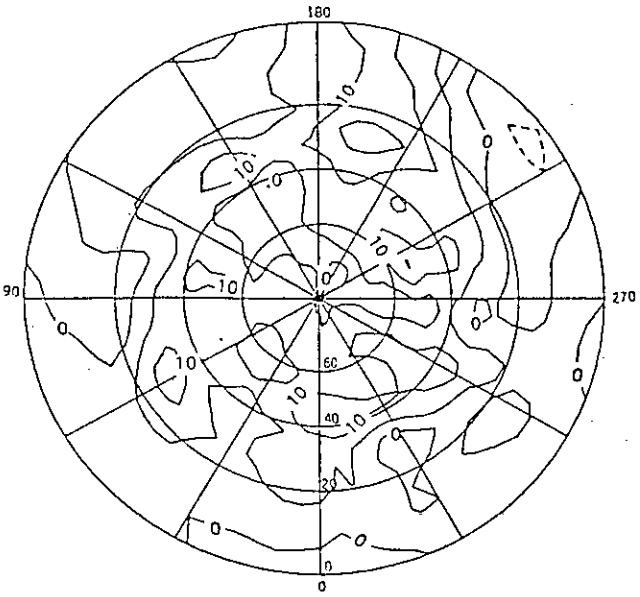
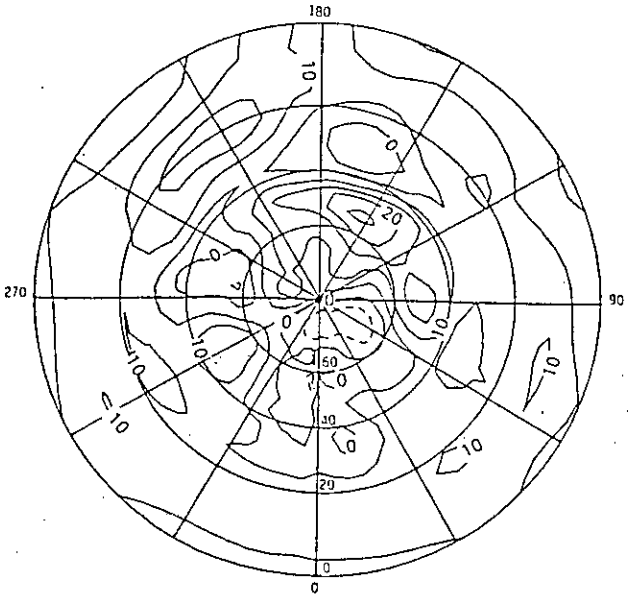


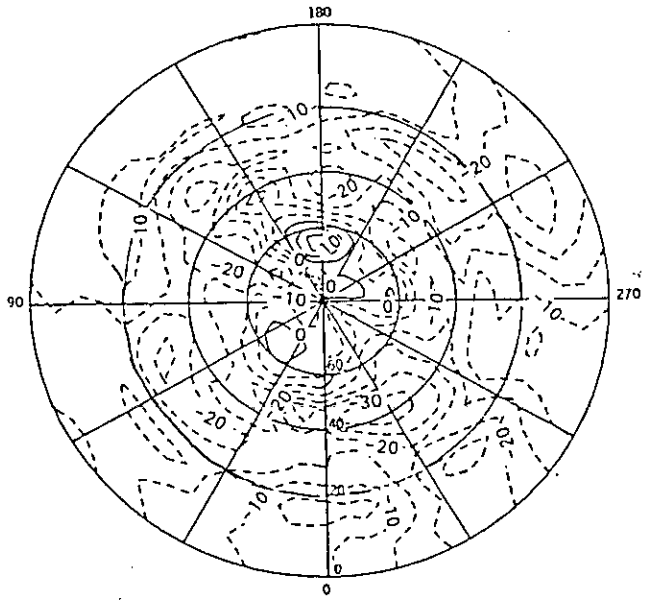
Figure 8.6 The Met Office model field on 18/1/87 at 1.5 mb analysed at t = 12 hrs using the sequential estimation of Fourier field coefficients. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing is 5 K. Negative values are dashed. a) analysed field obtained using initial values 1.1 times those calculated from the model field. b) as a), except initial values are 0.9 times those calculated from the model field.

c)





d)



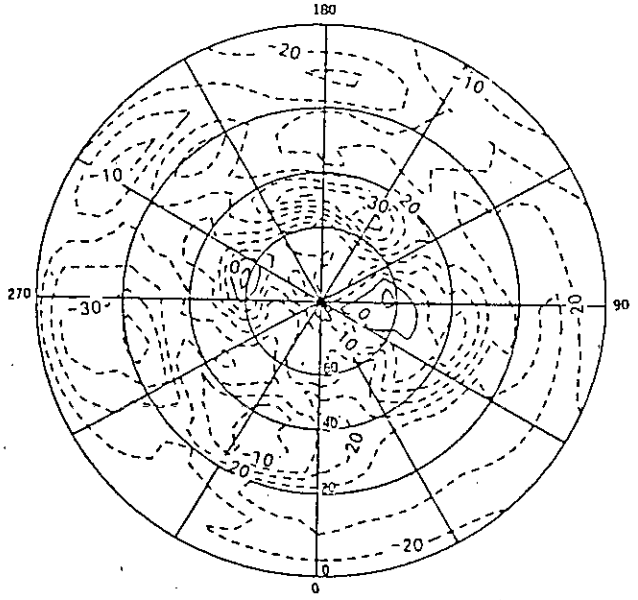


Figure 8.6 (cont.) c) difference between model field (Figure 7.3a) and field in b). d) as c), except initial values are 1.2 times calculated from the model field.

e)

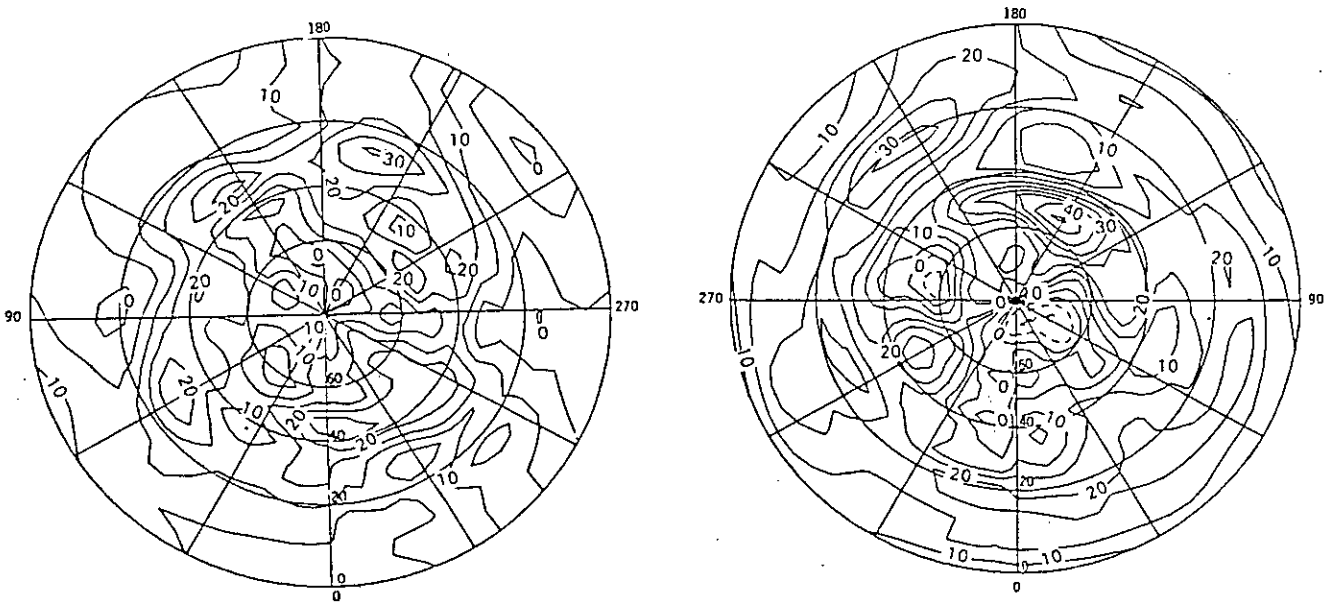


Figure 8.6 (cont.) e) as a), except initial values are 0.8 times those calculated from the model field.

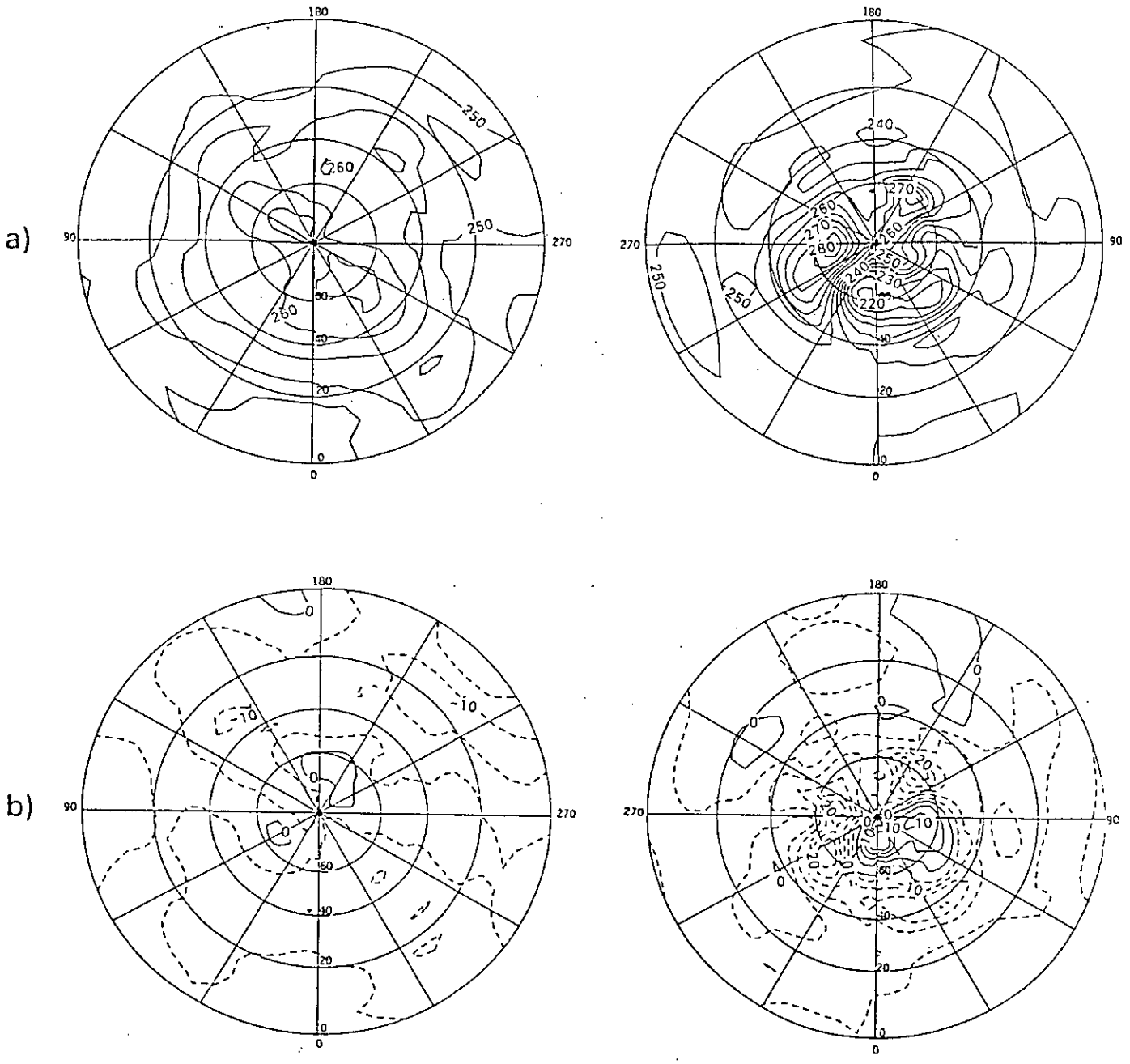


Figure 8.7 The Met Office model field on 18/1/87 at 5 mb analysed at t = 12 hrs using the sequential estimation of Fourier field components. The northern hemisphere field is plotted on the right and the southern hemisphere field is on the left. The contour spacing is 5 K. Negative values are dashed. a) analysed field obtained using initial values 1.1 times those calculated from the model field. b) difference between model field (Figure 7.1c) and field in a).

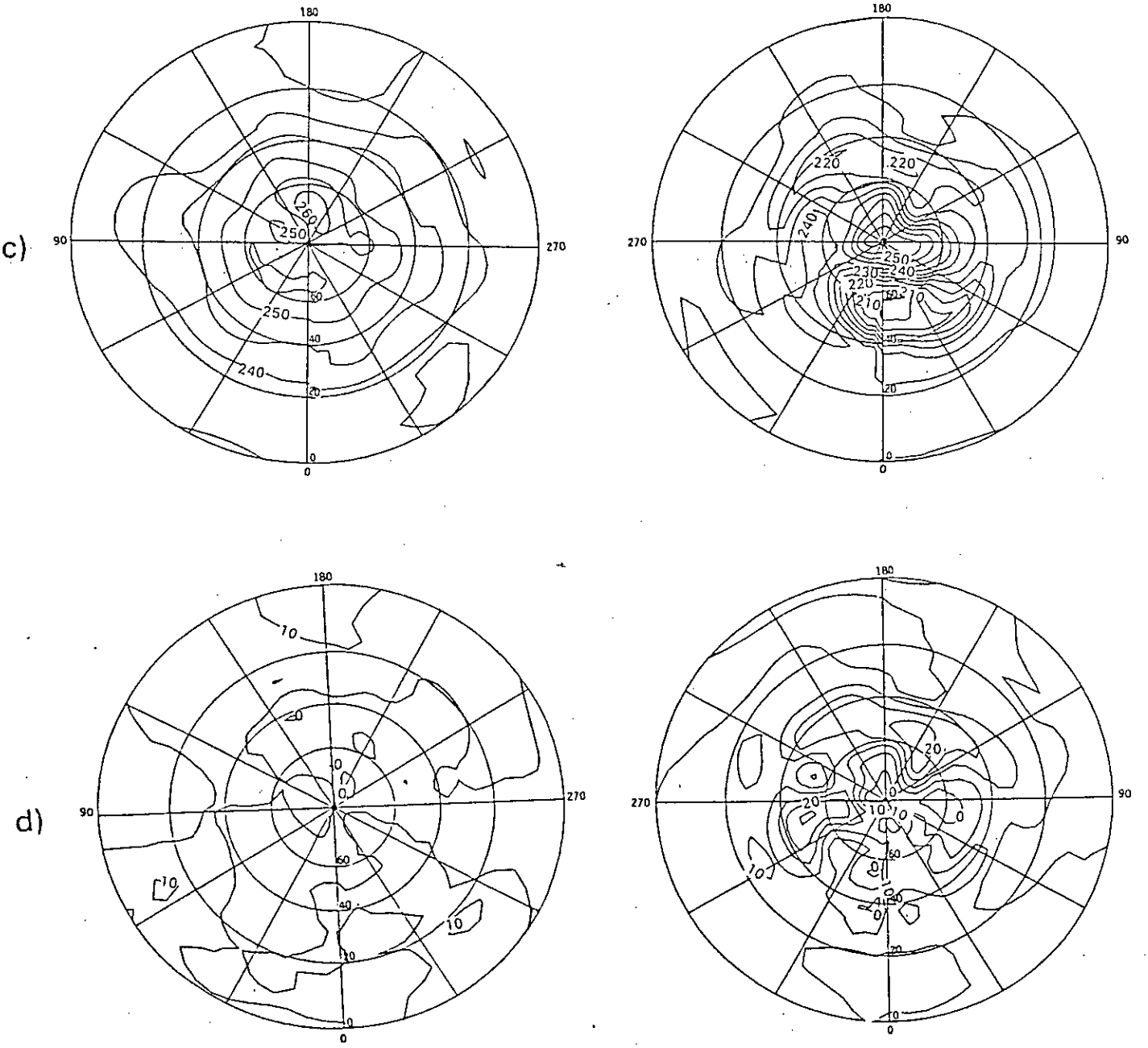


Figure 8.7 (cont.) c) as a), except the initial values are 0.9 times those calculated from the model field. d) difference between model field (Figure 7.1c) and field in c).

e)

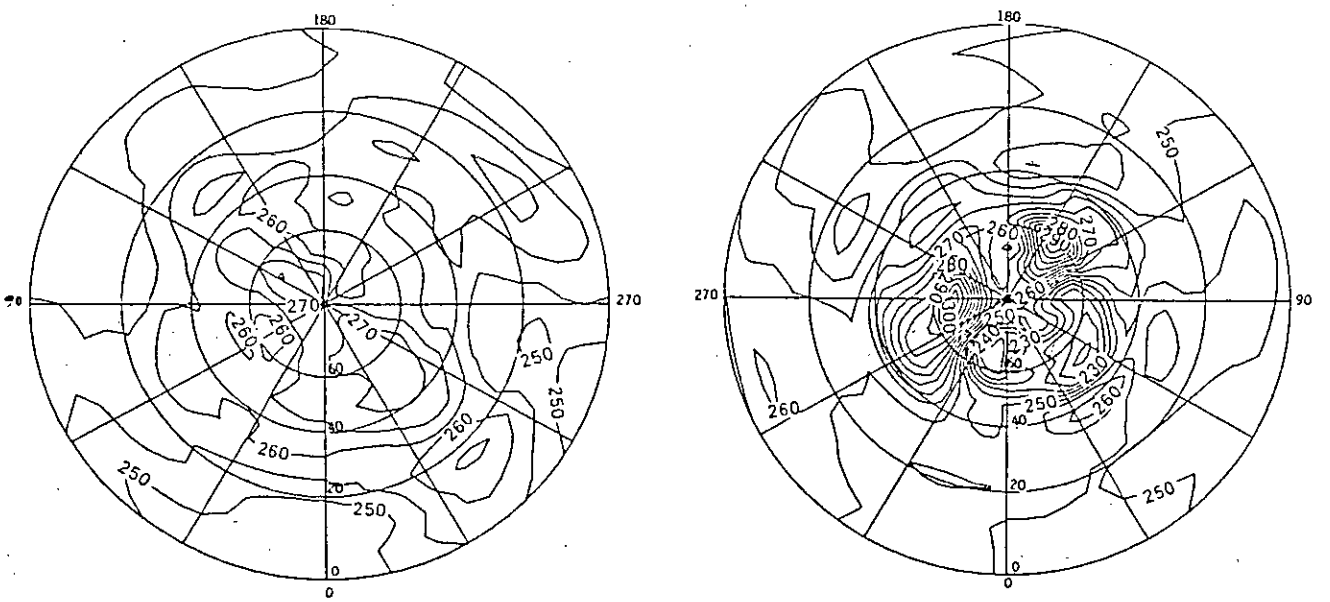


Figure 8.7 (cont.) e) as a), except the initial values are 1.2 times those d from the calculated from the model field.

field, and the field contains a considerable amount of small-scale structure that does not exist in the model field. Differences between the two fields are shown in Figure 8.7b; the field is overestimated everywhere, apart from a small region in the northern hemisphere where the field is affected by the sudden warming. Run CC (Figure 8.7c) also qualitatively reproduces many of the model features; the zonal flow in the southern hemisphere, and the major high and low in the northern hemisphere are clearly seen. However, both major high and major low are underestimated by between 10 to 25 K (see Figure 8.7d) and the major high is much smaller in area than the model major high.

Changing 'initial values' by 20 % results in estimates which are even more different from the Run AA estimates than are the Run BB and CC estimates. Figure 8.7e shows the Run DD estimate at 5mb. In very broad terms, the zonal flow in the southern hemisphere and the major high and low in the northern hemisphere have been reproduced, but the field contains a large amount of unrealistic small-scale features, and most model features are overestimated by up to 20 K (in the southern hemisphere) and up to 40 K (in the northern hemisphere). The estimate for Run EE (not shown) is broadly similar to that of Run DD, except that most model features are underestimated by up to 20 K (southern hemisphere) and 45 K (northern hemisphere).

These results show that when 'initial values' are multiplied by 1.1 or 1.2, the estimated field values are generally higher than the model field, whilst when the 'initial values' are multiplied by 0.8 or 0.9, the estimated values are usually less than the model values. This suggests that the quality of the analysis depends on the value of x_{ot} at t = 0 hrs, but very little on the value of ΔS or S_{ot} at t = 0 hrs. To test this, estimates with each of these values changed from Run AA values by 50 % were performed. Results appear in Tables 8.4a and 8.4b.

ZONE	RUN IDENTIFIER					
	FF	GG	HH	II	JJ	KK
1	22.86	24.45	4.33	4.32	4.29	4.40
2	40.78	44.11	4.70	4.74	4.84	4.50
3	43.66	41.70	3.76	3.61	3.66	3.82
4	43.66	39.58	3.75	3.63	3.66	3.79
5	45.42	44.07	2.25	2.17	2.22	2.22
6	26.62	28.42	3.14	3.12	3.12	3.15
7	20.34	21.88	1.53	1.49	1.56	1.46
Global	37.40	36.83	3.66	3.50	3.54	3.56

Table 8.4a R.m.s errors at 1.5 mb for the combined retrieval and analysis in degrees K. The analysis has been performed using sequential estimation of Fourier components. In each test the 'initial values' used are as Run AA, but with the following exceptions: Run FF – x_{ot} at t=0 hrs is multiplied by 1.5; Run GG – x_{ot} at t=0 hrs is multiplied by 0.5; Run HH – S_{ot} at t=0 hrs is multiplied by 1.5; Run II – S_{ot} at t=0 hrs is multiplied by 0.5; Run JJ – Δ S is multiplied by 1.5; Run KK – Δ S is multiplied by 0.5.

ZONE	RUN IDENTIFIER					
	FF	GG	HH	II	JJ	KK
1	33.97	31.95	6.90	6.97	6.75	7.03
2	62.51	61.58	4.88	4.77	4.94	4.76
3	45.00	47.47	4.04	3.91	4.28	3.83
4	29.29	29.63	1.87	1.82	1.85	1.86
5	28.96	26.77	1.73	1.64	1.71	1.70
6	21.04	20.77	1.31	1.29	1.26	1.33
7	25.58	25.85	0.59	0.57	0.54	0.63
Global	35.89	35.77	3.39	3.35	3.38	3.37

Table 8.4b Like Table 8.4a, except r.m.s errors at 5 mb are shown.

These results confirm that changing S_{ot} or Δ S has little effect on the quality of the estimate. R.m.s errors of Runs HH to KK are close to

corresponding Run AA errors, and the fields (not shown) are qualitatively similar. On the other hand, changing $\mathbf{x}_{ot}$ clearly influences the quality of the estimate. The r.m.s error of Runs FF and GG is over ten times the error for Run AA. Estimated field values for Run FF (not shown) are up to 100 K higher than model field values, whilst estimated values for Run GG (not shown) are up to 100 K lower.

8.3. Conclusions and Pointers to Future Research

In the tests described in Section 8.1, it has been shown that sequential estimation produces a qualitatively good estimate of an idealised stratospheric radiance field. It is particularly encouraging that the shape, location and magnitude of fast-moving large amplitude waves in the northern hemisphere middle latitudes have been reasonably estimated (ie to within about 4 r.u, compared to a variation over the hemisphere of 24 r.u). Further tests in Section 8.2 reveal that the sequential estimation technique can also produce a qualitatively good estimate of a model temperature field which has been affected by a sudden warming. Comparison of fields at 5 mb show that the sequential estimation scheme can estimate the large temperature gradients associated with the sudden warming better than a time/space interpolation scheme which uses a distance radius of 2000 km, but that time/space interpolated estimates which use distance radii smaller than 2000 km reproduce the shape and location of such a gradient better.

The best sequentially estimated analyses were made using 'initial values' - $\mathbf{x}_{ot}$ and $\mathbf{S}_{ot}$ at t = 0 hrs, and Δ S - which were calculated from the observations we were trying to analyse. Changing the calculated value of $\mathbf{x}_{ot}$ by 50 % led to a large decrease in the quality of the estimates of both analytical and model fields. However, corresponding changes to the values of $\mathbf{S}_{ot}$ and Δ S resulted in very little change in the analysis quality. This is presumably because, since a

random walk model is used, the evolving error covariance becomes much larger than S_{ot} and ΔS. Thus the importance of the selected values of S_{ot} and ΔS is diminished. We chose a change of 50 % as this value clearly shows the impact of the change in each of the initial values on the analysis. However, in practice the initial estimates of the mean Fourier field coefficients and its covariance would probably be within 10 or 20 % of the actual values, especially if they were calculated from the previous day's analysis, and hence further tests of sequential estimation were made with 'initial values' 10 and 20 % different from those values calculated from the field itself. Estimates of the analytical field made with such values are qualitatively only slightly poorer than those made with 'initial values' which were calculated from the model field. However, corresponding changes in the 'initial values' caused the global r.m.s errors of estimates made of the model field to increase by about 100 % (for a 10 % change in 'initial values') and by about 300 % (for a 20 % change in 'initial values').

The possible reasons why the increase in r.m.s error is much greater for model field estimates are related to the way we estimate the time evolution of Fourier field coefficients. Here we have used a random walk, in which the 'first guess' estimate at time t is assumed to be equal to the optimal estimate at time t-1. While this is reasonable when the time interval between data is around 16s, it is probably a poor assumption when the time interval between data switches to about 50 mins. The greater small-scale variation in the model field means that the random walk is probably an even poorer assumption for the model field than for the analytical field. Another important difference is that estimates of the analytical field are made at 10° latitude intervals, whilst those of the model field are made at 5° intervals. As mentioned in Section 8.1.1, the interpolated observations used in the scheme are in groups: each group is separated by about 50 mins, and observations within each group are about 16 s

apart. The ratio of interpolated observations spaced at 16 s to those spaced at 50 min when performing model field estimates is thus half the corresponding ratio for analytical field estimates. If we assume that estimates made with observations 50 min apart are poor, whilst estimates made with observations 16 s apart are good, then the model field estimate thus has half as many 16 s-spaced observations to 'recover' from a poor 50 min-spaced observation than does a corresponding analytical field estimate. An illustration of this idea is Figure 8.8 (from Ledsham and Staelin, 1978) which shows that a Kalman Filter used to retrieve temperature requires 3 or 4 observations to achieve acceptable performance. By analogy, after a poor estimate made by data 50 min apart, the scheme here may require about 3 or 4 16 s-spaced observations to re-achieve acceptable performance. The analytical field utilises groups of about 10 such observations, each group spaced by about 50 min, so should recover acceptable preformance before the next 50 min-spaced observation becomes available. The model field estimate uses 16 s-spaced observations in groups of about 5 (because of the higher latitudinal resolution), and there is thus the possibility that the estimate does not re-achieve acceptable performance before the next 50 min-spaced observation becomes available. The problem will be more acute if the 'initial value' for x_{ot} is not close to the 'true' value. The hypothesis described here could be further examined in a future study by performing sequential estimates at a variety of latitudinal resolutions.

In any event, a further investigation of the way we estimate the time evolution of Fourier coefficients is required. The random walk is a first-order autoregressive process with an autoregressive coefficient of unity, but sequential estimation can also be performed using 'first guess' models which have a non-unity autoregressive coefficient, or which are higher-order autoregressive models. It would thus be profitable to test these more sophisticated estimates of the time evolution of the Fourier coefficients, to

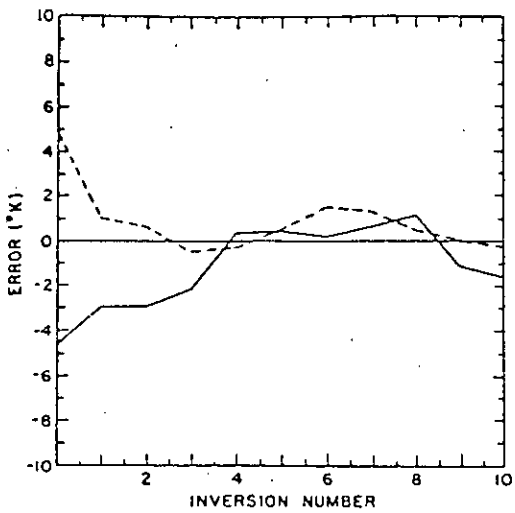


Figure 8.8 Two transient responses of a Kalman Filter at 700 mb showing initial poor performance (from Ledsham and Staelin, 1978).

assess the impact these have on the quality of the analysis.

Although the tests described in this chapter suggest that sequential estimation produces similar quality analyses to those produced by time/space interpolation only when 'initial values' are close to 'true' values, there may be situations where it is clearly advantageous to use sequential estimation. An example is the estimate of quantities which have a strong diurnal variation (e.g certain atmospheric constituents). In this example, a time/space interpolation of this field using a large time radius would lead to a large smoothing of the 'true' field, whilst use of a small time search radius would mean that at certain gridpoints there are no observations within the search radii. The sequential estimation method, on the other hand, would not have this problem and could produce a good estimate of the field, provided the model used to estimate the time evolution of the Fourier field coefficients was suitable.

CHAPTER 9

CONCLUSIONS

The aim of this thesis is to test schemes to retrieve and analyse temperature from satellite observations. To summarise: tests of a retrieval/analysis scheme made using a field calculated by a numerical model show that errors of the regression retrieval scheme are influenced by the first guess estimate, and are higher within the region affected by a sudden warming; on the other hand, tests made using both idealised and model fields show that the time/space interpolation scheme satisfactorily estimates stratospheric fields; the sequential estimation technique can produce adequate analyses of idealised and model fields, but these analyses are not as good as corresponding estimates made by the time/space interpolation method.

Temperature retrieval errors were highest within the region of a sudden warming. This is because the vertical temperature structure in the sudden warming was too small to be 'observed' by the satellite instrument, and also because the data used to calculate the regression coefficients were inevitably not representative of sudden warming conditions. The former errors, due to small-scale vertical structure, are inherent in the method of observation rather than the method of temperature retrieval. In contrast, the latter errors are caused by dissimilarities between the sonde dataset and sudden warming conditions. These errors can be described as 'first guess errors' since the mean of the dataset can be thought of as a first guess estimate of the retrieved temperature. Retrieval errors outwith the region of the sudden warming are also often caused by high first guess errors. Ideally one would wish to reduce such errors by using a first guess more appropriate to the atmospheric conditions we are trying to estimate. One possibility is to adopt techniques used in tropospheric retrievals. We could, for example, use 'stratified climatology' (eg

Uddstrom and Wark, 1985), in which radiances are used to 'point' to a class of atmospheres to which a profile probably belongs. Regression coefficients are then obtained from the statistics of the class. Another related technique is that of the library search (eg Chedin et al, 1985). Whilst these techniques have been used successfully in the troposphere, the shortage of stratospheric sonde observations limits their usefulness. For example, there may not be enough rocketsonde observations of a sudden warming to be able to define an atmosphere of that class. A more promising method of reducing first guess errors is to use a first guess based on a model forecast rather than on climatology, and future research may involve the development of a retrieval scheme which uses for the *a priori* information the output at the previous analysis time from a numerical weather prediction model (this has been tried in the troposphere by eg Susskind et al, 1984). At present such models do not produce operational forecasts for the stratosphere, but it is expected that a number of such models will be extended up to the stratosphere in the near future.

The time/space interpolation scheme was initially tested using idealised radiance fields and tests showed that most features of these fields were satisfactorily reproduced, including those of high temporal variability such as a 5-day planetary wave. It was found that reducing the size of distance radius in the scheme (down to certain limit) decreased analysis errors, but that varying the size of time radius altered the error values only very slightly. Further tests of the analysis scheme were made using a field calculated by a numerical model, which provides a more realistic representation of the observed stratosphere than do idealised fields. Conclusions drawn from the preliminary tests were found to be still valid when used to evaluate the results of tests made using the model field. To provide a stringent test of the scheme we used a field affected by a sudden warming. The region of the field affected by the

sudden warming was satisfactorily estimated, but small scale structure present in the upper stratospheric and lower mesospheric model fields (ie 0.2 and 1.5 mb) was not resolved by the analysis scheme. Hence, we conclude that users of analyses should be aware that the analysis scheme is unable to resolve such small-scales. However, we expect that most other stratospheric field features, including events such as sudden warmings, will be satisfactorily estimated by the time/space interpolation scheme.

It has been demonstrated that the sequential estimation of Fourier field components produces satisfactory analyses of an idealised radiance field. In particular, it is encouraging that fast-moving large amplitude waves in the northern hemisphere middle atmosphere were adequately estimated. However, corresponding analyses made using the time/space interpolation method were slightly better than the sequentially estimated analyses. In addition, the sequential estimation method method can satisfactorily estimate a model stratospheric field which is affected by a sudden warming. In particular, it is encouraging that the scheme can estimate the strong temperature gradient associated with the sudden warming. However, in general the sequential estimation scheme produces noticeably poorer estimates than the time/space interpolation method, especially when 'initial values' are different to those calculated from the field we are trying to analyse. Such poor estimates may be a result of the high number of latitudes at which estimates of the model field are made, or because the random walk model used to estimate the temporal evolution of the Fourier coefficients is not appropriate to the model field. Further research should thus include testing the effect of varying the number of latitudes at which estimates are made, and examining the effect of more sophisticated methods of estimating the time evolution of the Fourier coefficients. Another important area of future research is the development of a scheme which estimates spherical harmonics instead of Fourier coefficients.

Comparison of the time/space interpolation and sequential estimation schemes reveals that the sequential estimation method is capable of providing a good estimate of major model highs and lows, and there may be situations where this will lead to better analyses than those made by time/space interpolation. If the method of modelling the time evolution of the Fourier coefficients is adequate, then one would expect the error in the estimate of the magnitude of such highs and lows to be small, but on the other hand, more care is required when using the time/space interpolation method. This scheme produces essentially a weighted average of a set of observations which smooths out highs and lows, and it is thus important to choose search radii small enough to make this effect negligible. However, an advantage of time/space interpolation over sequential estimation is that it is easier to understand and is more versatile. Sequential estimation requires a lot of preliminary calculations, such as the interpolation of observations to grid latitudes, estimates of the first guess and error covariance of the Fourier coefficients at t = 0 hrs, and an estimate of how this covariance will increase per unit time. Time/space interpolation, on the other hand, does not require knowledge of such statistics, and is thus versatile enough to to analyse any sort of satellite data (eg composition measurements, temperature, radiance) easily and effectively.

I. Proofs to Optimal Solutions of the Retrieval Problem

I.Combination of Observations

Suppose we have two measurements, x_1 and x_2, of a quantity x, made by two different experimental techniques. We wish to combine the two observations to produce the optimal estimate of x. We form a combined estimate, $\hat{x}$, thus

$$\hat{x} = A\,x_1 + B\,x_2 \qquad \text{I..1}$$

where A and B are diagonal matrices. Let ε be the error in $\hat{x}$. Thus

$$\begin{aligned} \varepsilon &= A\,x_1 + B\,x_2 - x \qquad \text{I..2} \\ &= A\,(x_1 - x) + B\,(x_2 - x) - (I - A - B)\,x \\ &= A\,\varepsilon_1 + B\,\varepsilon_2 - (I - A - B)\,x \end{aligned}$$

where ε_1 and ε_2 are the errors in x_1 and x_2 respectively. We must ensure that $I - A - B = 0$, because we obviously want $\varepsilon = 0$ when $\varepsilon_1 = \varepsilon_2 = 0$. Thus $B = I - A$, and thus, equation (I..2) is rewritten as

$$\varepsilon = A\,\varepsilon_1 + (I - A)\,\varepsilon_2$$

If S_1 is the covariance of ε_1 and S_2 is the covariance of ε_2, then, provided ε_1 and ε_2 are independent, this relationship implies that $\hat{S}$, the error covariance of $\hat{x}$, is

$$\hat{S} = A\,S_1\,A^T + (I - A)\,S_2\,(I - A)^T \qquad \text{I..3}$$

The best approach is to choose A so that $\hat{S}$ is a minimum. Hence we find A such that

$$0 = d\hat{S} / dA = 2\,A\,S_1 - 2\,(I - A)\,S_2$$

So

$$A = S_2 (S_1 + S_2)^{-1} \qquad \text{and} \qquad \text{I..4}$$

$$I - A = S_2 (S_1 + S_2)^{-1} S_1 S_2^{-1}$$

Substituting these expressions into equation (I..1) gives

$$\hat{x} = S_2 (S_1 + S_2)^{-1} x_1 + S_2 (S_1 + S_2)^{-1} S_1 S_2^{-1} x_2 \qquad \text{I..5}$$

Next, we rearrange the $S_2 (S_1 + S_2)^{-1}$ term using the well-known matrix identity $\beta^{-1} \alpha^{-1} = (\alpha \beta)^{-1}$, viz

$$\begin{aligned} S_2 (S_1 + S_2)^{-1} &= ((S_1 + S_2) S_2^{-1})^{-1} \qquad \text{I..6} \\ &= (S_1 S_2^{-1} + I)^{-1} \\ &= (S_1 (S_2^{-1} + S_1^{-1}))^{-1} \\ &= (S_2^{-1} + S_1^{-1})^{-1} S_1^{-1} \end{aligned}$$

Substituting this into equation (I..5) gives

$$\hat{x} = (S_2^{-1} + S_1^{-1})^{-1} (S_1^{-1} x_1 + S_2^{-1} x_2)$$

Next we wish derive an expression for the error covariance $\hat{S}$. Inserting expressions (I..4) into equation (I..3) (and assuming the covariance matrices are symmetric) gives

$$\hat{S} = S_2 (S_1 + S_2)^{-1} \{ S_1 S_2 (S_1 + S_2)^{-1} + S_1 S_2 (S_1 + S_2)^{-1} S_1 S_2^{-1} \}$$

Using the result of equation (I..6), the error covariance is rewritten as

$$\hat{S} = (S_1^{-1} + S_2^{-1})^{-1}$$

II.Rearranging Optimal Estimate and Error Covariance Equations

Here, we wish to rearrange the expression for $\hat{x}$, equation (3.12), into the form of equation (3.14). Equation (3.12) is

$$\hat{x} = (S_x^{-1} + K^T S_\varepsilon^{-1} K)^{-1} (S_x^{-1} x_o + K^T S_\varepsilon^{-1} y) \qquad \text{I..7}$$

The first bracketed term on the right hand side of equation (I..7) is rearranged using the matrix identity $(A B)^{-1} = B^{-1} A^{-1}$, viz

$$\begin{aligned}
(S_x^{-1} + K^T S_\varepsilon^{-1} K)^{-1} &= ((I + K^T S_\varepsilon^{-1} K S_x) S_x^{-1})^{-1} \\
&= S_x (I + K^T S_\varepsilon^{-1} K S_x)^{-1} \\
&= S_x ((K^T + K^T S_\varepsilon^{-1} K S_x K^T) K^{T\,-1})^{-1} \\
&= S_x K^T (K^T + K^T S_\varepsilon^{-1} K S_x K^T)^{-1} \\
&= S_x K^T (I + S_\varepsilon^{-1} K S_x K^T)^{-1} K^{T\,-1} \\
&= S_x K^T (S_\varepsilon^{-1} (S_\varepsilon + K S_x K^T))^{-1} K^{T\,-1} \\
&= S_x K^T (S_\varepsilon + K S_x K^T)^{-1} S_\varepsilon K^{T\,-1}
\end{aligned}$$

Substituting (I..8) into (I..7) gives

$$\hat{x} = S_x K^T (S_\varepsilon + K S_x K^T)^{-1} (S_\varepsilon K^{T\,-1} S_x^{-1} x_o + y) \qquad \text{I..8}$$

rewriting $S_\varepsilon K^{T\,-1} S_x^{-1} x_o$ as

$$(S_\varepsilon + K S_x K^T) K^{T\,-1} S_x^{-1} x_o - K x_o$$

and substituting in equation (I..9) gives

$$\hat{x} = x_o + S_x K^T (S_\varepsilon + K S_x K^T)^{-1} (y - K x_o) \qquad \text{I..9}$$

Next, the error covariance equation (3.13) is rearranged into the form of equation (3.15). Equation (3.13) is

$$\hat{S} = (S_x^{-1} + K^T S_\varepsilon^{-1} K)^{-1}$$

Using the result of equation (I..8), the covariance is rewritten as

$$\hat{S} = S_x K^T (S_\varepsilon + K S_x K^T)^{-1} S_\varepsilon K^{T\,-1} \qquad \text{I..10}$$

If we rewrite $S_\varepsilon K^{T\,-1}$ as

$$(S_\varepsilon + K S_x K^T) K^{T\,-1} - K S_x$$

then equation (I..11) becomes

$$\hat{S} = S_x - S_x K^T (S_\varepsilon + K S_x K^T)^{-1} K S_x$$

II. References

Al-Ajmi, D.N, Harwood, R.S and Miles, T (1985) A sudden warming in the middle atmosphere of the southern hemisphere. *Quart. J.R Meteor. Soc* **111** 359-389

Andrews, D.G (1989) Some comparisons between the middle atmosphere dynamics of the southern and northern hemispheres. *Pure Appl. Geophys.* **130** 213-232

Barnett, J.J, Harwood, R.S, Houghton, J.T, Morgan, C.G, Rodgers, C.D and Williamson, E.J (1975) Comparison between radiosonde, rocketsonde and satellite observations of atmospheric temperatures. *Quart. J.R. Meteor. Soc* **101** 423-436

Barnett, J.J and Corney, M (1985) Middle atmosphere reference model derived from satellite data. pp 47-85 in Handbook for MAP 16: Atmospheric structure and its variation in the region 20 - 120 km: Draft of a new reference middle atmosphere. Edited by K Labitzke, J.J Barnett and B Edwards.

Bergthorsson, P and Doos, P.R (1955) Numerical weather map analysis. *Tellus* **7** 329-340

Boville, B.A and Randel, W.J (1986) Observations and simulation of the variability of the stratosphere and troposphere. *J. Atmos. Sci.* **43** 3015-3034

Brownscombe, J.L, Nash, J, Vaughan, G and Rogers, C.F (1985) Solar tides in the middle atmosphere. I: description of satellite observations and comparison with theoretical calculations at equinox. *Quart. J.R. Meteor. Soc.* **111** 677-689

Chapman, W.A, Cross, M.J, Flower, D.A, Peckham, G.E and Smith, S.D (1974) A spectral analysis of global atmospheric temperature fields observed by the Selective Chopper Radiometer on the Nimbus 4 satellite during year 1970-1.

Proc. R. Soc. A338 57-76

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

Chedin, A, Scott, N.A, Wahiche, C and Moulinier, P (1985) The Improved Initialization Inversion method: a high resolution physical method for temperature retrievals from satellites of the TIROS-N series. *J. Clim. App. Meteor.* 24 128-143

Clough, S.A, Grahame, N.S and O'Neill, A (1985) Potential vorticity in the stratosphere derived using data from satellites. *Quart. J.R. Meteor. Soc.* 11 335-358

Crane, A.J (1978) Uses of satellite data in studies of stratospheric dynamics. D Phil Thesis, University of Oxford.

Cressman, G.P (1959) An operational objective analysis system. *Mon. Wea. Rev.* 87 367-374

Drayson, S. R (1966) Atmospheric transmission in the CO_2 bands between 12 μm and 18 μm. *App. Optics* 5 385-391

Duck, K.I and King, J.C (1983) Orbital mechanics for remote sounding. in Manual of Remote Sensing Vol. I pp 699-718. Edited by Colwell,R.N, Simonett, D.S and Ulaby, F.T. American Soc. of Photogrammetry.

Eyre, J.R (1987) On systematic errors in satellite sounding products and their climatological mean values. *Quart. J. Roy. Meteor. Soc.* 113 279-292

Fairlie, T.D.A and O'Neill, A (1987) Aspects of dynamics of the middle atmosphere as inferred using data from a satellite and a numerical model *Phil.*

Trans. Roy. Soc. London A323 679-692

Fairlie, T.D.A and O'Neill, A (1988) The stratospheric major warming of winter 1984/85: observational and dynamical inferences. *Quart. J.R. Meteor. Soc.* **114** 557-578

Fisher, M (1987) The Met O 20 stratosphere / mesosphere model *UK Meteorological Office Met O 20 DCTN 52*

Fleming, H.E and McMillin, L.M (1977) Atmospheric transmittance of an absorbing gas. 2: a computationally fast and accurate transmittance model for slant paths of different zenith angles. *App. Optics* **16** 1366-1370

Forbes, J.M (1984) Middle atmosphere tides *J Atmos. Terr. Phys.* **46** 1046-1067

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

Grose, W.L and O'Neill, A (1989) Comparison of data and derived quantities for the middle atmosphere of the Southern hemisphere *Pure Appl. Geophys.* **130** 195-212

Gille, J.C and Russell, J.M (1984) The Limb Infrared Monitor of the Stratosphere: experiment description, performance and results. *J. Geophys. Res.* **89** 5125-5140

Grose, W.L and Rodgers, C.D (1986) Coordinated study of the behaviour of the middle atmosphere in winter: monthly mean comparisons of satellite and radiosonde data and derived quantities. MAP Handbook Vol. 21 pp79-111

Harwood, R.S (1975) The temperature structure of the southern hemisphere stratosphere: August - October 1971 *Quart. J. Roy. Meteor. Soc.* **102** 757-770

Hirooka, T and Hirota, I (1985) Normal mode Rossby waves observed in the upper stratosphere: Part II: second anti-symmetric and symmetric modes of zonal wavenumbers 1 and 2. *J. Atmos. Sci.* **42** 536-548

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

Hirota, I (1979) Kelvin waves in the equatorial middle stratosphere observed by the Nimbus 5 SCR. *J. Atmos. Sci.* **36** 217-222

Hirota, I (1980) Observational evidence of the semiannual oscillation in the tropical middle atmosphere - a review. *Pure Appl. Geophys.* **118** 217-238.

Hirota, I (1984) Climatology of gravity waves in the middle atmosphere. *J. Atmos. Terr. Phys.* **46** 767-773

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

Houghton, J.T (1972) The Selective Chopper Radiometer on Nimbus 4 *Bull. Amer. Meteor. Soc.* **53** 27-28

Houghton, J.T and Smith, S.D (1970) Remote sensing of atmospheric temperature from satellites. *Proc. Roy. Soc. London* **A320** 23-33

Kalman, R.D (1960) A new approach to linear filtering and prediction problems *J. Basic Eng.* **82** 35-45

Kalman, R.D and L Bucy (1961) New results in linear filtering and prediction theory *J. Basic Eng.* **83** 95-108

Kaplan, J.D (1959) Inference of atmospheric structure from remote radiation measurements. *J. Opt. Soc. Amer.* **49** 1004–1007

Koehler, T.L (1989) Limb correction effects on TIROS–N Microwave Sounding Unit observations. *J. App. Meteor.* **28** 807–817

Labitzke, K and Goretzki, B (1982) A catalogue of dynamic parameters describing the variability of the middle stratosphere during the northern winters. MAP Handbook No 5 188pp.

Ledsham, W.H and Staelin, D.H (1978) An extended Kalman–Bucy filter for atmospheric temperature profile retrieval with a passive microwave sounder. *J. App. Meteor.* **17** 1023–1033

Le Marshall, J.F and Schreiner, A.J (1985) Limb effects in satellite temperature sounding. *J. Clim. App. Meteor.* **24** 287–290

Lindzen, R.S (1979) Atmospheric tides *Annu. Rev. Earth Planet. Sci.* 7 199–225

London, J (1980) in Proceedings of the NATO Advanced Institute on atmospheric ozone (Portugal). US Dept. of Transportation, FAA – Washington, D.C, USA – No FAA–EE–80–20.

Lorenc, A.C (1981) A global three–dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.* **109** 701–721

McIntyre, M.E and Palmer, T.N (1983) Breaing planetary waves in the stratosphere *Nature* **305** 593–600

McIntyre, M.E and Palmer, T.N (1984) The 'surf zone' in the stratosphere *J. Atmos. Terr. Phys.* **46** 825–849

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

McMillin, L.M and Fleming, H.E (1976) Atmospheric transmittance of an absorbing gas: a computationally fast and accurate transmittance model for absorbing gases with constant mixing ratios in inhomogeneous atmospheres *App. Optics* **15** 358–363

McMillin, L.M, Fleming, H.E and Hill, M.L (1979) Atmospheric transmittance of an absorbing gas. 3: a computationally fast and accurate transmittance model for absorbing gases with variable mixing ratios. *App. Optics* **18** 1600–1606

McMillin, L.M, Gray, D.G, Drahos, H.F, Chalfant, M.W and Novak, C.S (1983) Improvements in the accuracy of operational satellite soundings. *J. Clim. App. Meteor.* **22** 1948–1955

Madden, R.E (1978) Further evidence of travelling planetary waves *J. Atmos. Sci.* **35** 1605–1618

Mechoso, C.R, Hartmann, D.L and Farrara, J.D (1985) Climatology and interannual variability of wave, mean-flow interaction in the Southern hemisphere. *J. Atmos. Sci.* **42** 2189–2206

Miller, D.E, Brownscombe, J.L, Carruthers, G.P, Pick, D.R and Stewart, K.H (1980) Operational temperature sounding of the stratosphere. *Phil. Trans. Roy. Soc.* **A296** 65–71

Nash, J and Brownscombe, J.L (1983) Validation of the Stratospheric Sounding Unit. *Adv. Space Res.* **2** No 6 59–62

O'Neill, A and Pope, V.D (1988) Simulations of linear and nonlinear disturbances in the stratosphere *Quart. J. Roy. Meteor. Soc.* **114** 1063–1110

Peckham, G (1974) The information content of remote measurements of atmospheric temperature by satellite infrared radiometry and optimum radiometer configurations *Quart. J. Roy. Meteor. Soc.* **100** 406–419

Pick, D.R and Barwell, B.R (1978) Analysis of the balloon flight of 17 March 1976 which measured atmospheric transmission using a pressure modulator equivalent to the high pressure channel of the Meteorological Office Stratospheric Sounding Unit. Met O 19 Branch Memo 45

Pick, D.R and Brownscombe, J.L (1979) The SSU: performance and products. *Met O 19 Tech Memo No. 52, UK Meteorological Office, Bracknell*

Pick, D.R and Brownscombe, J.L (1981) Early results based on the stratospheric channels of TOVS on the TIROS-N series of operational satellites. *Adv. Space Res.* **1** 247–260

Pick, D.R, Warner, D.E, Kent, P and Carter, R.D (1976) Stratospheric Sounding Unit (SSU) for TIROS-N: radiometric and spectroscopic performance of the D1 development model. SSU Technical communication No 98

Rodgers, C. D (1976a) Evidence for the five-day wave in the upper stratosphere *J. Atmos. Sci.* **33** 710–711

Rodgers, C.D (1976b) Retrieval of atmospheric temperature and compositions from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.* **14** 609–624

Rodgers, C.D (1976c) In: *Inversion methods in Remote Sensing* (ed Deepak) 117–138. New York: Academic Press

Rodgers, C.D (1984) Coordinated study of the behaviour of the middle atmosphere in winter (PMP–1) MAP Handbook Vol. 12 154pp.

Rodgers, C.D (1987) A general error analysis for profile retrieval. unpublished

Salby, M.L, Hartmann, D.L, Bailey, P.L and Gille, J.C (1984) Evidence for equatorial Kelvin modes in Nimbus 7 LIMS *J. Atmos. Sci.* **41** 220–235

Schlatter, T.W, Branstator, G.W and Thiel, L.G (1976) Testing a global multivariate statistical objective analysis scheme with observed data *Mon. Wea. Rev.* **104** 765–783

Schmidlin, F.J (1984) Intercomparisons of temperature, density and wind measurements from *in situ* and satellite techniques. *Adv. Space Res.* **4** No 6 101–110

Schwalb, A (1978) The TIROS-N / NOAA A – G satellite series. *NOAA Technical Memo NESS 95*

Seaman, R.S (1988) Some real data tests of the interpolation accuracy of Bratseth's successive correction method. *Tellus* **40A** 173–176

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

Smith, W.L, Hayden, M.C, Wark, D.Q and McMillin, L.M (1979) The TIROS-N Operational Vertical Sounder *Bull. Amer. Meteor. Soc.* **60** 1177–1187

Susskind, J, Rosenfeld, J, Reuter, D and Chahine, M.T (1984) Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS-N. *J. Geophys. Res.* **89** 4677–4697

Smith, W.L and Woolf, H.M (1976) The use of eigenvectors of statistical covariance matrices for interpretating satellite sounding radiometer observations *J. Atmos. Sci.* **33** 1127–1140

Taylor, F.W, Houghton, J.L, Peskett, G.D, Rodgers, C.D and Williamson, E.J (1972) Radiometer for remote sounding of the upper atmosphere. *App. Optics* **11** 135–141

Uddstrom, M.J and Wark, D.Q (1985) A classification scheme for satellite temperature retrievals *J. Clim. App. Meteor.* **24** 16–29

Venne, D and Stanford, J.L (1982) An observational study of high latitude stratospheric planetary waves in winter *J. Atmos. Sci.* **39** 1026–1084

Vincent, R.A and Reid, I.M (1983) HF Doppler measurements of mesospheric gravity wave momentum fluxes *J. Atmos. Sci.* **40** 1321–1333

World Meteorological Organisation (WMO) (1985) Atmospheric Ozone 1985: World Meteorological Organisation global ozone research and monitoring project. Report 16. Vol I, 392 pp.

III. Publication

Tests of a Scheme for Regression Retrieval and Time/Space Interpolation of Stratospheric Temperature from Satellite Measurements

by D.R Jackson, R.S Harwood and E Renshaw

submitted to the Quarterly Journal of the Royal Meteorological Society

Tests of a Scheme for Regression Retrieval and Time/Space Interpolation of Stratospheric Temperature from Satellite Measurements

by

D.R Jackson[1,3], R.S Harwood[1], and E Renshaw[2]

Summary

A scheme to retrieve and analyse stratospheric temperatures from satellite measurements is tested. Because of the lack of 'ground truth' in the stratosphere, the 'true' atmosphere is represented by an atmosphere simulated by a numerical model. Simulated observations are calculated by computing the radiance that would be observed from the 'true' atmosphere by a satellite instrument. The radiances are then retrieved and analysed and the resultant analyses compared with the corresponding 'true' fields. The tests are made using output from a day when a sudden warming was present. The retrievals are made by using a multiple linear regression model which regresses radiances against Planck function. The corresponding temperatures are then analysed on a grid using a linear time/space interpolation scheme.

The retrieval scheme is seen to perform less well within the area of the sudden warming than outside it. However, this may be expected as the vertical structure within the sudden warming is generally too small to be resolved by a satellite instrument. The analysis scheme analyses the stratospheric field well, even in the area of a sudden warming. The effect of varying the distance radius used to select observations for the scheme is also discussed.

[1] - *Department of Meteorology, University of Edinburgh*

[2] - *Department of Statistics, University of Edinburgh*

[3] - Present Affiliation: *Department of Meteorology, University of Reading*

1. Introduction

Electromagnetic radiation leaving the top of the atmosphere carries information about the distribution of temperature and of the emitting gases. If measurements are made at wavelengths at which the emission is by a gas of known mixing ratio such as carbon dioxide, then some details of the temperature distribution may be deduced (Kaplan, 1959); this is a principle exploited in several remote sensing satellites. However, the process of deducing the temperature structure from measurements is not necessarily straightforward. Usually there are two aspects to the problem, although it is possible to devise procedures in which they are combined. The two aspects are: 'retrieval', in which a single temperature profile is deduced from a more-or-less instantaneous set of measurements ; and 'analysis' in which the state of the atmosphere at a given instant is deduced on a regularly spaced grid of points from the retrieved profiles which are asynoptic and distributed according to the shifting satellite orbit. Without further information the retrieval problem is under-constrained because in general an infinite number of atmospheric profiles can yield the same finite set of measurements. Moreover the analysis problem can suffer from aliasing difficulties. Aliasing occurs in all Fourier analyses of discrete data: the time period of the data imposes a limit on the highest resolvable frequency, and hence any higher frequency present will be analysed falsely within the range of the lower, resolvable, frequencies.

The aim of this paper is to evaluate the performance of a retrieval/analysis scheme for obtaining stratospheric temperatures from the TIROS Operational Vertical Sounder (TOVS) instrument (Schwalb, 1978 ; Smith et al, 1979) on the TIROS-N series of polar-orbiting satellites. TOVS comprises three sounders, of which the Stratospheric Sounding Unit (SSU) (see Miller et al (1980), and references therein) is of most relevance to stratospheric studies.

The retrieval scheme which we have tested is based on a regression model similar to that used by the UK Meteorological Office (Pick and Brownscombe, 1981), the main difference being that their scheme uses the measured radiances to give thicknesses of fairly thick layers of atmosphere, whereas ours gives the temperature profile at 31 pressure levels from 0.2 to 570 mb. Many previous tests of retrieval schemes have compared retrievals with coincident rocketsonde measurements. Nash and Brownscombe (1983) and Pick and Brownscombe (1981) tested the Stratospheric Sounding Unit (SSU) on TOVS; Barnett et al (1975) tested the Selective Chopper Radiometer (SCR) on the Nimbus 5 satellite. Whilst the chief purpose of those tests was to assess the performance of the satellite instrument, here we pay particular attention to the regression/retrieval scheme itself. We examine the representativeness of the datasets used to calculate the regression coefficients, and test the ability to retrieve temperature in various atmospheric conditions.

The time/space analysis method is that used operationally by the UK Meteorological Office, so that this paper will help users of those analyses to evaluate the confidence which can be placed in them. The analysis scheme gives each observation a time and distance weight which decreases the further the observation is from the gridpoint or analysis time. Only observations lying within a specified time and distance (called 'search radii') of the gridpoint and analysis time are used in the scheme. Most tests of the scheme are initially made using idealised fields, and then on a field simulated by a numerical model. We concentrate chiefly on the way the quality of the analysis changes when the search radii are changed.

A difficulty in testing such schemes using real observations is the absence of adequate 'ground-truth' observations. Rocketsondes observe temperature in the upper stratosphere but rocket flights are infrequent and badly spaced.

Radiosondes observe in the lower stratosphere and give better global coverage, but even so there are few observations made over the oceans or in the southern hemisphere. Tests which have been done with real data compare retrieved profiles with coincident rocketsonde measurements (Nash and Brownscombe, 1983, Pick and Brownscombe, 1981, Barnett et al, 1975). Furthermore it is generally even more difficult to make comparisons with ground truth for *fields* analysed from satellite measurements than it is for retrieved profiles. Possible difficulties with analyses have been reported by Al-Ajmi et al (1985) and by Clough et al (1985). The former paper gives evidence that there may be temporal variations too rapid for proper resolution, and the latter provides evidence that the vertical temperature structure is not always adequately resolved. Accordingly we have chosen to test the schemes in a simulation experiment which uses an atmosphere calculated in a numerical model. Simulated observations are calculated by computing the radiances which would be observed from this model atmosphere by a TOVS-like instrument, including the effects of instrumental noise. These radiances are then retrieved and analysed and the resultant analyses compared with the corresponding model fields.

Section 2 gives details of the model and of the simulation method. Section 3 describes the retrieval scheme and the method of obtaining the regression coefficients, together with the results of tests of the retrieval scheme. Section 4 contains a description of the analysis scheme. Results of tests of the analysis scheme on analytical fields also appear in Section 4, whilst results of tests made on fields calculated in a numerical model appear in Section 5. Conclusions appear in Section 6.

2. Simulation of Atmosphere and Observations

a) The Stratosphere/Mesosphere Model

The retrieval and analysis schemes are tested below in a simulation experiment which uses an atmosphere calculated in a numerical model. Although the model gives a reasonable representation of the 'true' atmosphere, it is not capable of reproducing certain phenomena, such as tides, which exist in the real stratosphere. The model used is the U.K Meteorological Office stratosphere/mesosphere multi-level model (Fisher, 1987) based on the primitive equations. These equations are solved to fourth-order accuracy in the horizontal, and to second-order accuracy in both the vertical and in time, using energy conserving 'box' type finite differences and leapfrog integration. The model utilises a regular grid in spherical coordinates with gridpoints at intervals of 5° in latitude and longitude and 33 levels between 100 and 0.001 mb, which are equally spaced in log pressure, and are approximately 2 km apart. To avoid having to represent the troposphere, a lower boundary condition is imposed near the tropopause, namely the geopotential height of the 100 mb surface specified from analysed observations.

Our simulation uses one day's output at 1 hour intervals from a run with lower boundary heights corresponding to 18 Jan. 1987. On that day a phenomenon known as a 'sudden warming' was present in the modelled northern hemisphere. Such warmings cause large and rapid changes to the temperature structure of the stratosphere - there may be temperature rises of the order of 50 K over a few days. This phenomenon provides the most stringent circumstances for testing both the retrieval and analysis schemes.

b) Satellite Observation Pattern

The TIROS-N satellite views the atmosphere by scanning from one side of the vertical to the other at 8 scan angles. Adjacent observations are then averaged in blocks of four so that the 16 observations of two successive scans are combined to give four 'superobservations' at effective angles of -30, -10, 10 and 30 degrees from the vertical. Figure 1 shows the superobservations made by a TIROS-N series satellite (NOAA-7) in a 24 hour period. There is almost global data coverage, though some areas in the subtropics are free of observations. Model temperatures are linearly interpolated in time and space to these observation points, and the interpolated temperatures are used to calculate radiances that the satellite would 'observe'. In these tests no data loss due to calibration sequences is assumed.[1]

c) Radiative Transfer

As mentioned in the Introduction, the radiance measured by the satellite is a function of atmospheric temperature and of the distribution of the emitting gas. Radiation transfer theory on which temperature sounding is based (Kaplan, 1959 ; Houghton and Smith, 1970) relates the spectral radiance, R_ν, of an instrument channel centred at wavenumber ν, to the Planck function, and hence temperature. With negligible transmission from the Earth's surface this theory gives

$$R_\nu = \int_0^\infty K(y)\, B_\nu(T)\, dy . \qquad (1)$$

Here y is a vertical coordinate given by $y = -\ln(p/p_o)$ where p is pressure and p_o is a reference pressure, $K(y)$ is the channel weighting function, and $B_\nu(T)$ is the Planck function at temperature T given by

[1] Fig. 1 in here

$$B_\nu(T) = c_1 \nu^3 / [\exp(c_2 \nu / T) - 1] , \qquad (2)$$

where $c_1 = 1.19096 \times 10^{-5}$ mW m^{-2} cm^4 ster^{-1} and $c_2 = 1.43879$ cm K.

The weighting functions depend upon the angle of view. However, to simplify the calculations we have simulated all radiances using the weighting function appropriate to the vertical view. There is thus the possibility of an extra source of error in the real case arising from a misrepresentation of the radiative transfer for the slant path which is not simulated in the present study. The transmission profile, and hence weighting function, for each channel was calculated using a numerical technique developed by McMillin and Fleming (1976): first, transmission profiles are calculated for a small number of representative and extreme atmospheres using the line-by-line method (Drayson, 1966), and then these pre-computed profiles are interpolated to any arbitrary temperature profile. 'Observed' radiances can thus be calculated by evaluating equation (1) and then adding a randomly generated number to simulate the radiometric noise of the instrument.

The largest contribution to the radiance comes from pressure levels close to the peak of the weighting function. Indeed, channel radiance may be considered to give a measure of the temperature of a layer 10 - 15 km thick situated about the peak of the weighting function.[2] The weighting functions for the 8 channels used in the retrieval scheme are shown in Figure 2. Table 1 shows the pressure level at each channel's weighting function peak and its central wavenumber ν.[3]

[2] Fig. 2 in here

[3] Table 1 in here

3. Retrieval

a) The Regression Model

Planck function, and hence temperature, is retrieved from radiance measurements by regression. The temperature is calculated from the Planck function under the assumption that the Planck function has been calculated at a reference wavenumber U, which here is chosen to be 668 cm^{-1}. The regression model assumes that Planck function is linearly related to the radiances, which have also been standardised to the reference wavenumber U. This is done because the standardised radiances are more linearly related to the Planck function at 668 cm^{-1} than are 'observed' radiances. The standardised radiance, X_j, of channel j is easily calculated from R_j, the measured radiance of channel j via

$$X_j = c_1 U^3 / [\exp (c_2 U / r) - 1] , \qquad (3)$$

where $r = c_2 \upsilon / \ln (1 + c_1 \upsilon^3 / R_j)$.

The estimated deviation of profile Planck function, $\hat{\Delta B}_i$, from the mean is written as a linear combination of the deviation of the radiances from the mean radiance, namely

$$\hat{\Delta B}_i = a_{0,i} + \sum_{j=1}^{N} (X_j - \overline{X}_j) a_{j,i} \qquad (i = 1,....,31) . \qquad (4)$$

The $a_{0,i}$ and $a_{j,i}$ are predetermined in advance by least squares as described in Section 3b below. If there is no bias in the system then, of course, $a_{0,i}$ would be zero. It has been shown (eg Eyre, 1987) that this retrieval scheme is mathematically equivalent to other schemes which retrieve temperature by a minimum variance solution of the radiative transfer equation. The mean Planck function and the mean normalised radiance, $\overline{X}_j$, are calculated from a set of

rocketsonde temperature measurements (this dataset is also described in Section 3b). N is the total number of channels and X_j is the observed normalised radiance. Retrievals are performed at 31 pressure levels between 0.2 and 570 mb.

The retrieved Planck function, $\hat{B}_i$, for level i is then given by

$$\hat{B}_i = \overline{B}_i + \hat{\Delta B}_i \qquad (i = 1,....,31) \qquad (5)$$

where $\overline{B}_i$ is the mean Planck function for pressure level i. It is straightforward to calculate temperature from the retrieved Planck function.

b) Calculation of Regression Coefficients

The regression analysis is based on a dataset of 1200 temperature profiles. Each profile is calculated using a combination of a radiosonde measurement and a quasi-coincident rocketsonde measurement. Standardised radiances are calculated from these measurements using equations (1) and (3). The data are divided into 7 zones (Table 2) according to the latitude and season of each rocketsonde measurement, and regression coefficients are calculated for each of the zones. This is done to restrict the range of atmospheric conditions over which the regression analyses are applied.[1] Note that in using calculated radiances to determine the regression coefficients, we are implicitly assuming that the weighting functions in equation (1) are known correctly (i.e that the 'forward model' is accurate). In practice this may not be the case, leading to potential errors (see eg Nash and Brownscombe, 1981) of a type beyond the scope of the present investigation.

---

[1] Table 2 in here

Discussion of The Sonde Dataset

The dataset used to calculate the regression coefficients contains rocketsonde measurements of the *real* atmosphere. However, in this paper the retrieval and analysis schemes are tested in a simulation experiment which uses an atmosphere calculated in a numerical model. It is important that the model adequately reproduces the real atmosphere observed by the rocketsondes. This is examined by comparing the means and standard deviations of the two datasets.

We consider first the southern hemisphere, zone 5. The model field in this zone is similar to the climatology of the sonde data. Figure 3a shows means and standard deviations of sonde and model datasets in zone 5. The dashed lines show the mean of the sonde temperatures together with departures of one standard deviation. The solid lines show the corresponding statistics for the *model* dataset on 18 Jan. 1987 in this zone.[2] Both sonde and model standard deviations are low, indicating the lack of variation in the summer stratosphere. Moreover, at most pressure levels the means and standard deviations of both datasets are close to one another, implying that the model field in zone 5 is similar to that observed by rocketsondes.

Whilst sonde and model statistics are similar for zone 5, the same is not true for zones 1, 2 and 3. On 18 Jan. 1987 the model reproduces a sudden warming in these latter zones. In Table 2 'winter' is defined as the 6 month period between October and March (for the northern hemisphere) or between April and September (for the southern hemisphere). Within such a large time period only a

[2]Fig. 3 in here

small proportion (if any) of rocketsondes will observe a sudden warming, thus it is likely that sonde data in zones 1, 2 and 3 will not be representative of sudden warming conditions. Figure 3b parallels Figure 3a, except that data for zone 1 are shown. As one might expect in a winter stratosphere, both sonde and model datasets have high standard deviations. The two mean profiles differ greatly at a number of pressure levels. Between 3 and 50 mb the model mean temperature is greater than one sonde standard deviation away from the sonde mean temperature. Thus model conditions in zone 1 differ greatly from those observed by rocketsonde. This is also true in zone 2 (and to a lesser extent in zone 3). It is important to test the retrieval scheme in a sudden warming precisely because of this extreme difference between sudden warming conditions and mean sonde values. If the scheme performs well in a sudden warming, then it is reasonable to conclude that it will perform well in most conditions.

These results have important implications for the method of retrieval by regression, since the sonde dataset used to calculate the regression coefficients is used widely. As this dataset appears to be unrepresentative of sudden warming conditions, future retrieval research might profitably use datasets which are taken from a forecast model rather than from climatology. Other approaches, in the context of tropospheric retrievals, have been suggested by Uddstrom and Wark (1985) and Chedin et al (1985). The former use a 'stratified climatology' of several atmospheric classes; the radiances are used to identify from which class to take the appropriate regression coefficients (or their equivalents). The latter use a somewhat related 'library search' technique.

Testing the Regression Model

It is an assumption of the regression model that the Planck functions in the profile are linearly related to the measured radiances. This is suggested by the

form of the radiative transfer equation (1), provided the weighting functions are not temperature dependent. To test the validity of this assumption we have produced plots of Planck function at level k against radiance for channel j for a number of k,j pairs (not shown). No evidence of a need to depart from the linear model was found.

It is of interest to investigate if a certain amount of effort could be saved by using less than 8 channels in the regression. Also, it is of interest to know what the information content of each channel is (Peckham, 1974). Accordingly, we have produced scatter plots of radiances in pairs of channels (not shown) to discover to degree of correlation between them. Several channels show a high degree of association, but test of regression models using less than 8 channels in all cases produced significantly higher r.m.s errors than the 8 channel model, which has, in consequence, been used in the rest of the work here.

c) Retrieval Results

In this section we test the retrieval scheme by retrieving the profiles corresponding to the simulated observations. Temperatures are retrieved using equations (4) and (5), and biases and r.m.s retrieval errors are calculated under the assumption that the interpolated model temperature is the 'true' temperature. The standard deviation of the retrieval is also calculated. This is, of course, equal to the square root of the difference between the square of the r.m.s error and the square of the bias.

For all regression zones the r.m.s retrieval error at pressures greater than than that of the peak of the bottommost weighting function (300 mb) is generally higher than at 300 mb, and the r.m.s error at pressures less than that of the topmost weighting function peak (1.5 mb) is generally larger than the r.m.s error at 1.5 mb. This is of course to be expected, as away from the region containing

the weighting function peaks the satellite measurements provide little information about the temperature.

As anticipated, errors in the zones containing a sudden warming (zones 1 and 2) are high. Difficulties arise because of the small-scale vertical temperature structure present in a sudden warming and because the sonde measurements in these zones were made in conditions different to those in the sudden warming (see, for example, Figure 3b). [1] Figure 4a shows the r.m.s error (solid), bias (dashed) and standard deviation (dotted) profiles for zone 1. For pressures higher than 1.5 mb the error is between 2 and 6 K. The standard deviation has nearly the same values as the r.m.s errors at most levels and hence the bias is small, not exceeding 2 K. The corresponding profiles for zone 2 (not shown) are similar to those of zone 1. In zone 3 (also not shown) the standard deviation is close to the r.m.s errors but their values are smaller than in zones 1 and 2. In zone 4 the r.m.s error profile (Figure 4b) has a zig-zag behaviour. This is due to biases in the retrievals, since peaks in the error profile occur at the same pressure levels as high values in the bias profile. For example, at 7 mb the r.m.s error is 4.33 K and the bias is -4.13 K and at 100 mb the r.m.s error is 5.61 K and the bias is -5.20 K. The shape of the bias profile is closely related to the difference between the means for model and sondes and therefore seems to be related to an unrealistic structure in the modelled tropics. The low standard deviation of the sonde measurements used to calculate the regression coefficients (not shown) constrains retrievals to the mean of the sonde temperatures. R.m.s errors for zone 5 (Figure 4c) are lower than for zones 1 to 3 and the profile has a zig-zag pattern, which is also due to bias in the retrievals. However, these

[1] Fig. 4 in here

model low centred at 15 mb (observation 125) has been retrieved in the right place, but its retrieved depth is around 10 K greater than its model magnitude. The model maximum at 1.5 mb (observation 100) is well reproduced, but the nearby maximum at 0.4 mb (observation 125) is underestimated by 15 K. The retrieved temperature gradient between the highs and the major low is not as strong as in the model field, especially between 3 and 1 mb and observations 109 to 141 – differences between model and retrieved fields are greater than 20 K in places. This emphasises the difficulty in retrieving such vertical structure.

4. Time/Space Interpolation

a) The Method

In the Introduction we stated that the problem of inferring temperature from satellite measurements has two parts, namely 'retrieval' which was discussed in Section 3 and 'analysis', or 'interpolation', which forms the subject of this and the next section.

Figure 6 illustrates the problem of 'analysing' observations on a grid. The Figure shows satellite observations made on 18th Jan. 1987 between −20° and 20° latitude, −160° and −120° longitude and 1020 and 1340 GMT. The grid has a spacing of 5° in both latitude and longitude and we wish to interpolate the observations to the space/time gridpoints.[1]

In the interpolation scheme under test, each observation is given a time and distance weight. The weights decrease the further an observation is away from

[1] Fig. 6 in here

the gridpoint and analysis time. Here the weights are chosen to be linear; schemes tested using cosine and negative exponential weights produced essentially similar results. Only observations which lie within a specified time or distance of the analysis time or gridpoint (called 'search radii') are employed.

The time weight for the kth observation point within the search time interval is given by

$$w_t(k) = \begin{cases} (r_t - |t - t_o(k)|) / r_t & , |t - t_o| \le r_t \\ 0 & , |t - t_o| > r_t \end{cases} \quad (6)$$

where r_t is the search radius, t is the analysis time and t_o (k) is the time of the kth observation within the search radius. Similarly, the distance weight for the kth observation within the distance search circle is given by

$$w_d(k) = \begin{cases} (r_d - d(k)) / r_d & , d(k) \le r_d \\ 0 & , d(k) > r_d \end{cases} \quad (7)$$

where r_d is the distance search radius and d (k) is the physical distance between the gridpoint and the observation point. Using a standard geometric argument, d (k) is expressed as

$$d(k) = a \{ (\cos\rho\cos\lambda - \cos\rho_k \cos\lambda_k)^2 + (\cos\rho\sin\lambda - \cos\rho_k \sin\lambda_k)^2 + (\sin\rho - \sin\rho_k)^2 \}^{1/2}$$

where (ρ, λ) is the latitude and longitude of the gridpoint and (ρ_k, λ_k) is the latitude and longitude of the kth observation point within the search radii.

The combined weight w (k) is taken to be the product of the time and distance weights

$$w(k) = w_t(k)\, w_d(k) \quad . \tag{8}$$

We now form the normalised weights

$$w'(k) = w(k) / \sum_{j=1}^{n} w(j), \tag{9}$$

since we require weights which sum to 1, where n is the number of observations within both time and distance radii. The interpolated value, $\hat{T}$, at time t and gridpoint (ρ,λ) is then expressed as a linear sum of the product of the weight, $w'(k)$, and the kth observation within the search radii, $T_o(k)$, i.e

$$\hat{T} = \sum_{k=1}^{n} w'(k)\, T_o(k) \quad . \tag{10}$$

b) Tests with Analytical Radiance Fields

The sizes of the time and distance radii selected for the interpolation scheme are important. If a large radius is chosen, then the analysis may be oversmoothed. The magnitude of this oversmoothing depends on the temporal and spatial variability of the field. The task of simulating radiances from model temperatures and then retrieving and interpolating them requires considerable computer time. Accordingly the effect of varying the search radii used in the interpolation scheme was initially tested using an analytically calculated radiance field. Such tests are computationally inexpensive, and thus can be repeated using a large variety of time and distance radii. In addition, they involve no retrieval errors, since we are solely concerned with interpolating radiances from satellite observation points to grid points. In this section we describe these tests.

The radiance field is expressed as a sum of spherical harmonics; radiances

are easily calculated at either grid or 'observation' points. Satellite 'observation' points are determined using a model of a polar-orbiting satellite orbit and for computational economy are calculated at larger time intervals than the real satellite would actually observe at. Observations are calculated at the nadir point only, and the time interval is 256 s. The 'observations' are then interpolated in time and space using the time/space interpolation scheme and compared with the grid point analytical radiance field.

Tests were made on a number of fields. We present results for a field which resembles the stratosphere during a northern hemisphere winter in that there is a small amplitude wavenumber 1 wave in the southern hemisphere and tropics and there are higher amplitude wavenumber 1 and 2 waves in the middle and high latitudes of the northern hemisphere. To provide a stringent test of the scheme we have chosen high wave speeds. The wavenumber 1 wave, for example, completes a latitude circle in 5 days.

Results for a variety of time and distance radii are shown in Table 3. [2] Reducing the distance radius clearly decreases the r.m.s error, provided the distance radius is greater than 2000 km. In addition, detailed consideration of the original and interpolated fields (not shown) reveals that reduction of the size of the distance radius leads to an improvement in the estimate of the model field's wavenumber 1 and 2 features. In particular

1. The analysis made with a time radius of 12 hrs and a distance radius of 12000 km fails to estimate the wavenumber 1 and 2

[2] Table 3 in here

behaviour of the model field.

2. Changing the distance radius from 12000 to 5000 km reduces the r.m.s analysis error by about 50 %, and the resultant analysis determines most of the model field's wavenumber 1 behaviour, and also some of its wavenumber 2 behaviour.

3. Changing the distance radius from 5000 to 3000 km again reduces the r.m.s analysis error by about 50 % and both wavenumber 1 and 2 features of the model field are better estimated. The estimates are improved even further when the distance radius is reduced from 3000 to 2000 km.

When a distance radius of 1500 km is used, the r.m.s error is slightly higher than with a distance radius of 2000 km. All wavenumber 1 and 2 features are well estimated but the analysed field is less smooth than the model field because in many cases there is only one observation within the search radii. In tests where the distance radius is reduced to less than about 1200 km there are a number of gridpoints which have no observations within their corresponding search radii (note that when observations are simulated at actual TIROS-N orbit locations, see below, this missing observation problem occurs with distance radii of less than *1000 km*). Changing the time radius has little effect on the r.m.s error. Tests made using a 5000 km distance radius showed that changing the time radius from 12 to 18 hrs has negligible effect on the r.m.s error, and that this error increases only slightly when the time radius is further raised to 24 hrs. Similar conclusions may be drawn from tests made using distance radii of 3000 and 2000 km.

Other tests were made using fields which contained higher wavenumber waves. R.m.s errors changed with varying time and distance radii in a similar

manner to that described above. Despite the fact that the analytical fields only crudely mimic the real atmosphere, one can conclude that, within the range of radii tested, r.m.s errors decrease when the distance radius is reduced to the limit where some analysis points have no corresponding observations. In addition, despite the high wave speeds of the analytical field, the size of the time radius appears to have little impact on the r.m.s error.

5. Interpolation of Retrieved Temperatures

In this section we present the results of the tests of the combined retrieval/analysis scheme using the numerical model atmosphere. The simulated observations and retrievals are based on 24 hrs of model fields forced from below using the observed geopotential heights for 18th Jan. 1987. The 5 mb temperatures developed at the beginning, middle and end of that period by the model are shown in Figure 7. [1] During the 24 hr period there is a cooling by about 10 K in a region which extends from near 50°N 300°E to the north pole, indicating that the sudden warming which is developed by the model around this date has passed its peak at this height. In the test which follow the retrieval/analysis scheme is used to attempt to retrieve the field at the centre of this period, viz 1200 GMT.

Retrieved temperatures from 18th Jan. 1987 are interpolated using three combinations of search radii. One analysis is made with a time radius of 6 hrs and a distance radius of 2000 km, while another uses a time radius of 12 hours and a distance radius of 1000 km. The third analysis uses the search radii employed in the United Kingdom Meteorological Office's operational stratospheric

[1] Fig. 7 in here

analysis scheme, namely 12 hrs and 500 km. When analyses are performed with distance radii of 2000 and 1000 km every gridpoint has at least one observation within its corresponding search radii. However, when a distance radius of 500 km is used there are a number of gridpoints which have no observations within their search radii. These missing values are filled by linearly interpolating the nearest 'good' observations from gridpoints east and west of those with no observations, and then smoothing the whole field with a 3-point smoother with weights of the form (0.25,0.5,0.25). The temperatures are interpolated to the model grid and to the analysis time of 1200 GMT. For computational reasons, analyses are made only at selected pressure levels. These include some close to the SSU weighting function peaks (1.5, 5.0 and 15.0 mb), two in between these peaks (3.0 and 10.0 mb), and two outwith the range of the weighting function peaks (0.2 and 25.0 mb).

a) R.m.s Errors

The r.m.s error of the combined retrieval and analysis process is calculated, under the assumption that the model temperature at 1200 GMT is the 'true' temperature. The biases of the estimated temperatures are also calculated. Both r.m.s errors and biases were calculated for every one of the 7 latitude/season retrieval zones, and also for all latitudes. R.m.s errors for the analyses made using distance radii of 2000 and 1000 km are shown in Table 4a; and those for analyses made using distance radii of 1000 and 500 km are shown in Table 4b. [2]

Inspection of Table 4a shows that the r.m.s errors at 0.2 mb are generally higher than for any other pressure level. There are two possible explanations for this. Firstly, 0.2 mb is far away from a weighting function peak so one would

[2]Table 4 in here

expect the retrieval at this level to be poor. Secondly, the 0.2 mb field (unlike the fields at the other 6 levels) has considerable small-scale structure that the interpolation scheme might have difficulty resolving. R.m.s. errors at 0.2 mb are higher using a 1000 km distance radius than when using a 2000 km radius. This is the opposite of the results at other levels and tends to suggest that the high error is due to poor retrievals.

The model fields at 1.5, 3.0, 5.0, 10.0, 15.0 and 25.0 mb are quite similar to each other. With a distance radius of 2000 km, global r.m.s errors range between 2.98 and 3.73 K. Reducing the distance radius to 1000 km reduces the global r.m.s error to between 1.71 and 3.71 K. Only at 1.5 mb is the global r.m.s error not noticeably reduced when the distance radius is reduced to 1000 km. An inspection of the errors in each zone reveals that when the distance radius is reduced from 2000 km to 1000 km, the largest fall in r.m.s error generally occurs in the region of the sudden warming. Away from the sudden warming r.m.s errors change little when the distance radius is changed. This means that we require a small distance radius when analysing the high spatial variability within a sudden warming, but that a larger distance radius is adequate to analyse the field elsewhere.

Table 4b reveals that changing the distance radius from 1000 to 500 km and interpolating between gridpoints for those with no nearby observations produces little change in the r.m.s error, and maps of fields analysed with a 500 km distance radius (not shown) are little different from corresponding fields analysed using a 1000 km distance radius. This suggests that the use of a 500 km distance radius is unnecessary, as comparably good analyses can be obtained using a distance radius of 1000 km without problems caused by missing observations.

b) Comparison of Maps at 5 mb

To study the performance of the interpolation scheme more closely, we compare model and analysed fields at 5 mb. The analysed fields used in the comparison are those estimated using distance radii of 2000 and 1000 km. The results at 5 mb provide a good example of the improvement in the analysis caused by reducing the distance radius from 2000 to to 1000 km. The 5mb model field (Figure 7b) shows there is strong wavenumber 1 activity in the middle and upper latitudes of the northern hemisphere. There is a major high situated between 40° and 90°N and 90°E and 90°W with two peaks at 70°W, 70°N (266 K) and at 80°N, 30°E (262 K). There is also a large low situated between 50° and 70°N, and 30°E and 30°W, having a minimum value of 216 K. The high and the low are separated by a region of strong temperature gradients. Outwith the middle and upper latitudes in the northern hemisphere the flow is generally zonal.

A comparison of two analyses of the 5 mb field confirms that use of a smaller distance radius reduces the r.m.s error. Figure 8a shows the 5 mb field analysed with a time radius of 6 hrs and a distance radius of 2000 km; [3] model field features have been reproduced adequately at the correct geographical location. However Figure 8b (which shows the difference between model and analysed fields) reveals that the major high in the northern hemisphere has been underestimated by up to 14 K. In addition, the major low has been overestimated by 10 K, and in consequence the large gradient between the major high and low has been poorly estimated. The interpolation made with a distance radius of 1000 km (Figure 8c) produces a much better estimate of this large temperature

[3] Fig. 8 in here

gradient. In addition, the estimated magnitudes of the major high and low are improved. Figure 8d shows that the underestimate of the major high has been reduced to between 6 and 8 K and that the major low has been estimated to within 3 K of the model value. Again, comparison (not shown) between maps of analyses made with a 1000 km distance radius and of analyses made with a 500 km distance radius with gap-filling reveals that they are very similar.

We conclude that there is little advantage in the current operational distance radius of 500 km, which requires gap-filling where orbits fall far from gridpoints, over the 1000 km radius, which requires gap-filling only where observations are missing due to calibration sequences or drop-outs. On the other hand, the performance would be degraded by increasing the radius beyond 1000 km.

6. Conclusions

We have demonstrated that (given an accurate forward model) the stratospheric temperature field can be satisfactorily (generally within 2.5 K in the stratosphere) estimated using a retrieval/analysis scheme which uses the techniques of multiple linear regression and linear time/space interpolation. Although the scheme was tested using only one model field, this conclusion is robust since the model field contains a sudden warming, and so provides a stringent test of both retrieval and analysis schemes.

Temperature retrieval errors were highest within the region of a sudden warming. This is because the vertical temperature structure in the sudden warming was too small to be 'observed' by the satellite instrument, and also because the data used to calculate the regression coefficients were inevitably not representative of sudden warming conditions. The former errors, due to small-scale vertical structure, are inherent in the method of observation rather than the method of temperature retrieval. In contrast, the latter errors are caused

by dissimilarities between the sonde dataset and sudden warming conditions. These errors can be described as 'first guess errors' since the mean of the dataset can be thought of as a first guess estimate of the retrieved temperature. Ideally one would wish to reduce such errors by using a first guess based on a model forecast rather than on climatology, and future research may involve the development of a retrieval scheme which uses for the *a priori* information the output at the previous analysis time from a numerical weather prediction model. At present such models do not produce operational forecasts for the stratosphere, but it is expected that a number of such models will be extended up to the stratosphere in the near future.

The time/space interpolation scheme was tested using both idealised radiance fields and a stratospheric temperature field which included a sudden warming. Most features of these fields were satisfactorily reproduced, even those of high temporal variability such as a 5-day planetary wave, or of high spatial variability such as exist in a sudden warming. It was found that reducing the size of distance radius in the scheme decreased analysis errors, but that varying the size of time radius altered the error values only very slightly. A future refinement of the tests would involve the simulation of atmospheric tides (which are not calculated in the model) and the examination of possible biases in the analysis when data from certain satellite orbits are missing. Further research should also involve the comparison of the time/space interpolation scheme with other analysis schemes. As this time/space interpolation scheme produces essentially a weighted average of a set of observations, maximum values in the field are always underestimated and minimum values are always overestimated. Other analysis methods may not have this problem, and thus a comparison with other techniques, such as sequential estimation (Rodgers, 1976), would be valuable. However, the time/space interpolation method used in this paper has the

advantage that it is easy to understand, is computationally efficient and, as has been demonstrated, produces good results. Moreover it is versatile, since it can be used to analyse any sort of satellite data (eg composition measurements, temperature, radiance) easily and effectively without the need to estimate the noise characteristics of the observations.

I. Acknowledgements

DRJ was supported by the Natural Environment Research Council during the course of this work. We would like to thank the U.K Meteorological Office for their co-operation, especially Alan O'Neill and Mike Bailey.

II. Figure Legends

Figure 1 'Superobservation' points of the TOVS instrument on NOAA-7 for 18th Jan. 1987. Observations are made at the 4 scan angles -30°, -10°, 10° and 30° from the local vertical.

Figure 2 The 8 TOVS channel weighting functions expressed between 0.2 and 1000 mb for a vertical view.

Figure 3 Mean temperature profile (°K), and this profile plus and minus the standard deviation for the sonde dataset (dashed lines) and model (solid lines). a) zone 5; b) zone 1.

Figure 4 R.m.s temperature retrieval error (solid), bias (dashed) and standard deviation (dotted) profiles (°K). a) zone 1 (70° to 90° latitude, winter); b) zone 4 (30°N to 30°S, all seasons); c) zone 5 (30° to 50° latitude, summer); d) zone 6 (50° to 70° latitude, summer).

Figure 5 'Cross-section' of temperature (°K) at 189 observation points along a

satellite orbit. a) model; b) retrieved values; c) difference between model and retrieved values.

Figure 6 TOVS observation points on 18th Jan. 1987 between -20° and 20° latitude, -160° and -120° longitude, and 1020 and 1340 GMT. Also plotted is the model grid with a spacing of 5°.

Figure 7 Temperature field on 18th Jan. 1987 at 5 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. a) model field at 0000 GMT; b) as a), except the field at 1200 GMT is plotted; c) as a), except the field at 2400 GMT is plotted.

Figure 8 Temperature field for 1200 GMT on 18th Jan. 1987 at 5 mb. The northern hemisphere field is plotted on the right and the southern hemisphere field on the left. The contour spacing is 5 K. Dashed contours represent negative values. a) analysed values obtained using a time radius of 6 hrs and a distance radius of 2000 km; b) difference between model field (Figure 7b) and field in a); c) as a), except the analysis is made with a time radius of 12 hrs and a distance radius of 1000 km; d) difference between model field (Figure 7b) and field in c).

III. References

Al-Ajmi, D.N, Harwood, R.S and Miles, T (1985) A sudden warming in the middle atmosphere of the southern hemisphere *Quart. J.R. Met. Soc.* 111 359-389

Barnett, J.J, Harwood, R.S, Houghton, J.T, Morgan, C.G, Rodgers, C.D and Williamson, E.J (1975) Comparison between radiosonde, rocketsonde and satellite observations of atmospheric temperatures. *Quart. J.R Met. Soc.* 101 423-436

Chedin, A, Scott, N.A, Wahiche, C and Moulinier, P (1985) The Improved

Initialization Inversion method: a high resolution physical method for temperature retrievals from satellites of the TIROS-N series. *J Clim. App. Meteor.* 24 128-143

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

Drayson, S.R (1966) Atmospheric transmission in the CO_2 bands between 12 μm and 18 μm. *App. Optics* 5 385-391.

Eyre, J.R (1987) On systematic errors in satellite sounding products and their climatological mean values *Quart.J.R.Met. Soc* 113 279-292

Fisher, M (1987) The Met O 20 Stratosphere/Mesosphere Model. *UK Meteorological Office Met O 20 DCTN 52*

Houghton, J.T and Smith, S.D (1970) Remote sensing of atmospheric temperature from satellites. *Proc. Roy. Soc. Lond.* A320 23-33.

Kaplan, J.D (1959) Inference of atmospheric structure from remote radiation measurement. *J Opt. Soc. Amer.* 49 1004-1007.

McMillin, L.M and Fleming, H.E (1976) Atmospheric transmittance of an absorbing gas : a computationally fast and accurate transmittance model for absorbing gases with constant mixing ratios in inhomogeneous atmospheres. *App. Optics* 15 358-363

Miller, D.E, Brownscombe, J.L, Carruthers, G.P, Pick, D.R and Stewart, K.H (1980) Operational temperature sounding of the stratosphere. *Phil. Trans. R. Soc. Lond.* A296 65-71.

Nash, J and Brownscombe, J.L (1983) Validation of the Stratospheric Sounding Unit. *Adv. Space Res.* 2 No 6 59-62

Peckham, G (1976) The information content of remote measurements of atmospheric temperature by satellite infrared radiometry and optimum radiometer configurations. *Quart.J.R.Met.Soc* 100 406-419

Pick, D.R and Brownscombe, J.L (1981) Early results based on the stratospheric channels of TOVS on the TIROS-N series of operational satellites. *Adv. Spac. Res* 1 247-260

Rodgers, C.D (1976) In : *Inversion Methods in Remote Sensing* (ed Deepak) 117-138. New York : Academic Press.

Schwalb,A (1978) The TIROS-N / NOAA A - G satellite series. *NOAA Technical Memo. NESS 95*

Smith, W.L, Hayden, M.C, Wark, D.Q .and McMillin, L.M (1979) The TIROS-N Operational Vertical Sounder. *Bull. Amer. Meteor. Soc.* 60 1177-1187

Uddstrom, M.J and Wark, D.Q (1985) A classification scheme for satellite temperature retrievals *J Clim. App. Meteor.* 24 16-29

IV. Tables

CHANNEL	WEIGHTING FN. PEAK/mb	CENTRAL WAVENUMBER/cm^{-1}
HIRS 1	30	669
HIRS 2	60	679
HIRS 3	100	690
MSU 23	280	1.832
MSU 24	100	1.932
SSU 25	15	667
SSU 26	5	667
SSU 27	2	667

Table 1. TOVS weighting functions. The acronyms refer to the individual instruments mounted on TOVS. HIRS is the High Resolution Infrared Radiation Sounder, MSU is the Microwave Sounding Unit, and SSU is the Stratospheric Sounding Unit.

ZONE	LATITUDE / SEASON
1	70° – 90°, winter
2	50° – 70°, "
3	30° – 50°, "
4	30°N –30°S, all seasons
5	30° – 50°, summer
6	50° – 70°, "
7	70° – 90°, "

Table 2 Latitude/season zones for which regression coefficients are calculated. 'Winter' is the six months between October and March (for the northern hemisphere) or between April and September (for the southern hemisphere). 'Summer' is the six months between April and September (for the northern hemisphere) or between October and March (for the southern hemisphere).

	r_t / hrs			
r_d / km	6	12	18	24
12000	3.35	3.34	...	...
5000	...	1.67	1.67	1.73
3000	...	0.93	0.93	1.01
2000	...	0.64	0.64	0.73
1500	...	0.67	0.66	0.73

Table 3 R.m.s analysis errors in radiance units (= mWm^{-2} cm $ster^{-1}$).

PRESSURE / mb

	0.2	1.5	3.0	5.0	10.0	15.0	25.0
ZONE	---	---	---	---	---	---	---
1	5.88/7.59	4.28/4.14	3.36/4.07	5.96/4.37	6.33/2.73	6.78/2.46	7.82/3.36
2	8.54/7.63	4.73/4.71	4.88/2.87	5.49/2.38	4.94/2.35	5.13/2.76	4.37/3.43
3	5.27/6.13	4.31/3.88	2.72/2.24	3.16/2.46	2.94/2.21	2.09/1.24	1.98/1.89
4	4.99/5.16	4.01/3.95	3.21/3.05	1.96/1.99	3.01/3.03	1.74/1.78	2.63/2.52
5	3.47/3.98	2.08/1.97	1.44/1.58	1.35/1.23	1.31/1.26	0.97/0.83	1.27/1.17
6	3.74/4.75	3.26/3.82	1.98/2.10	1.49/1.58	0.80/0.91	0.65/0.67	1.22/1.33
7	3.50/3.56	1.60/1.98	0.89/0.97	0.67/0.59	0.59/0.64	0.50/0.63	2.12/1.78
Global	5.25/5.62	3.73/3.71	2.98/2.67	3.19/2.28	3.38/2.32	3.11/1.71	3.54/2.41

Table 4a R.m.s error of the combined retrieval and analysis in degrees K. Errors for the analysis made using a time radius of 6 hrs and a distance radius of 2000 km are shown to the left of the slash, errors for the analysis made using a time radius of 12 hrs and a distance radius of 1000 km are shown to the right of the slash.

PRESSURE / mb

	0.2	1.5	3.0	5.0	10.0	15.0	25.0
ZONE	---	---	---	---	---	---	---
1	7.59/8.17	4.14/4.26	4.07/4.74	4.37/4.66	2.73/2.76	2.46/1.82	3.36/2.65
2	7.63/7.32	4.71/4.80	2.87/2.73	2.38/2.02	2.35/2.48	2.76/2.39	3.43/3.30
3	6.13/6.45	3.88/3.83	2.24/2.29	2.46/2.37	2.21/2.17	1.24/1.14	1.89/1.96
4	5.16/5.20	3.95/3.95	3.05/3.02	1.99/2.08	3.03/3.14	1.78/1.90	2.52/2.46
5	3.98/4.30	1.97/2.00	1.58/1.67	1.23/1.24	1.26/1.28	0.83/0.82	1.17/1.17
6	4.75/5.06	3.82/3.94	2.10/2.08	1.58/1.58	0.91/0.98	0.67/0.72	1.33/1.43
7	3.56/3.81	1.98/1.98	0.97/1.21	0.59/0.73	0.64/0.73	0.63/0.73	1.78/1.68
Global	5.62/5.79	3.71/3.74	2.67/2.81	2.28/2.33	2.32/2.39	1.71/1.59	2.41/2.26

Table 4b R.m.s errors of the combined retrieval and analysis in degrees K. Errors for the analysis made using a time radius of 12 hrs and a distance radius of 1000 km are shown to the left of the slash, errors for the analysis made using a time radius of 12 hrs and a distance radius of 500 km are shown to the right of the slash.

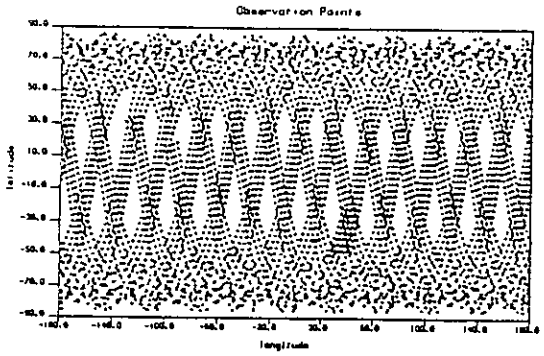


Figure 1

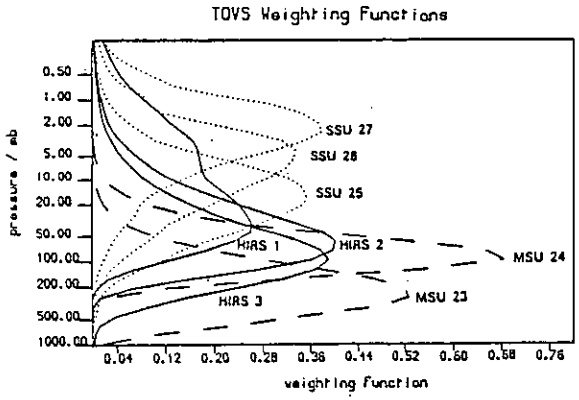
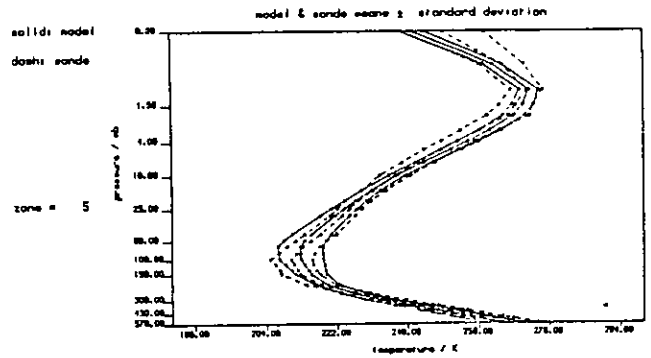


Figure 2

a



b

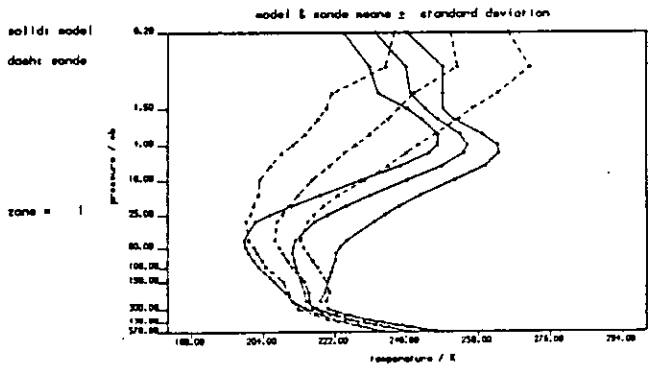
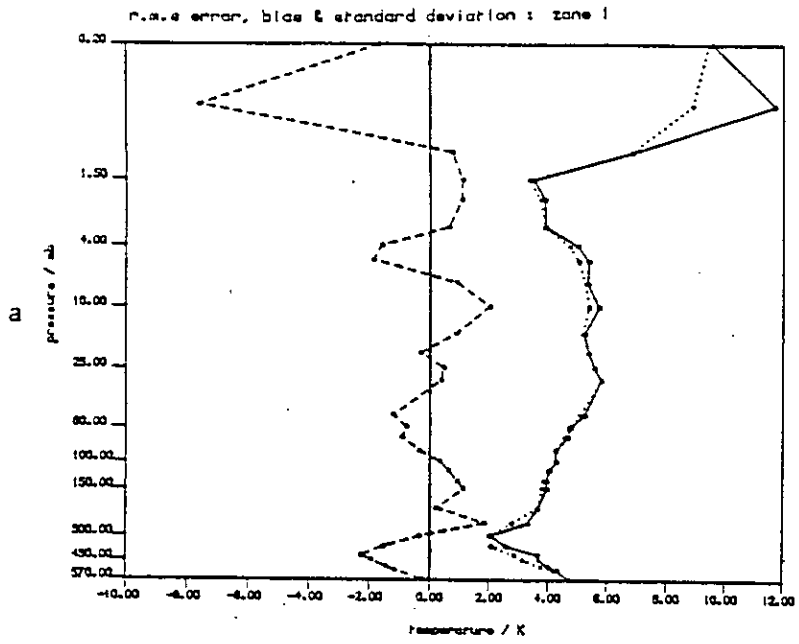
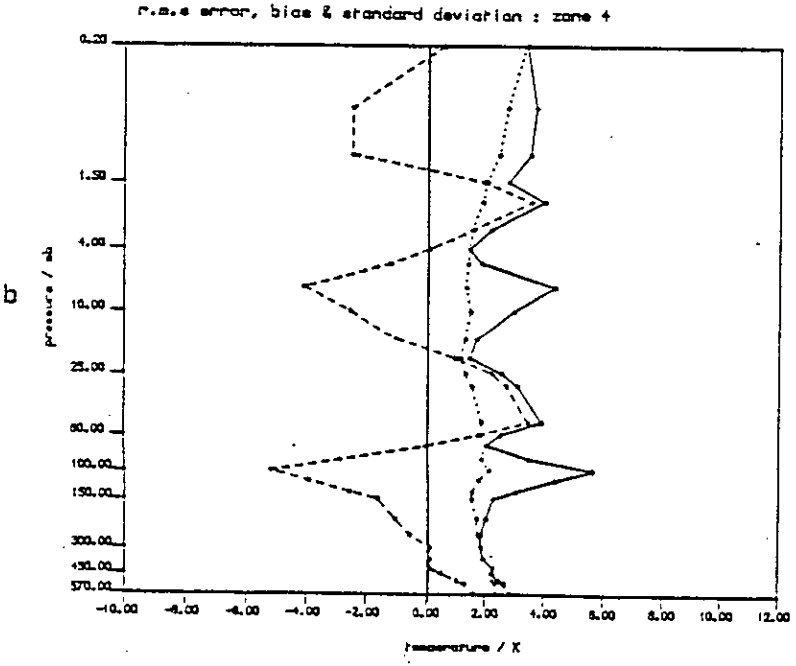


Figure 3

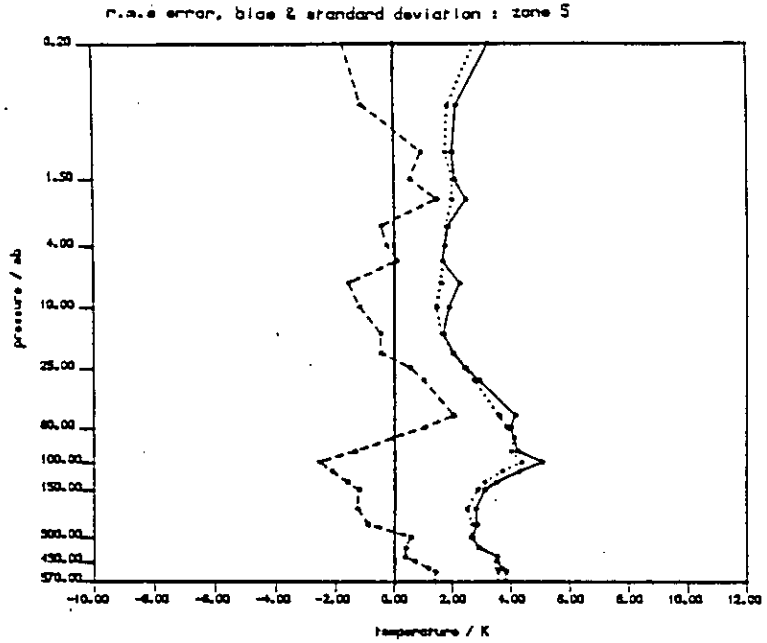


a

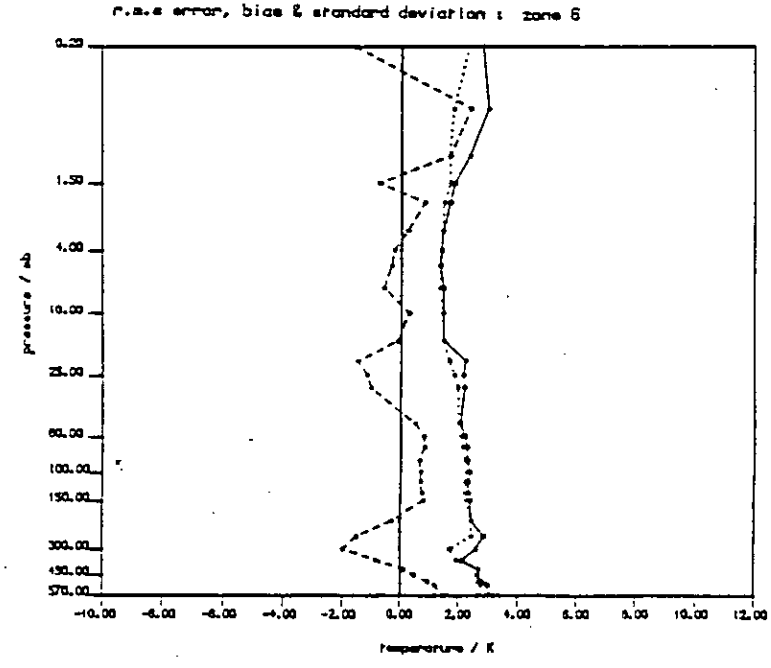


b

Figure 4



c



d

Figure 4

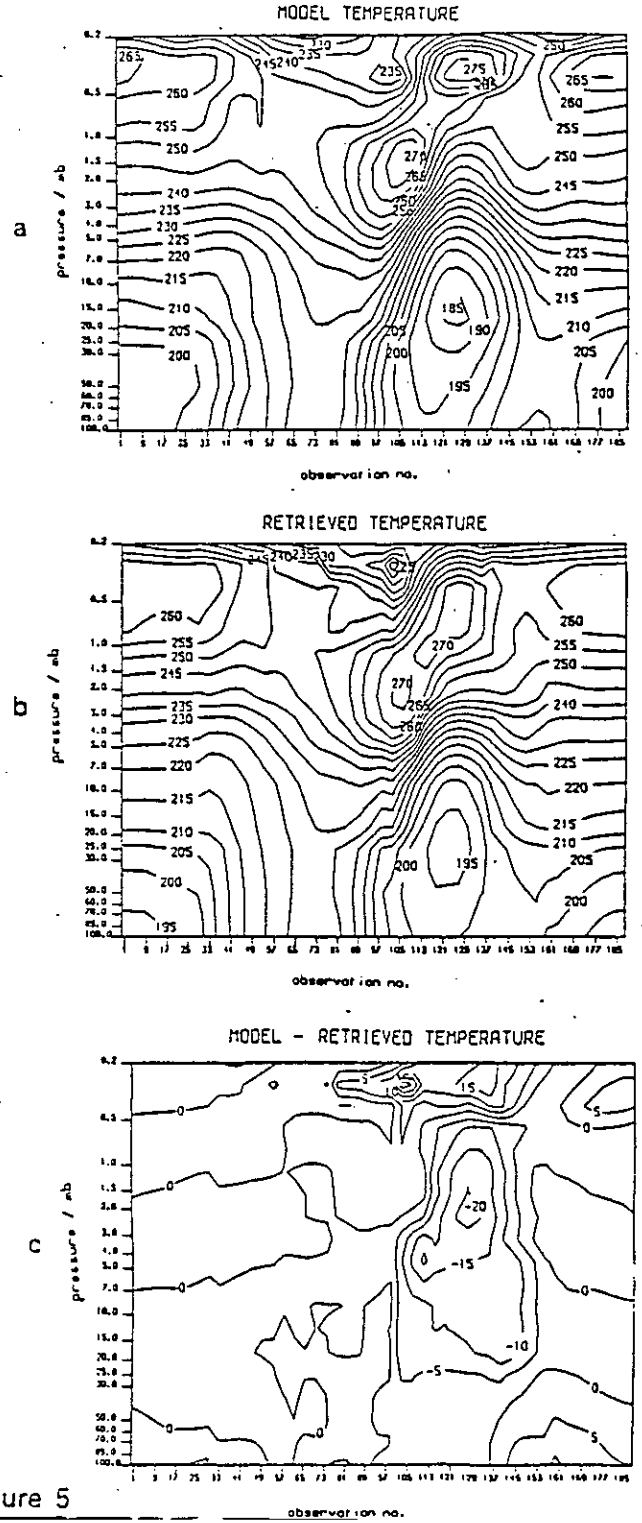


Figure 5

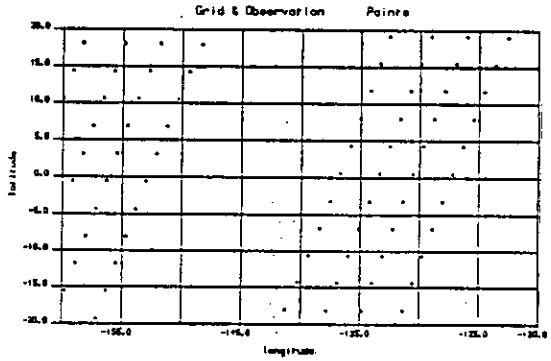


Figure 6

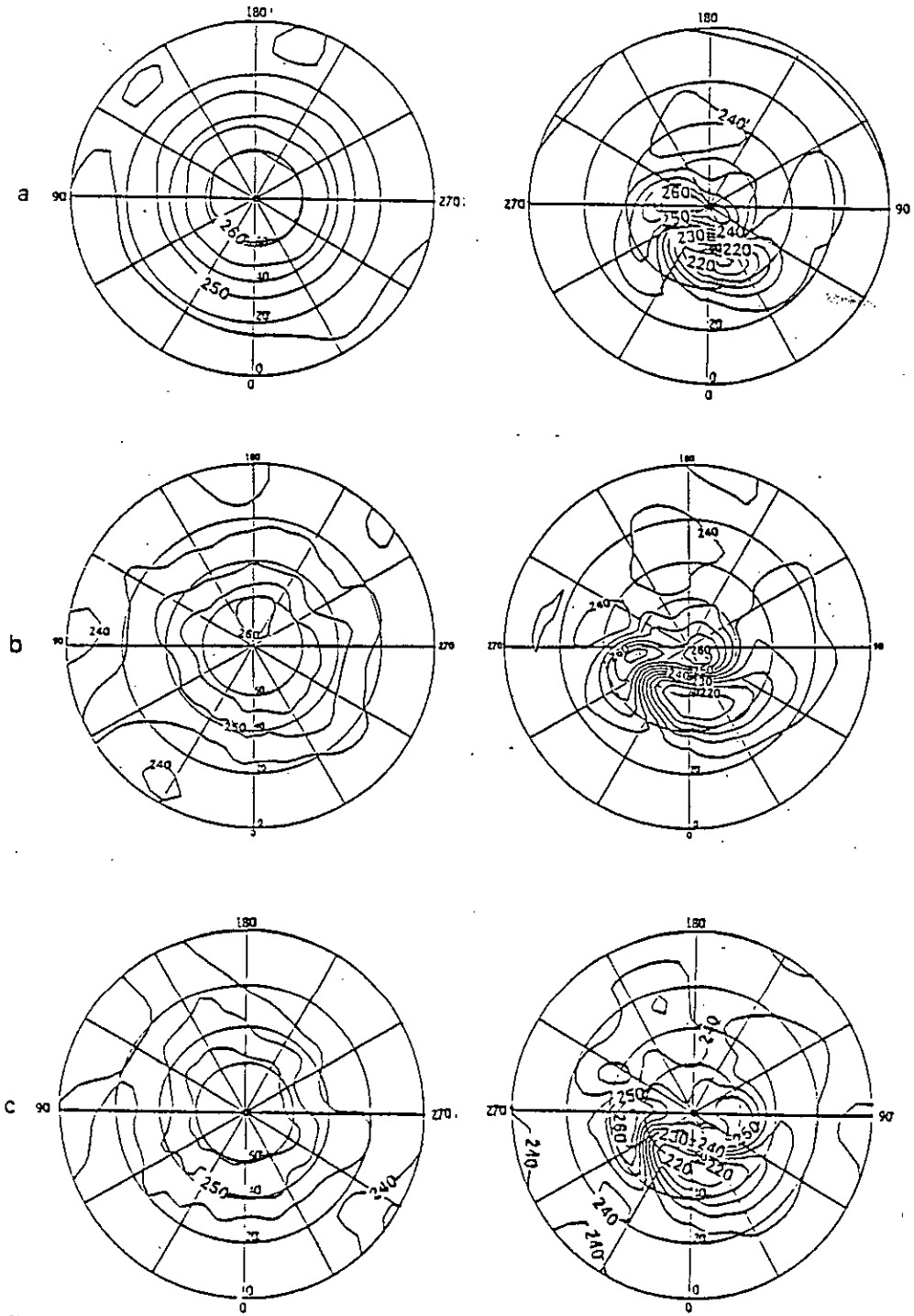


Figure 7

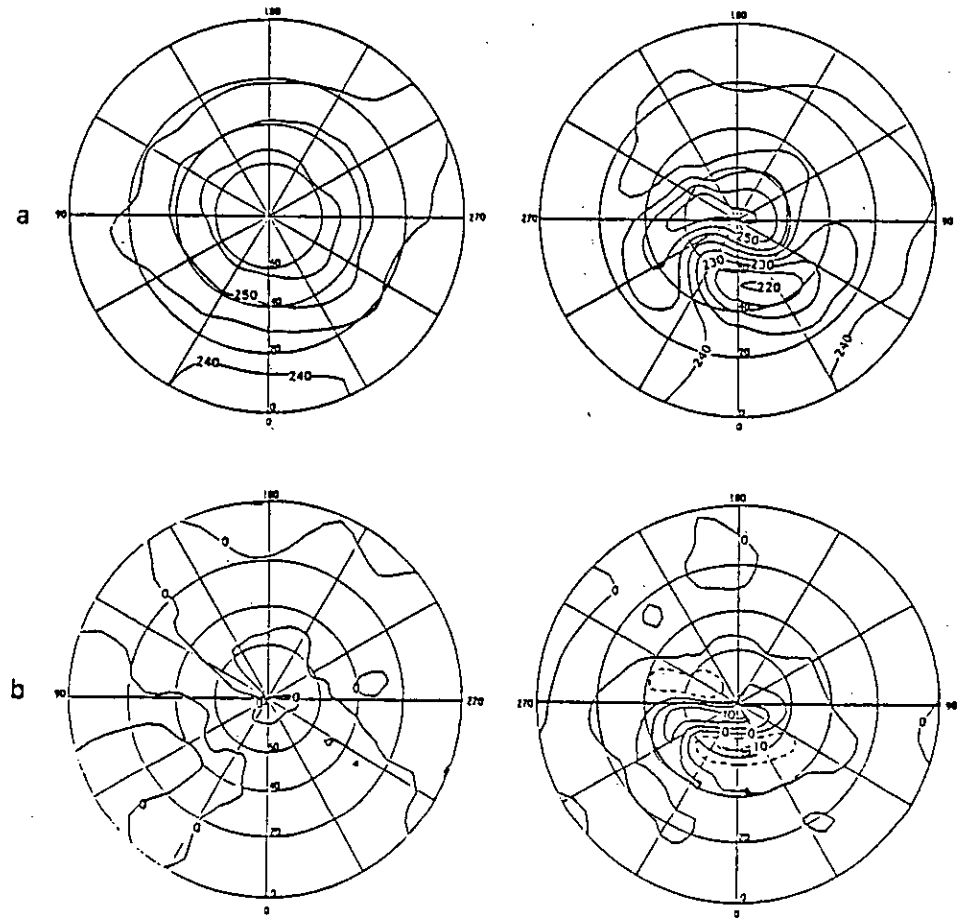


Figure 8

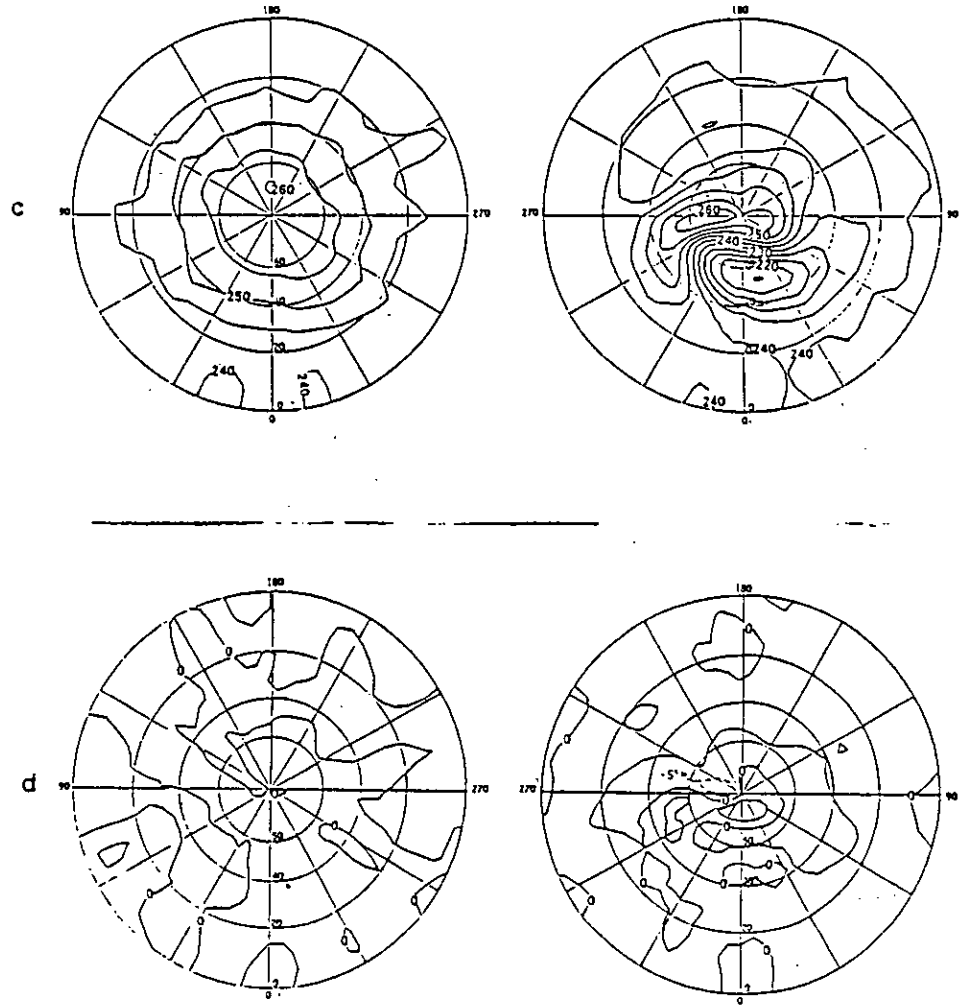


Figure 8