Analysis of atmospheric temperature and humidity from radio occultation measurements

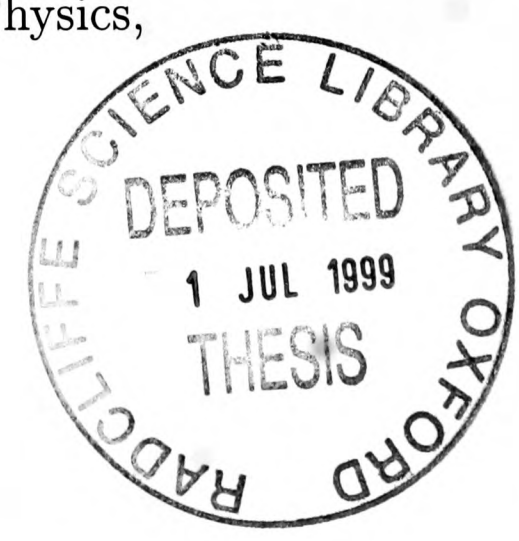
Paul Ian Palmer

Linacre College

A thesis submitted to the Faculty of Physical Sciences
for the degree of Doctor of Philosophy in the University of Oxford
Michaelmas Term, 1998



Atmospheric, Oceanic & Planetary Physics,
Clarendon Laboratory,
Oxford



Abstract

Radio occultation measurements from global navigation satellite systems, such as the Global Positioning System (GPS), represent a new source of numerical weather prediction information.

Conventionally, radio occultation measurements are inverted using an Abel integral transform to obtain a profile of refractivity, and subsequently pressure, on geometric height levels via the hydrostatic relation. Although accurate temperature/water vapour retrievals are possible with a background estimate of water vapour/temperature, it is not possible to retrieve these quantities simultaneously using this method. Also, using this method, refractivity retrieval errors are introduced by assuming no horizontal structure local to ray periapsis (the spherical symmetry assumption), and the method is subject to 'first-guess' errors when the hydrostatic relation is initialised.

Results from an investigation into how departures from spherical symmetry affect the performance of the Abelian inverse method are presented. It is shown that realistic horizontal humidity inhomogeneities can compensate for or reinforce horizontal temperature inhomogeneities, and therefore are important to consider in both the forward and inverse modelling. Using the results from this experiment, regression coefficients are fitted in an effort to predict temperature and refractivity retrieval errors from the horizontal temperature and humidity structure local to the measurement. The largest contribution to the predicted errors is shown to be from the local parabolic component of the horizontal structure, but is found to predict only a small fraction of the total error.

A non-linear optimal estimation inverse method is presented with which it is possible to retrieve simultaneously profiles of temperature, humidity and surface pressure. Using this method, the measurements are assimilated with *a priori* information utilising error estimates of the *a priori* information and the measurements. The method implemented is validated using an ensemble of numerical simulations.

Real observations from the GPS/MET pilot experiment are used to retrieve profiles of temperature, humidity and surface pressure which are validated using collocated ECMWF and NMC model analyses, radiosondes, and the Abelian inversion results. Temperature and refractivity comparisons between the optimal estimate and results from the Abelian inverse method show good agreement at high latitudes for all altitudes, resolving small-scale structure not shown by the model analyses. At low latitudes there is good agreement above the tropopause, below which a temperature bias ensues between the Abelian inversion and all other correlative data. Retrieved values for water vapour and surface pressure compare well with model analyses and collocated radiosondes. Biases between the UKMO and the ECMWF model analyses are consistent with known differences between the two models at the time of the dataset.

Acknowledgements

Like all great works of art, this thesis has involved contributions – large or small, directly or indirectly – from a great number of people. I have attempted to collate a list of such people here which is in no particular order of precedence.

I should like to convey my sincere gratitude to my supervisors John Barnett and John Eyre who have provided many useful discussions throughout my stint as a postgraduate. Their useful comments about draft copies of this document have been invaluable.

I thank Anu Dudhia for all those curt, erudite observations concerning my work; also for comments on my disgraceful misuse of grammar throughout the early drafts of this document.

I have benefited greatly from the collaboration I have had with the UKMO; in particular, Sean Healy, Juha-Pekka Luntama (now at EUMETSAT) and Dave Offiler who have provided assistance whenever I've needed it. Many thanks to Sean Milton who provided invaluable information about UKMO–ECMWF model differences.

My thanks to Clive Rodgers for many useful and interesting discussions about retrieval theory, and for taking the time and effort to sit down and proof-read several draft chapters.

Ta to Kam, Paul K-D, Dave, Rob, Stephan, Conor, Alan, Guillermo and Miguel, to name but a few of the people without whom tea-breaks just wouldn't have been the same.

Cheers to all the people in AOPP who have looked out for me and contributed to my general well-being throughout my three years, including Sarah, Debbie, Robert W., Fred R., Pat B., George and Roy.

For providing support and a forum for the work presented herein I should like to thank Luis Kornbleuh (Max-Planck Institute, Hamburg), Michael Gorbunov (Obukhov Institute for Atmospheric Physics, Moscow), Gottfried Kirchengast (University of Graz, Austria), and Stig, Mette and Georg (Danmarks Meteorologiske Institut). For many useful discussions about the radio occultation technique and providing the GPS/MET dataset used in the analysis shown herein my sincere thanks to Rob Kursinski and Stephen Leroy (Jet Propulsion Laboratory); for many useful discussions about the GPS I thank Peter Clarke (Department of Earth Sciences). For providing me with collocated NMC data I thank Bob Wells, and for allowing access to ECMWF fields I should like to thank the BADC (and Warwick Norton for assistance with obtaining ECMWF data).

For providing financial support throughout my three years, I should like to thank the NERC (for my studentship), the UKMO (for my CASE award), the Linacre House Fellowship Trust, AOPP, Oxford University Graduate Studies committee, the Institute of Physics and the Max-Planck Institute.

Special thanks and love to Adèle who has put up with me through the writing of this thesis, and has always been there for me.

For getting me drunk, playing silly buggers and making me forget how boring my life was during writing up I reserve a special thanks to Adèle, Augustín, Ingo, Ann, Helena, Pete Lewis, Peter Varnai, Andre, Jeff, Barry, Ian S., Karen, Ed and Jessica.

I am indebted to Mike, Eileen and Katie who have provided a very important source of encouragement and support.

To my parents.

"We are all in the gutter,
but some of us are
looking at the stars."

Oscar Wilde

1854 – 1900

Contents

1 Introduction **9**

1.1 Research impetus 10

1.2 Thesis outline 11

1.3 The Earth's atmosphere 12

1.4 The radio occultation technique and its application to the Earth 13

1.5 Other upper-air observation systems 16

1.6 Global Positioning System radio occultation measurements 20

2 The Global Positioning System and radio occultation **23**

2.1 System configuration 23

2.2 GPS signals 24

2.2.1 Observables 26

2.2.2 Military encryption 27

2.2.3 Decryption 28

2.3 Receiver technology 30

2.3.1 Loss of signal 31

2.4 GPS radio occultation 32

2.4.1 Atmospheric attenuation 35

2.4.2 Spatial resolution 35

2.4.3 Ionospheric effects 38

2.4.4 The GPS/MET experiment 39

3 The forward radio occultation problem **43**

3.1 Forward modelling of problem 43

3.1.1 Refractive medium 43

3.1.2 Path taken through the refractive medium 45

3.2 Factors that limit modelling of the problem 47

4 The inverse radio occultation problem **53**

4.1 The inverse refraction problem 53

4.2 Ionospheric correction 55

4.3 The Abelian inverse method 56

4.3.1 Refractive index determination 56

4.3.2 Pressure and temperature determination 58

4.3.3 Factors affecting the retrieval accuracy 63

4.4 The optimal estimation inverse method 65

4.4.1 Problem formulation 65

4.4.2 The maximum likelihood solution 69

4.5 Error Analysis Theory 76

4.5.1 Diagnostic Matrices 76

4.5.2 Contributory error covariance matrices 79

4.5.3 Systematic errors 80

5 The effects of line-of-sight horizontal inhomogeneities **81**

5.1 General forward model implementation 81

5.2 The Abelian inverse model implementation 84

5.2.1 The inverse Abel transform 84

5.2.2 The hydrostatic relation 87

5.3 Simple horizontal inhomogeneity structures 90

5.3.1 Types of inhomogeneity studied 90

5.3.2 Results 94

5.3.3 Fitting regression coefficients to temperature retrieval errors . . . 105

5.3.4 Horizontal smoothing 107

5.3.5 Inclusion of water vapour 110
5.3.6 Fitting regression coefficients to refractivity retrieval errors 114
5.4 Characterisation of the Abelian inverse method 115
5.5 Chapter summary 119

6 Optimal estimation: implementation and simulation 121
6.1 Implementation 121
6.1.1 The a priori vector and its error covariance matrix 121
6.1.2 The measurement vector and its error covariance matrix 124
6.2 Convergence and quality control 135
6.3 Characterisation of the optimal estimation inverse method 135
6.3.1 Matrix of weighting functions 135
6.3.2 Matrix of contribution functions 138
6.3.3 Matrix of averaging kernels 140
6.3.4 Contributory error covariance matrices 144
6.4 Ensemble of numerical simulations 146
6.5 Quality of surface pressure retrievals 150
6.6 Sensitivity of inverse model statistics 150
6.7 Chapter summary 154

7 Optimal estimation: validation 156
7.1 Description and preparation of the JPL GPS/MET dataset 156
7.1.1 Recomputing the ionospheric correction 158
7.2 Computing optimal estimates from GPS/MET observations 160
7.3 Description and preparation of correlative data 165
7.3.1 Radiosondes 165
7.3.2 NWP model analyses 168
7.3.3 Abelian inversion results 169
7.4 Validation 169
7.4.1 Theoretical retrieval r.m.s. errors 170
7.4.2 Global statistics 175

7.4.3 Latitude bands 177
7.4.4 Surface pressure validation 179
7.4.5 UKMO–ECMWF model differences in 1995 181
7.4.6 The importance of good temporal collocation 182
7.4.7 Particular occultations 184
7.5 Chapter summary 186

8 Conclusions and further work **190**
8.1 Summary of conclusions 190
8.1.1 How do typical line-of-sight horizontal meteorological gradients affect the Abelian inverse method of radio occultation observations? 190
8.1.2 Using a non-linear optimal estimation inverse method, could radio occultation observations improve numerical weather prediction? . 192
8.2 Suggested topics for further work 195

A Derivation of Fresnel zone radii **199**

B Equation for the radio refractive index of air **202**

C Information content **204**

D Cross correlative data statistics **207**

E Acronyms and Notation used **213**

Chapter 1

Introduction

There are two possible ways to improve the quality of forecasts from numerical weather prediction models: use a better model (i.e. improve the dynamics or physics of the model), or use better initial conditions (i.e. improve the observations or assimilation schemes).

Radio occultation measurements of the Earth's atmosphere represent a new source of data for numerical weather prediction. The Abelian inversion of these observations represents the conventional inverse method, which utilises an Abel integral transform. To investigate the potential of these observations two important questions are addressed:

1. How do typical line-of-sight horizontal meteorological gradients affect the Abelian inverse method of radio occultation observations?
2. Using a non-linear optimal estimation inverse method, could radio occultation observations improve numerical weather prediction?

Although retrieved temperature values using the Abelian inverse method have been found to show good agreement with collocated radiosondes and model analyses this inverse method has two main disadvantages. First, it is subject to a temperature–water vapour ambiguity which derives from refractivity (retrieved from the Abel integral transform) being proportional to both these quantities. This ambiguity means that although it is possible to retrieve a profile of temperature/water vapour with a background estimate of water vapour/temperature it is not possible to retrieve these quantities simultaneously. And second, it assumes no horizontal variation along the line of sight about the tangent point (the spherical symmetry assumption), which is the subject of the first question addressed in this thesis. Because there is often considerable horizontal variation along

the line of sight, the spherical symmetry assumption often causes the largest source of retrieved refractivity error and therefore should be investigated.

The second question involves an alternative formulation for the inverse problem. The non-linear optimal estimation inverse method requires a reasonable background estimate of the atmospheric state. Using the background estimate (and its estimated uncertainties) and a set of observations (with their estimated uncertainty), an optimal estimate of the atmospheric state can be computed. This inverse method has the distinct advantage of being able to 'separate' the effects of temperature and water vapour from the observations and thus simultaneously retrieve both quantities; furthermore, using this method it is possible to account for realistic horizontal variations along the line of sight about the tangent point. Although, in principle, this inverse method is preferable to the Abelian inverse method it is first necessary to investigate whether optimal estimation improves the prior knowledge of the atmosphere.

1.1 Research impetus

The high vertical resolution, all-weather, global nature of radio occultation observations means that they possess the ability to resolve small-scale vertical atmospheric structures not currently seen in weather prediction models. These small-scale atmospheric perturbations, which are not resolved by other satellite instruments, can lead to large-scale weather changes elsewhere on the globe. For this reason alone, it is essential that all reasonable atmospheric observations should be exploited to their full potential so that improved modelling and forecasting can be achieved.

The plans of several countries to fly the necessary instruments to obtain radio occultation observations represent the prospect of a wealth of data to contribute to numerical weather prediction. These observations are especially important for data sparse areas, e.g. polar regions and over oceans.

1.2 Thesis outline

The material included in this thesis includes the background knowledge required to understand the problems that were tackled, and to interpret the results from several investigations with simulated and real radio occultation measurements of the Earth.

The first chapter provides the reader with a broad idea of the impetus for this research, presents the basic ideas behind the radio occultation method (and some details of its Earth application), and a summary of other upper atmospheric observation systems that are currently assimilated by operational numerical weather prediction (NWP) centres.

An overview of the Global Positioning System and further details of radio occultation are presented in chapter 2. This information is vital to the understanding of the measurements obtained from a pilot experiment, which is also discussed in depth.

The forward radio occultation problem and its modelling are described in chapter 3. It is necessary to consider the fundamental physics involved with the radio occultation technique and address some of the factors that limit the accurate forward modelling of the problem so that subsequent errors can be assessed.

Details of the inverse radio occultation problem are presented in chapter 4. This chapter includes information about the conventional inversion of radio occultation observations (used by the Jet Propulsion Laboratory, amongst others) and an alternative non-linear optimal estimation inverse method.

The general implementation details of the forward model and the Abelian inverse method are described in chapter 5. Using these models the spherical symmetry assumption used by the inverse Abel transform is investigated by simulating measurements from atmospheres with horizontal structure and retrieving temperature and refractivity (temperature and water vapour) using the Abel integral transform. Details of the Abelian inverse method error and solution characterisation are also shown.

Implementation details of the non-linear optimal estimation inverse method are presented and discussed in chapter 6. This chapter also presents the solution and error

characterisation of the method which is vital for a detailed understanding of the optimal estimation retrievals. The method is validated using an ensemble of simulated retrievals.

Radio occultation observations from the GPS/MET pilot experiment are assimilated with collocated UKMO model analyses. The impact upon the UKMO model analyses is shown and analysed in detail. Optimal estimates are validated using several sources of correlative data and are shown in chapter 7.

Finally, chapter 8 summarises the main conclusions from the thesis and discusses possible topics of future work that should be undertaken.

1.3 The Earth's atmosphere

The Earth's neutral atmosphere can be thought of being composed of several different 'layers'. Figure 1.1 shows the variation of atmospheric temperature with altitude.

Temperature generally decreases with altitude throughout the troposphere to a minimum at the tropopause (10−15 km, low/high at high/low latitudes), primarily due to the emission of infra-red radiation by water vapour and clouds. Because of constant evaporation from oceans, the troposphere is abundant in water. This temperature minimum restricts the amount of water vapour that is transported from this region to the stratosphere.

In the immediate region above the troposphere, the lower stratosphere, temperature is fairly isothermal. Further into the stratosphere the temperature increases with altitude, to a maximum at the stratopause. Increased temperature here is primarily due to strong absorption of solar ultra-violet radiation by ozone.

Above the stratosphere lies the mesosphere, where temperature decreases with altitude to a minimum at the mesopause, primarily due to reduced amounts of ozone. Above this region, the lower thermosphere extends to the homopause (approximately 110 km), where molecular diffusion begins to dominate and atmospheric gases become separated vertically according to their differing molecular masses.

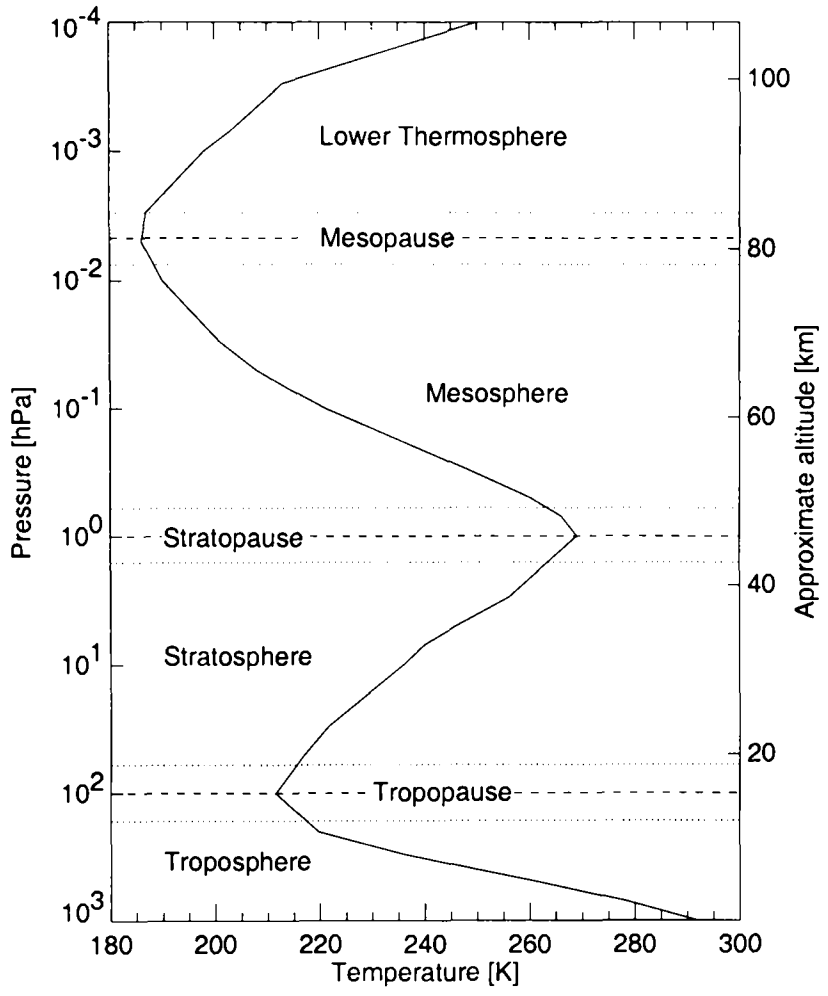


Figure 1.1: A schematic of the Earth's atmospheric temperature profile. The broken horizontal lines indicate the position of the pausal layers for the profile shown; and the dotted lines represent the possible variation in their vertical position with latitude. The data represents a model atmosphere for 40°N (Houghton (1977)).

1.4 The radio occultation technique and its application to the Earth

This section describes radio occultation and discusses its application to the Earth using the Global Positioning System.

Consider two satellites, T and R, radially separated by some distance; T carrying a transmitter and R a low Earth orbiting satellite with an onboard receiver (figure 1.2). Assuming a continuous radio link between T and R, when R passes behind the atmosphere with respect to T, the signal travels through the atmosphere and is refracted in response to variations of refractive index along its path. This refraction causes the

ray to travel over a longer path than it would in the absence of the atmosphere, which subsequently causes an atmospheric phase shift (time delay) in the received signal. This excess phase shift can be measured, from which a profile of refractivity as a function of geometric height at the ray's lowest approach to the Earth's surface can be inferred. Profiles of temperature and/or humidity can be calculated from refractivity.

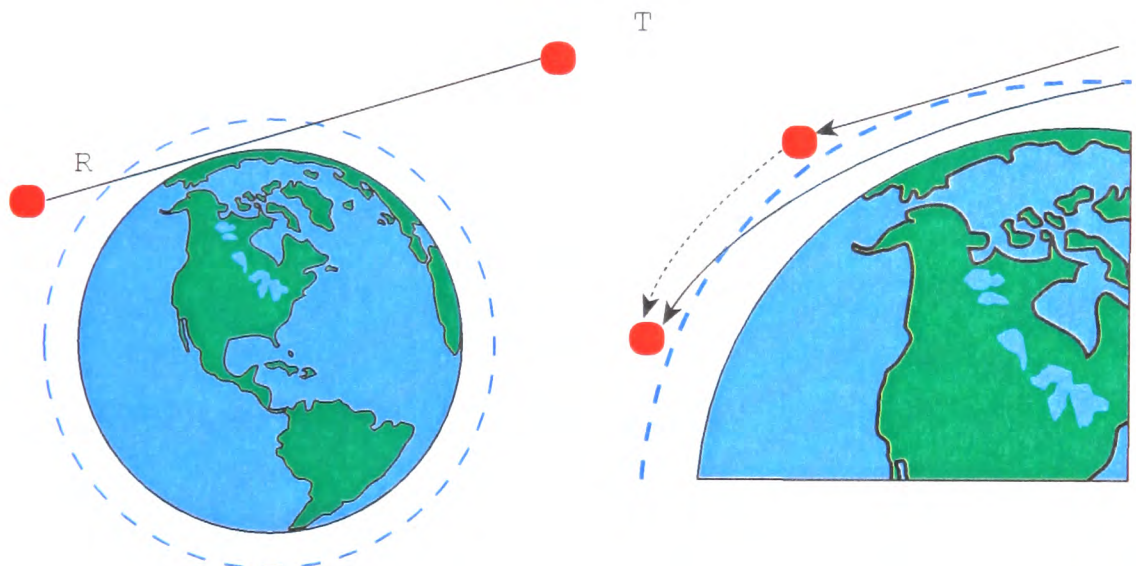


Figure 1.2: A schematic of the basic radio occultation geometry. The blue dashed line represents the atmosphere. Assuming a continuous radio link between the transmitter (T) and receiver (R), when R passes behind the planet's atmosphere with respect to T, the radio signal passes through the atmosphere and is refracted by variations of refractive index along the ray's path.

This method of reconstructing atmospheric temperature and pressure from phase delay data is based upon a technique that was developed at Stanford University and the Jet Propulsion Laboratory (JPL) and applied to studies of other planetary atmospheres (chapter 4).

Radio occultation (RO) experiments have played a prominent role in the NASA programme for solar system exploration for more than two decades and have contributed to studies of the atmosphere of Mars (Fjeldbo and Eshleman 1968), Venus (Fjeldbo and Kliore 1971), Jupiter (Kliore et al. 1975), Saturn (Lindal et al. 1985), Uranus (Lindal et al. 1987) and Neptune (Lindal 1992).

The general method used to sound other planetary atmospheres involves only one

spacecraft (flying past the planet to be sounded) and is shown by figure 1.3. The spacecraft's signal is tracked by a receiver station on Earth.

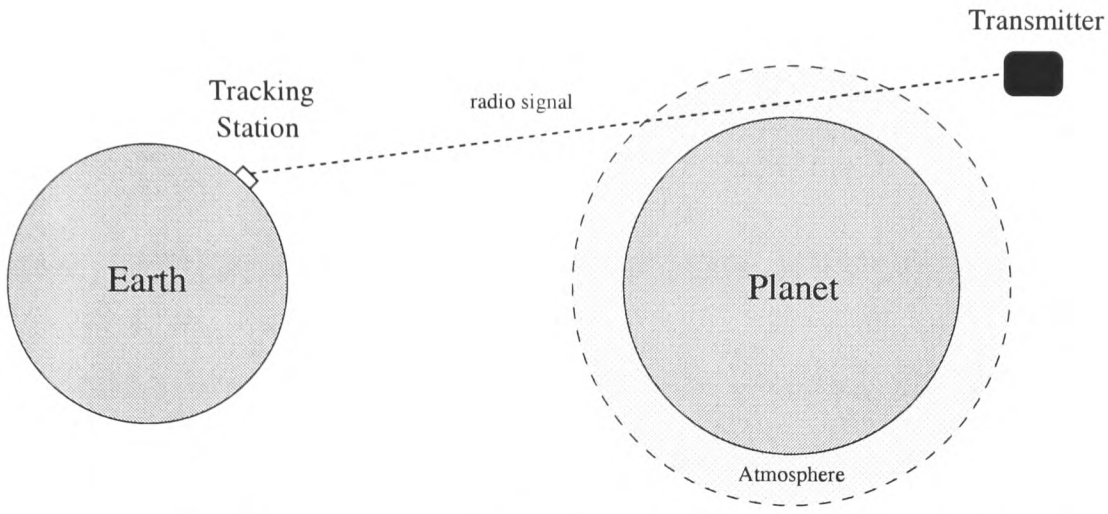


Figure 1.3: The radio occultation geometry used by astronomers to analyse both the neutral atmosphere and the ionosphere of other planets.

Radio occultation studies of the Earth

Radio occultation was not successfully applied to the Earth until some time after the first planetary experiments, mainly because the meteorological parameters retrieved from this technique are already known within a reasonable degree of accuracy from other sources; radio occultation measurements would only be of any value if they were of a much higher quality.

For many years space-based measurements were insufficiently accurate to solve this type of problem. It was not until the advent of the Global Positioning System that accurate measurements have became possible.

The Global Positioning System (GPS) is the U.S. military Global Navigation Satellite System (GNSS). The GPS can be described by three main elements: the satellites, the monitoring stations and the receivers. The satellite network comprises 24 satellites that provide complete coverage of the Earth at any given time. Tracking and monitoring stations across the globe detect radio signals from the satellites and relay the information to the master control station at Colorado Springs in the U.S.. This master station computes the coefficients of the precise orbital model for the entire GPS constellation and then sends an updated navigation message back to each satellite.

A pilot study involving one receiver (GPS/MET) in low Earth orbit was launched in 1995 and has since provided a wealth of data which is exploited in this thesis. The fundamentals of this experiment are presented in §1.6; a more detailed account of the experiment may be found in §2.4.4.

1.5 Other upper-air observation systems

At present, temperature and humidity observations available for climate research and numerical weather prediction are from various sources. The main ones are discussed below. The pictorial information included in this section has been kindly provided by the European Centre of Medium-range Weather Forecasts (ECMWF). The data represents possible sources of information for the ECMWF NWP model on the 9th June 1998, and is typical of all global NWP centres.

Radiosondes

Since the 1940s, radiosondes have been the main source for upper-atmosphere measurements of wind, temperature, pressure and humidity; in 1991, there were over 1000 radiosonde stations operating in 92 nations, using 39 different types of radiosondes. Under ideal atmospheric conditions and careful calibrations, radiosondes measure temperature to about ± 0.5 kelvin. However, due to radiation errors in the upper troposphere and stratosphere, temperature measurements can be degraded with errors of 1 kelvin above 250 hPa. A humidity sensor (hygristor) is located within the instrument package at a place where the outside air passes the hygristor. For many instruments the hygristor is designed to record the ambient relative humidity in the range from 15% to 100%. The accuracy of radiosonde humidity measurements can approach $\pm 1\%$ relative humidity using the latest capacitive thin-film humidity sensors.

Figure 1.4 shows that the radiosonde coverage is very limited, and is generally uneven over the globe, with concentrations in the U.S. and mainland Europe. In particular, polar and oceanic coverage is sparse, consequently there is a large disparity in coverage between the northern and southern hemispheres.

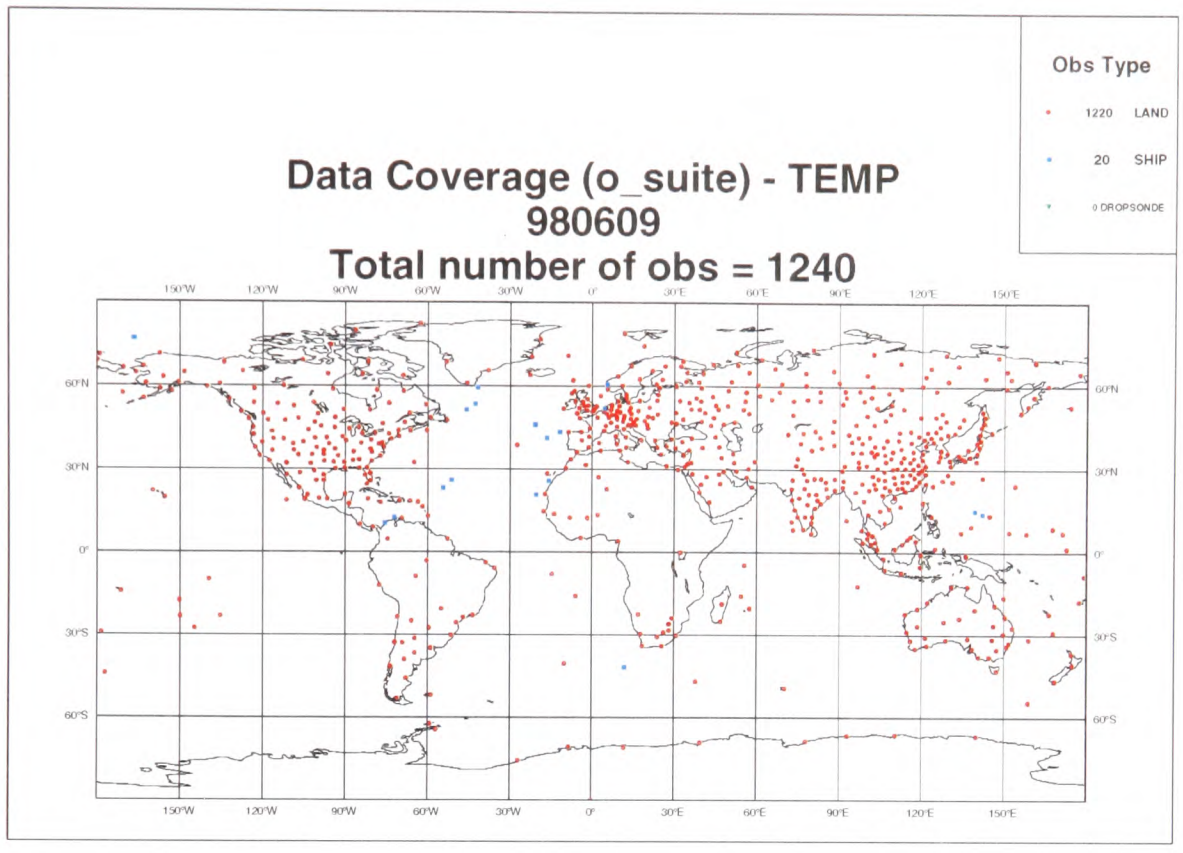


Figure 1.4: The global distribution of radiosondes on the 9th June 1998. There are three types of radiosondes shown: radiosondes that are launched from land (red), from a ship (blue) and dropsondes (green).

Commercial Aircraft

Some commercial aircraft provide accurate wind and temperature observations which are reported in real time by radio. However, the observations are usually at the cruising height (9–12 km) and are restricted to the main commercial air corridors. Limited vertical information is also obtained when aircraft ascend and descend. The information from this source is summarised by figure 1.5.

Three different types of measurements are shown: AIREPS, AMDAR and ACARS. AIREPS represents manual observations from aircraft (usually transmitted by the pilot), and AMDAR and ACARS represent different methods of automated aircraft measurements. It can be seen that the majority of the information is from central and northern America, where a large number of internal flights occur: the North Atlantic air corridors are also well covered.

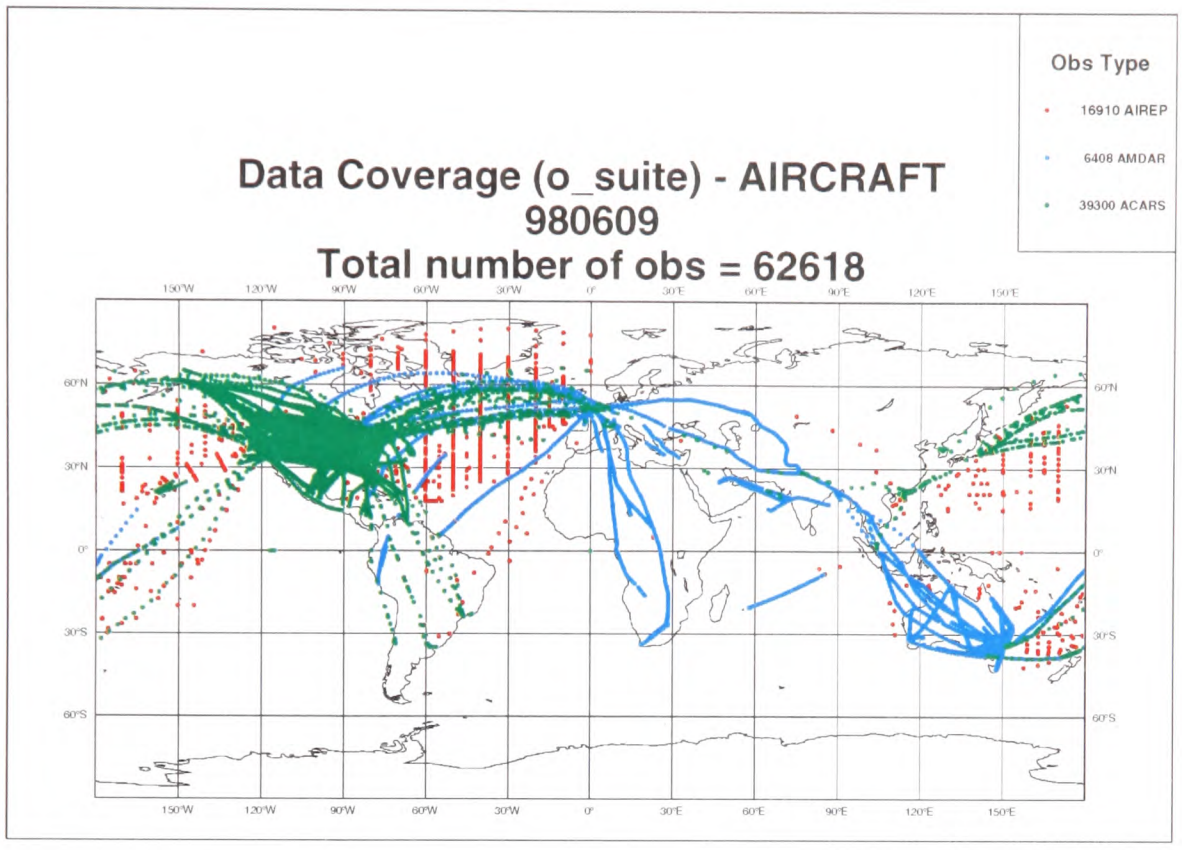


Figure 1.5: The global distribution of atmospheric information from commercial aircraft on the 9th June 1998.

Satellites

Satellite observations have the obvious advantage of global coverage. The current operational system on the TIROS-N/NOAA series of polar-orbiting satellites consists of the High-resolution Infrared Radiation Sounder (HIRS), the Microwave Sounding Unit (MSU) and the Stratospheric Sounding Unit (SSU). They are collectively known as TIROS Operational Vertical Sounder (TOVS) which provides tropospheric and stratospheric temperature and tropospheric humidity with global coverage at a horizontal resolution of about 50–100 km. Vertical resolution is poor, typically 3–4 km in the troposphere for cloud-free observations (worse in the stratosphere), and poorer for cloud-contaminated views. The accuracy of the present system for temperature and relative humidity is of the order 1 kelvin and 15%, respectively. The successor to TOVS is the Advanced TOVS (ATOVS), which consists of HIRS plus the Advanced Microwave Sounding Unit (AMSU).

The current series of geostationary satellites includes GOES-E, GOES-W,

METEOSAT-7, INSAT and GMS. One advantage of this type of satellite is that it is possible to observe the same location as often as needed, which is especially important when viewing hurricanes, etc. The disadvantages include that these satellites see little poleward of 60°, and to view all around the Earth involves the use of more than three satellites.

Figures 1.6 and 1.7 show the global distribution of information available from polar-orbiting and geostationary satellites in a typical six-hour period, respectively: over longer periods of time overlapping of information occurs. Both figures clearly show the global coverage that satellite observations provide.

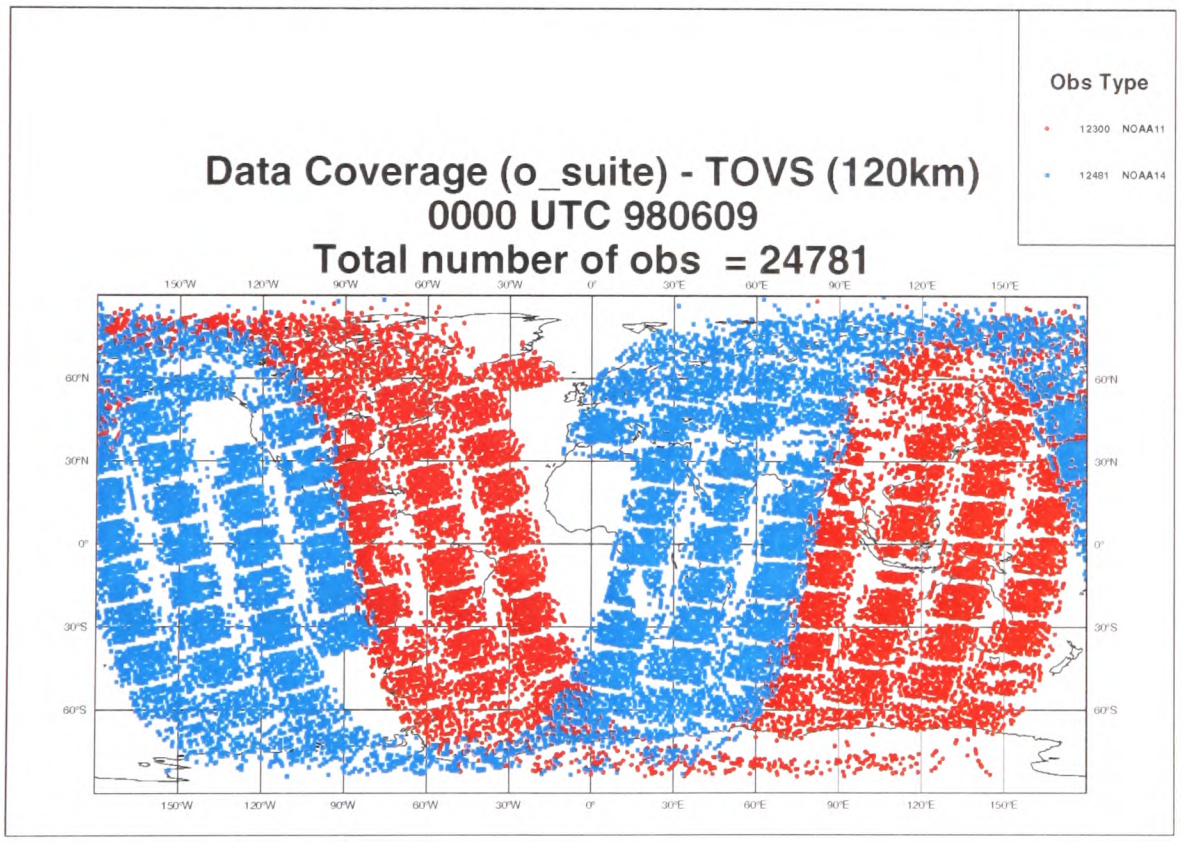


Figure 1.6: The global distribution of atmospheric information from the polar-orbiting satellites NOAA-11 (red) and NOAA-14 (blue), both with TOVS instruments onboard, during a six hour period on the 9th June 1998.

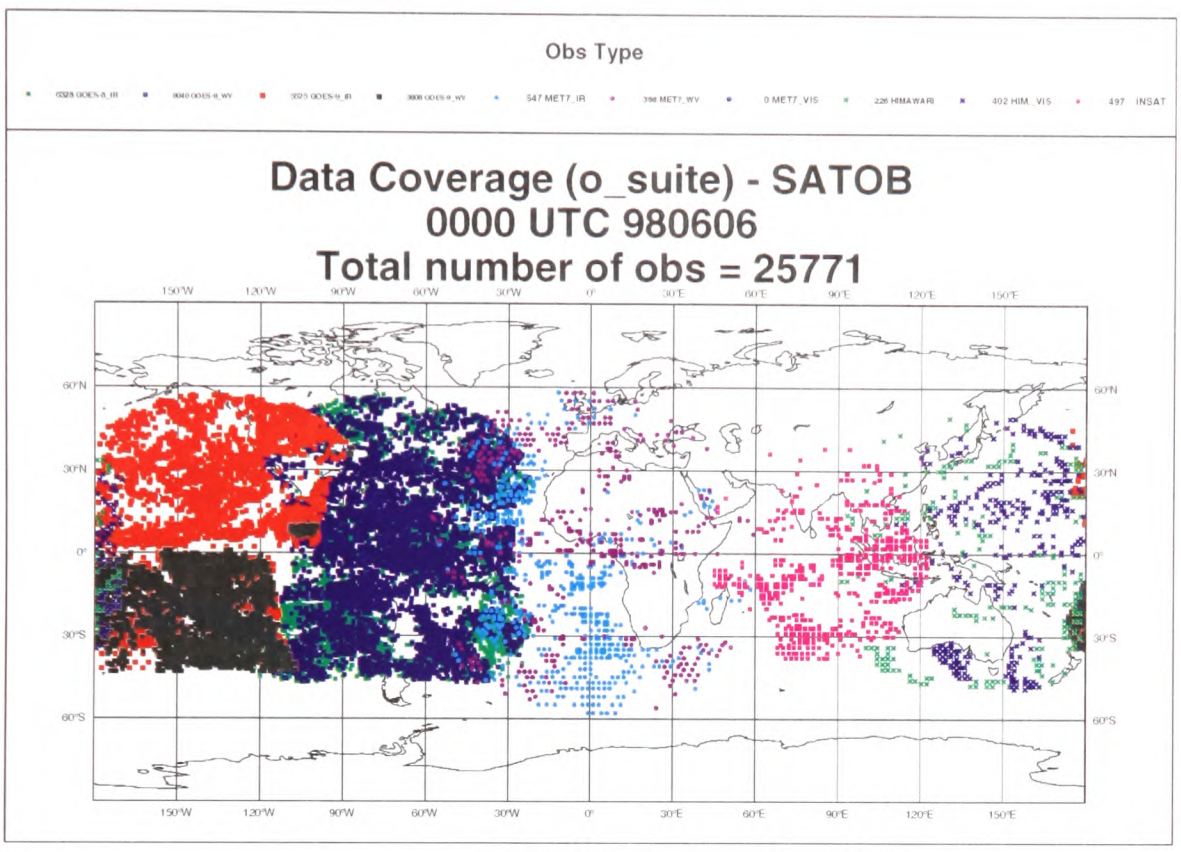


Figure 1.7: The global distribution of atmospheric information from geostationary satellites from GOES-E, GOES-W, METEOSAT-7, INSAT and GMS (also known as HIMWARI) during a six hour time period. The green, dark blue, red, black and pink squares represent the data from GOES-8/IR, GOES-8/WV, GOES-9/IR, GOES-9/WV and INSAT. The light blue, purple and dark blue circles represent data from MET7/IR, MET7/WV, MET7/VIS. The green and dark blue crosses represent data from HIMAWARI and HIM/VIS.

1.6 Global Positioning System radio occultation measurements

GPS radio occultation observations have a number of advantages compared to other atmospheric systems: the measurements can be taken at any time of the day or night, regardless of weather conditions (except during extreme ionospheric disturbances §3.2): the measurement is the change in frequency of the signal as it is refracted through the atmosphere, and is measured by time. This particular advantage means that the observation is extremely stable and intrinsically self-calibrating (as opposed to complicated radiometric calibration), which makes it very attractive for measuring long-term changes in the atmosphere. The GPS signal frequencies (§2.2) are, for most of the atmosphere, well below any absorption lines of significant atmospheric constituents: and

the wavelengths are sufficiently long that aerosols and clouds have little effect on the propagation of the radio signal. Combined, these advantages mean that this technique can be used from the upper stratosphere down to near the Earth's surface.

However, there are weaknesses: the refracted signal is subject to an ambiguity arising from the signal refraction being proportional to both air density (dependent of pressure and temperature) and water vapour density. Although the measurement contains information about both, there is nothing in the measurement to separate their effects. This problem is not significant except in the lower atmosphere at mid to low latitudes where there are relatively high levels of humidity. In extreme cases, water vapour in the lower atmosphere can also introduce multi-path effects (§2.4.1), which occur when the transmitted signal splits into two or more near-parallel rays which may lead to conflicting received signals. This effect is primarily due to sharp refractive gradients typical of water vapour distribution along the line of sight (Kursinski 1997). A point which is common to all satellite limb-sounding observations is that any meteorological quantities obtained from the measurements represent a weighted mean of the atmosphere over the atmospheric limb path taken (in this case, the transmitter–receiver distance) which is of the order of 250 km.

The GPS Meteorology experiment (GPS/MET), operated by the U.S. University Corporation for Atmospheric Research, consists of a GPS receiver (developed at the JPL) carried by the Microlab-1 satellite. In April 1995, Microlab-1 was launched into a low Earth orbit (at an altitude of approximately 700 km) by a Pegasus rocket of the Orbital Sciences Corporation.

The GPS/MET experiment has only been able to fulfill its potential during short periods of time because GPS signals are heavily encrypted by the U.S. military (§2.2.2). These short periods of time have involved the military disabling their encryption. Using the Abelian retrieval scheme, initial profiles retrieved from radio occultation measurements during one such period agreed well with collocated radiosondes and NWP analyses; results were accurate to within one kelvin, at a vertical resolution of 0.5–1 km in the upper troposphere and lower stratosphere (Kursinski et al. 1996). Utilising *a priori* temperature information, theoretical retrieval error estimates of water vapour at

low latitudes are reported to be accurate to 10−20% below the 6−7 km altitude, and ∼5% or better within the boundary layer (Kursinski et al. 1995).

Chapter 2

The Global Positioning System and radio occultation

There are currently two Global Navigation Satellite Systems: the U.S. Global Positioning System (GPS) and the eponymous Russian GLONASS; the latter is not discussed here but the general principles are similar.

The development of GPS, also known as NAVSTAR[1], began in the early seventies. Initially intended for military use, it was designed so that a user at any point on or near the Earth could obtain three-dimensional coordinates within a few seconds, using a suitable receiver.

This chapter gives a brief description of the GPS, including details of its signals; the methods used to receive signals; the radio occultation technique; and the GPS Meteorology (GPS/MET) experiment.

Further details on the GPS may be obtained from Leick (1990) or Dixon (1991); further details on its many applications may be obtained from Palmer (1997).

2.1 System configuration

The GPS comprises three segments: the space segment, the control segment and the user segment. The space segment is the constellation of 24 GPS satellites (of which three are operational spares) that encircle the Earth at an approximate altitude[2] of 20,200 km; which are distributed equally over six orbital planes with an inclination of 55°. The orbital period of each satellite is approximately 12 hours. A minimum of four satellites

[1] Navigation System using Timing and Ranging

[2] At this altitude level air drag and the effects of gravity anomalies are usually insignificant.

is required to compute the latitude, longitude, altitude (with reference to mean sea level), and GPS time. The system is configured so that at least five satellites can be viewed at any one time.

The control segment consists of a worldwide system of tracking and monitoring stations. The monitoring stations measure the signals from the GPS satellites and relay the information on to the Master Control Station at Colorado Springs, U.S.. The collected data are then used to help compute the coefficients of precise orbital models for the entire GPS constellation. This information is then formatted into updated navigation messages for each satellite.

The user segment comprises the users and receivers used by the military and civil communities.

2.2 GPS signals

Each satellite transmits on two L-band radio frequencies: L1 (1.57542 GHz; $\lambda \sim$0.19m) and L2 (1.2272 GHz; $\lambda \sim$0.24m). These carrier frequencies are coherent multiples of a 10.23 MHz on-board atomic clock[3], which provides a frequency standard on each satellite. Each GPS satellite carries several Caesium[4] clocks, where the 'clock' is defined by an atomic resonance using the energy difference between two states.

The Caesium clocks aboard the GPS satellites have instabilities approximately an order of magnitude smaller than the time scales involved with radio occultation (Kursinski et al. 1997).

These clocks are continually monitored with the 'master' clock located at the Falcon Air Force Station near Colorado Springs and can be easily updated as required; an error of a millionth of a second translates to an error of about 300 metres in position. The master clock is a Hydrogen Maser clock[5]. However, the receiver clocks are usually of comparatively poor quality, typically using a quartz crystal; this has to be corrected for

[3]This clock frequency is actually set slightly lower than this nominal frequency to account for relativistic effects, so that the observer on the ground 'sees' 10.23 MHz almost exactly.

[4]Caesium has an oscillation frequency of 9.192 631 770 GHz, and has an accuracy of 1 in 10^{13} seconds

[5]Hydrogen Maser has an oscillation frequency of 1.420 405 751 GHz, and has an accuracy of 1 in 10^{14} seconds

(§2.2.3). The quartz oscillator used in the GPS/MET experiment has an accuracy of 1 in 10^{10} during a 1 second integration time. This level of stability is 2–3 orders of magnitude worse than a GPS satellite clock, and is generally large compared to other error sources (Melbourne et al. 1994).

The L1 signal consists of two components: a high-frequency precise ranging code (P code of frequency 10.23 MHz), and a lower frequency coarse acquisition code (C/A code of frequency 1.023 MHz); the two codes are modulated 90° out of phase relative to each other, so they can be identified independently. The L2 signal is modulated only with a P code.

Higher accuracy observations are available by utilising the P codes rather than the C/A code because the frequencies are higher, and thus contain more information. Another advantage is that the L1 and L2 P codes can be 'combined' to remove ionospheric effects (§4.2).

All three carriers are modulated with a 50 bit s^{-1} data stream. The data stream is divided into 30-second frames (1500 bits of information). Each frame is divided up into 6-second sub-frames (300 bits of information). The first subframe contains the correction for the onboard clock, including the current error and drift rate of the clock. The second and third subframes contain the satellite ephemeris which consists of precise information about the position of the host satellite, including satellite clock time, and some ionospheric data. The fourth subframe contains the navigation messages and satellite health status information; and the fifth subframe contains the constellation almanac giving the approximate positions of all the other satellites in the system. In each 30-second frame, the almanac for a different satellite is contained in subframe 5. Thus, over a transmission interval of 12 minutes, each satellite tells the user where all the other satellites in the constellation are located in space. The almanac information is not used directly to calculate the navigation solution, it is used to determine which satellites are currently above the horizon and which are most favourably located to provide the smallest error in the calculated position solution (§2.2.1).

These codes, $P(t)$, $C(t)$ and $D(t)$ (P, C/A and 50 bs^{-1} data stream, respectively) can be considered as square waves with values of $\pm$ 1 and are termed as pseudo random

noise (PRN) because they have sufficiently long repeat times that they appear random to a user without knowledge of the code structure, e.g. the P codes have a repeat time of approximately 37 weeks. Each satellite is assigned a unique portion of the code with a repeat time of one week, and, since the number of active satellites will not exceed 24, each satellite can be uniquely identified by its PRN number, corresponding to the code portion transmitted. The L1 and L2 signals can be represented as:

$$S_1(t) = A_{P1}P(t)D(t)\cos(\omega_1 t) + A_C C(t)D(t)\sin(\omega_1 t), \quad (2.1)$$

$$S_2(t) = A_{P2}P(t)D(t)\cos(\omega_2 t) \quad (2.2)$$

where ω_1 and ω_2 are the angular frequencies associated with the L1 and L2 carriers, and A_P and A_C are the relative amplitudes of the P and C/A codes, respectively. The amplitude of the C/A is higher to facilitate lock-on.

One effect of the PRN code modulation on the carriers is to spread the energy of the P code signal ± 10.23 MHz around the carrier centre frequency; because of this wide bandwidth the GPS signal is often referred to as a spread spectrum.

2.2.1 Observables

There are two important types of observable associated with the GPS signals: pseudo ranges and carrier phases (Leick 1990).

The pseudo range is a measure of the distance between the satellite and the receiver at the epochs of transmission and reception of the signals, respectively. This itself is subject to an iterative calculation since the satellites will move in the time it takes the signal to be transmitted and received. The PRN code from each satellite is compared with a replica of the PRN code which can be generated from a receiver with knowledge of the code structure and an internal clock. A code-tracking loop is then used to look at the two codes and find the point of maximum correlation. This process is subject to unavoidable errors due to satellite and receiver clocks, and atmospheric effects; consequently, the

measured value (pseudo range) will differ from the true geometric distance. This method can be applied both to C/A and P codes.

Utilising observations from four satellites, giving four unknowns (the three spatial dimensions and the satellite time), yields the receiver position and satellite clock error leading to the evaluation of the navigation solution. These pseudo range measurements are used in differential GPS (DGPS) which will be discussed later in §2.2.3. The positional accuracy achieved by using one receiver and no corrections is approximately 10−20 metres if the P codes are used and 100 metres if the C/A code is used (Leick 1990).

The carrier phase measurement is the difference between the phase of the carrier signal of the satellite, measured at the receiver, and the phase of the local oscillator within the receiver at the epoch of measurement t. Phase measurements are inherently more accurate than pseudo ranges since the wavelength involved with phase is much smaller than the wavelengths associated with the codes.

Carrier phase is not measured directly, as this would require very high sampling rates. Rather, the signal is heterodyned with a signal generated by the receiver's internal clock (local oscillator) and, after band-pass or low-pass filtering, the resulting lower frequency 'carrier beat phase' is sampled. Positional accuracy using the phase as the observable is approximately 1 metre using the P codes.

Observation geometry also affects the quality of the resulting 3-d positioning. Satellites that appear close to one another in the sky provide correlated (and therefore redundant) range information: an effect known as the geometric dilution of precision (GDOP). Geometric effects can be minimised by choosing satellites which maximise the volume of a tetrahedron, defined by the points of intersection on a unit sphere centred on the user, using vectors between the satellites and the ground receiver.

2.2.2 Military encryption

In practice, the accuracy achieved in using the GPS signals is degraded due to encryption imposed by the U.S. military. Two techniques are employed to limit access and prevent

jamming of GPS receivers from false signals, and limit accuracy of the data: anti-spoofing (AS) and selective availability (SA), respectively.

The AS technique is used to encrypt the P codes (used for high precision positioning measurements), reducing the system's accuracy from 10 metres to 100 metres. With the imposition of AS, the P codes are combined with a secret 500 kHz W-code, to create the encrypted Y-code; only receivers with knowledge of the W-code can make use of the P codes. The military can also introduce an error in the satellite clock data, so that transit times cannot be calculated precisely and/or alter the data message so that satellite positions are computed incorrectly: this process is known as SA.

2.2.3 Decryption

There are several methods available to overcome the loss of accuracy imposed by the military encryption, some in real-time and others post-measurement. Many of the methods involve more than one receiver.

Differential GPS makes use of the codes (C/A, or P if possible) at a roving receiver and corrections to the pseudo ranges transmitted by radio from a nearby fixed base station (assuming that the roving receiver utilises the same GPS satellites as the fixed receiver uses) to calculate an absolute position of the roving receiver in real time (figure 2.1). The actual pseudo ranges at the fixed receiver are unimportant; only their apparent errors are transmitted, which vary relatively slowly. The accuracy involved with careful measurements can approach 1 metre. Making use of the carrier phase measurements is also possible with this method.

Relative GPS (RGPS) measurements usually consist of phase measurements. Differencing range measurements with respect to satellite A and two receivers (i and j) improve the knowledge of errors in the orbit and satellite clock. This combination of measurement is called a single difference (SD) and represents the simplest form of RGPS (figure 2.2).

If an SD is then formed with respect to satellite B and the same two receivers, and differenced with the SD from satellite A, a double difference is formed. This combination

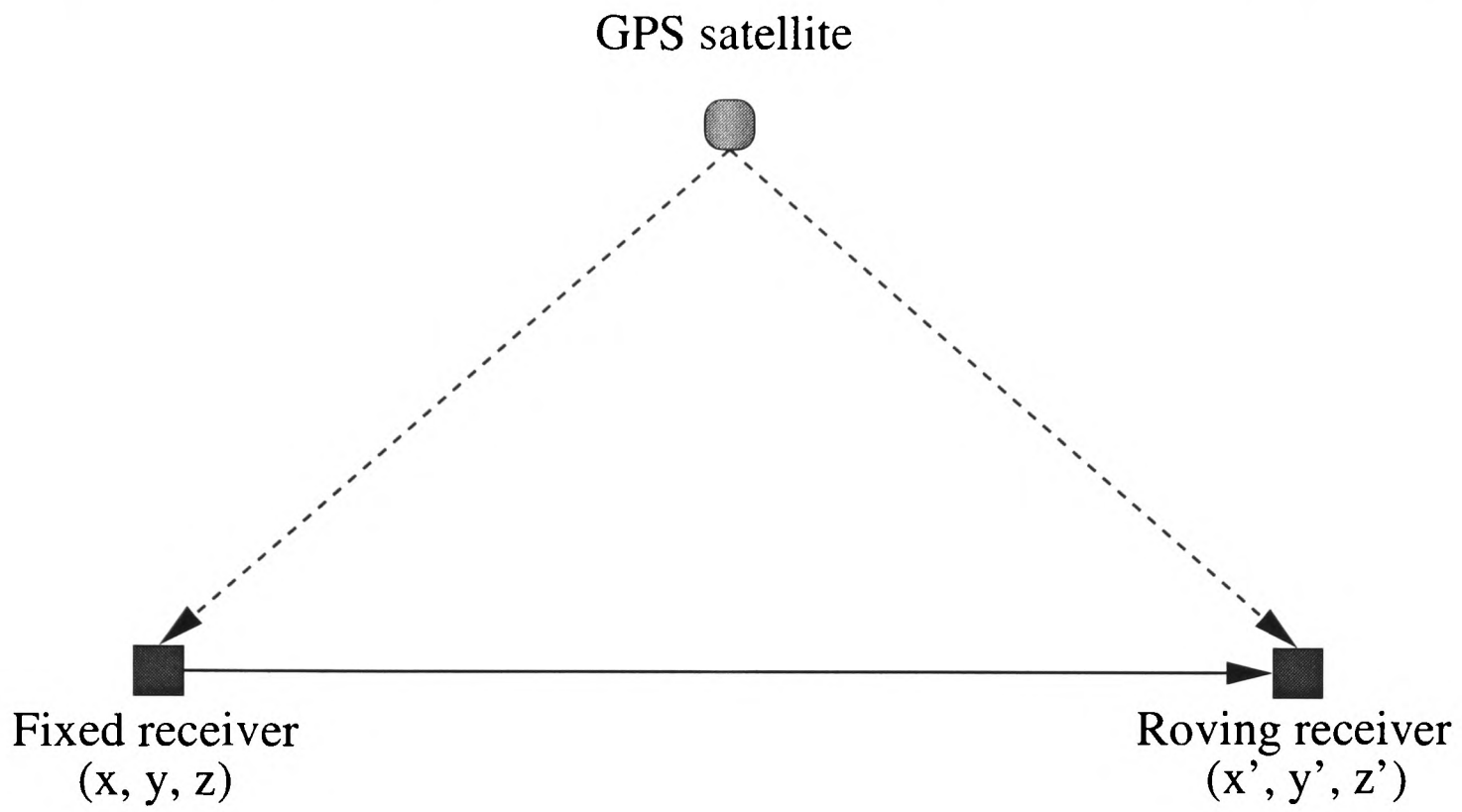


Figure 2.1: The basic principle of differential GPS geodesy. The fixed receiver (of known position) calibrates the signal by calculating its position exclusively from GPS data, thus computing the encryption errors on the GPS signals. The roving receiver which is located nearby, i.e. close enough so that it uses the same satellite signals as the fixed receiver, can utilise the errors computed from the fixed receiver to correct its position. The broken lines represent satellite–receiver signal and the unbroken lines represent the propagation of error information from the fixed to the roving receiver.

of measurements from satellites A and B, and receivers i and j, effectively cancel errors common to the receiver clocks. If this method is used in conjunction with a receiver which has been synchronised with the clocks on board the GPS satellites, then the SA encryption is effectively cancelled. The accuracy achieved using this method can be centimetres.

By utilising a local network of receivers, relative accuracy achieved can be millimetres[6]. This level of accuracy is achieved by leaving the roving receiver at the same location for a suitable period of time, typically several hours [Dr. Peter Clarke, Dept. Earth Sciences: private communication]. Since there is an error attached to the signal, the measured position will vary in a random manner around the true position. A weighted average of the measured position over time can lead to an accurate estimate of the true position.

[6]This level of accuracy is required for measuring crustal deformation which often occurs on the millimetre scale

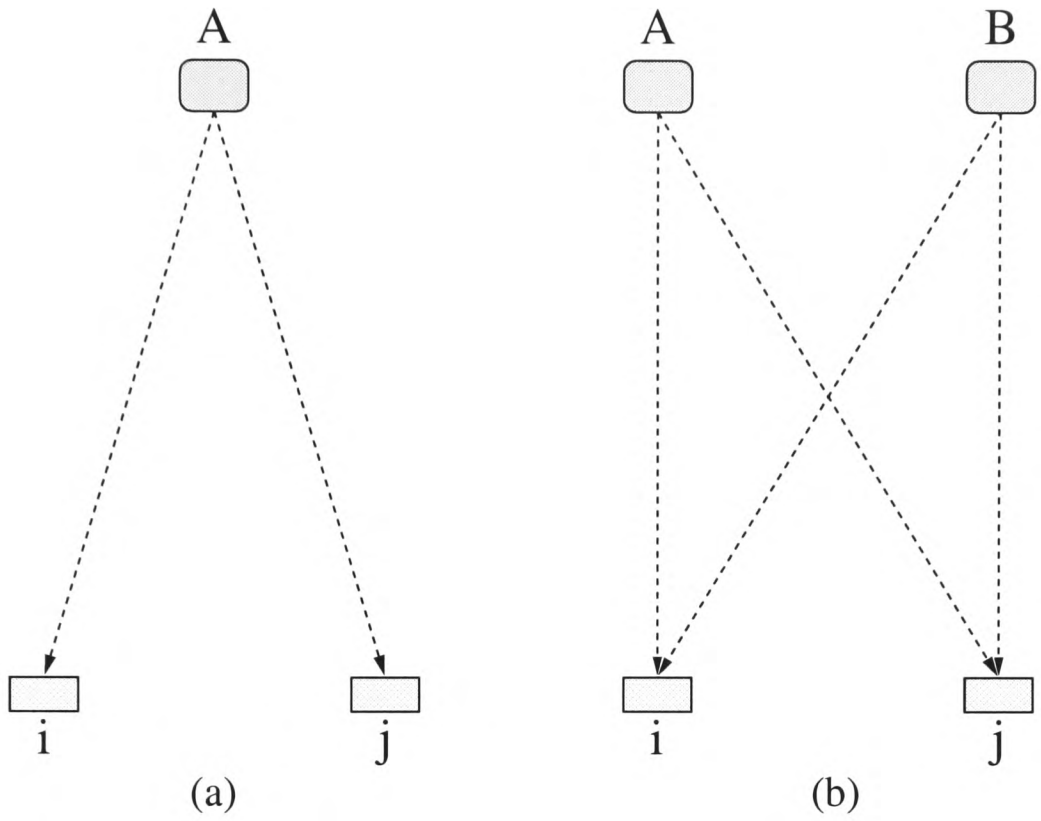


Figure 2.2: The basic principle of relative GPS measurements. (a) A single difference is formed by differencing range measurements with respect to satellite A and the two ground receivers i and j. This process can be used to improve the knowledge of errors in the orbit and satellite clock. (b) If a second single difference is formed by utilising satellite B and the same two receivers, and differencing it with the single difference from satellite A, a double difference is formed. This process effectively cancels the errors common to the receiver clock.

This method involves transferring and comparing large quantities of data, so is usually a post-measurement processing method.

2.3 Receiver technology

If AS encryption is present on the GPS signals, receivers may still obtain the L1 phase as the C/A code is unrestricted. However, there are several strategies available in order to obtain the L2 phase from the Y-code encrypted signal on L2. These strategies are usually in the form of receiver circuitry.

Older receivers usually work by multiplying the incoming L2 signal by itself. As the PRN code is essentially a series of polarity changes, squaring the signal removes all traces of the code. The resulting signal is known as the L2-squared ($L2^2$) signal. Using this method, the frequency of the $L2^2$ carrier is double that of L2 (i.e. the wavelength is halved, which makes ambiguity resolution more difficult), and the noise of the new

signal is increased by a factor of $\sqrt{2}$. Also, this method does not yield either P1 and P2 pseudoranges.

A receiver with knowledge of the code structure and an internal clock can recover an estimate of signal transit time by cogenerating the code sequence and performing cross-correlation between the received signal and its internal code, determining the time delay necessary to match the two sequences. Modern cross-correlating receivers utilise the fact that the P code is present on both L1 and L2 carriers. Cross-correlating the P code between the L1 and L2 yields the phase lag between L1 and L2 phase. A proxy for the $P2$ pseudorange can be reconstructed given the C/A code pseudorange and L1 – L2 phase lag, but the resultant information will only have the precision of C/A and not the precision of P code.

Other modern receivers track the P code on L2 in short windows between W-code polarity reversal and use this tracking to get $P2$ pseudorange. The L2 signal is then squared to get the $L2^2$ phase.

The most advanced receivers correlate the P code in short time windows between W-code polarity reversal on both L1 and L2 signal, and thereby obtain a real-time estimate of the W-code. This estimate can then be removed from the signals to recover true $P1$ and $P2$ pseudoranges, and full-wavelength L2 phase, with much less signal degradation than other techniques.

2.3.1 Loss of signal

If a receiver momentarily loses lock on the signal from a satellite (e.g. because the signal path is temporarily blocked by an obstruction), it attempts to extrapolate the signal using its internal clock during the break. However, the internal clock is relatively inaccurate so that the projected phase measurement may differ from the true value. When phase lock is achieved again, the projected phase measurements are 'snapped' back to the actual phase, but the number of complete cycles during this outage may differ from the true number by an integer amount. This is known as cycle slip and must

be eliminated from carrier phase measurements for the GPS to work properly (Leick 1990).

Results from the GPS/MET receiver show that when the signal is lost momentarily the associated SNR goes to zero until the signal is picked up again [Dr. R. Kursinski: private communication]. The general consensus is that once the receiver has lost the signal, atmospheric data below that level is not reliable and therefore should be omitted from any analysis.

2.4 GPS radio occultation

The basic observable of GPS radio occultation is the phase shift of the L1 and L2 signals (time delay) imposed by the atmosphere (both ionosphere and neutral). The Doppler shifts of the signals are the derivatives of the uncorrected atmospheric phase shifts centred about the time given as epoch. The cumulative effect of the atmosphere on the ray path can be expressed in terms of the total refractive bending angle ε, as a function of the tangent point impact parameter a, i.e. the rays' nearest approach to the Earth. The impact parameter may be defined as the perpendicular distance between the centre of local curvature at the tangent point of the ray and the asymptotic straight line followed by the ray as it approaches the atmosphere (figure 2.3).

Profiles of bending angles can be constructed from both sets of Doppler shifts (§4.1) and then corrected for the ionosphere (§4.2). Further details of the measurements from GPS/MET may be found in §2.4.4.

The first proposals to apply refractometric occultation methods for remote sensing of the Earth's atmosphere from space were made in the mid-seventies. For many years the capabilities of space-based techniques for positioning and Doppler measurement were insufficiently accurate[7]. However, during these years some experiments on remote sensing of refractivity from space were carried out, e.g. Rangaswamy (1976). As part of the Apollo-Soyuz Test Project, Apollo was tracked in the satellite-to-satellite

[7]The refraction effect of the Earth's atmosphere is very small: the maximum bending angle is of the order of 0.02 radians (1°) (Hardy et al. 1992)

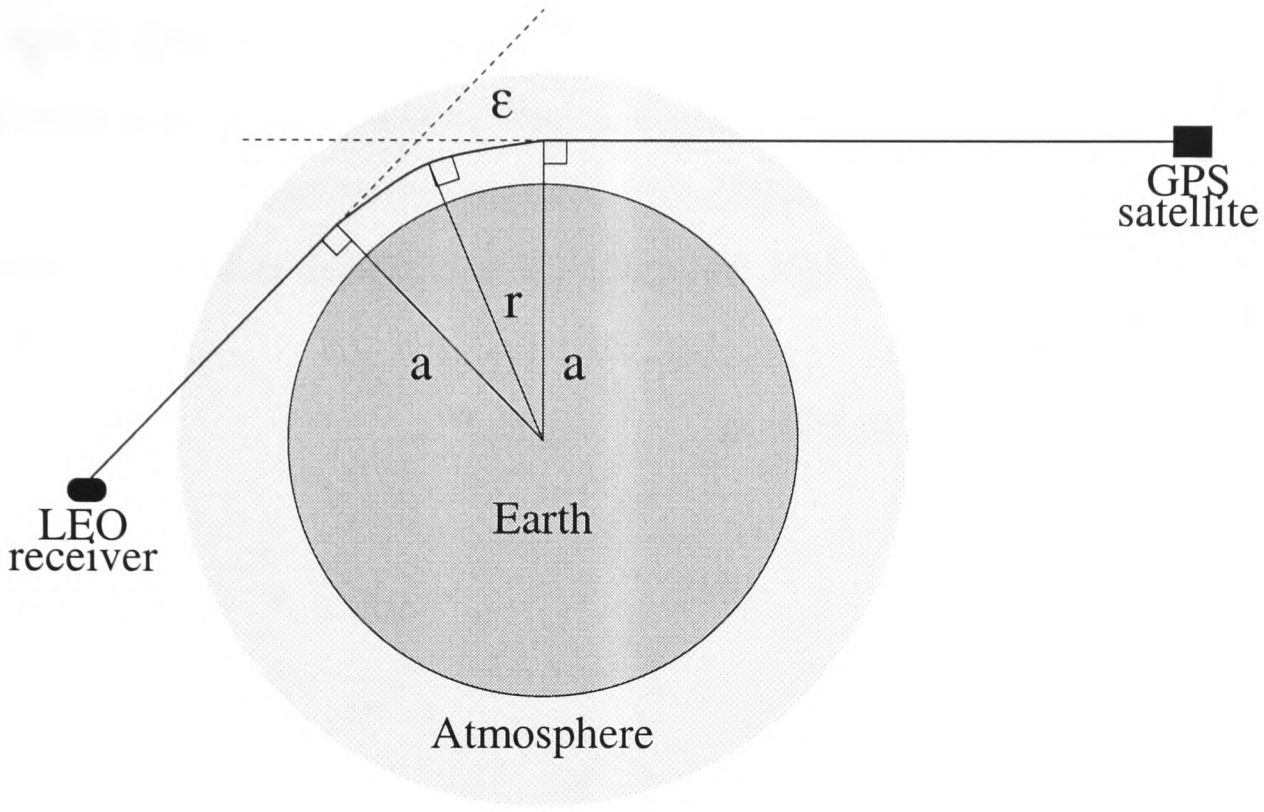


Figure 2.3: A schematic of the GPS radio occultation geometry, defining the occultation bending angle ε, the impact parameter a and the ray tangent radius r. The impact parameter follows the asymptotic dashed line. The true shape of the Earth is more complex than is depicted here, which is reflected in the paths that the GPS signals take; in practice these effects have to be considered.

tracking mode by the NASA synchronous ATS-6 satellite. The tracking data obtained during occultation of Apollo by the Earth was used to retrieve temperature on pressure levels, which was subsequently compared with a nearby radiosonde station. Although this confirmed that the deviation of refractivity could be determined, it was not possible to assess the overall accuracy of the method due to insufficient accuracy of direct measurements, e.g. precise orbit determination. However, the experiments did demonstrate the feasibility of determining pressure levels on geometric height levels in the atmosphere from occultation measurements.

It was not until the advent of the Global Positioning System that accurate measurements became possible. Even now, because of its military origin the P codes are still heavily encrypted (§2.2.2), and it was not until the early 1980s that the potential of this technique with reference to the GPS was realised.

The satellite configuration required to apply radio occultation principles to the Earth's atmosphere includes a receiver in Low-Earth orbit (LEO); two GPS transmitters, of which one is required to be occulted by the atmosphere with respect to the LEO

receiver; and a ground control station, which both transmitters can see. The reason for the ground control station is so that a double difference (§2.2.3) can be formed to eliminate transmitter clock errors and selective availability. The receiver clock error is also corrected since it is usually not of sufficiently high accuracy. The configuration is shown in figure 2.4.

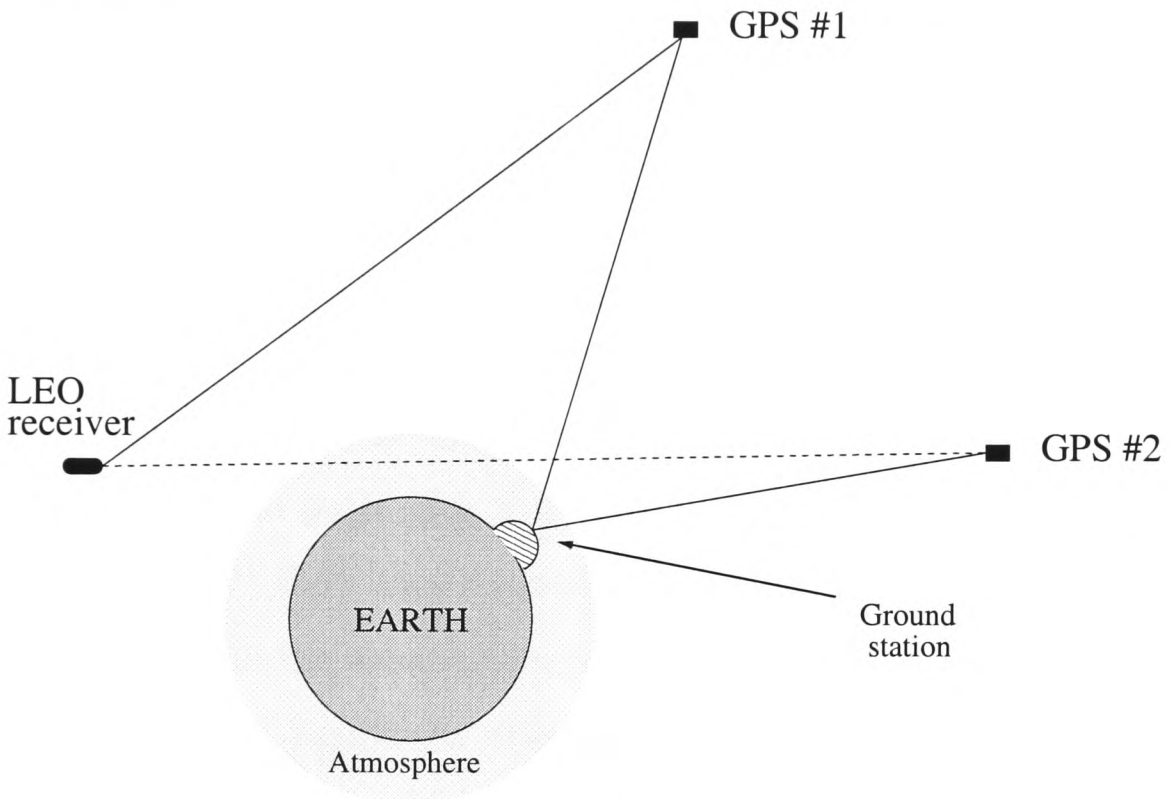


Figure 2.4: The GPS radio occultation configuration. An occultation event requires a receiver in Low-Earth orbit; two GPS transmitters, of which one is being occulted by the atmosphere with respect to the receiver (indicated by the broken line); and a ground control station, which both transmitters can see.

Once the receiver has lock with the transmitter signal, a complete occultation takes approximately one minute. The extent of the vertical range that the occultation sounds is subject to several factors which will be discussed later (§3.2).

Measurements of the time delay become useful for neutral atmospheric sounding when the GPS signal begins to transect the mesosphere at an altitude of about 85 km; at this altitude the atmospheric phase delay is sufficiently large such that a detectable 1 mm (3×10^{-12} s) delay in the dual-frequency carrier wave can be observed by the LEO GPS receiver (Ware et al. 1996).

For the receiver to obtain useful information, the signal-to-noise intensity ratio (SNR) of the received GPS signals must exceed a critical value of approximately 10 (Kursinski et al. 1997); if the SNR value falls below this value, phase lock is lost. The value of SNR

is determined by the GPS transmitter power[8], receiver noise performance, observational geometry and signal attenuation.

2.4.1 Atmospheric attenuation

The primary atmospheric attenuation mechanisms that affect the signal propagation through the limb-path are: absorption by atmospheric constituents; extinction by particle scattering; and defocusing by sharp refractive gradients. Atmospheric absorption at L-band wavelengths is dominated by the far wings of pressure-broadened molecular oxygen and water vapour lines. The contribution from molecular oxygen is at least an order of magnitude greater than water vapour (Kursinski et al. 1997); Chahine et al. (1983) calculated that absorption by molecular oxygen reduces GPS signal intensities by a factor of 2 for a limb path with a tangent height near the surface. Calculations by Kursinski et al. (1997) show that extinction caused by realistic suspensions of water droplets or ice particles in the atmosphere is negligible.

Defocusing is the most significant attenuation mechanism and is caused by rapidly changing horizontal refractivity structures, which lead to adjacent, near-parallel rays entering the atmosphere to bend differentially and diverge (multi-path effects). This mechanism causes the signal intensity to reduce at the receiver. Complete attenuation of the signal from this mechanism, where the refractivity gradient becomes large, effectively causes the radius of curvature of the ray to be smaller than the local radius of curvature of the atmosphere. Under these conditions the ray is unable to leave the atmosphere and ultimately transects the surface (Kursinski et al. 1997).

2.4.2 Spatial resolution

Kursinski et al. (1997) defines vertical resolution as the vertical interval above the tangent height that contributes 50% to net bending. The horizontal resolution of a single ray may be defined by the distance traversed as it enters and exits a layer having a vertical resolution of ΔZ. The resolutions discussed here are primarily determined by

[8]Since GPS was designed for small, hand-held receivers, the signals are relatively powerful.

the geometry of the occultation, i.e. the GDOP (§2.2.1). In this case the vertical and horizontal resolution are related by the approximate expression (Kursinski et al. 1997):

$$\Delta L = 2(2R\Delta Z)^{1/2} \tag{2.3}$$

where ΔL and ΔZ are the horizontal and vertical resolution, and R is the radius of the atmosphere at the ray path tangent height.

The variation of the vertical resolution of an individual bending angle measurement (one in a sequence taken during an occultation) is illustrated in figure 2.5 as a function of tangent altitudes.

The vertical resolution for individual measurements vary from 0.5km (at an altitude of approximately 4 km) to 2 km (at the altitude level corresponding to the tropopause region), which corresponds to horizontal resolutions of 160 km and 320 km, respectively, using equation 2.3. Some of the more complicated structure seen in the lower atmosphere (particularly in the lower troposphere) can be attributed to relatively large fluctuations in the water vapour mixing ratio (also shown by figure 2.5). Figure 2.5 also illustrates the magnitude of the bending at each tangent height. It is clear that generally the magnitude of the bending increases as the tangent height decreases, but increases more slowly in the lower troposphere.

The vertical resolution of the complete occultation is limited by the sampling rate of the instrument but in practice for GPS radio occultation measurements it is limited by Fresnel diffraction. The underlying physics of this phenomena is presented in chapter 3 and a brief summary of the methods currently used to reduce the effects of diffraction are presented in chapter 4. The diameter of the first Fresnel zone is plotted in figure 2.6 as function of tangent altitude (Kursinski et al. 1997).

Figure 2.6 shows that the overall size of the first Fresnel zone diameter increases with height; in particular, there are fluctuations of the diameter in the troposphere which are associated with sharp vertical changes in the refractivity gradient caused primarily by vertical water vapour mixing ratio fluctuations.

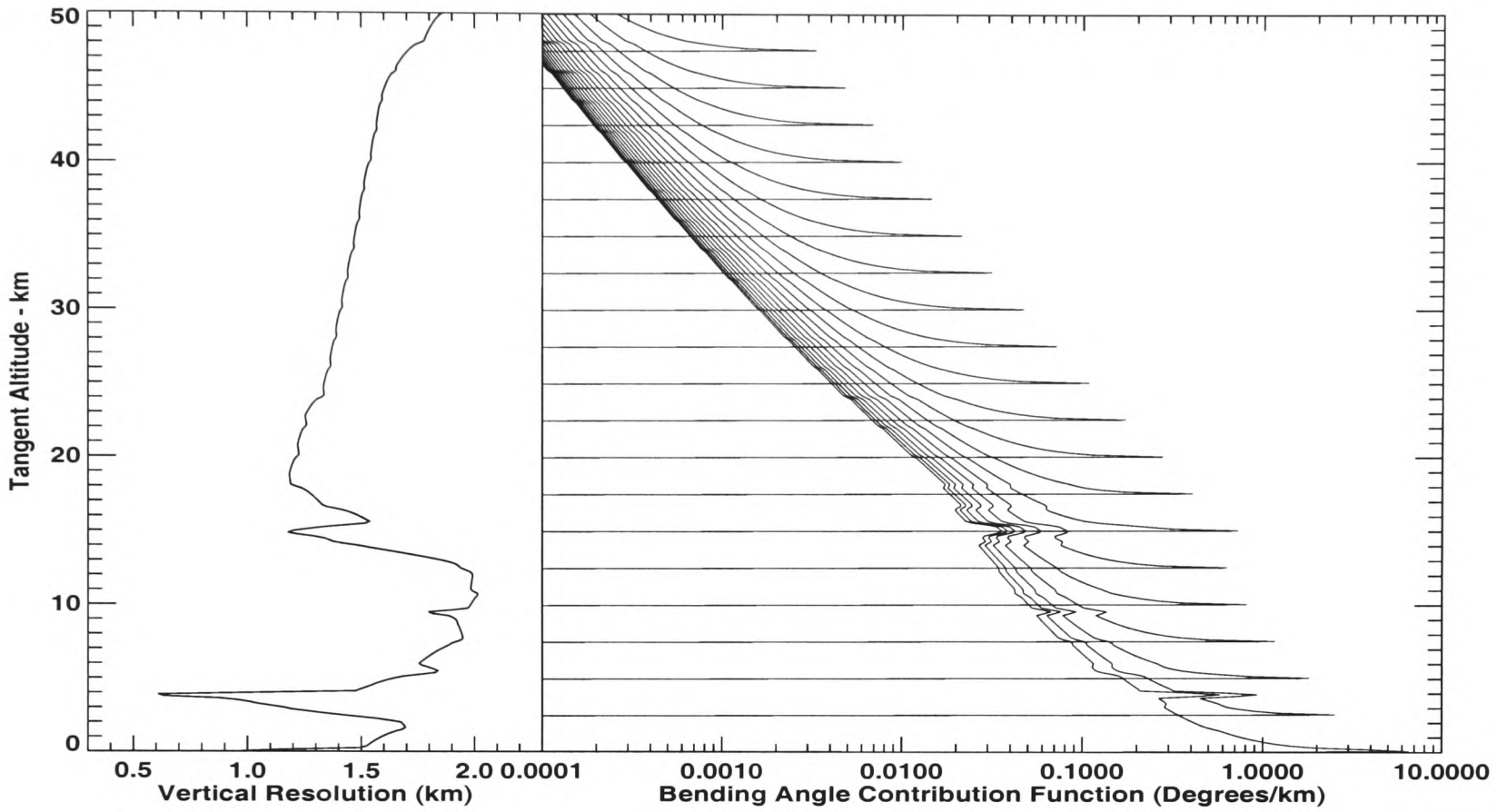


Figure 2.5: Bending contribution functions [sic] (solid lines) and vertical resolution profile for individual radio occultation bending angle measurements plotted as a function of altitude (upper panel). Resolution is defined here as the vertical interval between the ray path tangent heights and the height at which half of the total bending has been accumulated. This profile has been calculated for a refractivity profile derived from radiosonde temperature and humidity profiles (bottom panels), assuming spherical symmetry. Radiosonde ascent 12:00 UT, 11 July 1991, Hilo, Hawaii (19.72°N, 204.93°E). Reproduced, with permission, from Kursinski et al. (1996).

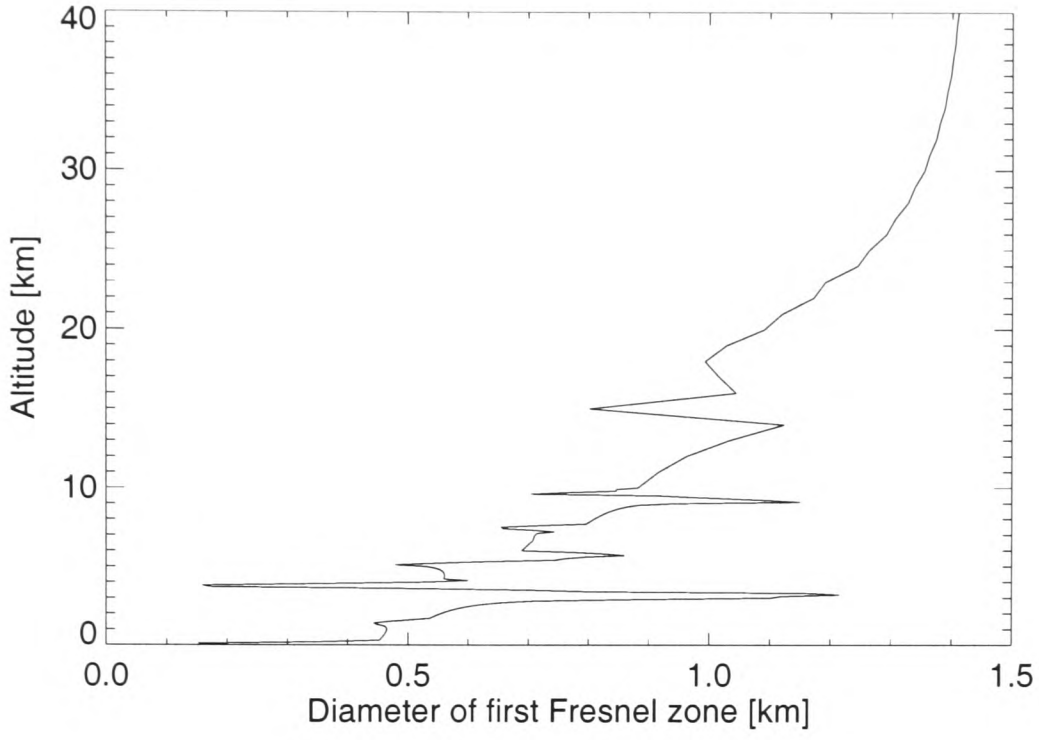


Figure 2.6: The diameter of the first Fresnel zone, defined by equation 3.11, plotted as a function of ray path tangent height for a refractivity profile derived from radiosonde temperature and humidity profiles, assuming spherical symmetry. Radiosonde ascent 12:00 UT, 11 July 1991, Hilo, Hawaii. Reproduced, with permission, from Kursinski et al. (1996).

2.4.3 Ionospheric effects

The ionosphere is a region of the Earth's upper atmosphere in which free electrons arising from ionization occur, mainly as a result of ultra-violet radiation and X-rays from the sun. The GPS signals propagating through the Earth's atmosphere are also affected by the ionosphere. However, the dual-frequency signal can be combined in such a way that this effect that can be reduced to second-order (§4.2). This represents one of the main reasons why the P codes are important.

Sudden fluctuations in the solar wind, the stream of charged particles from the Sun, can cause sudden fluctuations in the density of the ionosphere, which cause geomagnetic storms. It is important to predict such storms since they can cause problems to satellite communication (including GPS) and disruption to short-range communications: power grids break down as the changing fields induce direct currents in power lines built for alternating currents. The errors incurred by these ionospheric disturbances are due to reach a peak in the start of the next century (2001), as the Sun enters the most active phase of its 11-year sunspot cycle. This is one of the reasons why the U.S.

government have been considering the introduction of a third signal to the system. This will effectively increase the accuracy of the system to the present military precision[9].

Further information about the physics of this region are presented in chapter 3 and the methods used to correct for the effects of signal propagation through this region are presented in chapter 4.

2.4.4 The GPS/MET experiment

The GPS/MET experiment, operated by the University Corporation for Atmospheric Research, consists of a GPS receiver developed at the JPL aboard the satellite Microlab-1. In April 1995, Microlab-1 was launched into an LEO by a Pegasus rocket of the Orbital Sciences Corporation, and placed in an Earth orbit of approximately 700 km.

The receiver is a JPL-developed TurboRogue, which has been commercialised by Allen Osborne Associates, California. This unit has a mass of approximately 3 kg and requires less than 18 W of power. The SNRs obtained by GPS/MET are of the order of $(1\text{–}20)\times10^3$ for the present radio occultation experiments (Kursinski et al. 1997).

The receiver has an on-board clock which is driven by a temperature-compensated, voltage controlled oscillator (VCO). This oscillator has a fractional frequency stability of roughly 1 in 10^{10}, which is 2–3 orders of magnitude worse than that of the GPS satellite clocks. There is no significant long-term drift in the oscillator because the receiver is constantly adjusting the VCO to match the GPS constellation time, which it estimates from concurrently tracking 8 GPS satellites. This receiver clock instability is reduced during an occultation by simultaneously tracking a second satellite (at a rate of 50 Hz) which is not involved with the occultation event (§2.4). The occulted and non-occulted signals are then differenced at 50 Hz eliminating the receiver clock noise (Kursinski et al. 1997). The ground network is then utilised to estimate and reduce the SA encryption on each of the 2 GPS satellites tracked during the occultation.

The flight receiver has knowledge of the locations of the high-rate ground sites which is updateable from the ground. The receiver will only initialise an occultation if the

[9]The frequency of the new signal is still being decided upon at the time of writing.

occulted GPS satellite and a second satellite can be viewed from one of the high-rate ground receivers.

With AS enabled, the TurboRogue has a codeless operating mode that cross-correlates the Y-code on L1 with the Y-code on L2 to produce a measure of the ionospheric delay on the signal from which the measured L1 carrier phase can be corrected, and the L2 carrier phase can be inferred (Kursinski et al. 1997).

For GPS/MET the atmospheric phase shifts (at both carrier frequencies) are measured at a 50 Hz sampling rate as the ray path traverses the limb. It may be defined as 'the total path length of the ray less the vacuum path length between the transmitter and the receiver'. To reduce diffraction effects, these data are smoothed with time constants slightly less than the time required by the ray to vertically traverse the local Fresnel diameter at the limb. The atmospheric phase shifts for L1 and L2 are obtained by sampling the smoothed curve at the time given by epoch [from the JPL literature written to accompany the GPS/MET dataset, also see chapter 7].

The differencing scheme used to calculate an accurate profile of bending angles introduces, or magnifies, errors already present on the signals. The level with which the errors increase due to differencing is summarised in table 2.1:

Difference scheme	No of GPS Satellites	No of receivers	Reduces errors in	Thermal & multipath error magnification
Single satellite difference	1	2	SA and satellite clock error	$\sqrt{2}$
Single receiver difference	2	1	Receiver clock error	$\sqrt{2}$
Double difference	2	2	SA and all clock errors	2

Table 2.1: A summary of the error magnification factors incurred by several differencing schemes.

It can be seen that the thermal and multipath errors are both magnified by

differencing. These factors are due to the fact that there are both thermal and multipath effects on each of the difference links, and since these error sources are uncorrelated, they combine in such a way as to increase the total error on the signal.

The GPS/MET experiment control segment

Because of the control segment requirement during an occultation (§2.4), the GPS radio occultation concept requires a globally distributed ground network of approximately 12 GPS receivers which can concurrently track the dual-frequency signals. There are more than 100 globally distributed permanent tracking stations and several operational data and analysis centres (ESA 1996). The majority of these stations were originally built for geodetic research and ionospheric monitoring.

The importance of clock noise, which limited many of the original planetary measurements (§1.4), was analysed and it was decided by the JPL to supplement the existing geodetic ground network of GPS receivers with a small network of high-rate ground stations. Utilising these high-rate stations, the errors introduced by clock noise (from both transmitter and receiver) are dramatically reduced (Kursinski et al. 1997).

In the absence of the AS encryption, the primary sources of clock noise include the instability of the receiver clock and the SA encryption. It is possible to reduce the receiver clock instability by improving upon the existing on-board reference oscillator. The JPL decided that the high-rate ground network should sample the signal at 1 Hz, and the space-borne receiver sample at a rate of 50 Hz. The sample rate of the high-rate ground network is adequate to eliminate the SA encryption (Kursinski et al. 1997).

The high-rate GPS receiver network consists of 6 ground stations (at the time of writing) located in: Australia (Tidbinbila), Alaska (Fairbanks), Hawaii (Kokee Park, Kauai), Germany (Pottsdam), Antartica (McMurdo), and the JPL (California, U.S.) [Dr. R. Kursinski: private communication]. The geographical positions of these receivers were primarily chosen to maximise the global coverage possible, but other constraints such as the number of spare high-rate receivers available, the fact that the placed receivers had to be under the sole control of the JPL (except the receiver placed in Germany) and placed just before the launch of GPS/MET, also constrained the receiver locations.

The McMurdo receiver was added sometime after the GPS/MET launch in an attempt to increase the southern hemisphere coverage (particularly over the Indian Ocean), and therefore it was not present for the early data [Dr. R. Kursinski: private communication].

The original geodetic network of GPS receivers sample at 1/30 Hz; if the SA encryption is disabled, as is being suggested at the time of writing, the supplementary high-rate GPS receiver network will no longer be needed.

The 1 Hz data from the high-rate receivers are transferred to the JPL daily. Although this preliminary receiver network provides adequate coverage for a demonstration occultation mission, it is does not provide uniform global coverage as would an ideal operational system. There are plans for an additional 3–4 high-rate ground GPS receivers to be added to the network to remove 'holes' over Northern Africa, the mid-Atlantic and the Indian Ocean, in support of the Danish Ørsted and the South African SUNSAT space-borne receivers planned for launch sometime in late 1998/early 1999.

A receiver capable of observing both rising and setting occultations, given the current distribution of satellites and ground stations, would obtain approximately 670 occultations profiles per day (ESA 1996), globally distributed; however, for GPS/MET the number of soundings is less than the theoretical maximum due to: unfavourable geometry suitable for useful soundings (§2.2.1); antenna pointing and memory limitations for the MicroLab-1 satellite; and the requirement for simultaneous tracking of the occulting GPS satellite by the LEO receiver and by one of the high-rate ground-based GPS tracking stations (Ware et al. 1996). In practice, approximately 250 occultations occur each day; however, with these limitations the number of useful occultations per day, using the GPS/MET receiver, is 150 (Rocken et al. 1997).

Chapter 3

The forward radio occultation problem

This chapter will describe the physics of the forward radio occultation problem and discuss some of the factors that may limit the accuracy of its modelling. The forward model is used to map from geophysical parameters to measurement space. The numerical formulation of the forward model and its implementation are presented in chapter 5.

3.1 Forward modelling of problem

This section will present the general theory assuming that the atmosphere is horizontally homogeneous (hereafter known as spherical symmetry). Details on the necessary modifications to this theory for the non-spherically symmetric case may be found in Eyre (1994) and chapter 5.

3.1.1 Refractive medium

Given knowledge of the atmospheric state (in this work described by temperature and humidity on pressure levels, surface pressure and height above the geoid), a profile of refractive index is derived.

In general, the refractive medium described by the Earth's atmosphere can be split into several different components of differing origins when considering radio frequencies, Smith and Weintraub (1953) (for first two terms) & Kursinski et al. (1997).

$$n - 1 = \frac{c_1 P_a}{T} + \frac{c_2 P_W}{T^2} - \frac{f_p^2}{2f^2} + c_3 \times 10^{-6} W \tag{3.1}$$

where n is refractive index, P_a is the partial pressure of dry air (hPa); T is the temperature (K); P_W is the partial pressure of water vapour (hPa); c_1, c_2, and c_3 are constants of proportionality with values equal to $(7.76 \pm 0.02\%) \times 10^{-5}$ (K/hPa), $(3.73 \pm 1\%) \times 10^{-1}$ (K^2/hPa), and 1.4×10^{-6} (m^3/g); f_p is the plasma frequency (Hz); f is the transmitter frequency (Hz); and W is the liquid water content in grams per m^3.

The constants used, c_1 and c_2, have been most recently measured by Thayer (1974). The relatively large uncertainty attached to c_2 is due to the difficulty of measuring atmospheric water vapour density. These constant values are valid for frequencies below 20 GHz and normally encountered values of pressure, temperature, and humidity. Further discussion of these parameters can be found in appendix B

The first term of equation 3.1 is the dry term, the second is the moist term, the third is the ionospheric term, and the fourth is the scattering term.

The dry term is due to the polarizability of molecules in the atmosphere, that is, the ability of an incident electric field to induce a charge separation within the molecules. This term dominates for altitudes between 0–90 km, with the moist term becoming equally important in the lower troposphere. The P/T terms are used to define refractivity because they are proportional to the number density of the atmospheric medium; Bean and Dutton (1968) provide a detailed explanation.

The delay due to the water vapour occurs because it possesses a permanent dipole moment; it is the only common atmospheric constituent to do so for radio frequencies. The dipole moment is caused by an asymmetric charge distribution and it is this which affects the propagation of the electromagnetic ray.

The ionospheric term is a first-order approximation of the Appleton-Hartree equation (equation 3.13). It is primarily due to free electrons in the ionosphere at altitudes higher than 80 km.

The scattering term is due to water droplets suspended in the atmosphere, and

is found to be negligible compared to the dry and moist terms for radio frequencies (Kursinski et al. 1997).

3.1.2 Path taken through the refractive medium

Snell's law can be used to describe adequately the path taken by electromagnetic radiation (when described by a ray) when traversing a region of varying refractive index. It can be written as

$$n \sin \alpha = k \tag{3.2}$$

where α is the angle between the path of the radiation and refractivity gradient vector, and k is a constant.

If it is assumed that the path taken by the ray remains in the plane described by the GPS and LEO satellites, the path taken can be described by the geometry defined in figure 3.1, and the change in direction can be expressed in polar coordinates (see Born and Wolf (1980) for full derivation).

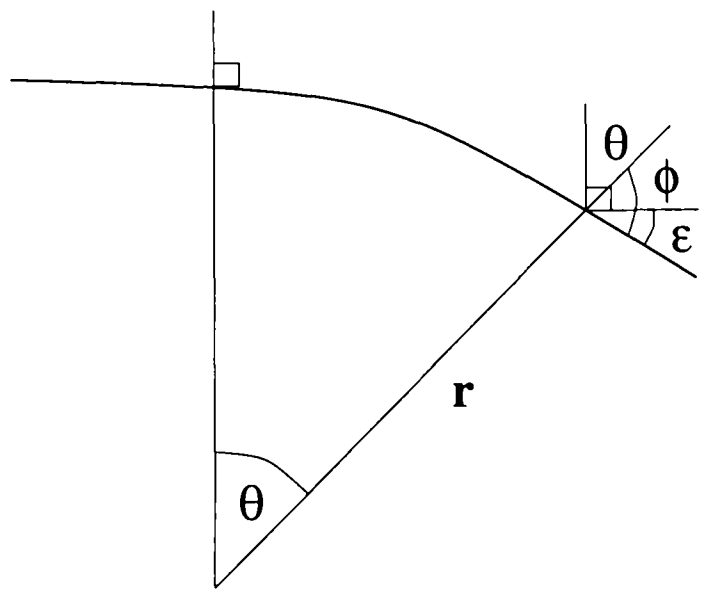


Figure 3.1: Defining the geometry of the forward refraction problem.

$$d\varepsilon = -\frac{1}{n}\left(\frac{\partial n}{\partial r}\right)_{\theta} r d\theta + \frac{1}{n}\left(\frac{\partial n}{\partial \theta}\right)_{r} \frac{dr}{r} \tag{3.3}$$

If ϕ is the angle between the path and the local radial, then

$$\tan\phi = r\frac{d\theta}{dr} \tag{3.4}$$

and

$$\varepsilon = \theta + \phi - \frac{\pi}{2} \tag{3.5}$$

If equations 3.3, 3.4, and 3.5 are combined

$$d\varepsilon = d\theta + d\phi = \tan\phi\ \frac{dr}{r} + d\phi = -\frac{1}{n}\left(\frac{\partial n}{\partial r}\right)_\theta\ \tan\phi\ dr + \frac{1}{n}\ \left(\frac{\partial n}{\partial \theta}\right)_r\ \frac{dr}{r} \tag{3.6}$$

If spherical symmetry is used to approximate the problem ($(\frac{\partial n}{\partial \theta})_r = 0$), equation 3.6 reduces to

$$\frac{d\phi}{\tan\phi} + \frac{dr}{r} + \frac{dn}{n} = 0 \tag{3.7}$$

which, when integrated gives

$$rn\sin\phi = constant = a \tag{3.8}$$

where a is the impact parameter (defined in §2.4). This equation is called Bouguer's formula (Born and Wolf 1980).

Defining $x = rn$, the refractive radius, and combining equations 3.6 and 3.8

$$d\varepsilon = -\left(\frac{\partial \ln n}{\partial x}\right)_\theta a(x^2 - a^2)^{-\frac{1}{2}}dx \tag{3.9}$$

The total refracted angle is obtained by integrating over the full limb path, noting that $x = a$ at the tangent point.

$$\varepsilon(a) = \int d\varepsilon = -2a \int_a^{+\infty} \left(\frac{\partial \ln n(x)}{\partial x} \right)_\theta (x^2 - a^2)^{-1/2} \, dx \tag{3.10}$$

This equation linking the total refracted angle over the limb path taken as a function of impact parameter and the refractive index is an integral which can be transformed using the Abel transform.

3.2 Factors that limit modelling of the problem

There exist a number of factors that limit the accurate modelling of the physics involved with the problem and subsequently the retrieval of meteorological parameters (chapter 4). Factors specific to the modelling are presented and discussed in this section.

Modelling the Earth's shape

In three dimensions, the Earth's shape can be approximated as an oblate spheroid, with addition high-order spherical harmonics.

Because the shape of the Earth, local to the occultation, can affect the accuracy of the retrieved atmospheric parameters, it is important to reproduce the shape as accurately as possible.

Local curvature may be parameterised by the local radius of curvature (LRC), defined as the radius of a circle with curvature equal to that of the given curve at that point. The LRC is illustrated by figure 3.2 for the two-dimensional case.

In general, the LRC is not equal to the distance from the Earth's centre of gravity to its surface. Note that the LRC varies as a function of the occultation plane, as well as the latitude.

Fitting a curve to a three-dimensional shape is subject to greater problems, but is possible by using accurate information of the satellites' position throughout an occultation event. Syndergaard (1998) has shown that if the Earth's oblateness is not

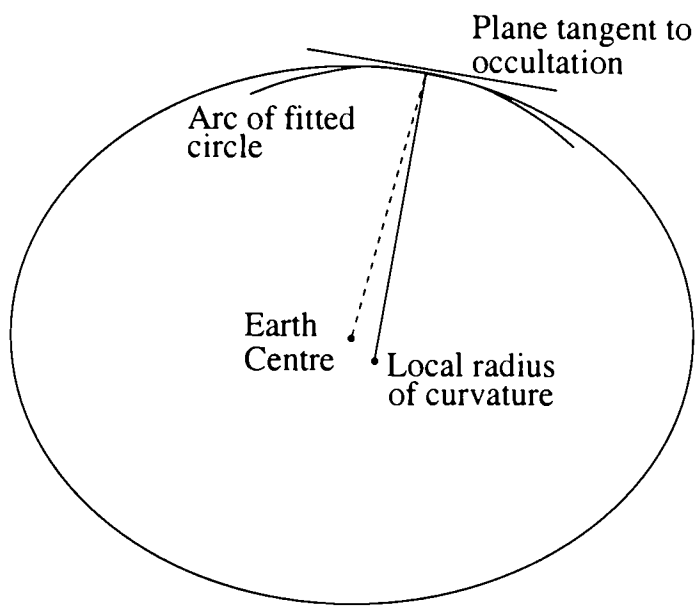


Figure 3.2: A schematic to illustrate the radius of curvature. By fitting a circle to an elliptical Earth it is possible to obtain the local radius of curvature.

correctly dealt with, in extreme cases, the errors in temperature may be as large as 3 K at 10 km altitude and 6 K at the surface.

At present, the method used by the JPL (for GPS/MET data) to calculate the LRC is the following. The ellipsoid (describing the Earth's shape) is expanded until it intersects the line describing the tangent to the radio signal path at a representative value of impact parameter. The normal to the point of intersection and the line describing the path taken by the radio signal defines a plane; this plane section of the original ellipsoid defines an ellipse. A circle is then fitted to the calculated ellipse at a representative time during the occultation and the LRC is calculated.

There are problems with the representative time at which to calculate the LRC at present, mainly due to the JPL taking it to be ~50 seconds before the end of the occultation, which means that the value calculated for the local radius of curvature is a nominal value.

Diffraction effects

As with many other remote sounding techniques, e.g. some forms of limb-sounding, radio occultation is limited by diffraction effects. Since these effects are not included in the geometric optics forward model only a brief description is necessary; further details may be found in Hecht (1987) (general) or Melbourne et al. (1994)(radio occultation specific).

The theory describing how the ray path is refracted when passed through the Earth's atmosphere relies upon the fact that the transmitted signal follows a dimensionless ray path to the receiver, i.e. no transverse dimension. This is the path of minimum phase delay relative to the delay associated with alternative paths. However, the concept of a geometrical ray path is an approximation to Fresnel diffraction[1] in the asymptotic limit of the wavelength approaching zero, i.e. as the wavelength tends to zero the diffraction effect is zero.

Because the signals originate from a point source and are detected by a receiver that is phase-locked to the minimum phase delay, the cross-beam sampling at the limb takes the form of concentric Fresnel zones (appendix A) of positive and negative response centred on the ray path. The contribution of adjacent Fresnel zones cancel so that the effective sampling can be represented by the first Fresnel zone (Hecht 1987).

In the absence of significant atmospheric bending, the radius of the first Fresnel zone, R_F is given by equation 3.11, which is derived in appendix A.

$$R_F = \left[\frac{\lambda L_T L_R}{L_T + L_R} \right]^{1/2} \tag{3.11}$$

where λ is the GPS signal wavelength; L_T is the distance from the transmitter to the limb; and L_R is the distance from the receiver to the limb. This equation reduces to approximately $(\lambda L_R)^{1/2}$ if $L_T >> L_R$.

If we assume L_T is approximately 28,500 km, for GPS transmitters; L_R is 3,100 km for the receiver; and λ is ~19cm, R_F is evaluated to be 1.4 km. From equation 2.3 the horizontal resolution corresponding to a vertical resolution of 1.4 km is 270 km.

Diffraction-limited vertical resolution is also influenced by the vertical refractivity gradients: when the refractivity gradient increases the radius of the first Fresnel zone decreases (vertical resolution improves) in accordance with the expression

[1] A class of diffraction phenomena in which the transmitter or the receiver, or both, are at a finite distance from the diffracting system.

$$R_F = R_{F\ (freespace)} M^{1/2} \tag{3.12}$$

where M is defined to be $(1 + L_R \frac{\partial \epsilon}{\partial a})^{-1}$ (Fjeldbo and Eshleman 1969), and all other variables are as previously defined.

The factor M decreases approximately exponentially with altitude. Corresponding values of R_F decrease from 1.4 km in the stratosphere to 0.5 km in the lower troposphere (figure 2.6). Within the troposphere, rapid fluctuations in R_F are mainly due to the effects of rapidly varying refractivity gradients. See Kursinski et al. (1997) for further details.

The diffraction-limited vertical resolution of GPS radio occultation measurements is comparable with the vertical resolution imposed by atmospheric structure on a single occultation measurement. Therefore there is no particular impetus to deconvolve the signal to obtain improved resolution.

Ionospheric effects

The dual-frequency signals from the GPS are also affected by the ionosphere. The ionosphere is a region of the Earth's upper atmosphere in which free electrons arising from ionization occur, mainly as a result of ultra-violet radiation and X-rays from the sun. The ionosphere is usually divided into three layers: the D region between 50 and 90 kilometres above the Earth's surface; the E region (the Heaviside-Kennelly layer) between 90 and 150 km; and the F region (the Appleton region) above 150 km. This vertical structure is illustrated by figure 3.3.

The ionospheric refraction effect is given by the Appleton-Hartree formula:

$$n \backsimeq 1 - \frac{f_p^2}{2f^2} \tag{3.13}$$

where n is the refraction index, f_p is the plasma frequency (Hz), and f is the operating frequency (Hz).

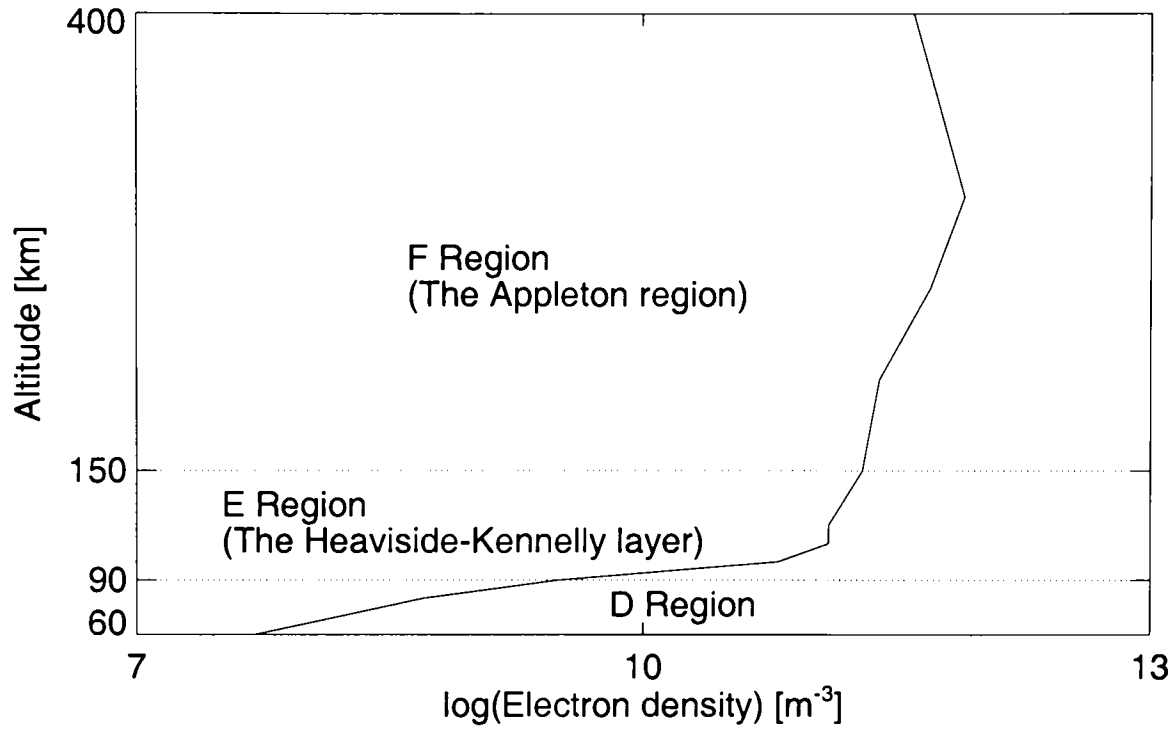


Figure 3.3: The vertical structure of the ionosphere expressed in terms of electron density [m^{-3}], as a function of altitude [km]. The ionosphere can be split into three major regions (D, E and F), each demonstrating their own characteristics. Data is from C. W. Allen (1976).

This refraction effect caused by the ionosphere can be used to build up distributions of this vertical region (Hajj et al. (1994) & Hajj and Romans (1996)). For neutral atmosphere research, the ionosphere degrades the information of temperature and water vapour lower in the atmosphere. However, the dual-frequency signal can be combined in such a way that this effect that can be reduced to second-order (§4.2).

The forward model used for the work herein does not include an ionosphere, but it is important to understand its effect on the GPS signals and subsequently the methods used to reduce it (§4.2).

Variation in the acceleration due to gravity

The acceleration due to gravity at the Earth's surface, g_S, varies with latitude ϕ in a manner that can be approximated by the International Gravity Formula, (U.S.-Air-Force 1961):

$$g_{\phi,S} = g_{E,S}(1 + 5.2884 \times 10^{-5} \sin^2 \phi - 5.9 \times 10^{-8} \sin^2 2\phi) \tag{3.14}$$

where the $g_{E,S}{}^2$ term is the equatorial value of gravity as was measured by Heiskanen in 1928, (U.S.-Air-Force 1961).

The variation of g with altitude, z, above a spherical Earth of radius R is approximated by equation 3.15.

$$g_\phi(z) = \frac{g_{\phi,S}}{(1 + z/R)^2} \tag{3.15}$$

Non-ideal gas behaviour

Van der Waals' equation of state gives an adequate description in the case of non-ideal behaviour by including the effect of mutual attraction between molecules, and accounting for the volume taken up by the molecules themselves.

$$\left(P + \eta^2 a\right)(1 - \eta b)\, V = \zeta RT \tag{3.16}$$

Where η is the number density in moles/m^3; a=0.0014 m^6 hPa/mole^2; and b=3.75$\times$10^{-5}.

In the ideal gas approximation, a and b are assumed to be negligible, and equation 3.16 reduces to the equation for a perfect gas: $PV = \zeta RT$, where ζ is the number of moles.

[2]The equatorial value of gravity used is equal to 9.780490 ms^{-2}

Chapter 4

The inverse radio occultation problem

Once the modelling is complete, or the measurements have been taken, some form of retrieval theory must be applied to obtain the meteorological parameters sought. This chapter addresses the inverse problem: given the measurements how to estimate the atmospheric state.

4.1 The inverse refraction problem

The inverse refraction problem has many formulations some of which make approximations about the motions of the transmitter and receiver during an occultation event. The work presented herein uses the formulation where the transmitter and receiver move such that during an occultation the horizontal tangent point displacement is much smaller than the characteristic horizontal scale of the meteorological fields. The geometry defining this problem is illustrated in the previous section (figure 3.1).

The effect of the atmosphere on the two GPS radio signals can be measured as an additional phase delay relative to the expected phase delay using Fermat's principle of least time. The Doppler shift of the signal is calculated from the additional atmospheric delay (the derivative of the phase delay). Subject to the assumption of local spherical symmetry, Doppler shift can be combined with the satellites' position and velocity vector measurements to determine the bending angle as a function of impact parameter a. By combining the two bending angle estimates from the two GPS signal frequencies, it is possible to reduce ionospheric effects to second-order (§4.2).

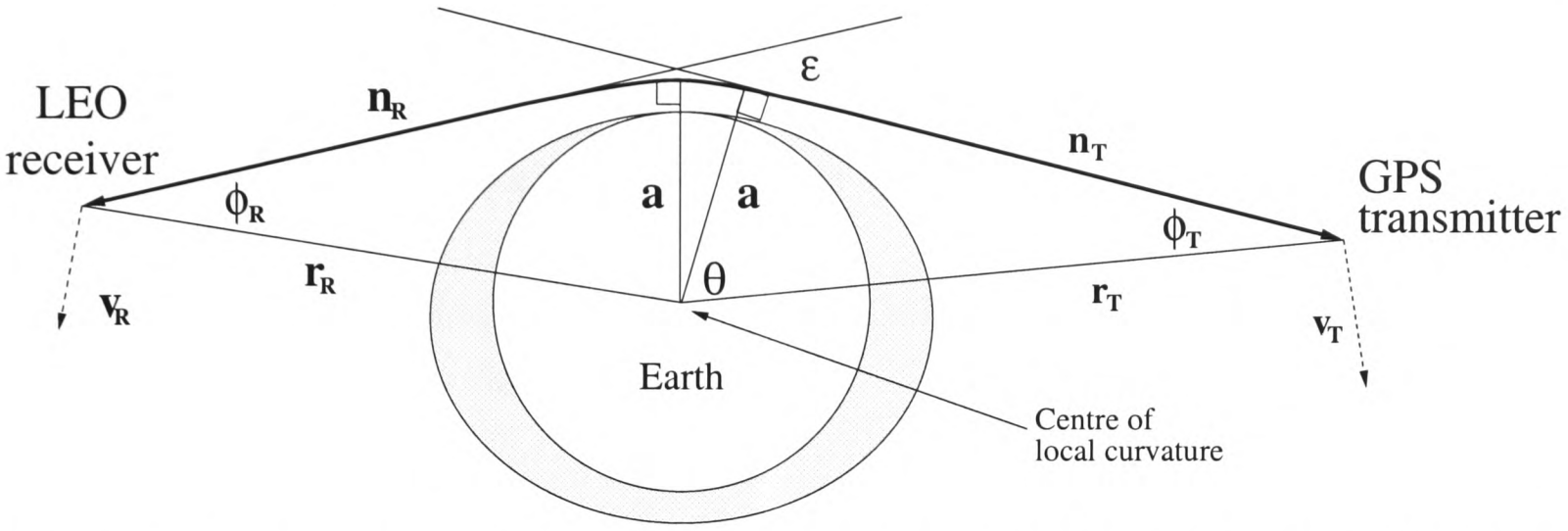


Figure 4.1: Defining the geometry for radio occultation needed to obtain information of bending angle from the time delay caused by the Earth's atmosphere. The dashed lines indicate motion.

Using the geometry and notation of figure 4.1, the Doppler shift f_d of the carrier frequency f_0 measured by the receiver is given by:

$$f_d = f_0 \left[\frac{(\mathbf{v}_T \bullet \mathbf{n}_T + \mathbf{v}_R \bullet \mathbf{n}_R)}{c} + \frac{(\Phi_T - \Phi_R)}{c^2} + \frac{(v_T^2 - v_R^2)}{2c^2} + \mathcal{O}\left(\frac{V}{c}\right)^3 \right] \tag{4.1}$$

where $\mathbf{v}_{T,R}$ are transmitter and receiver velocity vectors, $\mathbf{n}_{T,R}$ are path direction vectors of the transmitter and receiver, $\Phi_{T,R}$ are gravitational potentials at the transmitter and receiver, and c is the velocity of light. The first term represents the Doppler shift produced by the projection of the spacecraft velocities on to the ray paths, the second term is a general relativity correction for the different gravitational potentials, and the third and fourth terms represent higher order relativistic corrections for spacecraft velocity. Note that the relative positions and velocities of the two satellites can be calculated very accurately using available tracking data, which is independent of radio occultation data.

The relativistic terms are eliminated based on knowledge of orbital geometry and the Earth's gravity field, leaving the first term on the right-hand side of equation 4.1 (Kursinski et al. 1997).

By specifying radial and tangential components of the velocity of satellite i in the plane coinciding with the ray trajectory as $\mathbf{v}_i^r$ and $\mathbf{v}_i^t$, and taking into account Snell's law (equation 3.2), the angles ϕ_R and ϕ_T can be calculated from the following relations:

$$f_d = f_0 c^{-1}(\mathbf{v}_R^r \cos\phi_R + \mathbf{v}_T^r \cos\phi_2 + \mathbf{v}_R^t \sin\phi_R - \mathbf{v}_T^t \sin\phi_T) \tag{4.2}$$

From Bouguer's rule (equation 3.8), i.e. spherical symmetry, and the geometry defined by figure 4.1, $\varepsilon(a)$ can be calculated.

$$r_R \sin\phi_R = r_T \sin\phi_T = a \tag{4.3}$$

$$\varepsilon = \phi_R + \phi_T + \theta - \pi \tag{4.4}$$

For further details on this method see Kursinski et al. (1997).

4.2 Ionospheric correction

Data from the GPS/MET receiver does contain ionospheric effects, so it is useful to review the method used to correct the GPS signals, L1 and L2.

The method used for the work herein was proposed by Vorob'ev and Krasil'nikova (1994); the method forms a linear combination of the refracted angles, ε.

$$\varepsilon(a) = \frac{\varepsilon_1(a) f_1^2 - \varepsilon_2(a) f_2^2}{f_1^2 - f_2^2} \tag{4.5}$$

The basic approximations for this method includes assuming that $\varepsilon(a)$ is a linear functional of $n(r)$, and neglecting higher-order terms of the frequency dependence on refractivity. Although this method removes the majority of the effect some residual remains which is mainly due to the non-linearity of the refraction angle with respect to the refractivity. However, the residual error for this method is smaller than the traditional method which involves forming a linear combination of the observed phase delays for both L1 and L2 signals at the same time t, even in the presence of the strong horizontal gradients (Melbourne et al. 1994). The traditional method is based upon the

assumption that the ray paths are coincidental for the two GPS frequencies. Although Bengtsson et al. (1995) show this assumption to be acceptable, the difference in the ray paths can be significant especially in the vicinity of the tangent points, whose vertical separations may be of the order of several hundred metres (Hardy et al. 1993).

Small-scale ionospheric disturbances may appear as bursts of noise in the ionosphere-corrected bending angle values (Kursinski et al. 1997).

4.3 The Abelian inverse method

The Abelian inverse method outlined in this section is an approach which was developed by Stanford University for planetary missions, and adopted by the JPL and the UCAR, to retrieve meteorological parameters from radio occultation measurements of the Earth.

4.3.1 Refractive index determination

If the measured or modelled atmosphere is assumed to be horizontally homogeneous, the inversion of the bending angle data, to determine the refractive index, may be simplified to an integral transform. The following section describes the two possible transforms that can be used.

The Abel transform

The forward problem as represented by equation 3.10 can be solved directly using the Abel transform (equation 4.7) if spherical symmetry of the atmosphere is assumed. Information in bending angle space may be calculated from refractive index space by one of the Abel transform pair; the other transforms from bending angle space to refractive index space. A full derivation of these transforms may be found in Phinney and Anderson (1968) and are given by

$$\varepsilon(a) = -2a \int_a^{+\infty} \left(\frac{\partial \ln n(x)}{\partial x} \right) (x^2 - a^2)^{-1/2} dx, \tag{4.6}$$

$$\ln n(x) = \frac{1}{\pi} \int_x^{+\infty} \varepsilon(a)(a^2 - x^2)^{-1/2} da \tag{4.7}$$

where x is called the refractive radius, and is equivalent to rn (§3.1.2).

For the purpose of the work described herein, one transform is used in the forward modelling of the radio occultation problem and the other represents a method of inverting the bending angle measurements. The transform used to invert the measurements will hereafter be referred to as the inverse Abel transform as a way of distinguishing between them.

Modelling the infinite upper limit is a potential source of error in the inversion, and as such is discussed in §4.3.3.

The mapping from refractive index as a function of refractive radius x to a function of radius r and subsequently as a function of geometric height z is given by

$$r = \frac{x}{n} \tag{4.8}$$

$$z = r - R_E \tag{4.9}$$

where R_E is the local radius of curvature.

Fresnel transform

As mentioned in chapter 2, the vertical resolution of the GPS measurements is limited by Fresnel diffraction. Vertical spatial structure superior to that imposed by Fresnel diffraction produces characteristic patterns on the GPS signal measured by the receiver, and are called phase and amplitude scintillations.

Melbourne et al. (1994) presents a method of enhancing the vertical resolution of GPS/MET measurements by deconvolving the diffraction effects of the atmosphere. However, since the forward model used in the work herein is based upon geometric

optics theory (which assumes no diffraction) only a brief description of the method will be included.

In formulating this transform, a thin-screen model of the atmosphere is assumed, i.e. in occultation geometry the total influence of the atmosphere on the GPS signals can be approximated by a thin horizontal section of the atmosphere. In other words, the majority of the 'bending' of the signal takes place in a very small horizontal region. For occultation geometry the Helmholtz-Kirchoff integral theorem applies (Born and Wolf 1980), which expresses the electric field at the observation point in terms of the field at the thin screen.

If the GPS electric field arriving at the thin screen and the subsequent electric field after the signal has passed through the thin screen are considered, accounting for obliquity and for attenuation of the signal by the atmosphere, the following transform pair may be derived (Melbourne et al. 1994) and used to invert atmospheric phase delay profiles (Mortensen and Høeg 1998).

$$\phi(h) = -2\int_{x/n(a)}^{\infty} \frac{1}{n}\frac{dn}{dr}\sqrt{n^2r^2 - a^2}\,dr, \tag{4.10}$$

$$\ln n(x) = -\frac{1}{\pi}\frac{d}{dx}\left[\int_{x}^{\infty} \frac{a\phi(a - R_E)}{a\sqrt{a^2 - x^2}}da\right] \tag{4.11}$$

where the height $h = (a/n(a)) - R_E$; $\phi(h)$ [m] is the ionosphere-corrected phase delay and is a measure of the phase delay due to the neutral atmosphere; and R_E is the local radius of curvature.

4.3.2 Pressure and temperature determination

After the ionospheric effects have been corrected for, only the dry and moist terms are significant, and so equation 3.1 reduces to:

$$N = \frac{c_1 P_a}{T} + \frac{c_2 P_W}{T^2} \tag{4.12}$$

where N is the refractivity and all other variables are as defined in equation 3.1. Since refractive index is close to unity, it is useful to work with refractivity which may be defined as $(n-1) \times 10^6$, adjusting the constants of proportionality accordingly.

The contribution from each of the terms in equation 4.12, assuming a realistic atmospheric profile is shown in figure 4.2. It can be seen that the moist term (dotted line) does not dominate the equation, but is significant lower down in the troposphere, with values comparable with the dry term.

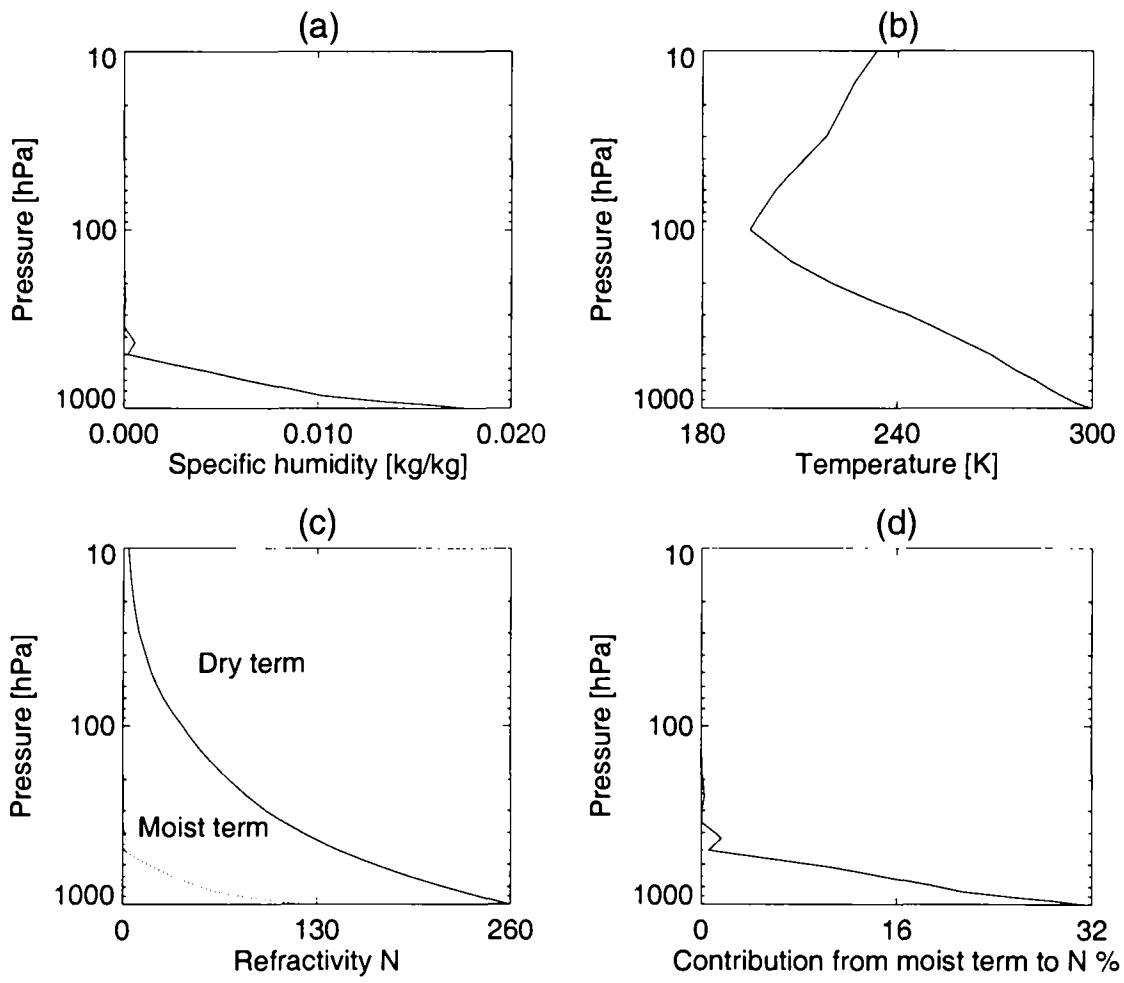


Figure 4.2: Illustrating how significant the dry and moist terms are to equation 4.12, using a UKMO unified model profile at 0.16°S, 343.56°E. Panel (a) and (b) show the profiles of specific humidity and temperature, respectively; panel (c) shows the dry and moist terms of the refractivity; and panel (d) shows the moist contribution to refractivity, expressed as a percentage.

Absence of water vapour

For a water vapour volume mixing ratio of 10^{-4}, the moist term contributes approximately 0.2% of refractivity (Kursinski et al. 1997); in the regions which are much drier the moist term can be neglected and equation 4.12 reduces to

$$N = \frac{c_1 P_a}{T} \tag{4.13}$$

Assuming the Earth's atmosphere is an ideal gas, and making use of the relation $m/V = \rho$, equation 4.13 is equivalent to:

$$\rho = \frac{Nm}{c_1 R} \tag{4.14}$$

Vertical accelerations of the atmosphere are small, which means that it is possible to assume static equilibrium in a vertical column of air of unit area. Moving from a height $z + \delta z$ to a height z, there is an increase in pressure due to a greater quantity of air above it. The increase in pressure δP will be the weight of the extra segment of the air column, i.e.

$$\delta P(z) = -g(z)\rho(z)\delta z \tag{4.15}$$

where ρ is atmospheric density which may be calculated using equation 4.14 and g is the acceleration due to gravity. This is known as the hydrostatic relation.

By integrating equation 4.15, it is possible to obtain a pressure profile from a density profile, as a function of altitude – provided that the pressure at one particular level is known (a 'first guess' for the integration). Starting the integration from a pressure high in the atmosphere is preferable, because the exponential increase with depth means that an error in the starting pressure becomes negligible lower in the atmosphere.

The temperature profile is derived from the equation of state of an ideal gas:

$$P(z) = R\rho(z)T(z) \tag{4.16}$$

The approach adopted by the JPL is to start the integral at the 50 km altitude level with a density calculated from equation 4.14 assuming a pressure corresponding to

a temperature of 260 K at this level. The JPL chose this height because it is where the receiver SNR of the GPS/MET receiver starts to be favourable for the neutral atmosphere; the temperature value was chosen because it represents a typical value at this height.

This method yields accurate profiles of temperature throughout the stratosphere down to the middle troposphere, and below, depending on latitude.

Inclusion of water vapour

As discussed, in the tropospheric regions the amount of water in the atmosphere can contribute a significant amount to refractivity; if this quantity is neglected when inverting the radio occultation data (often known as a dry retrieval) a cold temperature bias ensues in the lower atmosphere. In the troposphere, when water vapour becomes significant, both terms in equation 4.12 should be used.

For low water vapour concentrations we can use an estimate of water vapour content to make a correction in order to retrieve temperature as before. Equation 4.17 provides a good estimate of $\rho(r)$, given measurements of N, an estimated profile of P_W[1], and an initial guess of T.

$$\rho = \frac{m}{c_1 R}\left(N - \frac{c_2 P_W}{T^2}\right) \tag{4.17}$$

where m is now the mean molecular mass of moist air.

The pressure and temperature profiles derived from the density profile using the hydrostatic relation are then used in equation 4.17, to update the estimate of ρ. This process is iterated until the derived temperature profile converges. Water vapour concentrations can be obtained from independent observations, weather analyses, or climatology.

Where water vapour amounts become greater, e.g. in the tropical lower-troposphere,

[1]This quantity is obtained from the relation $P_w = \frac{PqM_a}{M_w}$, where M_a and M_w are the molar mass of air and water, respectively, q is specific humidity [kg/kg] and P is atmospheric pressure.

the bending of the ray paths become more pronounced due to water vapour possessing a smaller density scale height than the atmosphere as a whole (Kursinski et al. 1997). In this case, equation 4.12 can be rearranged so that the partial water vapour pressure is the subject:

$$P_W = \frac{NT^2 - c_1 PT}{c_2} \tag{4.18}$$

where all variables are as defined in equation 4.12.

The recovery of water vapour from refractivity measurements then requires knowledge of temperature and pressure derived from climatology or independent observations or analyses.

Because the GPS signals are refracted in both air and water vapour, there is no direct method of separating their effects. As such, water vapour limits the accuracy of temperature recovery below a certain altitude because of the uncertainty in its abundance and distribution.

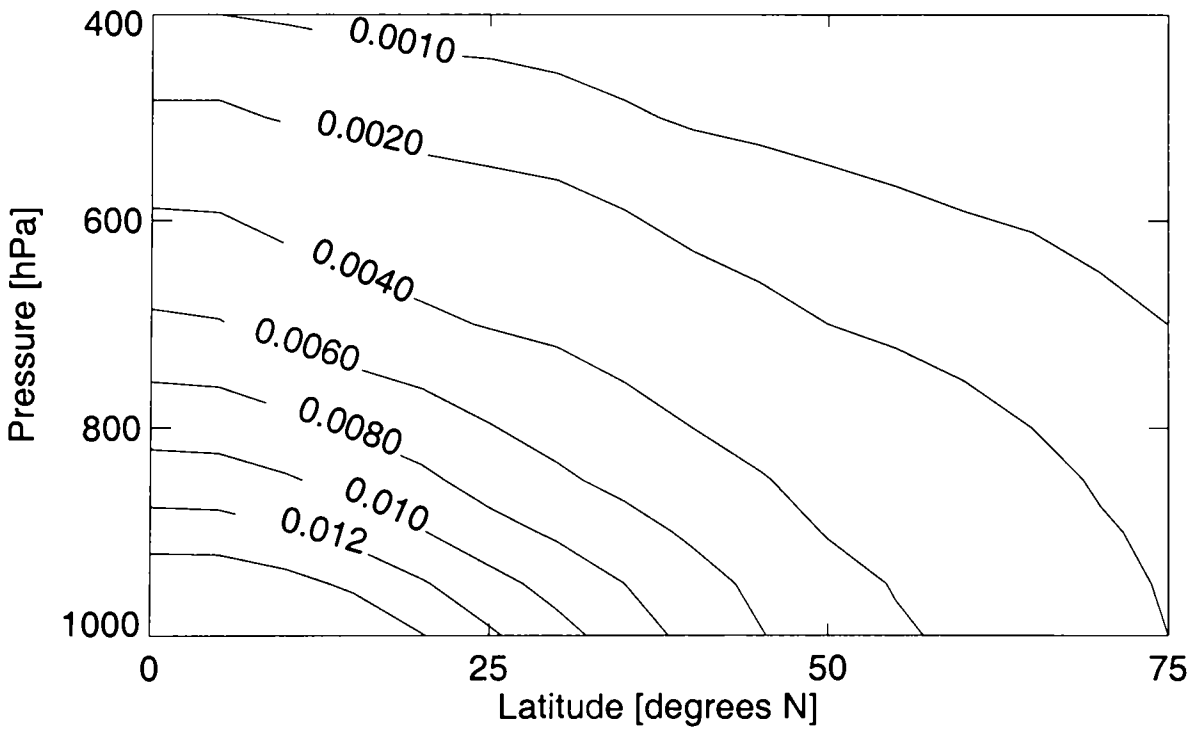


Figure 4.3: The climatological distribution of water vapour, in terms of specific humidity [kg/kg], as a function of latitude. At lower latitudes the 'wet' component of refractivity makes a significant contribution to total refractivity below 600 hPa (equation 4.12); towards the dry, polar atmosphere, retrievals may be evaluated down to near-surface altitudes. Data from Oort & Rasmusson (1971).

Figure 4.3 shows the distribution of specific humidity [kg/kg], as a function of latitude and vertical coordinate of pressure. The figure clearly shows a decrease in the amount of water vapour in the higher latitudes from the equatorial region. This enables GPS/MET temperature retrievals to be taken down near to the surface at high latitudes.

Using equation 4.18, Kursinski et al. (1995) estimates that in low latitudes water vapour could be retrieved with accuracies of 10 to 20% below the 6−7 km altitude, and ∼5% or better within the boundary layer.

4.3.3 Factors affecting the retrieval accuracy

All of the factors affecting the forward modelling of radio occultation data also affect the accurate inverse modelling. Factors such as modelling the Earth's shape, variation in the acceleration due to gravity and ionospheric effects are therefore discussed in §3.2. Factors specific to the inversion of the data are discussed here and include modelling the upper limit of the inverse Abel transform and line-of-sight variations.

Line-of-sight horizontal variations

Bouguer's rule (equation 3.8), used to compute bending angle from atmospheric phase delay, invokes spherical symmetry. However, bending angle and impact parameter errors due to this assumption have a small effect on the retrieved refractive index profile compared with the use of the inverse Abel transform (Kursinski 1997).

The effect of assuming spherical symmetry can be investigated by using the Abel transform pair (equations 4.7). The impetus for this research is that the Earth's atmosphere typically has considerable structure in the line of sight direction, and it is therefore important to investigate the effects of assuming spherical symmetry in the presence of horizontal variation. This effect is investigated in chapter 5, where further details may be obtained.

Inverse Abel transform upper limit

Incorrect modelling of the infinite upper limit for the inverse Abel transform can introduce errors in the inverted radio occultation data. Errors that are mapped from bending angle space onto refractivity space are described by equation 4.19 (Melbourne et al. 1994).

$$\frac{\delta n(x)}{n(x)} = \frac{1}{\pi} \int_x^\infty \delta\varepsilon(a)(a^2 - x^2)^{-1/2}\, da \tag{4.19}$$

The conventional method of accounting for the infinite upper limit is to define a height level to be the 'top' of the atmosphere. This height is chosen such that it can be safely assumed that the value of refractivity at that height is zero and therefore no useful atmospheric information can be deduced using radio occultation methods. By considering how the bending angle decreases with increasing height at lower levels, the bending angle profile can be extrapolated to the 'top' height level. The signal-to-noise ratio of GPS/MET data above 40 km is low (Hocke 1997), and it is therefore necessary to extrapolate carefully, i.e. not using data from this region.

Hocke (1997) outlines a statistical method of obtaining an upper limit to the inversion. The method calculates the most likely bending angle profile from the observations and climatology, using a 'weighting function' defined by their respective error statistics. This approach results in a smooth transition from the mean atmospheric state at high altitude (where observations are typically noisy) to observations at lower altitudes, minimising error propagation downwards. 'Weighting functions' are calculated from observational noise in the 60 to 80 km range. The 'weighting' function is equal to 0.5 when the observation noise equals the standard deviation of the atmospheric state from the mean atmospheric state. This happens typically at an altitude of 50 km. Below an altitude of approximately 40 km the climatology is weighted such that it has no significant influence.

The temperature results retrieved using this method were compared with the temperature profile using the uncorrected bending angle profiles. It was found that large errors at high altitudes (50–60 km), incurred by using uncorrected bending angles

in that region, propagated downwards in the atmospheric retrieval: an error of 2 kelvin at 20 km is quoted from Hocke (1997).

4.4 The optimal estimation inverse method

Often it is necessary to infer the state of a system from observations of its physical variables. In many situations, the observations are not complete, too few in number, or are noisy. In such an event, the most 'likely' state of the system can be calculated. Optimal estimation is a branch of mathematics that is devoted to such problems.

4.4.1 Problem formulation

In finding the optimal solution to a problem it is first necessary to formulate the problem properly; this is achieved by fulfilling a number of conditions.

We consider an error-free linear measurement, equivalent to solving a linear equation:

$$\mathbf{y} = \mathbf{K}\mathbf{x} \tag{4.20}$$

where $\mathbf{x}$ is called the state vector and represents the atmospheric state, $\mathbf{y}$ is called the measurement vector and represents a vector of measurements from which the optimal solution will be obtained, and $\mathbf{K}$ is a matrix. The rows of matrix $\mathbf{K}$ map from state space to the measurement space in a way described by

$$\mathbf{y_i} = \mathbf{k_i}\mathbf{x} \tag{4.21}$$

These vectors span a subspace called the row space, of dimension equal to the rank of $\mathbf{K}$. Only those components of $\mathbf{x}$ in the row space can be measured. The null space is the part of the state space which is not in the row space.

In general, if a problem is ill posed there exists null space, a subspace of state space

that cannot be measured by the observing instrument. It may also be ill conditioned if a small region of measurement space (comparable to error bounds) maps onto large region of state space.

To make the problem well conditioned, it is necessary to apply some constraint to the null space component of the state; this is achieved either explicitly by defining an *a priori*, or implicitly by setting all null-space components to zero.

In the case of the radio occultation problem (in particular the GPS/MET receiver), there are two main differences compared to other observing systems. First, the measurement noise is sufficiently low compared with other error sources in regions of primary interest, although to exploit fully the radio occultation technique it is still necessary to allow for it. Second, it possesses high vertical resolution, limited only by Fresnel diffraction (§3.2).

Consequently, the approach adopted by both the JPL and the UCAR has been able to use very elementary retrieval methods (outlined earlier this chapter), that for other forms of data would be unsatisfactory. It should be noted that because measurement noise has been neglected whilst using the Abelian inverse method, the method effectively assumes the noise to be zero.

The inverse problem

In the case of the radio occultation problem, $\mathbf{y}$ represents a vector of bending angles (§2.4), and $\mathbf{x}$ is a vector describing an atmospheric state: temperature (as a function of pressure), water vapour (expressed in terms of specific humidity as a function of pressure), and surface pressure.

The two vectors can be related conceptually by a formula

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \mathbf{b}) \tag{4.22}$$

where $\mathbf{f}$ is called the forward function which maps from state space onto measurement space.

To describe fully the physics, it may be necessary to include additional parameters (represented here by the vector $\mathbf{b}$) which are assumed to be known perfectly and are not the subject of the retrieval. In the case of modelling the radio occultation problem the elements of $\mathbf{b}$ include the local radius of curvature (§3.2) which is unique to any one occultation.

The modelling of the physics is subject to unavoidable errors (e.g. the discrete representation of the atmosphere as opposed to a continuous representation), and it is important to consider these errors in the mapping. In this case, the relationship between the state vector and the measurement vector is called the forward model $\mathbf{F}$ and represents the best that can be achieved in modelling the physics, given the constraints of computer resources and available knowledge of the physical process. This can be described by

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}}) + \boldsymbol{\epsilon}_{\mathrm{y}} + \boldsymbol{\epsilon}_{\mathrm{b}} + \Delta\mathbf{F} \tag{4.23}$$

where $\hat{\mathbf{b}}$ represents an approximation to the vector $\mathbf{b}$, $\boldsymbol{\epsilon}_{\mathrm{y}}$ is the error associated with the measurement, $\boldsymbol{\epsilon}_{\mathrm{b}}$ is the error associated with the $\mathbf{b}$ parameter, and $\Delta\mathbf{F}$ is the difference between the forward function and the forward model physicss.

It is necessary to account for the fact that a continuous profile is being represented by a discrete vector so that the *a priori* information does not overconstrain the retrieval. This is called a representative error which essentially describes the inability of the forward model using a discretised state vector to represent a continuous physical state. This difference between the forward function $\mathbf{f}$ and the forward model $\mathbf{F}$ is represented here by $\Delta\mathbf{F}$ and is described by the covariance matrix $\mathbf{S}_{\Delta\mathbf{F}}$.

Although elements of the $\mathbf{b}$ vector may vary within a set of measurements, the vector $\hat{\mathbf{b}}$ is constant and represents an estimate of $\mathbf{b}$. The effect of approximating $\mathbf{b}$ by $\hat{\mathbf{b}}$ in the forward model is to introduce an error $\boldsymbol{\epsilon}_{\mathrm{b}}$, and is described by the covariance matrix $\mathbf{S}_{\mathrm{b}}$. The effect of instrument noise (and possibly pre-processing of data) takes the form of an error on the measurements $\boldsymbol{\epsilon}_{\mathrm{y}}$ and is described by the covariance matrix $\mathbf{S}_{\mathrm{y}}$.

To simplify the algebraic formulation of the problem, the measurement noise, the error associated with the **b** parameter and the representative error can be accounted for in one vector which will be denoted as $\boldsymbol{\epsilon}_\Sigma$ and described by the covariance matrix $\mathbf{S}_\Sigma$, i.e. $\mathbf{S}_\mathrm{y} + \mathbf{S}_\mathrm{b} + \mathbf{S}_{\Delta\mathbf{F}}$. This effectively represents the total measurement error associated with the statistical estimate.

Although bending angle measurements are thought to be independent (§6.1.2), derived values of refractivity are not, due to being an integrated quantity. If it is considered that the bending angle values are statistically independent, i.e. no covariance between errors in each measurement, then the covariance matrix describing the measurement errors is diagonal. In the case of derived parameters, the error structure becomes more complex, and so will contain off-diagonal elements.

For the moment, it is assumed that all noise can be represented as zero bias random vectors that are uncorrelated between one retrieval and the next, and not a source of systematic error. The effect of systematic errors is discussed in §4.5.3.

The inverse model **I** describes how the best estimate of the state $\hat{\mathbf{x}}$ is obtained from the measurements:

$$\hat{\mathbf{x}} = \mathbf{I}\left[\left(\mathbf{F}(\mathbf{x}, \hat{\mathbf{b}}) + \boldsymbol{\epsilon}_\Sigma\right), \hat{\mathbf{b}}, \hat{\mathbf{c}}, \mathbf{x}_\mathrm{a}\right] \tag{4.24}$$

where $\hat{\mathbf{c}}$ is a vector of parameters used in the retrieval but not appearing in the forward model, and are unconnected with the measurements, and all other terms are as before.

There are many values of state vector that, when mapped through the forward model, will produce the same measurement vector equal to the actual measurements within the measurement noise. The *best* estimate of $\mathbf{x}$ can be obtained by various means, but some solutions will be obviously incorrect because they will be inconsistent with the *a priori* information[2] of the atmospheric state.

Probability density functions (PDF) can be used to describe the measurement space

[2]The *a priori* knowledge of the state vector is represented by $\mathbf{x}_\mathrm{a}$ and the covariance matrix $\mathbf{S}_\mathrm{a}$. This may be based upon a climatology (an ensemble of measurements or model outputs) where $\mathbf{x}_\mathrm{a}$ is an average, or NWP analysis, and $\mathbf{S}_\mathrm{a}$ represents the variation about that average.

and the *a priori* information; the solution of the inverse problem should be described by a PDF that is consistent with the measurement and *a priori* PDFs.

Furthermore, there are still several criteria that can be used to obtain the optimal solution of the atmospheric state: a representative parameter of the state vector PDF for the mean or the maximum likelihood; the solution that has the smallest error variance; the solution which has the highest spatial resolution; or the solution with the smallest weighted sum of resolution and error variance. The radio occultation retrieval (described below) uses the the maximum likelihood solution.

The general idea of optimal estimation is to minimise a penalty function $\xi(\mathbf{x})$ which describes the errors associated with the measurements and the *a priori* information (and any other constraints), and effectively measures the degree of fit of the solution to the measurements and the *a priori* information. This idea of a penalty function will be retained throughout this chapter.

4.4.2 The maximum likelihood solution

The most accessible approach to explaining this technique is to use Bayes' theory of probability, where knowledge of a particular quantity is expressed in terms of PDFs. It is the PDF of the state given the measurements that is sought, i.e. $P(\mathbf{x}|\mathbf{y})$.

The PDF of the state, describing what is known about the state before the measurement is taken, is denoted by $P(\mathbf{x})$; the *a priori* PDF of the measurement is denoted by $P(\mathbf{y})$; and the PDF of the measurement given the state (which depends on experimental error and forward model error) is denoted by $P(\mathbf{y}|\mathbf{x})$.

Bayes' theorem states

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \tag{4.25}$$

By combining equation 4.25, assuming that all statistics are Gaussian, and with the assumption that $P(\mathbf{y})$ is a constant, it is possible to show that the value of $\mathbf{x}$ that maximises $P(\mathbf{x}|\mathbf{y})$ will also minimise

$$[\mathbf{y} - \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}})]^{\mathrm{T}} \mathbf{S}_{\Sigma}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}})] + (\mathbf{x} - \mathbf{x}_{\mathrm{a}})^{\mathrm{T}} \mathbf{S}_{\mathrm{a}}^{-1} (\mathbf{x} - \mathbf{x}_{\mathrm{a}}) \tag{4.26}$$

where all terms are as before (Rodgers 1998).

Differentiating with respect to $\mathbf{x}$ and setting equal to zero gives

$$\mathbf{S}_{\mathrm{a}}^{-1} (\mathbf{x} - \mathbf{x}_{\mathrm{a}}) - \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right)^{\mathrm{T}} \mathbf{S}_{\Sigma}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}})] = 0 \tag{4.27}$$

which will yield, when solved for $\mathbf{x}$, the optimum solution $\hat{\mathbf{x}}$.

The method used to find the solution of equation 4.27 depends upon the nature of $\mathbf{F}(\mathbf{x})$: whether it is linear or non-linear.

Linear forward operator

If the forward model is a linear function of $\mathbf{x}$, i.e. $\mathbf{F}(\mathbf{x}) = \mathbf{K}\mathbf{x}$, where $\mathbf{K}$ is often known as the matrix of weighting functions[3] then equation 4.27 gives the optimal solution as:

$$\hat{\mathbf{x}} = (\mathbf{S}_{\mathrm{a}}^{-1} + \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\Sigma}^{-1} \mathbf{K})^{-1} (\mathbf{S}_{\mathrm{a}}^{-1} \mathbf{x}_{\mathrm{a}} + \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\Sigma}^{-1} \mathbf{y}). \tag{4.28}$$

This formulation of the linear optimal estimation is conceptually easy to understand; it can be shown that $\hat{\mathbf{x}}$ is simply the weighted mean of two estimates of $\mathbf{x}$: $\mathbf{x}_{\mathrm{a}}$ and $\mathbf{D}\mathbf{y}$, where $\mathbf{D}$ is a matrix described as the matrix of contribution functions (§4.5.1).

Equation 4.28 can be rearranged by using a matrix identity (Rodgers, 1976):

$$\hat{\mathbf{x}} = \mathbf{x}_{\mathrm{a}} + \mathbf{S}_{\mathrm{a}} \mathbf{K}^{\mathrm{T}} (\mathbf{S}_{\Sigma} + \mathbf{K} \mathbf{S}_{\mathrm{a}} \mathbf{K}^{\mathrm{T}})^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}_{\mathrm{a}}) \tag{4.29}$$

Note that although equation 4.28 involves a greater number of matrix inverse operations than equation 4.29, the *a priori* covariance matrix can often be pre-computed

[3] Also known as the Frechet derivative or Jacobian, and represents how sensitive the measurement vector is to a perturbation in the state vector.

since it is not updated by the optimal estimation (this can also apply to the total error covariance if it does not alter with different retrievals). The number of elements in $\mathbf{x}$ and $\mathbf{y}$ should also be considered when deciding which formulation of the optimal estimation to use: if, say, the number of elements in the state vector is greater than the number of measurements, then equation 4.28 may become unfavourable since the number of elements involved with each matrix operation is proportional to the number of elements squared.

Non-linear forward model

If the physics of the problem is non-linear, then an iterative approach should be implemented to calculate for the optimal solution. If the problem is moderately non-linear it may be possible to linearise it such that an iterative method can be used to converge to a solution. Often a simple Newtonian iterative method is all that is needed.

Linearising $\mathbf{F}(\mathbf{x})$ about the state $\mathbf{x}_n$ yields

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_n) + \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}} \bigg|_{\mathbf{x}_n} \right) (\mathbf{x} - \mathbf{x}_n) + \mathcal{O}(\mathbf{x} - \mathbf{x}_n)^2 \tag{4.30}$$

If the higher order terms are ignored, and the terms $\mathbf{K}_n$ (the matrix of weighting functions evaluated at the linearisation point $\mathbf{x}_n$) and $\mathbf{y}_n = \mathbf{F}(\mathbf{x}_n)$ are introduced, then it is possible to write equation 4.30 as

$$\mathbf{F}(\mathbf{x}) = \mathbf{y}_n + \mathbf{K}_n(\mathbf{x} - \mathbf{x}_n) \tag{4.31}$$

Insertion into equation 4.27 gives

$$\mathbf{S}_a^{-1}(\mathbf{x}_{n+1} - \mathbf{x}_a) - \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} [\mathbf{y} - \mathbf{y}_n - \mathbf{K}_n(\mathbf{x}_{n+1} - \mathbf{x}_n)] = 0 \tag{4.32}$$

which can be rearranged into the following forms

$$\mathbf{x}_{n+1} = (\mathbf{S}_a^{-1} + \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} \mathbf{K}_n)^{-1} [\mathbf{S}_a^{-1} \mathbf{x}_a + \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} (\mathbf{y} - \mathbf{y}_n + \mathbf{K}_n \mathbf{x}_n)] \tag{4.33}$$

This equation may be rearranged to give $\mathbf{x}_{n+1}$ as a departure from either $\mathbf{x}_n$ or $\mathbf{x}_a$ using the same matrix identity used to obtain equation 4.29. These equations are called the n-form since they involve manipulating matrices of size $n \times n$ (n being the number of unknowns)

$$\mathbf{x}_{n+1} = \mathbf{x}_n + (\mathbf{S}_a^{-1} + \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} \mathbf{K}_n)^{-1} (\mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} [\mathbf{y} - \mathbf{y}_n] + \mathbf{S}_a^{-1} [\mathbf{x}_n - \mathbf{x}_a]) \quad (4.34)$$

$$= \mathbf{x}_a + (\mathbf{S}_a^{-1} + \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} \mathbf{K}_n)^{-1} \mathbf{K}_n^T \mathbf{S}_\Sigma^{-1} ([\mathbf{y} - \mathbf{y}_n] - \mathbf{K}_n [\mathbf{x}_n - \mathbf{x}_a]) \quad (4.35)$$

If the number of elements in $\mathbf{y}$ is greater than the number of elements in $\mathbf{x}$ by a considerable factor (as often is the case) then a more computationally efficient form can be derived from the second of these equations and is called the m-form since it involves manipulating matrices of size $m \times m$ (m being the number of measurements)

$$\mathbf{x}_{n+1} = \mathbf{x}_a + \mathbf{S}_a \mathbf{K}_n^T (\mathbf{K}_n \mathbf{S}_a \mathbf{K}_n^T + \mathbf{S}_\Sigma)^{-1} (\mathbf{y} - \mathbf{y}_n - \mathbf{K}_n [\mathbf{x}_a - \mathbf{x}_n]) \quad (4.36)$$

Unlike the case for the linear forward model (equation 4.28), the vector $\mathbf{x}_{n+1}$ is not the optimum solution until the iterated state vector has converged. This is because $\mathbf{x}_{n+1}$ is the optimum solution for the $\mathbf{y} = \mathbf{K}_{\mathbf{x}_n}\mathbf{x}$ problem, and not the optimum solution for the $\mathbf{y} = \mathbf{K}_{\hat{\mathbf{x}}}\mathbf{x}$ problem. It is desirable that $\mathbf{K}(\mathbf{x})$ in the region considered is linear enough for $\mathbf{x}_{n+1}$ to be a closer estimate of the optimum solution ($\hat{\mathbf{x}}$) than $\mathbf{x}_n$ was. A new value of $\mathbf{K}$ can be calculated at $\mathbf{x} = \mathbf{x}_{n+1}$ and the procedure repeated.

Repeated calculation of the $\mathbf{K}$ matrix can require a large amount of computer resources, as the change in the measurement vector $\mathbf{y}$ needs to be calculated when each element of the state-vector $\mathbf{x}$ is perturbed. The 'brute force' method[4] involves computing the forward model $i+1$ times, where i is the number of elements in the state-vector $\mathbf{x}$; however if the tangent linear[5] of the forward model is calculated (numerically

[4] $\mathbf{K}$ can be calculated by perturbing each element of the state vector in turn and looking at the difference between the perturbed simulated measurements and the unperturbed simulated measurements.

[5] The first derivative of the forward model w.r.t. to the state vector elements.

or analytically) the number of operations required to calculate the matrix is reduced significantly.

The non-linearity of the forward model can be investigated by iterating the solution and recalculating the $\mathbf{K}$ matrix and solving for the optimal solution. After each iteration, the solution and its associated $\mathbf{K}$ matrix is displayed. If the problem turns out to be nearly linear then a constant matrix can be used for $\mathbf{K}$ until a solution convergence occurs. This solution will be non-optimum, but can be used as a new starting point for a new value of $\mathbf{K}$ to be computed. This approach requires a larger number of iterations, but fewer calculations of $\mathbf{K}$, which in the work herein, results in less overall CPU time[6].

Equation 4.34 can also be obtained by considering the penalty function formulation. If we assume that the errors involved with the optimal estimate have Gaussian distributions, then the maximum likelihood penalty function can be shown to be a sum of quadratic terms.

$$\xi(\mathbf{x}_n) = (\mathbf{y} - \mathbf{y}_n)^T \mathbf{S}_\Sigma^{-1} (\mathbf{y} - \mathbf{y}_n) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x}_n - \mathbf{x}_a) \tag{4.37}$$

where $\mathbf{y}$ is a vector of the measured bending angles; $\mathbf{y}_n$ is a vector of simulated measurements corresponding to the atmospheric state, i.e. involving forward model calculations; $\mathbf{S}_\Sigma$ is primarily the expected error covariance of the measured bending angles, but should also include the expected error covariance of the forward model; $\mathbf{x}_a$ is the *a priori* atmospheric state (e.g. short-range forecast), and $\mathbf{S}_a$ is its expected error covariance.

Using the Newtonian iteration method the nth estimate of $\mathbf{x}$ is updated as follows:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - [\mathbf{H}_x \xi(\mathbf{x}_n)]^{-1} \nabla_x \xi(\mathbf{x}_n) \tag{4.38}$$

where $\mathbf{H}_x$ is the Hessian operator. $\nabla_x \xi(\mathbf{x})$ is obtained by differentiating equation 4.37:

[6]In some cases, the inverse radio occultation problem is found to be non-linear enough to warrant the calculation of $\mathbf{K}$ at each iteration.

$$\nabla_{x_n}\xi(\mathbf{x}) = \mathbf{S}_a^{-1}(\mathbf{x}_n - \mathbf{x}_a) - \mathbf{K}_n^T(\mathbf{x}_n)\mathbf{S}_\Sigma^{-1}(\mathbf{y} - \mathbf{y}_n) \tag{4.39}$$

The Hessian matrix is obtained by differentiating again:

$$\mathbf{H}_x\xi(\mathbf{x}_n) \simeq \mathbf{S}_a^{-1} + \mathbf{K}^T(\mathbf{x}_n)\mathbf{S}_\Sigma^{-1}\mathbf{K}(\mathbf{x}_n) \tag{4.40}$$

Note that relation 4.40 is approximate since terms involving the gradient of $\mathbf{K}$ have been omitted. If it is assumed that the problem is not *too* non-linear, then this assumption is valid. By comparing equations 4.34 and 4.38 we can match up terms so that the gradient of the penalty function if given by $(\mathbf{K}_n^T\mathbf{S}_\Sigma^{-1}[\mathbf{y} - \mathbf{y}_n] - \mathbf{S}_a^{-1}[\mathbf{x}_n - \mathbf{x}_a])$ and the inverse Hessian is given by $-(\mathbf{S}_a^{-1} + \mathbf{K}_n^T\mathbf{S}_\Sigma^{-1}\mathbf{K}_n)^{-1}$

Convergence and quality control

A useful convergence indicator is the gradient of the penalty function since it is this variable that measures the degree of fit of a set of measurements. The general idea of optimal estimation is to fit the solution to the measurements and the *a priori* information by minimising the sum of errors associated with the measurements and the background information, i.e. the total error associated with the solution, so the gradient of the penalty function should tend towards zero as the solution converges.

In addition to using the gradient of the penalty function as an indication of convergence, once the solution has converged it is possible to use it as a tool to determine whether the solution converged upon is satisfactory. An ensemble of retrievals should follow a χ^2 distribution with m degrees of freedom because there are $m + n$ measurements and virtual measurements to which n parameters have been fitted, m is the number of measurements and n is the number of elements in the state vector (Marks and Rodgers 1993). On average, a moderately good retrieval should have a χ^2 value of the order of m. If χ^2 is too large (the retrieval does not fit the measurements) there is a number of

possible causes: the forward model approximations are too crude, the measurement noise has been underestimated, the *a priori* profile is inappropriate for the measurements, or the *a priori* errors under-estimated. If the χ^2 value is too small (if the retrieval fits the measurements and *a priori* too well) there is again a number of possible causes: the measurement noise have been overestimated or the *a priori* errors over-estimated.

Furthermore, this quantity may be used to construct a form of quality control. If the solution given a set of measurements and the appropriate *a priori* information does not converge within a set number of iterations then the set of measurements should be investigated. Alternatively, the number of iterations the solution takes to converge should be noted — if the solution has converged using a larger number of iterations than efficient then that particular set of measurements should be omitted.

The convergence criteria used for the radio occultation retrieval and further analysis is investigated in more detail in chapter 6.

The Marquardt-Levenberg method

The Marquardt-Levenberg method represents a modification to the approach outlined above (equations 4.34–4.36). If the solution is too far away from the current iteration point, it is quite possible that the quadratic assumption used to derive the equation that in turn is used to calculate the optimal solution represents the surface so poorly that the retrieval increments away from the desired minima. In this case, a steepest descent method is preferable which is safe but slow:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \gamma^{-1} \nabla_x \chi_i^2 \tag{4.41}$$

where γ^{-1} determines the size of the increment at each iteration. The value γ^{-1} should be varied numerically for a minimum in the direction of $\nabla_x \chi_i^2$.

The Marquardt-Levenberg combines the two approaches: if χ^2 increases (outside the bounds of numerical error), increase γ – do not update the state vector; if χ^2 decreases,

decrease γ – do update the state vector. Note that only the n-form can be easily converted to include this modification.

$$\mathbf{x}_{i+1} = \mathbf{x}_a + ((1+\gamma)\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\Sigma^{-1}\mathbf{K})^{-1}[\mathbf{K}^T\mathbf{S}_\Sigma^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) + (\gamma\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\Sigma^{-1}\mathbf{K})(\mathbf{x}_i - \mathbf{x}_a)] \tag{4.42}$$

In practice, the value of γ is set initially to zero. If subsequent retrieval χ^2 values start to diverge γ is set to one; if non-convergence persists then γ is increased by a factor of 10 until the χ^2 value at that particular iteration is less than the value of the previous iteration. If a solution continues to converge and γ has a value other than 0 then γ is reduced by a factor of 10.

4.5 Error Analysis Theory

Any measurement is of limited worth if it is not accompanied by an associated estimate of its error. This section is concerned with the theory of the propagation of errors and uncertainties through the retrieval algorithm, to produce an error that can be quoted alongside each element of the retrieved quantity.

The error analysis in this section is based upon that of Rodgers (1976, 1990), and will introduce the error diagnostics that will be used to characterise the radio occultation retrieval.

4.5.1 Diagnostic Matrices

A general inversion method can be expressed by the formula:

$$\hat{\mathbf{x}} = \mathbf{I}\left[\left(\mathbf{F}(\mathbf{x}, \hat{\mathbf{b}}) + \epsilon_\Sigma, \right)\hat{\mathbf{b}}, \hat{\mathbf{c}}, \mathbf{x}_a\right] \tag{4.43}$$

where $\mathbf{I}$ is the inversion algorithm and all other terms are as before.

This can be linearised about $\mathbf{x} = \mathbf{x}_a, \mathbf{b} = \hat{\mathbf{b}}$ and assuming $\Delta\mathbf{F}$ is constant to give

$$\hat{\mathbf{x}} - \mathbf{x}_a = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_a) + \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}}\left(\frac{\partial \mathbf{y}}{\partial \mathbf{b}}\boldsymbol{\epsilon}_b + \Delta\mathbf{F} + \boldsymbol{\epsilon}_y\right), \tag{4.44}$$

where $(\hat{\mathbf{x}} - \mathbf{x}_a)$ is the retrieval increment arising from a real profile increment $(\mathbf{x} - \mathbf{x}_a)$ and errors $\boldsymbol{\epsilon}_y$, $\boldsymbol{\epsilon}_b$ and $\Delta\mathbf{F}$. This equation enables the following terms to be defined.

The matrix of contribution functions

This is defined by the formula

$$\mathbf{D}_y = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}} \tag{4.45}$$

where the ith column of $\mathbf{D}_y$ is the contribution to the solution due to a unit change in the ith element of $\mathbf{y}$.

Differentiating equation 4.29 w.r.t. $\mathbf{y}$ gives

$$\mathbf{D}_y = \mathbf{S}_a\mathbf{K}^T(\mathbf{S}_\Sigma + \mathbf{K}\mathbf{S}_a\mathbf{K}^T)^{-1} \tag{4.46}$$

The matrix of averaging kernels

As mentioned previously, if the measuring system, (instrument and retrieval) is perfect, i.e. permits retrieval of atmospheric variables with zero error, then $\mathbf{D}_y\mathbf{K} = \mathbf{I}$. However if the system is not perfect then the resulting matrix will contain information on how the retrieval scheme uses the measurements.

The matrix of averaging kernels is defined by the formula

$$\mathbf{A} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} = \mathbf{D}_y\mathbf{K} \tag{4.47}$$

where $\mathbf{D}_y$ is the matrix of contribution functions, and $\mathbf{K}$ is the matrix of weighting functions. The rows of $\mathbf{A}$ are known as the averaging kernels (after Backus and Gilbert (1970)) because the retrieval at any location is an average of the whole profile weighted

by this row. The averaging kernels are often plotted, an equally meaningful quantity are the columns of $\mathbf{A}$, known as the delta-function responses. If the real atmosphere were perturbed by a delta function at a particular height, then the perturbation in the retrieved profile will be reflected by the shape of the corresponding column of $\mathbf{A}$. An ideal retrieval would have delta functions as the delta-function responses ($\mathbf{A} = \mathbf{I}$, a one-to-one relationship between the true profile and the retrieval). The delta-function responses for a real observing system form peaked functions, in other words the retrieved profile is a smoothed version of the real atmospheric profile. The width of these peaks give an indication of the resolution of the retrieval.

The sum of all the elements in each averaging kernel represent how the retrieval increments the *a priori* information if the entire atmosphere were shifted by a unit amount. This is a useful diagnostic and as such will be presented alongside results calculated.

The forward model parameter sensitivity matrix $\mathbf{K_b}$

This is defined by the formula

$$\mathbf{K}_\mathrm{b} = \frac{\partial \mathbf{y}}{\partial \mathbf{b}} \tag{4.48}$$

and is analogous to the matrix of weighting functions $\mathbf{K}$. A column of the forward model parameter sensitivity matrix $\mathbf{K}_\mathrm{b}$ is the change in the measurement vector $\mathbf{y}$ due to a unit change in the corresponding forward model parameter.

The solution error covariance matrix $\hat{\mathbf{S}}$

Uncertainty in the solution is a consequence of uncertainty in the measurements, and is given by the solution error covariance matrix $\hat{\mathbf{S}}$. The diagonal elements of this matrix contain variances of the individual components of the state vector. The principal diagonal is thus the simplest method of presenting the errors associated with the solution.

The solution covariance may be calculated in many ways (which are only exact in the linear limit); here, three formulations are presented:

$$\hat{\mathbf{S}} = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\Sigma^{-1}\mathbf{K})^{-1} \tag{4.49}$$

$$= \mathbf{S}_a - \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\Sigma)^{-1}\,\mathbf{K}\mathbf{S}_a \tag{4.50}$$

$$= (\mathbf{I} - \mathbf{A})\mathbf{S}_a \tag{4.51}$$

It is also noted that although the second of these equations is algebraically more complex, it only contains one inverse operation, and as such may be computationally more efficient, but is subject to the same criteria as equation 4.29.

4.5.2 Contributory error covariance matrices

Equation 4.44 can be rearranged to give the error in the retrieval

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{I} - \mathbf{A})(\mathbf{x}_a - \mathbf{x}) + \mathbf{D}_y[\mathbf{K}_b\boldsymbol{\epsilon}_b + \Delta\mathbf{F} + \boldsymbol{\epsilon}_y]. \tag{4.52}$$

This form of equation 4.44 allows the calculation of the solution error covariance matrix to be split into its constituent parts. The error terms can then be calculated to give:

- $\mathbf{S}_{xa}$, the contribution to the retrieval error covariance from the *a priori* error covariance.

$$\mathbf{S}_{xa} = (\mathbf{I} - \mathbf{A})\mathbf{S}_a(\mathbf{I} - \mathbf{A})^T \tag{4.53}$$

  $\mathbf{S}_{xa}$ is non-zero when the matrix of delta-function responses is not equal to the identity matrix. If there are off-diagonal elements in $\mathbf{A}$ then this error will cause the retrieved profile to be a smoothed version of the true profile. For this reason it is often known as the smoothing error covariance matrix.

- $\mathbf{S}_{ay}$, the contribution to the retrieval error from the measurement noise;

$$\mathbf{S}_{ay} = \mathbf{D}_y \mathbf{S}_y \mathbf{D}_y^T \tag{4.54}$$

- $\mathbf{S}_{ab}$, the contribution to the retrieval error from the error associated with the forward model parameters;

$$\mathbf{S}_{ab} = \mathbf{D}_y \mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T \mathbf{D}_y{}^T \tag{4.55}$$

- $\mathbf{S}_{a\Delta F}$, the contribution to the retrieval error from the forward model modelling error;

$$\mathbf{S}_{a\Delta F} = \mathbf{D}_y \mathbf{S}_{\Delta F} \mathbf{D}_y^T \tag{4.56}$$

4.5.3 Systematic errors

The retrievals performed using $\mathbf{S}_\Sigma$ include only zero bias random errors that are uncorrelated between one retrieval to the next. However, there may be errors in the forward model, or in the forward model parameters that do not have a zero bias, i.e. systematic errors.

The treatment of this type of error can be very *ad hoc*. However, if any information involved with the retrieval, whether it be *a priori* information, real observations or forward model parameters, is found to possess a bias then it should be corrected in a consistent manner. Two such methods are illustrated in Eyre (1987).

The effect of biases in any of the forward model or inversion parameters should also be considered by investigating how small perturbations to these parameters are mapped through to the retrieved solution.

Chapter 5

The effects of line-of-sight horizontal inhomogeneities

In this chapter general details of the forward model implementation and the Abelian inverse method are presented. The error and solution characterisation of the Abelian method is also shown and discussed.

An investigation of the spherical symmetry assumption used by the Abelian inverse model is undertaken. By defining an atmosphere with horizontal structure and using the forward model (with knowledge of horizontal structure) and its inverse (which uses the spherical symmetry assumption), it is possible to determine quantitatively the effects of the assumption.

5.1 General forward model implementation

The forward model, as discussed in chapter 4, maps from state space to measurement space accounting for the physics of the problem.

The forward model takes as input temperature (T) and humidity (q) on fixed pressure (P) levels; surface pressure (P_0); height above the geoid (z_0); the local radius of curvature (R_E); and a profile of tangent points, expressed on levels of impact parameter (a), the vertical position of measurement (§2.4). These inputs may take the form of a 2-d field of values or a single profile.

The input fields are linearly interpolated to give profiles of P, T and q at tangent point locations on some specified set of levels. The hydrostatic relation (equation 5.1)[1]

[1]This form of the relation can be constructed if temperature is assumed to vary linearly across a layer.

is used to map the pressure levels on to geopotential height levels ϕ, making use of z_0. If a single layer of the atmosphere is considered, the change in altitude over that layer is given by values of its mean temperature, and the pressure values directly above and below the layer.

$$\Delta\phi = \frac{-R}{Mg}\frac{(T_2^* - T_1^*)}{\ln(T_2^*/T_1^*)}\ln\left(\frac{P_2}{P_1}\right) \tag{5.1}$$

where g is the acceleration due to gravity, R is the gas constant; $\Delta\phi$ is the change in the geopotential height; T_2^* and T_1^* are the virtual temperatures[2] directly above and below the layer; and P_2 and P_1 are the pressures directly above and below the layer.

To improve numerical accuracy, these levels are subdivided further; and corresponding T, P and q values are calculated. At each of these levels, refractivity N is calculated using equation 3.1, where the partial pressure of water vapour P_W is calculated by equation 5.2.

$$P_W = \frac{Pq}{E} \tag{5.2}$$

where E is the ratio of the molecular weights of water vapour and dry air, and all other variables are as before.

The geopotential height surfaces are mapped on to geometric height surfaces z. Given the local radius of curvature, the radius r at each of these levels may be found by $r = R_E + z$, and hence the refractive radius, $x = rn$.

To find the bending angle at a particular value of impact parameter, it is necessary to find the position of the tangent point ($x = a$) on the refractivity profile, and to integrate the ray path for all heights above this point. For a spherically symmetric atmosphere, the contribution to the bending angle for a ray path between x_{upper} and x_{lower}, is given by the integral

[2]Virtual temperature definition: $T^* = T\frac{M_{ra}}{M_{rw}}$, where M_{ra} and M_{rw} are the molecular weights for dry air and air containing water vapour,respectively.

$$\Delta\varepsilon = -a \int_{x_{lower}}^{x_{upper}} \left(\frac{\partial \ln n(x)}{\partial x} \right) (x^2 - a^2)^{-1/2} \, dx \tag{5.3}$$

This is solved by assuming that refractivity decreases exponentially with between height levels, reducing equation 5.3 to a standard integral. The total angle is given by $\varepsilon = 2 \sum \Delta\varepsilon$, where the sum is taken from the tangent point to the 'top' of the model atmosphere, and the factor 2 arises from the fact that each path length, between x_{upper} and x_{lower}, is effectively sampled twice in spherical symmetry. Further details of this method can be found in Eyre (1994).

Non-spherically symmetric case

If there are departures from spherical symmetry the equivalent of equation 3.7 is (Eyre 1994)

$$\frac{d\phi}{\tan\phi} + \frac{dr}{r}\left(1 - \frac{1}{n}\left(\frac{\partial n}{\partial \theta}\right)_r \cot\phi\right) + \frac{dn}{n} = 0 \tag{5.4}$$

Assuming $K = \left(1 - \frac{1}{n}\left(\frac{\partial n}{\partial \theta}{}_r\right)\cot\phi\right)$ may be approximated as constant over a layer, the generalised form of Bouguer's formula (equation 3.8) is obtained

$$r^K n \sin\phi = x^* \sin\phi = \text{constant} = a^* \tag{5.5}$$

Using equations 3.3 & 5.5 the following is obtained

$$\Delta\varepsilon = -a^* \int_{x_{lower}}^{x_{upper}} \left(\frac{\partial \ln n(x)}{\partial x^*} \right) (x^{*^2} - a^{*^2})^{-1/2} \, dx + \frac{1}{r}\left(\frac{\partial \ln n}{\partial \theta}\right)_r \, dr \tag{5.6}$$

where x^* and a^* replace x and a in equation 5.3. It is assumed that $\frac{1}{r}\left(\frac{\partial \ln n}{\partial \theta}\right)_r \, dr$ is constant for a single layer.

This calculation must be performed separately for each side of the path which, in the non-spherical case, provides a different contribution to the refracted angle at the tangent point.

Further details of this numerical representation of the forward Abel transform can be found in Eyre (1994).

5.2 The Abelian inverse model implementation

The inverse model maps from measurement space on to state space. By forming a loop containing the forward model and its inverse, it is possible to compare the 'truth' with the inverted measurement vector and subsequently estimate the errors incurred by the inverse model (and forward model).

The inverse model comprises two main equations: the inverse Abel transform and the hydrostatic relation (equation 4.15). This section isolates each of the numerical representations of these equations and investigates their characteristics.

5.2.1 The inverse Abel transform

The inverse Abel transform is used to map from bending angle space to refractivity space (as a function of refractive radius x). The equation describing the transform is given by equation 4.7.

If values of bending angle decay exponentially with increasing impact parameter, the inverse Abel transform reduces to two Bessel functions which can be approximated by polynomial expansions [Prof. G. J. Pert, University of York: private communication]; see Abramowitz and Segun (1968) for further details.

By defining bending angles at two height levels in a specified impact parameter range and fitting an exponential curve between the two points, it is possible to compare the numerical routine for the inverted Abel transform with the tabulated Bessel functions.

The numerical representation for the inverted Abel transform used in some of the work herein was partially written by Dr. S. Healy.

A bending angle profile was defined at two altitudes (0 and 80 km) and an exponential curve was fitted between them (defined using 150 data points). The refractivity profile calculated with the numerical representation of the inverse Abel transform was compared with the tabulated Bessel function. The results from this experiment are shown by figure 5.1.

Panels (a) and (b) show the defined bending angle profile and the difference between the numerical integration and the tabulated Bessel functions. The largest percentage errors associated with the control run are of the order of $5\times10^{-5}\%$ which are situated in the upper atmosphere and can be attributed to computer truncation error.

An investigation into how bending angle errors are mapped to refractivity via the numerical transform can be performed by adding structure to the defined bending angle values, typical of noise, and calculating the associated refractivity profile. By comparing the retrieved refractivity profile with the 'truth', bending angle errors from the numerical inverse Abel transform can be estimated.

The results from perturbing bending angle elements with an alternating value (about the unperturbed values) are shown in panels (c)−(h) of figure 5.1.

In particular, panels (c) and (d) show the results from perturbing the control run by an alternating value of magnitude 0.5 %; and panels (e) and (f) show the results from perturbing the control run in a similar manner but with a magnitude of 5 %. This form of perturbation in bending angle space is reduced significantly when it is mapped to refractivity space, with the percentage refractivity errors an order of magnitude smaller than the percentage bending angle errors. This is expected since refractivity is an integrated quantity and thus will smooth unbiased errors.

Another, more realistic, form of error structure can also be used to simulate instrument, or atmospheric, noise. A normally distributed random number sequence can be used to form a percentage bending angle perturbation (panels (g) and (h)) to the control run. As with the alternate perturbations of constant amount the integral quantity of refractivity smooths the sharp differences in bending angle.

The main conclusion from this study is that unbiased bending angle structure (typical of noise) are smoothed by the Abel transform integral. Using a number of integral

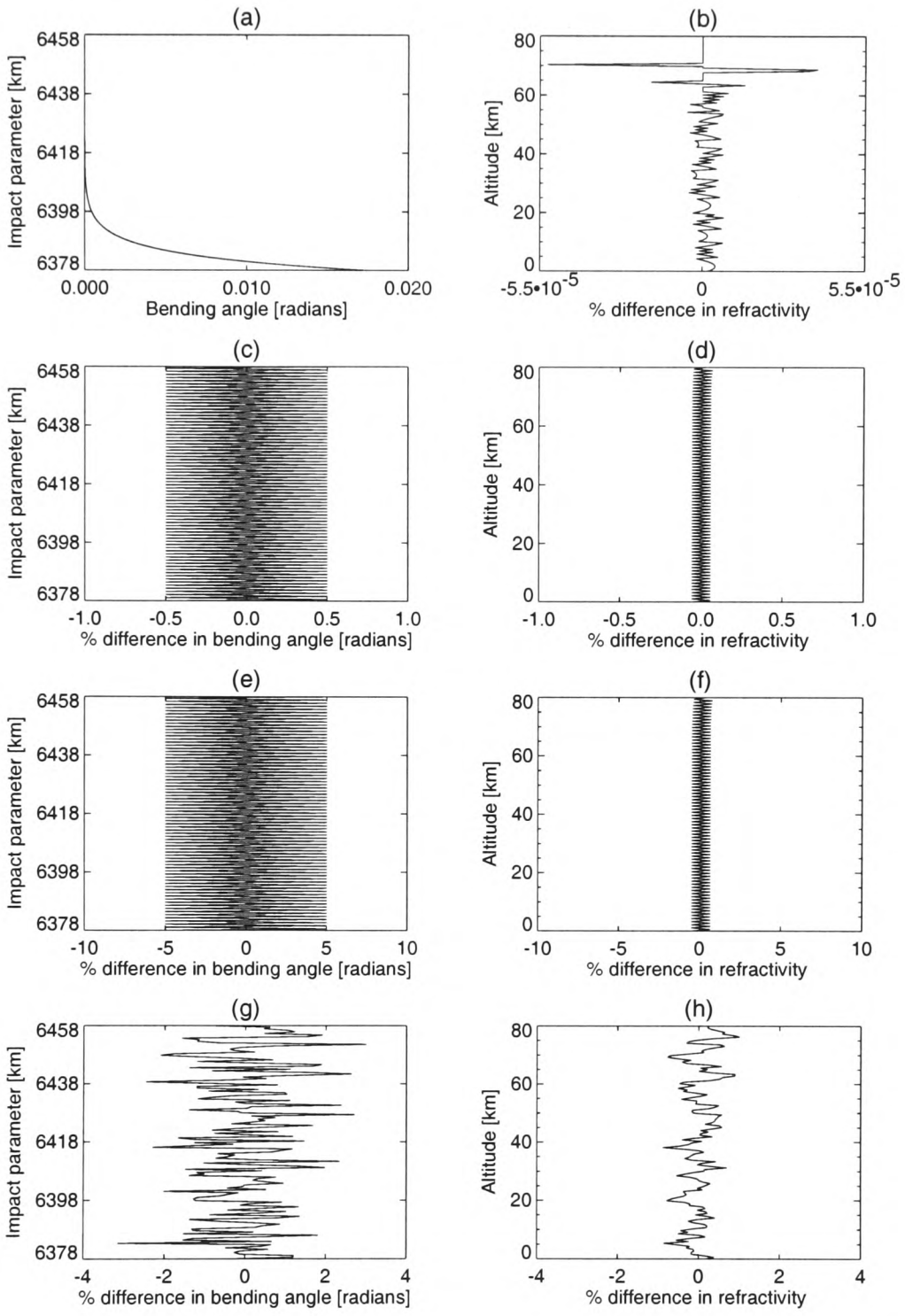


Figure 5.1: Assuming an exponentially decaying bending angle profile, the inverse Abel transform is equivalent to two Bessel functions. This can be used to assess the numerical inverse Abel transform performance. Panel (a) shows the unperturbed bending angle profile used to retrieve refractivity and panel (b) shows the percentage difference between the numerical inverse Abel transform and the tabulated Bessel functions, i.e. the control run. The left panels (c), (e) and (g) show the difference between the perturbed and unperturbed bending angle profiles; and the right panels (d), (f) and (h) show the difference between the perturbed and unperturbed retrieved refractivity profiles. In particular, panels (c) and (d) show the results from perturbing the control bending angle profile by an alternating value of magnitude 0.5 %; panels (e) and (f) show the results from perturbing the control run in a similar manner but with a magnitude of 5 %; and panels (g) and (h) show the results from perturbing the control bending angle profile with a vector of normally distributed random numbers.

definition points typical of the number of filtered bending angle observations (Rocken et al. 1997) has the effect of reducing the bending angle error by approximately an order of magnitude in refractivity.

5.2.2 The hydrostatic relation

Geometric height levels z are calculated from values of x using the local radius of curvature thus:

$$z = \frac{x}{n} - R_E \tag{5.7}$$

Equation 4.14 is used to calculate density from refractivity. The hydrostatic relation is used to calculate pressure from density as a function of height, assuming a starting pressure for the integration.

$$dP = -\int \rho \, g \, dz \tag{5.8}$$

where ρ is the atmospheric density, and all other variables are as before.

The method used to integrate this relation assumes density decreases linearly in ln-pressure space, which leads to the following analytic form for the relation

$$\Delta P = g \left[\frac{\phi_{i+1} - \phi_i}{\ln \rho_i - \ln \rho_{i+1}} \right] (\rho_i - \rho_{i+1}) \tag{5.9}$$

where i is the level number and all other variables are as before. Temperature is calculated via the ideal gas equation.

The density assumption used to derive this form for the hydrostatic relation is equivalent to assuming a mean layer temperature. Although this is not consistent with the assumption made in the forward model (§5.1), it can be justified if the number and

thickness of forward model levels is considered. In the forward model representation the number of levels describing the vertical atmospheric range sounded by radio occultation is a fraction of the typical number of radio occultation measurements, which means that the assumption of a mean temperature is used only for a comparatively small vertical region.

The representation of the hydrostatic relation used (equation 5.9) can be investigated by taking a typical atmospheric profile (here represented by a modified UKMO model profile, as described in §6.1.1) to calculate a refractivity profile, omitting atmospheric water vapour[3]. Density is calculated as a function of refractive radius (using equation 4.14) then as a function of height (equation 5.7). Equation 5.9 is used to calculate pressure (and subsequently temperature) on height levels, respectively. This experiment is essentially a reduced form of forward and inverse model loop which does not use the Abel transform pair (equations 4.7).

If considerable vertical structure is present in the region where the hydrostatic integration is initialised, the first-guess for the pressure may lead to biases lower in the atmosphere. This is due to the nature of the hydrostatic relation: the first-guess represents information about a single infinitesimal point in height, however the hydrostatic relation considers the vertical density gradient over a small region. If the first-guess and the vertical density gradient about the position of the first-guess conflict, pressure (and subsequently temperature) errors are introduced which decay exponentially with depth.

For purposes of illustration this problem can be avoided by defining a temperature (260 K) at a suitable height (50 km, say), above which the atmosphere is isothermal. If the hydrostatic relation is initialised in this isothermal region, where the density falls off uniformly, then there should be no conflicting information. For this experiment the integration is initialised at an altitude of 55 km. The retrieved temperature profile on retrieved geopotential height levels are compared with the original temperature profile as a function of geopotential height. Figure 5.2 shows results from this investigation.

[3]Omitting water vapour from the forward model avoids the temperature/water vapour ambiguity (§4.3.2).

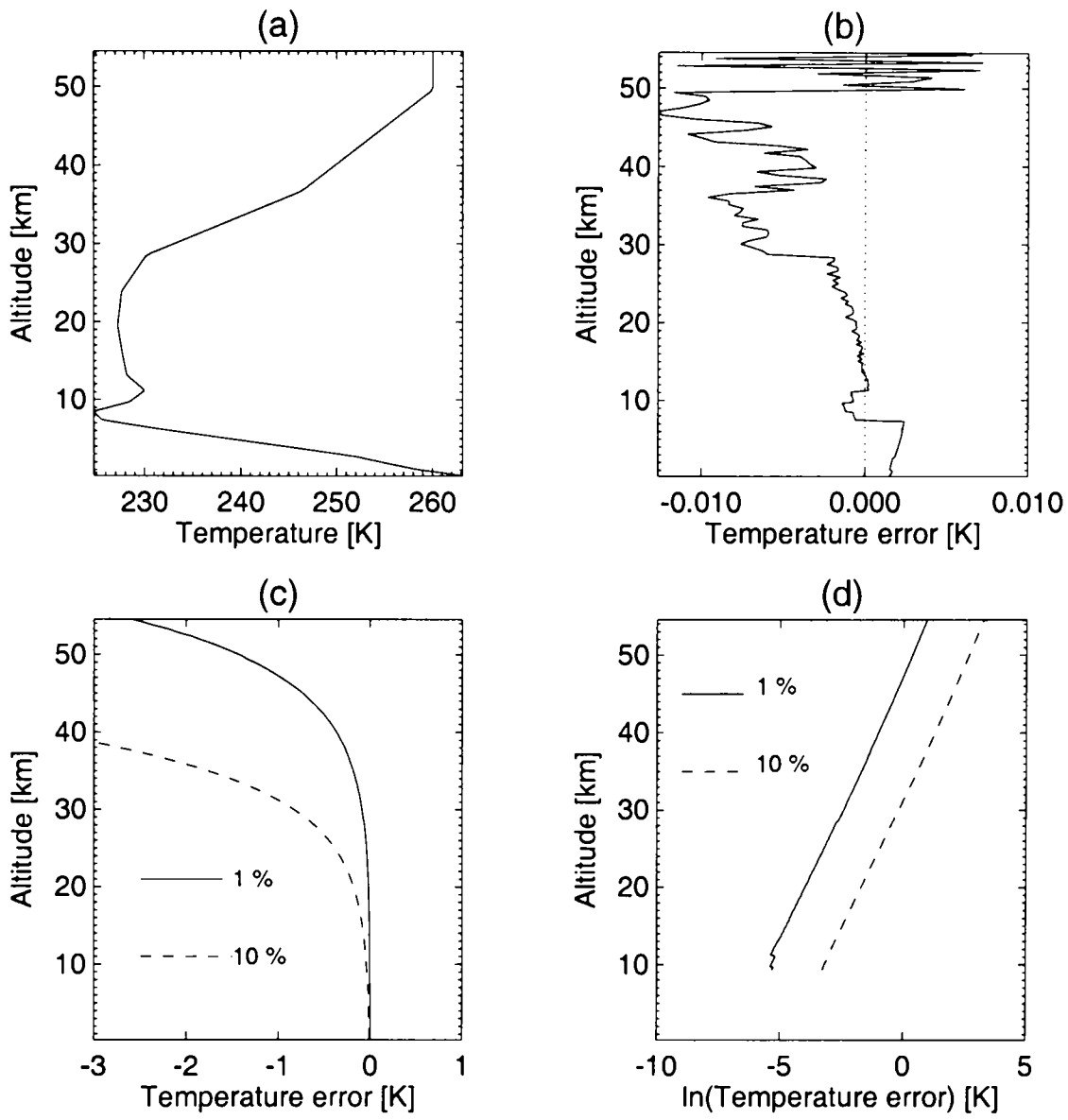


Figure 5.2: Quantitatively assessing errors introduced by the analytic representation of the hydrostatic relation. Panel (a) shows the modified UKMO model temperature profile which is used for the experiment; panel (b) shows the errors introduced by the analytic hydrostatic relation with a perfect first-guess for the integration; panel (c) shows the errors introduced by perturbing the first-guess for the hydrostatic integration at 55 km; and panel (d) shows the ln(temperature error) confirming that the error reduction is exponential.

Panel (b) of this figure is the temperature error introduced by the hydrostatic representation[4] using perfect first-guess of the isothermal state for the temperature profile shown in panel (a). The errors from this are of the order expected from numerical error. Small error structure can be attributed to sharp vertical temperature gradients, e.g. the tropopause.

Panel (c) shows that perturbing the first-guess for the hydrostatic relation introduces an error which decreases exponentially with depth. The exponential decrease in error is confirmed by panel (d). Lower levels of panel (d) are omitted because the retrieval

[4]It should be noted that interpolation will also introduce errors, but are comparatively small with those from the numerical hydrostatic representation.

temperature error is of the order of numerical error and does not add any information to the figure.

In this section it has been shown that the errors introduced by the numerical representation of the inverse model are small (relative to values of temperature) and unbiased. It is important to show this because they will be used as 'black boxes' for further experiments shown in this chapter.

5.3 Simple horizontal inhomogeneity structures

Although Bouguer's formula (equation 3.8) assumes local spherical symmetry, errors in derived values of bending angle and impact parameter due to this assumption have a small effect on retrieved refractive index profiles compared with the use of the inverse Abel transform (Kursinski 1997).

Since the Earth's atmosphere often has considerable departures from spherical symmetry it is necessary to investigate quantitatively the error incurred when applying this assumption to an atmosphere with a realistic geoid and horizontal structure.

Using the forward model, a meteorological field (omitting the effects of atmospheric water vapour) is defined from which to calculate a profile of simulated bending angle observations. These observations can then be inverted, using the inverse Abel transform and hydrostatic relation, to retrieve a profile of refractivity, pressure and temperature as a function of geometric height. This retrieved profile can then be compared with the information from the field used to calculate the observations.

5.3.1 Types of inhomogeneity studied

The different types of inhomogeneity studied include a linear form, a second-order form, and a realistic atmospheric section. The first two forms are obviously idealised, but can be used to study extreme conditions in which occultation measurements may be inverted: the errors can then be quantified and studied. It should be realised that horizontal structure is unlikely to be consistent throughout the vertical.

Linear form

The first, and simplest, form is the linear structure; it is a temperature field with a uniform linear horizontal gradient (figure 5.3, panel (a)). This form is used to simulate a sloping temperature gradient.

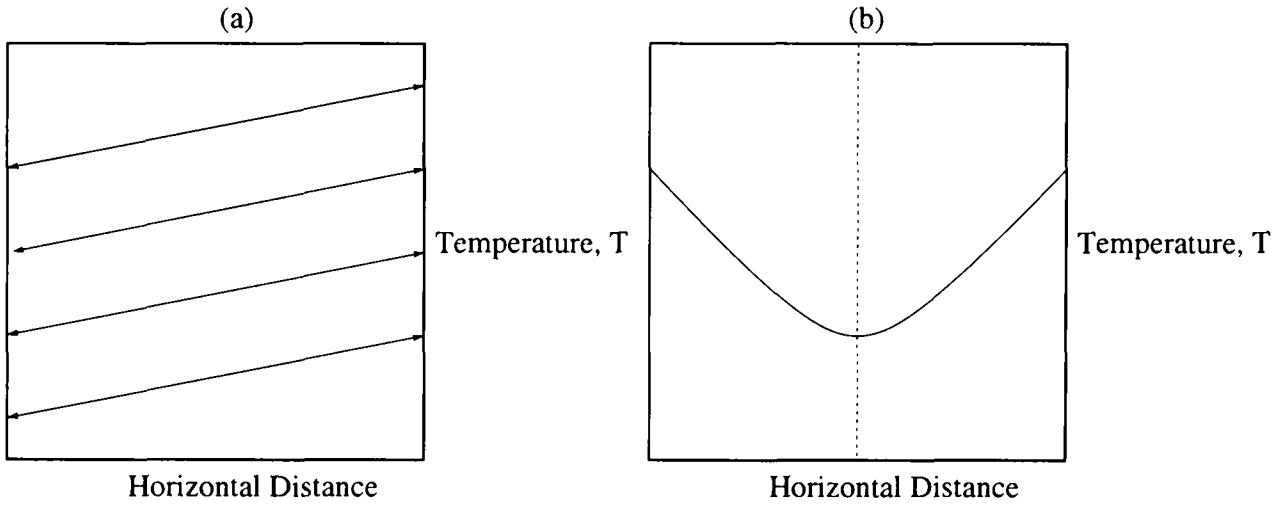


Figure 5.3: A schematic of the different horizontal inhomogeneities studied: (a) the linear horizontal inhomogeneity and (b) the second-order horizontal inhomogeneity. The vertical axis for both plots is height or ln(pressure).

Second-order form

The second-order form is a natural extension to the linear form. The curved structure studied here is symmetrical about the horizontal midpoint (figure 5.3 (b)). This form is used to simulate temperature retrieval errors from observations taken in a rapidly varying atmospheric field. By altering the magnitude of the curve it is possible to investigate how the curvature of the inhomogeneity affects the retrieval error.

Realistic atmospheric 2-d field

A realistic atmospheric 2-d field can be obtained from the output of an operational NWP model, which can be interpolated as required, and used to compute the errors incurred by typical atmospheric horizontal structure along the line of sight.

The fields used were obtained from the ECMWF archives at the British Atmospheric Data Centre (BADC), at the Rutherford Appleton Laboratory. The data are from operational analysis fields with the model resolution spectrally truncated from T213 to

T106, for storage purposes. The model has 31 vertical levels (surface−10 hPa), which are expressed in hybrid-sigma pressure coordinates, and has 320 elements in the East−West direction. For this experiment, a field of constant latitude (40°N), over all longitudes on the 15th May, 1992 is used (figure 5.4). The horizontal spacing of the measurements is calculated by considering the circumference of the circle described by 40°N on the Earth, divided by the number of model values in the East−West direction (an approximate distance of 96 km).

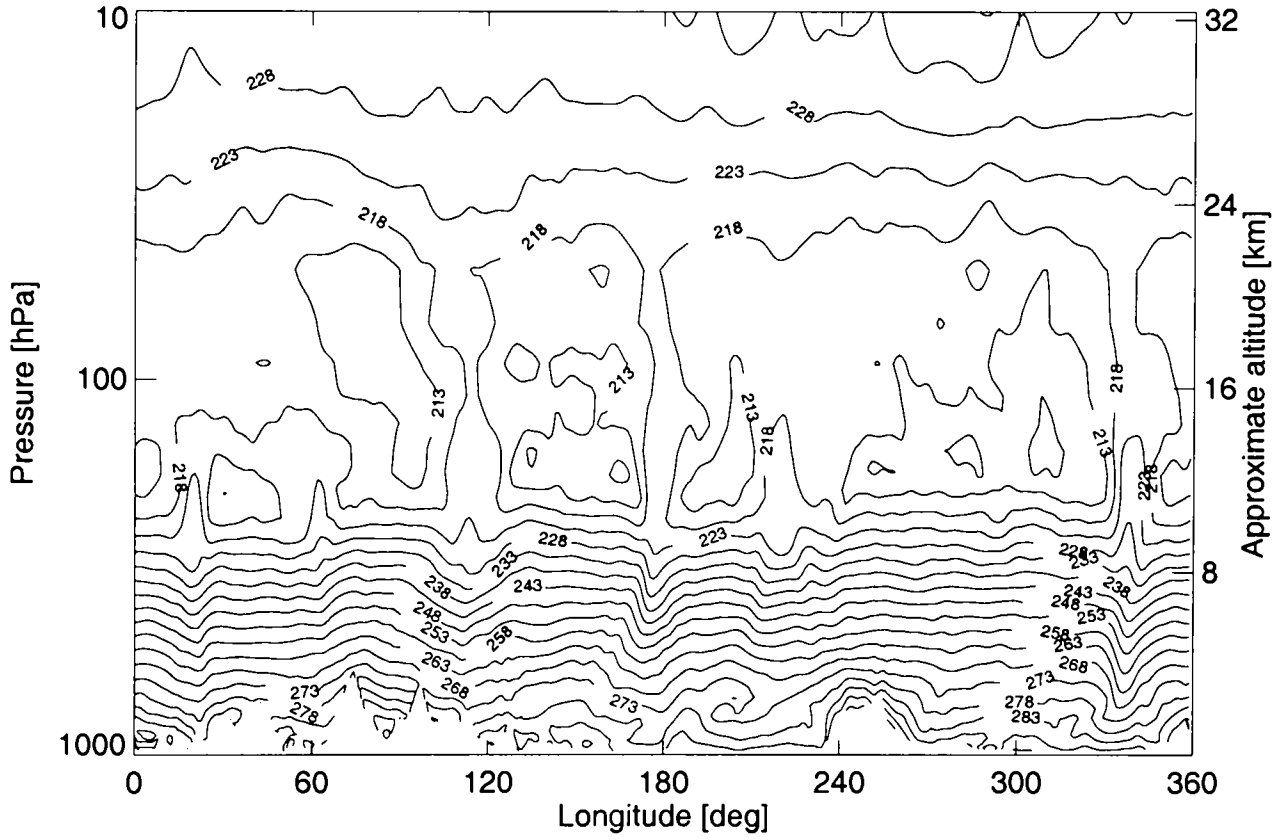


Figure 5.4: Cross-section of temperature (K), as a function of pressure (hPa), through a mid-latitude temperature field as represented by spectrally truncated operational analysis from the ECMWF archives at the BADC. The section is for 40°N, over all longitudes on the 15th May, 1992.

To investigate fully how horizontal variations in meteorological parameters along the line of sight affect retrieval errors, a 'window' is defined which can move along the horizontal field; the width of the window being the same as the width of the section used in each forward model run (9 horizontal elements). The centre of the forward model section is used to define the horizontal position of the measurement vector, and subsequently the inverted measurement vectors. By moving the window along the

ECMWF field, the location of the bending angles are moved along the horizontal. Once this procedure is completed for all horizontal nodes in the ECMWF field, an error contour map on the height levels can be plotted. This 'map' can be used to display how retrieval errors are related to the original atmospheric section.

For the 'dry' experiment (i.e. neglecting the water vapour field, figure 5.16), the forward model (§5.1) utilises the temperature, height above the geoid and surface pressure sections. Since the model has a top boundary level of 10 hPa, the temperature section is extrapolated to pressure levels corresponding to height levels of approximately 50 km, first with the top model level values and second with an isothermal (231 K) atmosphere (the magnitude of which represents the mean top layer value of the field) to eliminate horizontal gradients that exist at the top of the model field. This field is used as the forward model input. The Earth's radius and the local radius of curvature in the East–West direction (constant for a particular latitude) at this latitude are computed using software written by Mr. D. Offiler (UKMO) using data from the U.S. Defense Mapping Agency. The impact parameter profiles are chosen carefully such that the initial values are higher than the surrounding orography; subsequent impact parameter levels are separated by a constant amount (500 metres), a distance comparable to GNSS radio occultation measurements. The geographical positioning of sequential observations in a profile are the same, i.e. they lie on top of each other. Although, this is a seemingly crude assumption to make (often sequential real observations 'drift', forming a skewed profile), the extent of the 'drift' in real observations should be such that this assumption is valid. The 2-d assumption made, i.e. each refracted ray path taken can be described by a 2-d plane, is valid since the theory used to invert radio occultation measurements makes the same assumption.

The inverse Abel transform maps the bending angle measurements (as a function of impact parameter) on to values of refractivity as a function of refractive radius. This grid is subsequently mapped on to a height grid, using the local radius of curvature (perfect knowledge of this parameter is assumed). Refractivity is used to calculate density, and subsequently temperature on retrieved pressure levels, using the hydrostatic relation (assuming a fixed pressure at a certain height). Because the original field is extrapolated

with an isothermal temperature value, the first guess for the hydrostatic relation is the isothermal temperature, which guarantees relatively small retrieved pressure errors (§5.2.2).

Using this procedure, several experiments are performed:

- First, the difference between the measurement vector computed assuming no horizontal refractivity structure (the 'forward' Abel transform, equation 4.7) and with horizontal refractivity structure (equation 5.6) over the complete field is computed on a set of fixed height levels[5], allowing an estimate of the forward modelling error to be obtained. These estimates can then be used by the optimal estimation inverse method to account for typical line-of-sight refractivity structures.

- Second, temperature retrieval errors incurred from assuming spherical symmetry can be investigated by comparing the inverted measurement vector with the true temperature at the window midpoint on the set of fixed height levels.

- Third, refractivity retrieval errors incurred from assuming spherical symmetry can be investigated by combining typical temperature and water vapour gradients (figure 5.4). The method used to investigate refractivity retrieval errors is similar to the temperature retrieval error investigation.

- Fourth, regression coefficients are fitted to both temperature and refractivity retrieval errors by considering the horizontal atmospheric structure local to the measurements. This represents an attempt to explain realistic retrieval errors using linear and second-order forms.

5.3.2 Results

This section presents and discusses results from the horizontal inhomogeneities described in §5.3.1. The effects of typical pressure inhomogeneities are also investigated.

[5]This allowed easy computation of the mean and the r.m.s. of the differences.

The retrieval error is defined as the departure of the retrieved profile from the vertical profile at the nominal mid-point of the forward model window. In general, the temperature retrievals are initialised at 55 km with perfect knowledge of the atmospheric state, i.e. zero error, unless otherwise stated.

Linear form

A linear temperature horizontal inhomogeneity is defined as the forward model input, with the measurements defined at the horizontal midpoint of the field. By comparing the retrieved temperature at the middle of the section with the actual temperature profile, the error incurred from this horizontal structure can be computed.

Results are shown by figure 5.5. The maximum gradient considered is 3 K/100 km because this represents a typical maximum temperature gradient within the atmosphere.

The background error (errors incurred from no horizontal variation) is eliminated from the results with linear horizontal structure such that the error shown represents an estimate of the error incurred by the horizontal inhomogeneity alone. The results show that the relationship between the temperature retrieval error incurred from a linear inhomogeneity and the magnitude of the inhomogeneity has a quadratic form.

The quadratic error structure common to all plots is caused by the overestimate of the refractivity on one side of the horizontal field midpoint and underestimate of the refractivity on the other side. The net value of refractivity is an underestimate and is a symptom of the non-linearity of the problem. Similar results are obtained when negative gradients are considered. It is noted that the vertical structure of resulting errors is similar to the background error but of opposite sign, which is of similar origin.

Second-order form

Several shapes were used when studying the horizontal curved temperature structure. In each case, the curve is described by considering the distance from the field midpoint and adding this increment to the original isothermal field (250 K). Four second-order curves have been studied.

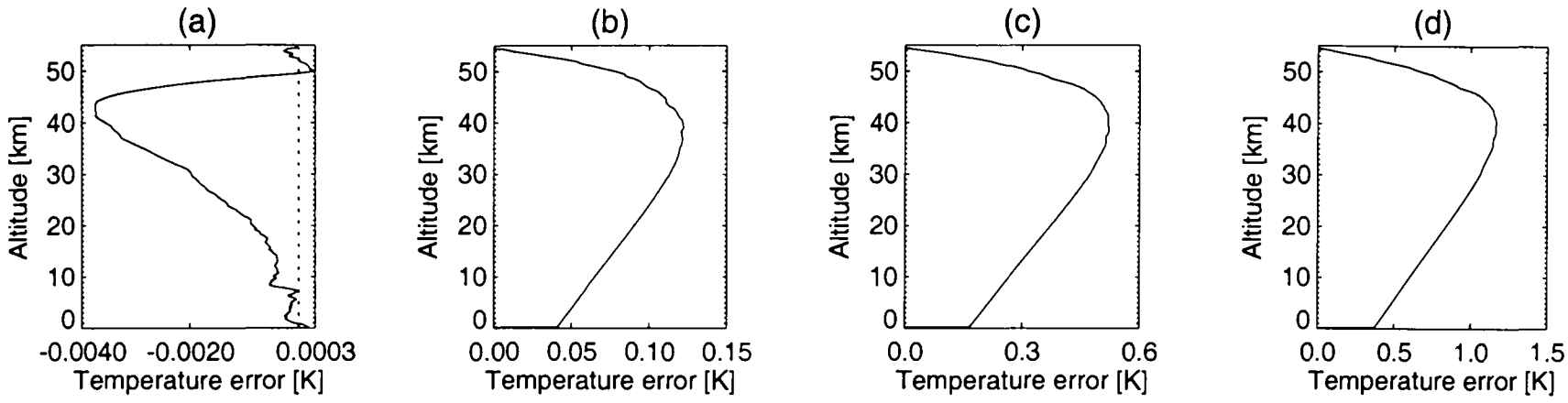


Figure 5.5: Error incurred when considering linear horizontal inhomogeneities superimposed on an isothermal field (250 K): (a) gradient=0, the 'background' retrieval error; (b) gradient=0.5 K/100 km; (c) gradient=2 K/100 km; and (d) gradient=3 K/100 km. The 'background' retrieval error has been removed from errors shown in panels (b), (c) and (d).

By defining the measurements to be positioned at the horizontal midpoint of the field, the error incurred by such a field can be assessed by comparing the retrieved temperature at the centre of the curved structure with the input temperature profile at the centre of the structure.

The experiment defined a field with 9 horizontal elements, with a horizontal separation of 100 km. Results are shown by figure 5.6.

The errors involved with such horizontal structures are, in general, larger than the linear case. This can be explained by the retrieval of refractivity: an excess contribution of refractivity from one side of the centre will add to the excess contribution from the other side; there is no cancelling of contribution so the retrieved temperature is overestimated.

Results from considering the $-x^2$ curves are of the same magnitude and structure but of opposite sign.

Pressure effects

Because pressure and temperature are closely inter-linked, it is interesting to study the effects of horizontally varying pressure, with temperature remaining constant (both vertically and horizontally), with the value of 250 K. A pressure variation of $\pm$ 4% over 200 km is used which is large enough to simulate a typical atmospheric situation [Dr. J. J. Barnett: private communication].

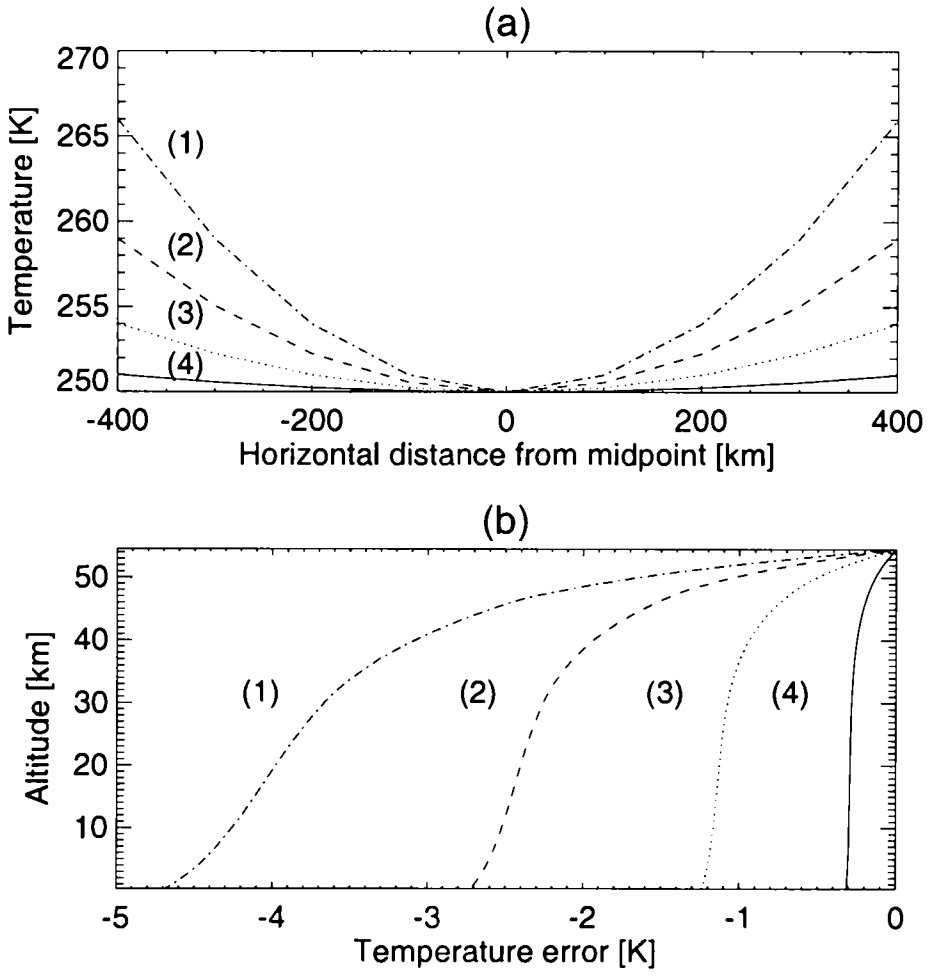


Figure 5.6: Errors incurred from second-order curved horizontal inhomogeneities. Panel (a) shows the horizontal temperature variation (as a function of distance from the field midpoint) for curves 1, 2, 3 and 4 used to define different temperature fields; and panel (b) shows the corresponding temperature retrieval errors as a function of altitude.

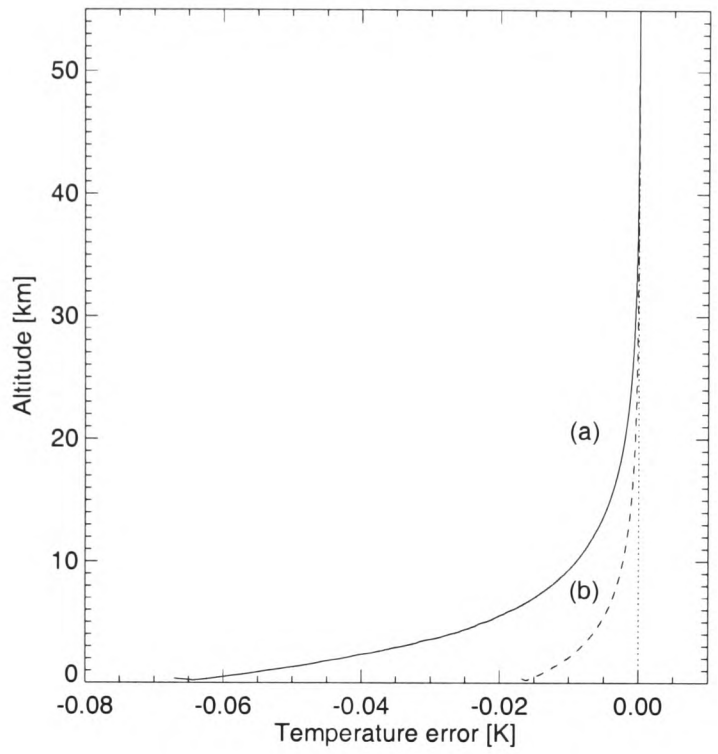


Figure 5.7: Results from realistically varying the horizontal pressure field by 4% (curve a) and 2% (curve b) over 200 km about the horizontal position of the ray tangent points.

Figure 5.7 shows the results from this experiment. The pressure perturbation described above is represented as $\Delta\ P$; the effect from perturbing the profile by $\Delta\ \frac{P}{2}$ is also investigated.

The results of this particular experiment indicate that altering the pressure by a realistic amount largely affects the retrieval near the surface which decreases as a function of altitude. Halving the pressure variation more than halves the error associated with the retrieved temperature error, implying that the relationship between pressure and bending angle is non-linear. The maximum error from a realistic pressure horizontal inhomogeneity is found to be comparable with a small linear temperature horizontal gradient. Therefore temperature horizontal gradients will provide the largest contribution to temperature (and 'dry' refractivity) retrieval errors.

Realistic atmospheric field

The ECMWF field used is shown by figure 5.4, and the method used to study the field is discussed in §5.3.1.

A useful method of studying the errors from an ensemble of retrieved profiles is to

look at the standard deviation (σ) of the temperature (or refractivity) field and the r.m.s. values of the retrieval errors at each altitude level.

Bending angle errors

The modelling of measurement vectors using the ECMWF model field can be used to estimate the error incurred from not accounting for horizontal variations in temperature and water vapour (§5.3.5). Figure 5.8 shows the gross statistics derived from figure 5.4.

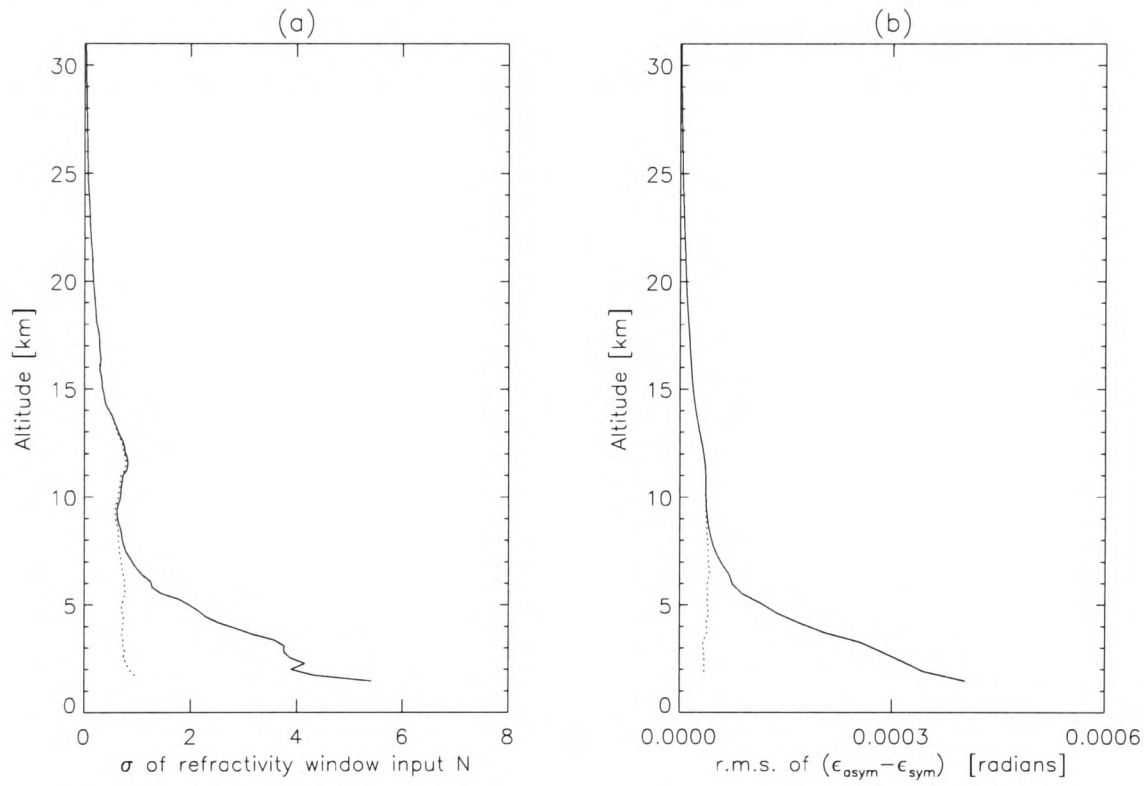


Figure 5.8: Gross bending angle statistics of figure 5.4. Panel (a) shows the horizontal standard deviation structure of the 'wet' refractivity field; and panel (b) shows r.m.s. values for the difference between the measurement vectors computed using a forward model with and without knowledge of the horizontal structure of the field (denoted by $\varepsilon_{\text{asym}}$ and ε_{sym}, respectively). Dotted lines show the statistics for the 'dry' refractivity field.

Figure 5.8 shows that the largest forward modelling errors from neglecting horizontal atmospheric variations (temperature and water vapour, shown by figures 5.4 & 5.16 respectively) are located in the lower atmosphere. Comparing the statistics for the 'wet' refractivity field with results from the 'dry' refractivity field shows that the largest errors are below 10 km, suggesting that horizontal water vapour variations in the lower atmosphere are important to consider when forward modelling.

Figure 5.9 shows the horizontal variation bending angle error estimates from figure 5.8 compared with the JPL version of this error[6].

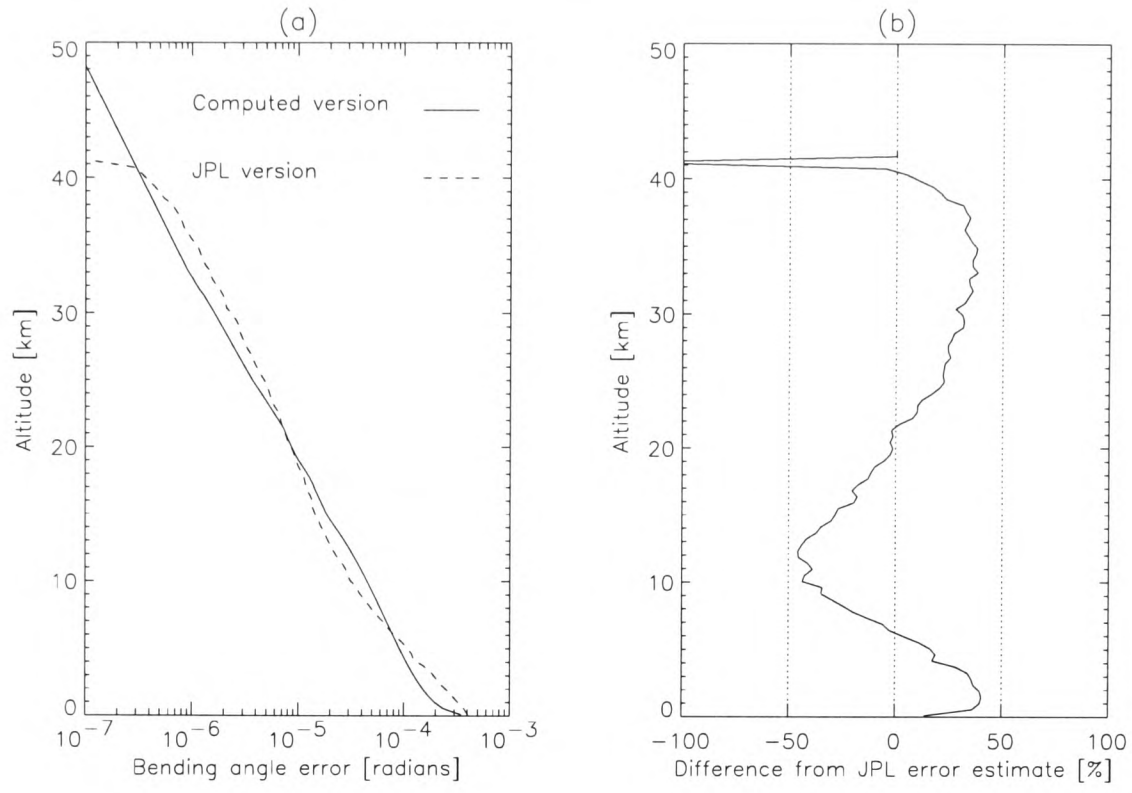


Figure 5.9: Bending angle error estimates incurred by assuming spherical symmetry. Panel (a) shows the estimates shown in figure 5.8 with the JPL error estimate; and panel (b) shows the percentage difference from the JPL error estimate.

The two error estimates show good agreement below 40 km with differences of less than 50%. The magnitude of the differences between the two estimates is expected since the fields used to compute them are conceivably very different. The error estimates computed in this work are preferred since they reflect errors from the same forward model used in further work.

Temperature retrieval error

Gross statistics of retrieved temperature corresponding to figure 5.4 are shown by figure 5.10. For each forward model run, the horizontal variation over a subwindow (i.e. temperature values and retrieval errors immediately surrounding the centre of the full forward model window) is noted.

Generally, the horizontal r.m.s structure of the temperature retrieval errors is similar to the standard deviation of the temperature field. The largest horizontal

[6]This error estimate is published in Kursinski et al. (1996), expressed as refractivity error, which was subsequently expressed as bending angle errors using a forward model by Luntama (1997) (chapter 6).

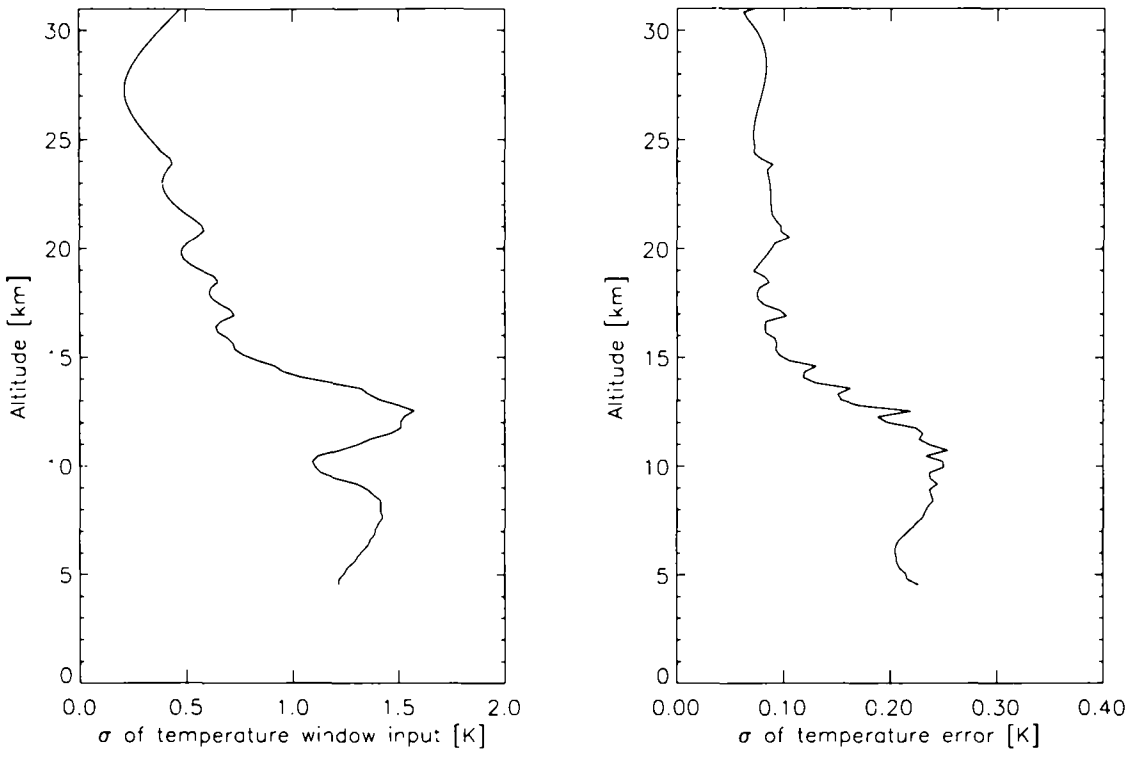


Figure 5.10: Gross temperature statistics of figure 5.4. Panel (a) shows the σ structure of the temperature window input and panel (b) shows the horizontal r.m.s structure of the retrieval error.

temperature variations are found to be situated at altitude levels corresponding to the lower stratosphere and the tropopause region, this is also shown by the temperature retrieval errors. The temperature retrieval errors are shown to be approximately an order of magnitude smaller than the values shown for the temperature variation.

In particular, the largest positive and negative retrieval errors (of the order of ± 1 kelvin) can be traced back to areas of local horizontal temperature gradients. For the temperature field, the largest positive and negative errors are associated with the same area of horizontal structure ($325°-350°$) and is shown by figure 5.11.

The largest retrieval errors centre themselves on the sharpest local temperature structure. The main positive error contour is flanked by two negative error lobes. This is expected: for any one local area of curvature there must be an area of curvature beside it. The magnitude of the main temperature horizontal inhomogeneity (panel (a)) is approximately 4 K/100 km, and the magnitude and sign of the retrieval error (of the order of 1 kelvin) agrees reasonably well with the analysis of simple linear inhomogeneities in the previous section.

A second region is investigated that exhibits similar horizontal structure, ($10°-35°$

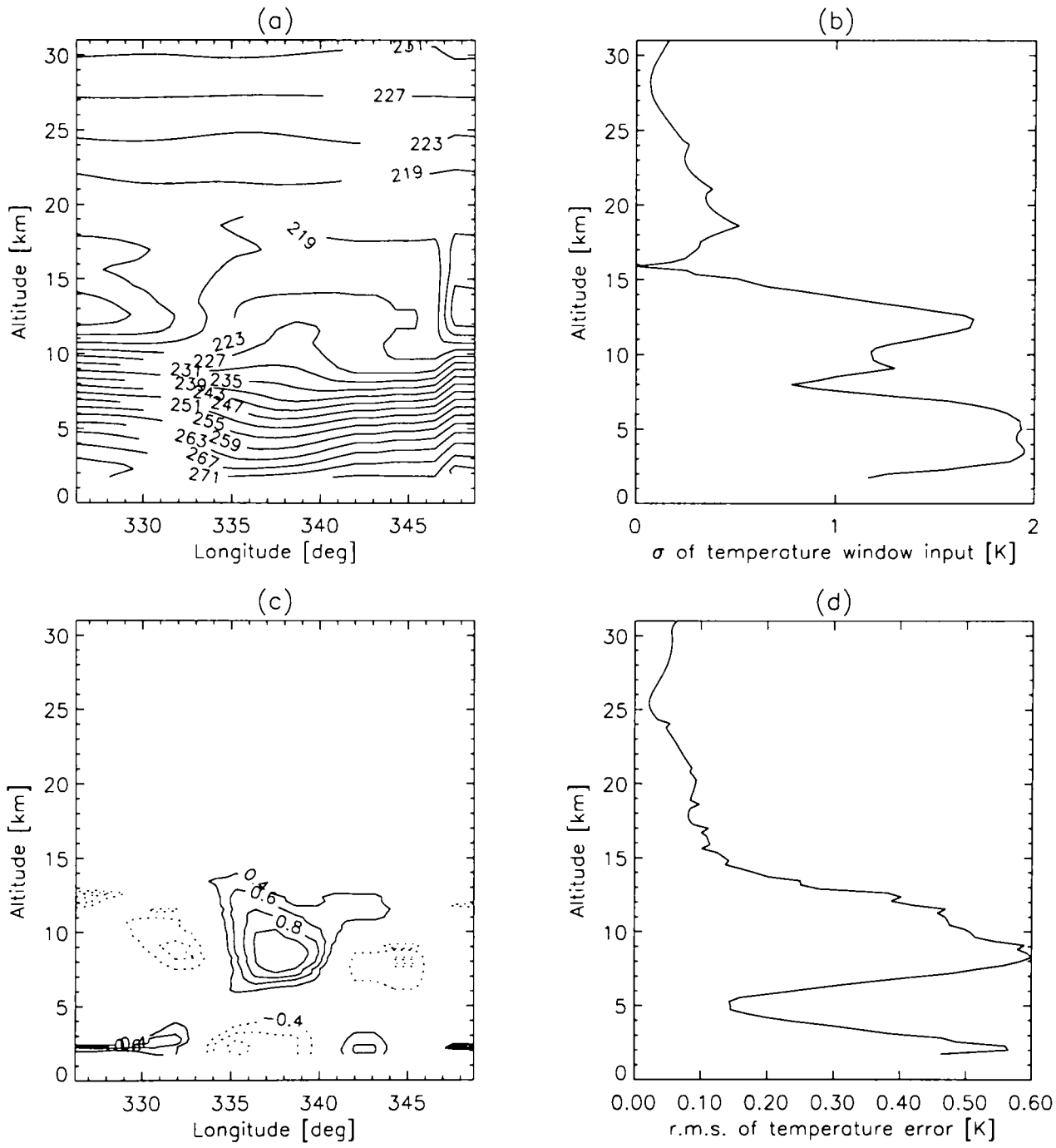


Figure 5.11: A summary of the temperature field in the region corresponding to the largest positive and negative temperature retrieval errors. Panel (a) shows the area of the original temperature field that corresponds to the largest retrieval errors (plotted with a contour interval of 4 kelvin); panel (b) shows the horizontal σ structure from the temperature field; panel (c) shows the retrieved temperature error contour pattern (with a contour interval of 0.2 kelvin); and panel (d) shows the horizontal r.m.s. structure for the retrieval errors. Solid and broken line contours represent positive and negative errors, respectively.

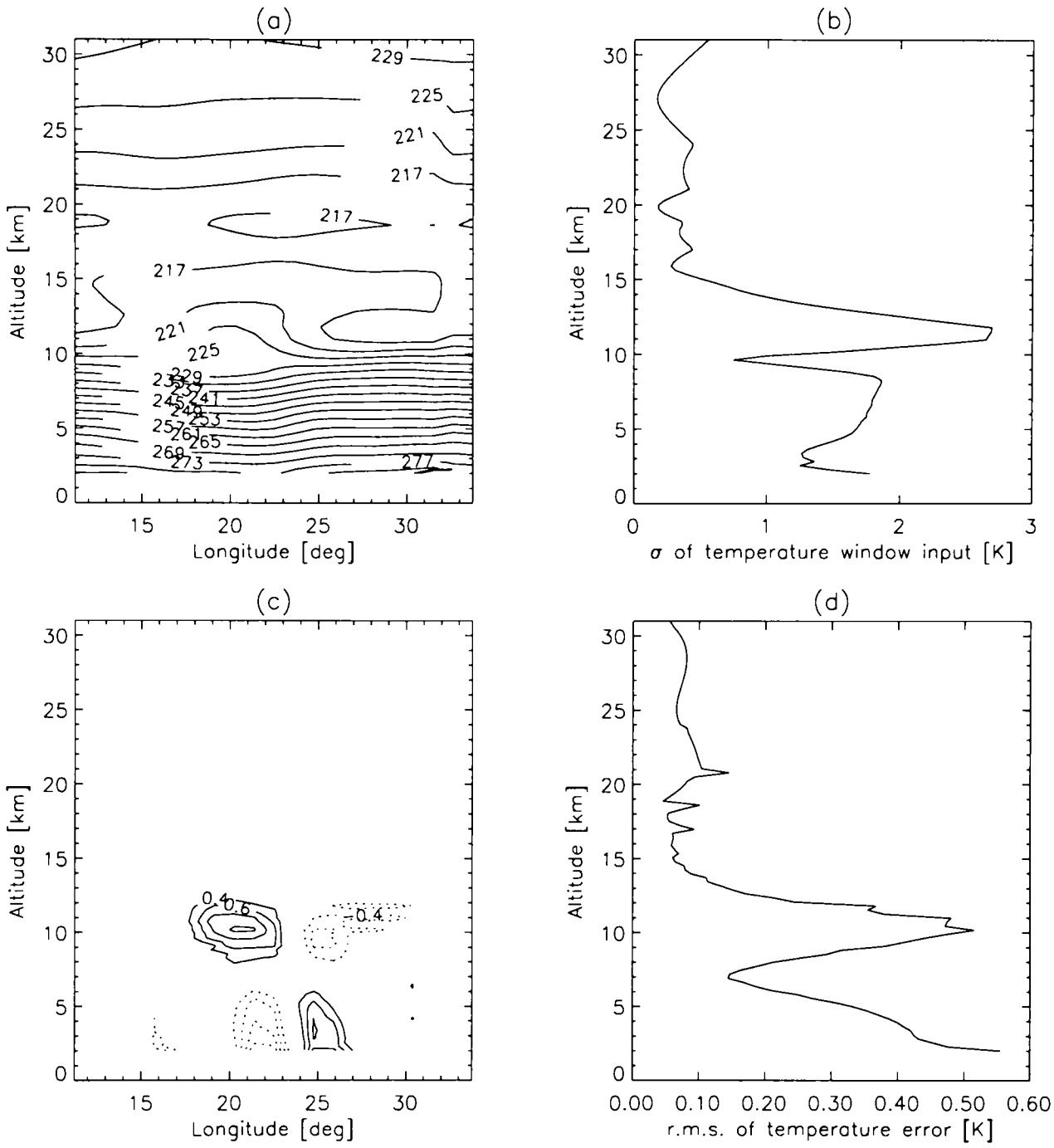


Figure 5.12: The local horizontal region that corresponds to the second largest temperature retrieval errors. Panels are as in figure 5.11.

longitude) and is shown by figure 5.12. Although the horizontal temperature structure is not as pronounced as shown by figure 5.11, the largest errors in this region (0.5—1 kelvin) can be attributed to identifiable horizontal temperature structures.

In general, it is noticed that although the vertical regions of large retrieval error correspond to the vertical regions with large values of horizontal variation, the retrieval errors peak are at a slightly lower altitude.

This can be explained if we consider a vertically stratified temperature field as a function of altitude Z, which has no horizontal variation, except for a finite region S (figure 5.13). Three rays are considered: **a**, **b** and **c**.

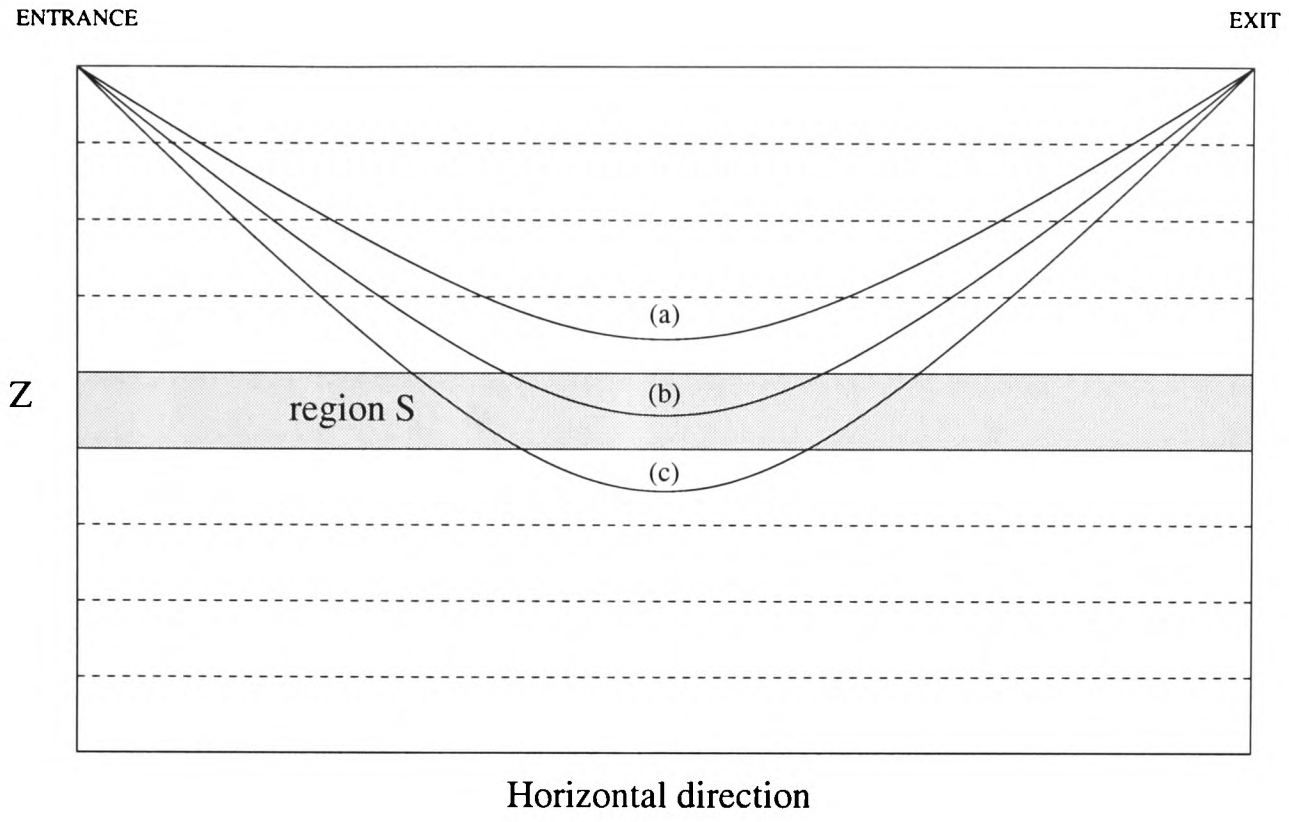


Figure 5.13: A figure to illustrate why the largest retrieval errors from a region of horizontal inhomogeneity are always below the region. The box shown represents the atmosphere sounded by sequential occultation observations; the dotted lines represent vertical tangent layers; the shaded region S represents a tangent layer of horizontal variation; and the curves (a), (b) and (c) represents GNSS radio signals propagating through the atmosphere.

If a ray is sent through the field, using geometry described by a transmitter and receiver orbiting the Earth, its path is described by a curve. If the ray's tangent point is above the top of S (ray **a**) the signal will be unaffected by S; however, if the ray's tangent point is inside S (ray **b**) then the ray will experience the horizontal variation described by this region, from where the ray enters the tangent layer S to where it exits.

As sequential ray paths sound more of S, the horizontal variation experienced increases. The ray path with its tangent point just below S experiences the largest horizontal variation: although ray paths with tangent points below S pass through S twice, the angle of entrance (and exit) to this region becomes more oblique (spending less time in this region), thus experiencing less of the variation in S.

This is an explanation for why the largest retrieval errors incurred from regions of horizontal variation are slightly below these regions.

Null test

The magnitude of errors associated with a temperature field without any horizontal variation in temperature or pressure can be used to assess how much of the error incurred in the experiment is due to computational error.

The field used for this experiment is constructed using one profile from the field shown in figure 5.4; further details of the experiment are identical to the experiment described in §5.3.1.

Errors associated with the null experiment are found to be several orders smaller that the experiment with horizontal variations. It can thus be concluded that errors in figure 5.10 are, indeed, from horizontal variations in temperature and pressure.

5.3.3 Fitting regression coefficients to temperature retrieval errors

The experiment using a realistic atmospheric field provides enough information such that it is possible to comment on the relationship between temperature retrieval errors and the local horizontal temperature structure.

By taking the field and fitting a quadratic curve ($ax^2+bx+c=0$, where x represents an independent variable) to every 3 horizontal nodes for every retrieved height, a matrix of a, b and c coefficients is obtained. Physically, the a and b coefficients represent the second and the first-order components of the temperature field fitted.

The regression equation used to model the temperature retrieval errors is linear in both quadratic and linear coefficients

$$\delta T = \boldsymbol{\alpha} a + \boldsymbol{\beta} b \tag{5.10}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent fitted regression coefficients for the second and first-order curve fitting coefficients. No offset term is included since δT should be zero when a and b are zero, i.e. no horizontal variation.

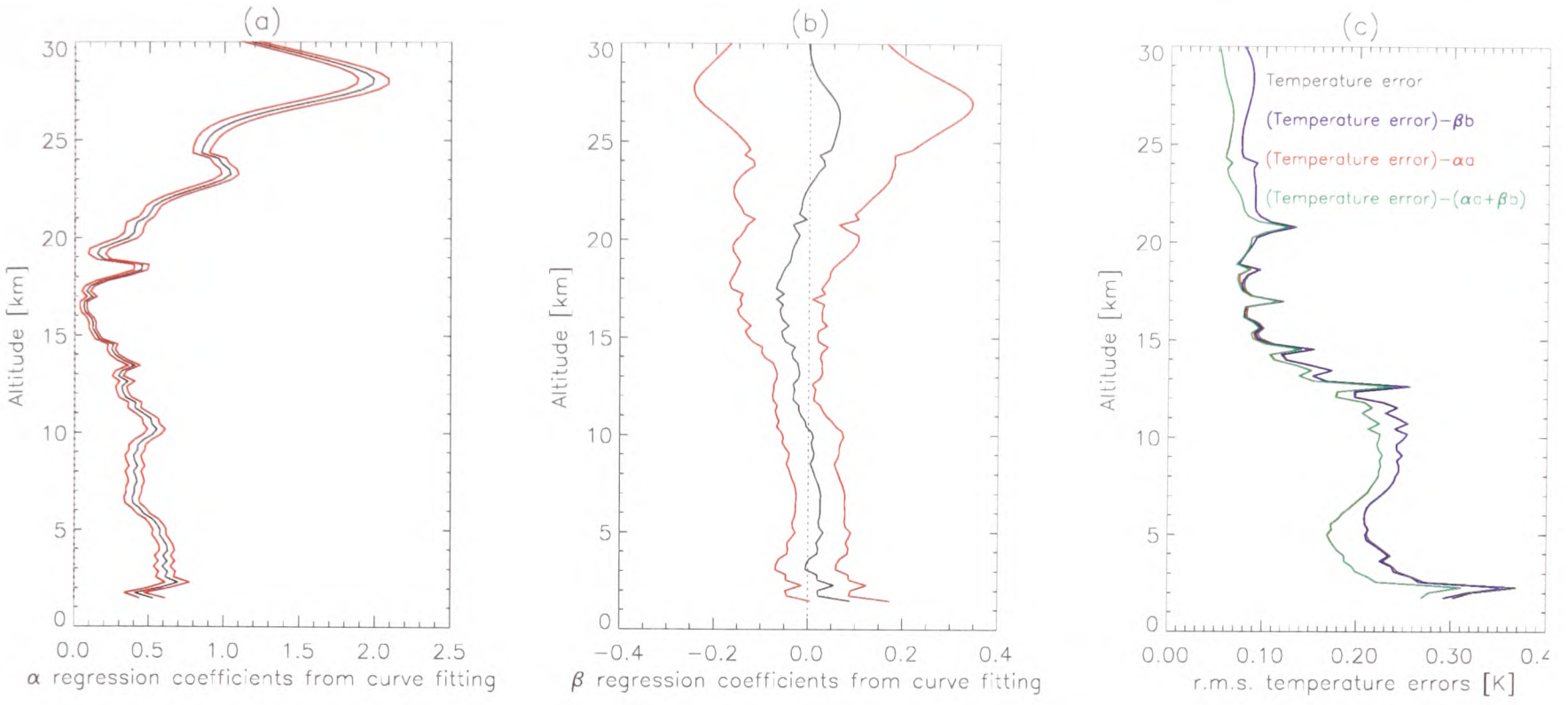


Figure 5.14: The regression coefficients $\boldsymbol{\alpha}$ (panel a) and $\boldsymbol{\beta}$ (panel b) as function of altitude modelled using $\delta T = \boldsymbol{\alpha} a$ and $\delta T = \boldsymbol{\beta} b$. The red lines represent the associated standard deviation values of the coefficients. Panel (c) shows the r.m.s. statistics from comparing the predicted temperature retrieval errors (using the regression coefficients shown in panels (a) and (b)) with the actual temperature retrieval errors.

Because the quadratic and linear curve fitting coefficients are independent variables it is possible to split this regression equation into: $\delta T = \boldsymbol{\alpha} a$ and $\delta T = \boldsymbol{\beta} b$ to obtain $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as a function of altitude.

Figure 5.14 shows that the $\boldsymbol{\alpha}$ regression coefficient is an order of magnitude larger than the $\boldsymbol{\beta}$ coefficient but cannot be compared directly because they have different dimensions. It is also noticed that $\boldsymbol{\alpha}$ remains positive throughout the vertical range considered. The standard deviation for the $\boldsymbol{\beta}$ coefficient is comparable with the coefficient itself implying that this is the most variable coefficient of the two shown. In general, below 15 km, the vertical structure of $\boldsymbol{\alpha}$ reflects the levels of horizontal temperature structure shown by figure 5.10. Above 15 km, where the level of horizontal temperature variation declines, the value of $\boldsymbol{\alpha}$ increases to a maximum, implying that the quadratic coefficient of the curve-fitting is the dominant source of retrieval error.

To investigate the extent to which the realistic temperature field can be explained in terms of first and second-order forms, the predicted temperature retrieval errors are

compared with the actual values. First, using the linear component ($\delta T = \beta b$); second using the second-order component ($\delta T = \alpha a$); and finally using equation 5.10.

Panel (c) shows that the second-order form provides the largest contribution to the predicted temperature error. The r.m.s. values of predicted−actual temperature retrieval errors show that the predicted values are most successful in the regions with large values for horizontal temperature variation. Below the tropopause 15% of the temperature retrieval is predicted using the regression coefficients.

The conclusion from this experiment is that it is possible to explain a small percentage of the total temperature error by considering linear and second-order shapes. Better prediction of retrieval errors may be possible if the temperature structure at altitudes above the ray tangent point is considered, since the magnitude of the refraction depends on levels above the tangent point.

5.3.4 Horizontal smoothing

Because of the horizontal smoothing inherent in the measurements, retrievals will not in general agree perfectly with profiles at single points. However it may be possible to smooth the field horizontally and obtain better agreement. It is also important to investigate this point because if retrieved profiles are assimilated by comparing them with appropriately smoothed NWP model equivalents, there may be less scientific justification for assimilating bending angle data, as opposed to inverting the measurements using the inverse Abel transform or optimal estimation.

The input field for the smoothing experiment is a sub-field of figure 5.4. To assess the impact of horizontal smoothing, the weighted average of the input field, about the temperature window midpoint (horizontal position of tangent points), is compared with the profile of retrieved temperatures. The weighted average is calculated by interpolating the information from the nearest neighbours onto the height grid of the temperature window midpoint and weighted as required. Only nearest neighbours of the window midpoint are considered because the width of the resulting window is of the order of the horizontal resolution of radio occultation observations. The triangular smoothing

schemes tested include 1-2-1, 1-3-1, 1-5-1, 1-10-1 and 1-50-1, where the numbers represent the weights attached to each horizontal node.

Results from various horizontal smoothing schemes

Figure 5.15 shows the results from the different triangular smoothing schemes. Results from the 1-2-1 weighting (panel (b)) show decreased retrieval errors in the upper stratosphere and for most of the troposphere, where the horizontal variation of temperature is large (panel (a)). The largest reduction in retrieval error is approximately 0.1 kelvin which is in the lower tropopause where the r.m.s. of the unsmoothed temperature retrieval is at a maximum. At the 9 km altitude level, the resulting errors from the smoothed fields are larger than the results using the unsmoothed field.

The results from the 1-3-1 smoothing scheme are very similar to the results obtained from the 1-2-1 scheme. The smoothing at the 9 km altitude level has decreased the retrieval error, indicating that this smoothing scheme is less crude than the 1-2-1 scheme.

In general, reducing the weighting attached to the nearest neighbours from 1/4 (1-2-1) to 1/5 (1-3-1) to 1/7 (1-5-1) to 1/12 (1-10-1) to finally 1/52 (1-50-1) reduces the difference between the smoothed field and the unsmoothed field (dotted line). This is expected: as weighting to the window midpoint is increased the overall weighting approaches a delta function and therefore the unsmoothed profile.

An optimum value of weighting for the triangular smoothing scheme is achieved by a trade-off between a sensible weighting value and a delta function. The value of the weighting will reflect the true horizontal function of the measurement. Although the 1-2-1 scheme is most successful in reducing errors for much of the atmospheric range investigated, over-compensation can increase the apparent retrieval error. Comparing the weighting schemes shown, the optimum weighting is found to be 1-5-1 which decreases the retrieval error throughout the atmospheric range investigated.

The distribution of weights, and indeed the form of smoothing, will depend on the horizontal scale of the atmospheric field used. A finer horizontal grid should also be able

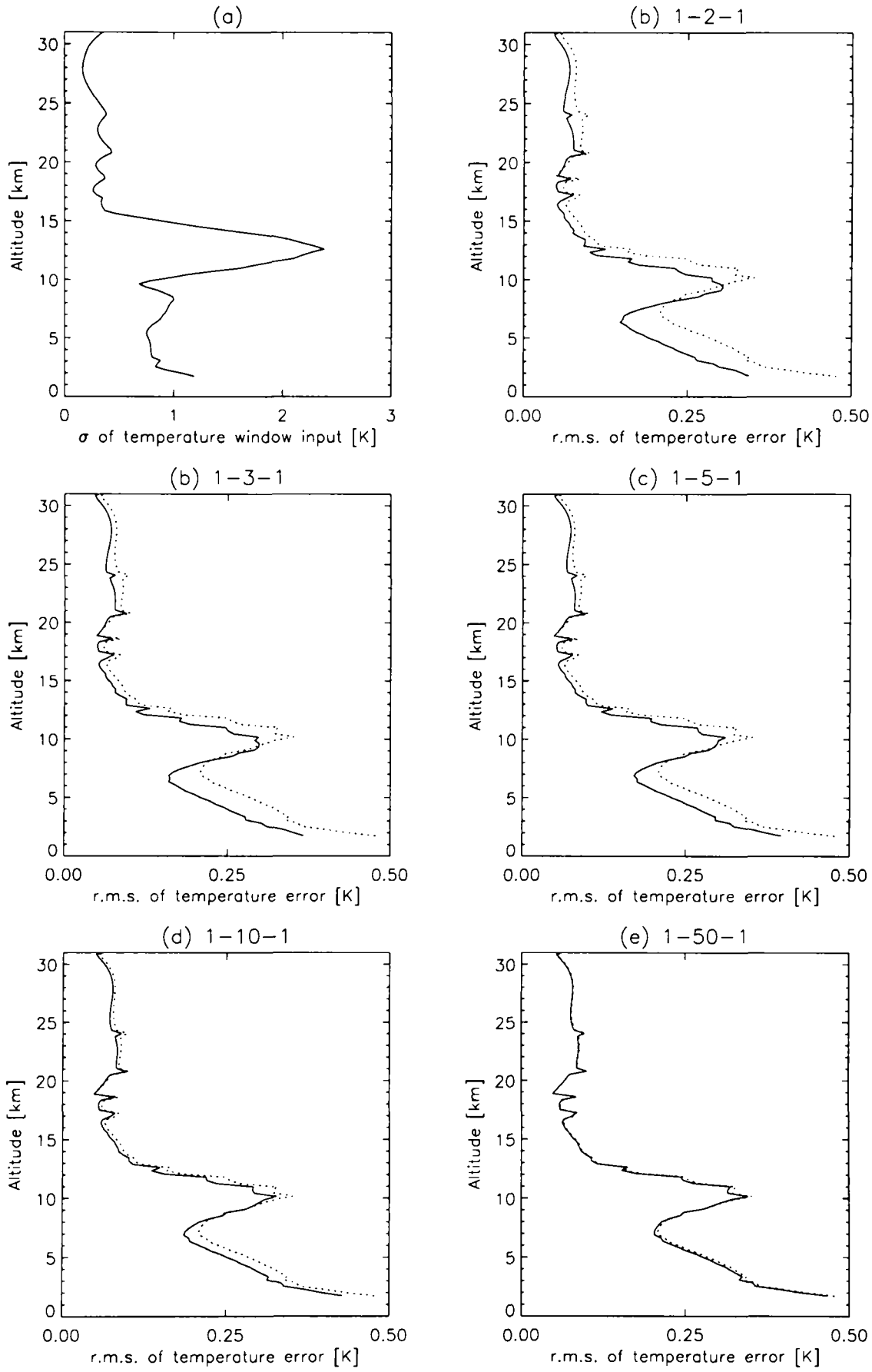


Figure 5.15: Results from smoothing the input temperature field with which to compare with the retrieved temperature profile. By decreasing the weights attached to nearest neighbours of the horizontal field midpoint, the difference between the smoothed and unsmoothed comparison become negligible. Panel (a) shows the standard deviation for the temperature field shown by figure 5.4 and all other panels show the effect of smoothing, with the weights used printed on top of each panel. The dotted lines superimposed on panels (b)−(e) denote the unsmoothed r.m.s. values of temperature retrieval error.

to represent smaller scale atmospheric structure which will also affect the results from the smoothing.

The results from this experiment show that simple linear filtering of the original field does not reduce the error significantly when a horizontally smoothed version of the truth is compared with the retrieved state.

5.3.5 Inclusion of water vapour

Considering water vapour is important because horizontal variations in this quantity may compensate or reinforce horizontal variations in temperature.

In the case when water vapour is considered alongside temperature, the inversion maps the bending angle in to refractivity using the inverse Abel transform. Using refractivity as the retrieved product by-passes the temperature-water vapour ambiguity (chapter 4). The ECMWF water vapour field (shown by figure 5.16, corresponding to the temperature field shown by figure 5.4), is expressed as specific humidity (kg/kg).

Figure 5.17 shows the computed refractivity statistics with and without the humidity field. The standard deviation of the 'wet' and 'dry' refractivity fields shown (panel (a)) are similar except in the lower atmosphere (below 10 km). In this vertical region water vapour becomes more prominent and begins to represent an increasing contribution to refractivity. Larger refractivity retrieval errors for the 'wet' refractivity field imply that the water vapour field used has reinforced the 'dry' refractivity gradients (also shown by figure 5.8). This is also reflected in the refractivity retrieval errors (panel (b)) which also increase in the lower atmosphere relative to retrieval errors from the 'dry' refractivity field.

Similar to the 'dry' experiment, where only horizontal temperature variations are considered, the largest positive and negative refractivity retrieval errors can be traced back to local horizontal structures in refractivity. The area associated with largest positive error is also where the largest negative error is situated (335°−355° longitude) (figure 5.18, panels (a) and (c)). This region is not the same as the largest temperature (and 'dry' refractivity) retrieval errors, implying that the horizontal water vapour

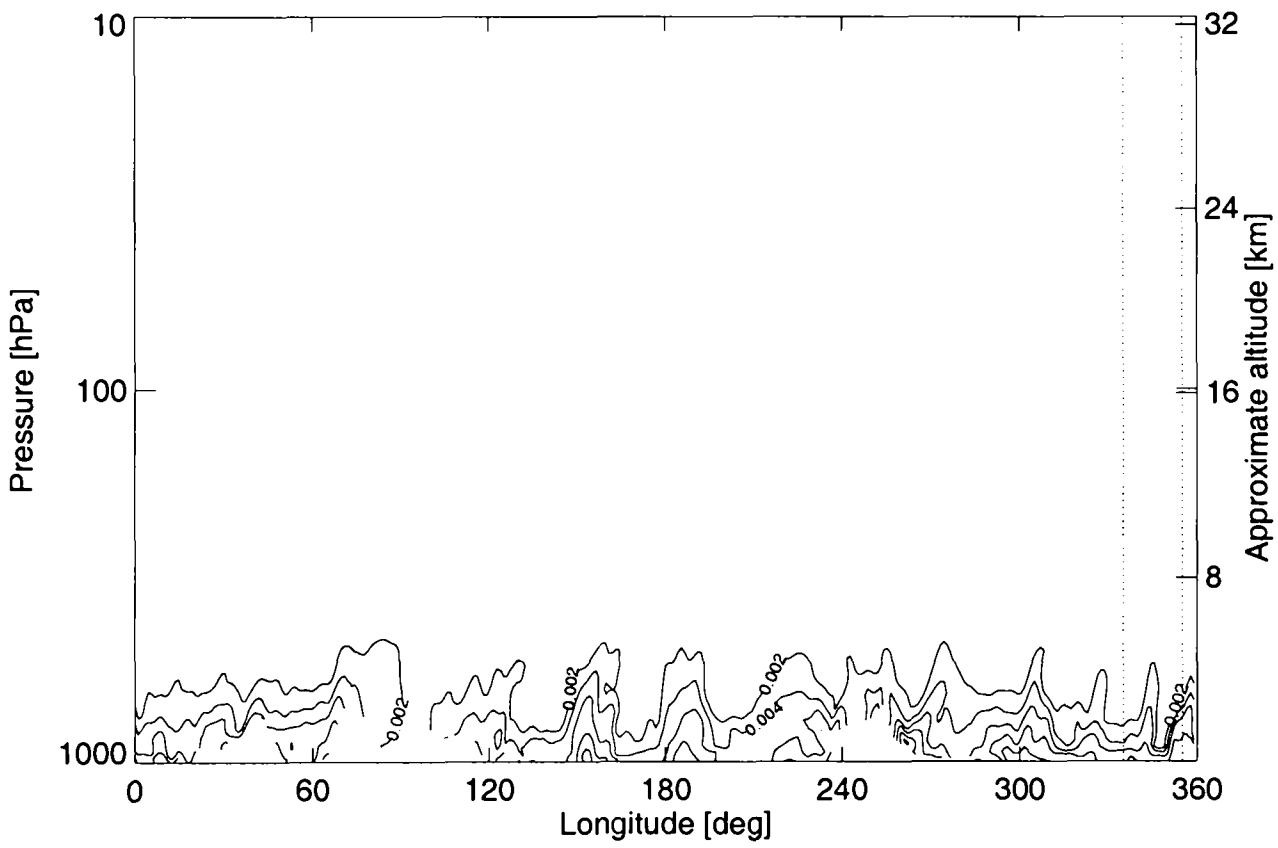


Figure 5.16: Cross-section of specific humidity (kg/kg), as a function of pressure (hPa), through a mid-latitude temperature field as represented by spectrally truncated operational analysis from the ECMWF archives at BADC. The field is for 40°N, over all longitudes on the 15 May, 1992. The dotted lines shown define a box from 335—355°E corresponding to the horizontal region of the largest refractivity retrieval errors. The contour interval is 0.002 (kg/kg).

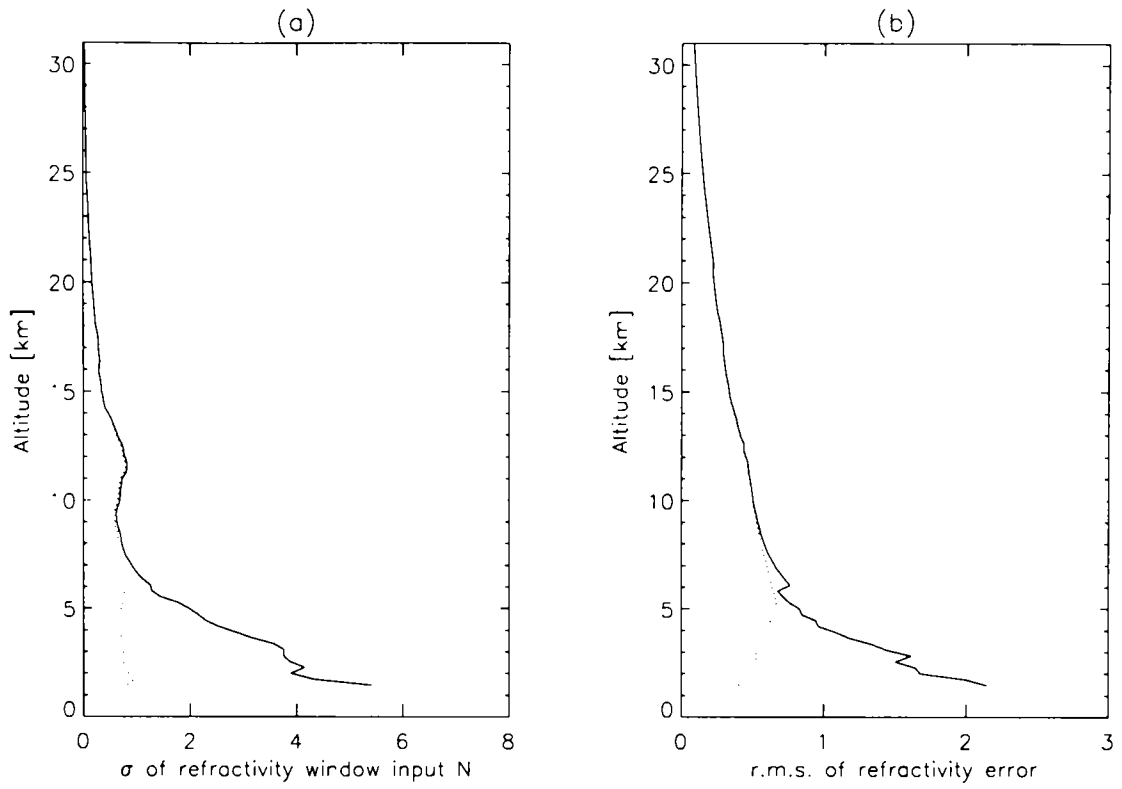


Figure 5.17: Gross statistics of the 'wet' refractivity calculated from figures 5.4 and 5.16. Panel (a) shows the horizontal standard deviation structure of the refractivity field; and panel (b) shows the horizontal r.m.s. structure of the refractivity retrieval errors. The gross statistics of the 'dry' refractivity are shown as broken lines.

structure has compensated for horizontal temperature variations in the region shown by figure 5.11 and reinforced temperature variations in the region shown by figure 5.18.

The contours of constant refractivity shown in panel (a) are proportional to contours of constant density (where water vapour is negligible) therefore there is a fairly uniform structure throughout the field. The contours of refractivity retrieval errors show similar structure to temperature errors caused by horizontal gradients, i.e. they can be attributed to sharp horizontal structure. The main positive error contour (of 9 units) is flanked by smaller positive errors similar to the temperature retrieval errors.

The standard deviation values for the 'wet' and 'dry' fields (panel (b)) are similar above the tropopause (~15 km). Below the tropopause, horizontal water vapour gradients reinforce horizontal temperature gradients, more than doubling the values for the 'dry' case. This disparity between the 'wet' and 'dry' fields can also be seen in the refractivity retrieval errors (panel (d)).

It can be concluded from this section that water vapour horizontal gradients play an important role in the forward and inverse modelling of horizontal refractivity inhomogeneities and should therefore be included in any analysis. Of course, the NWP

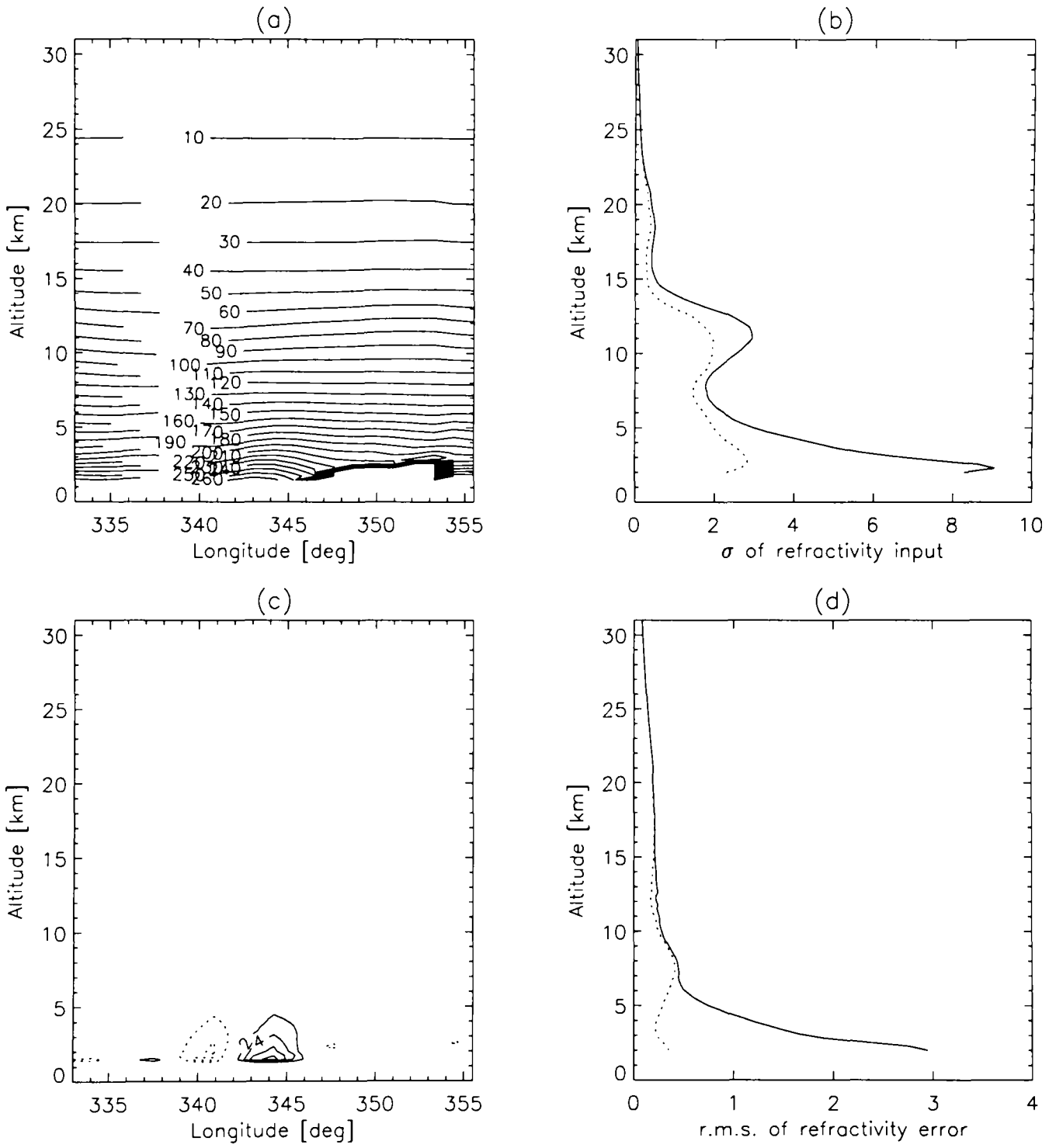


Figure 5.18: A summary of the refractivity field in the region corresponding to the largest positive and negative retrieval errors. Panels (a) and (b) show the refractivity field (with a contour interval of 10 units) and the standard deviation of the field, respectively. Panels (c) and (d) show the temperature retrieval error corresponding to the panel (a) (with a contour interval of 2 units) and the r.m.s. of the temperature retrieval errors. The dotted line superimposed on panels (b) and (d) show the statistics from the 'dry' refractivity field in the same region.

data used in this example is mid-latitudinal and does not represent the possible extremes of water vapour content. In the polar regions, the water vapour concentration is much less and will not contribute a significant amount to refractivity so the analysis of this region will be similar to the 'dry' statistics shown; and in the tropics, water vapour horizontal gradients can cause other problems, e.g. far-field multipath, which is unsuitable for the radio occultation technique.

5.3.6 Fitting regression coefficients to refractivity retrieval errors

Using a similar method to obtain the regression coefficients for temperature retrieval errors, regression coefficients were obtained for refractivity retrieval errors (figure 5.19).

Panel (a) and (b) that the α regression coefficient is dominant throughout the vertical region, peaking at the tropopause region. Below this maximum, the α and β 'wet' regression coefficients become smaller and larger, respectively, than their 'dry' counterpart. Above 15 km, the 'wet' and 'dry' regression coefficients are similar, reflecting the refractivity retrieval error statistics.

Panels (c) and (d) show that the regression coefficients for the second-order form provide the largest contribution to the predicted refractivity retrieval errors, which is consistent with results shown by figure 5.14. Both the 'dry' and 'wet' statistics show that predicted values represent more than 30% of the actual retrieval errors. These results suggest the prediction of refractivity retrieval errors is more accurate at lower altitudes where there are large horizontal refractivity gradients. At high altitudes, the prediction becomes less accurate where the fitted regression coefficients have relatively large associate standard deviations.

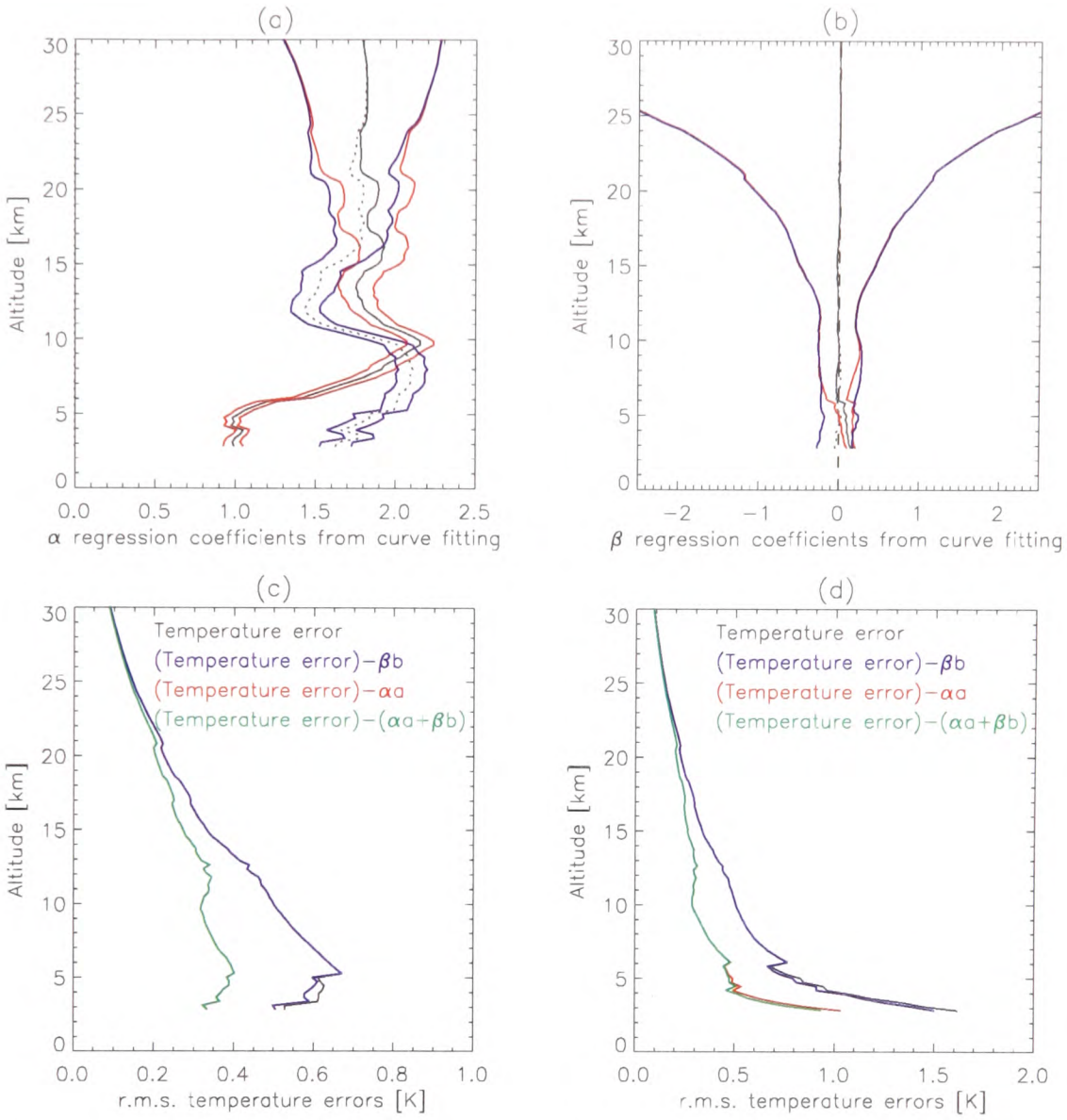


Figure 5.19: The regression coefficients $\boldsymbol{\alpha}$ (panel a) and $\boldsymbol{\beta}$ (panel b) as function of altitude modelled using $\delta N = \boldsymbol{\alpha} a$ and $\delta N = \boldsymbol{\beta} b$. The solid and dotted black lines represent the coefficients for the 'wet' and 'dry' refractivity field; and the red and blue lines represent the standard deviations of the 'wet' and 'dry' regression coefficients. Panels (c) and (d) show the r.m.s. statistics from comparing the predicted 'dry' and 'wet' refractivity retrieval errors (using the regression coefficients shown in panels (a) and (b)). respectively, with the actual refractivity retrieval errors.

5.4 Characterisation of the Abelian inverse method

For the Abelian inverse method to be fully understood, it is essential for the errors involved with the retrieval to be characterised. It can be argued that a retrieval method without an error analysis and characterisation is of little value.

Error sources

The method used to characterise the errors is described in §4.5.2. The retrieval error can be split into its main sources: the *a priori* constraint, measurement noise, the

forward modelling and forward model parameters. The error contributions from the *a priori* constraint and forward modelling are not used in the Abelian inverse method, and therefore are not considered.

The implementation of the forward model for this error characterisation is similar to the general forward model implementation. The difference is that the comparison between the forward model temperature input and the retrieved temperature profile is calculated on the input pressure levels; this is to be consistent with the error characterisation of the optimal method. The data used in the forward model is a hybrid UKMO model profile at (-39.7°S, 282.8°E), the construction of which is described in §6.1.1.

The state vector for the Abelian method is temperature as a function of pressure (40 elements, §6.1.1). The forward model parameters considered in this characterisation include specific humidity (first 15 of 40 temperature levels), surface pressure and the local radius of curvature (none of which are retrieved quantities).

The matrix of averaging kernels (§4.5.1) is calculated by perturbing each element of the state vector by a unit amount and retrieving temperature using the inverse Abel transform. The retrieved values are compared with the retrieved temperatures corresponding to the unperturbed state vector, interpolated linearly (in ln pressure) on to the state vector pressure levels. This provides an indication of the retrieval sensitivity to changes in the state vector. The matrix of contribution functions (§4.4) is calculated by using the unperturbed state vector to calculate the measurement vector. Each element of the measurement vector is perturbed by a constant amount after which temperature is retrieved. The retrieved values are compared with the retrieved temperatures corresponding to the unperturbed state vector, interpolated on to the state vector pressure levels.

The measurement noise estimates are from Luntama (1997) (described in §6.1.2) and from the horizontal inhomogeneity results shown by figure 5.8; and forward model parameter errors are discussed in §6.1.2. The error value for the local radius of curvature error (1 metre) has been chosen after much consideration and discussion [Dr. J. J. Barnett: private communication].

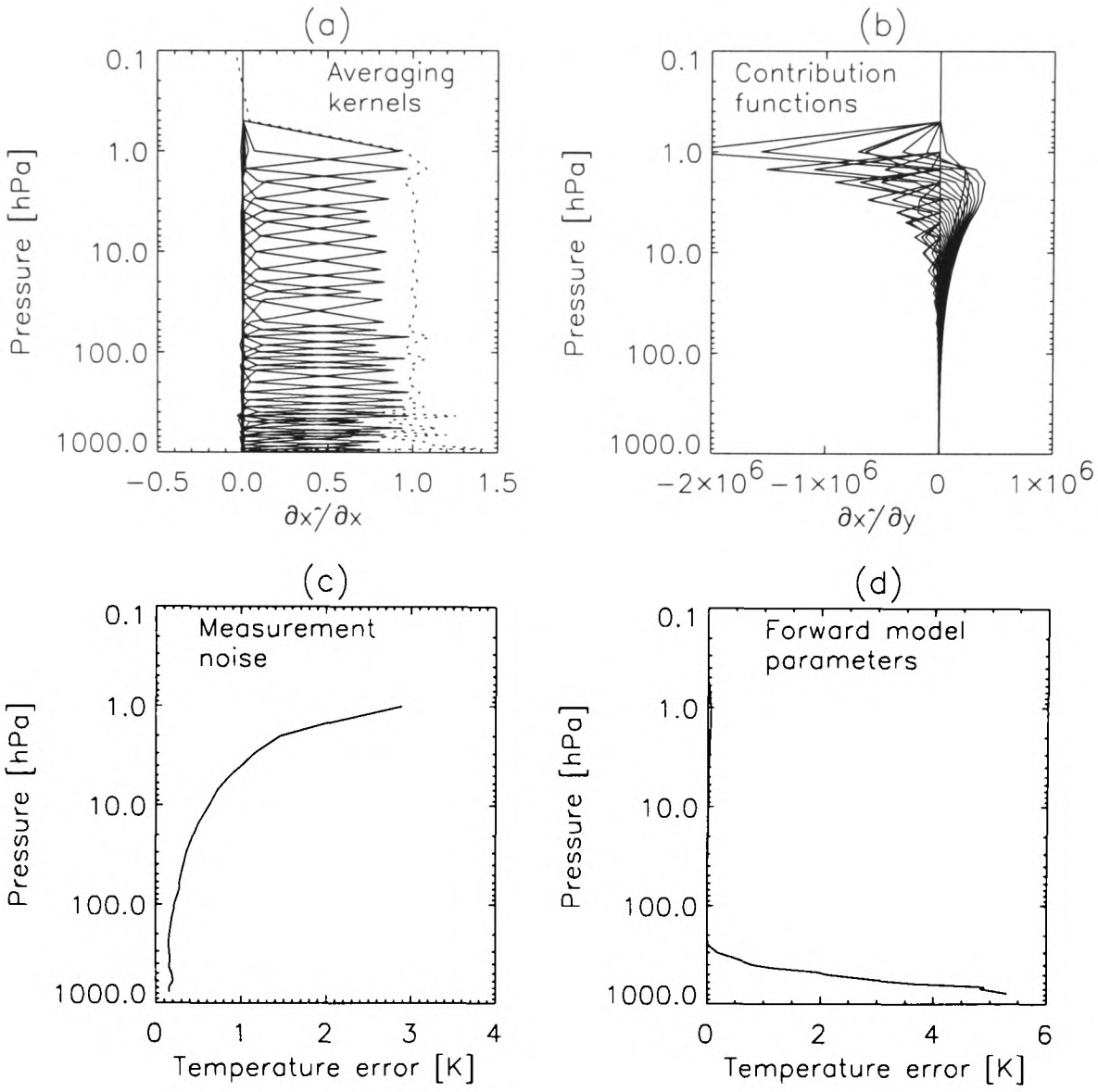


Figure 5.20: Error and solution characterisation of the Abelian inverse method. Panels (a) and (b) show the averaging kernels (the broken line represents the area beneath each kernel) and the contribution functions, respectively; panels (c) and (d) show the contributory retrieval errors from the measurement noise and forward model parameters, respectively. The errors shown represent square-root values of the principal diagonals of the covariance matrices which are expressed in kelvin.

The measurement vector was chosen because it is of typical length which does not fully describe the upper atmosphere state vector pressure levels. The upper end of the measurement vector is ~1.0hPa which is where a reduction in retrieval sensitivity is seen in all plots. It should be noted that the overall shape of the curves is a direct consequence of the state vector used.

The averaging kernels represent the retrieval sensitivity to state vector perturbations. The product of the true atmosphere state vector and the matrix of averaging kernels represents the retrieved state (in addition to errors). The averaging kernels for the Abelian retrieval are essentially delta functions, which is expected for an analytic method.

The kernel spacing varies proportionally to the thickness of the state vector pressure levels. The dotted line enveloping the kernels in figure 5.20(a) represents the area underneath each kernel as a function of pressure. This curve effectively represents the retrieval response to a unit shift in the entire state vector, and gives an indication of the retrieval performance. The area underneath each kernel is approximately 1, implying that there is a good correspondence between the *a priori* state vector and the retrieval state vector.

The contribution functions (panel (b)) represent the retrieval response to perturbations in the measurement vector. These functions can be used to calculate the contributory retrieval error covariance matrices. The magnitude of these functions increase with height, with the dynamic range of the function spanning several orders of magnitude. The Abelian inverse method relies upon measurements above the tangent point and the hydrostatic relation is integrated downwards, therefore an error in any one bending angle measurement will affect all measurements below the perturbation, decreasing exponentially. In general, the pressure level of the perturbation in measurement space is reflected directly in state vector space.

The retrieval sensitivity due to measurement noise (panel (c)) reflects the realistic values used for this error source (§6.1.2). This source of error increases in the upper atmosphere because of a lower limit used in the construction of the measurement constraint, originating from noisy upper stratospheric measurements.

The retrieval sensitivity to forward model parameters (panel (d)) is dominated by the ln(specific humidity) elements, which can be seen below 300 hPa (the upper altitude limit of the state vector humidity representation). The temperature error increases with depth and with the magnitude of the humidity value, and represents the largest contribution to the retrieval error in the lower atmosphere.

5.5 Chapter summary

The work presented in this chapter represents a relatively simple analysis of the effects of horizontal inhomogeneities on radio occultation data. The simplicity of the 2-d analysis[7] leads to errors that can be readily understood and quantified.

Errors incurred from the numerical implementation of the Abelian inverse method are shown to be small and unbiased. It is also shown that the numerical representation of the hydrostatic relation used is sensitive to first-guess errors, but incurred errors decrease exponentially with depth.

Using idealistic line-of-sight horizontal temperature and pressure variations, temperature retrieval errors are quantified. The largest temperature retrieval errors are shown to correspond to horizontal variations of temperature which are symmetric about the tangent point, and can be explained by considering the excess contributions to the Abel transform either side of the tangent point. Realistic pressure variations are found to have a relatively small effect on the Abelian inverse method.

Applying the ideas formed from considering simple horizontal structures, a realistic mid-latitude 2-d field of temperature, pressure and humidity is used to compute several estimates involved with horizontal inhomogeneities. First, the difference between the measurement vector assuming no horizontal structure and with horizontal structure over the complete field is computed. This provides an estimate of typical forward modelling errors caused by assuming spherical symmetry. The differences are shown to be proportional to horizontal refractivity variations, as expected. In particular, it is shown that the largest errors are below the tropopause which are attributable to horizontal water vapour variations. Second, temperature retrieval errors incurred from assuming spherical symmetry are investigated. The largest temperature retrieval errors are found to be due to sharp horizontal variations in the temperature field; the size and magnitude of which agree reasonably well with those from considering simple horizontal inhomogeneities. Third, refractivity retrieval errors incurred from assuming spherical symmetry are investigated. Retrieval errors from the 'wet' and 'dry' refractivity fields

[7]The missing third dimension should not be significant since much of the theory used by the Abelian inversion assumes that each ray path taken can be described by a 2-d plane.

are compared. Gross statistics show that horizontal humidity variations can compensate for or reinforce horizontal temperature gradients, agreeing with the forward modelling error analysis.

Regression coefficients are fitted to the temperature and refractivity retrieval errors by considering first and second-order curve fitting coefficients of the local horizontal structure. For both the retrieved temperature and refractivity errors it is shown that the second-order curve provides the largest contribution to the prediction of retrieval error. In particular, the most successful predicted values are found to be in vertical regions where horizontal variation is largest, e.g. the tropopause. However, predicted values for retrieval error only account for a small percentage of the total error. Better predictions may be possible if the atmospheric structure above the tangent point is also considered.

The error and solution characterisation of the Abelian inverse method is shown. Averaging kernels and contribution functions have very 'clean' shapes, reflecting the analytic nature of the method and the absence of realistic statistics. Theoretical error estimates for the contribution from measurement noise to the total retrieval error suggest that this error increases with altitude, reflecting the realistic statistics used; the retrieval error contribution from forward model parameters shows that the largest values are in the lower atmosphere, reflecting the lack of use of water vapour by the method when retrieving temperature.

Chapter 6

Optimal estimation: implementation and simulation

This chapter presents a detailed description of the information used to implement the optimal estimate inverse method; and presents and discusses the results from simulated retrievals, which contain a considerable amount of information concerning the nature of the computed optimal estimate.

6.1 Implementation

This section describes the *a priori* vector $\mathbf{x}_a$ and its error covariance matrix $\mathbf{S}_a$, and the measurement vector $\mathbf{y}$ and its error covariance matrix $\mathbf{S}_y$.

6.1.1 The a priori vector and its error covariance matrix

The *a priori* profile $\mathbf{x}_a$ is essentially obtained from the UKMO unified model (Cullen 1993). The model from which the data is derived has 19 levels, and is expressed on hybrid-sigma pressure coordinates (surface−10 hPa). The global model has a resolution of 0.833° (180°/217) latitude and 1.25° (360°/288) longitude. The data used are 6-hour forecasts which have been interpolated to occultation event positions, using the mean latitude and longitude of each occultation[1].

These 19 levels are linearly interpolated (in ln-pressure) onto the state vector levels used for TOVS retrievals (Eyre 1990), assuming Cospar International Reference Atmosphere climatology (CIRA-86 1990) above the UKMO model accounting for the

[1]The mean latitude and longitude of an occultation typically corresponds to altitudes in the lower stratosphere/upper troposphere.

latitudinal and the seasonal variation of the profile. This climatology should provide reasonable *a priori* and first guess information in the upper stratosphere, and is thought to be more accurate than assuming temperatures at key points in the atmosphere.

The *a priori* error covariance matrix $\mathbf{S}_a$ was obtained from the UKMO (used for TOVS retrievals), where lower atmospheric values (surface–50 hPa) were generated from radiosonde–forecast difference statistics; upper stratosphere values were found by regression from the levels provided (Eyre 1989).

The full UKMO error covariance matrix is a 67-element square matrix. The radio occultation retrieval uses the 40 temperature elements, 15 ln(specific humidity) elements[2] and the surface pressure element[3]. The fixed pressure levels used are [hPa]: 0.1, 0.2, 0.5, 1., 1.5., 2., 3., 4., 5., 7., 10., 15., 20., 25., 30., 50., 60., 70., 85., 100., 115., 135., 150., 200., 250., 300., 350., 400., 430., 475., 500., 570., 620., 670., 700., 780., 850., 920., 950., 1000.

The temperature and ln(specific humidity) inter-quantity covariance values have been set to zero. The reasoning behind this is that these inter-quantity covariances are not known to any great extent and assuming zero covariance between them is more conservative than an erroneous covariance. The surface pressure element is independent of both temperature and ln(specific humidity).

Since CIRA climatology is used to form the first-guess it is prudent to consider the errors that may be attached to such information. In general, if the r.m.s. values from the diagonal elements of the UKMO *a priori* error covariance matrix are smaller than the latitudinal-dependent (θ) climatological errors then the climatological errors are assumed at the levels in the upper atmosphere described by the CIRA climatology: for $|\theta| \geq 45°$ σ=15 K (winter) and σ=5 K (summer); and for $|\theta| \leq 45°$ σ=5 K [Dr. J. J. Barnett: private communication].

The r.m.s. values for the diagonal elements of the modified UKMO *a priori* error covariance matrix are shown by figure 6.1.

---

[2]Specific humidity is expressed as the natural logarithm of specific humidity since forecast errors in this quantity are more constant than in specific humidity.

[3]Since radio occultation observations often reach altitudes near to the surface, it is thought they could provide surface pressure information.

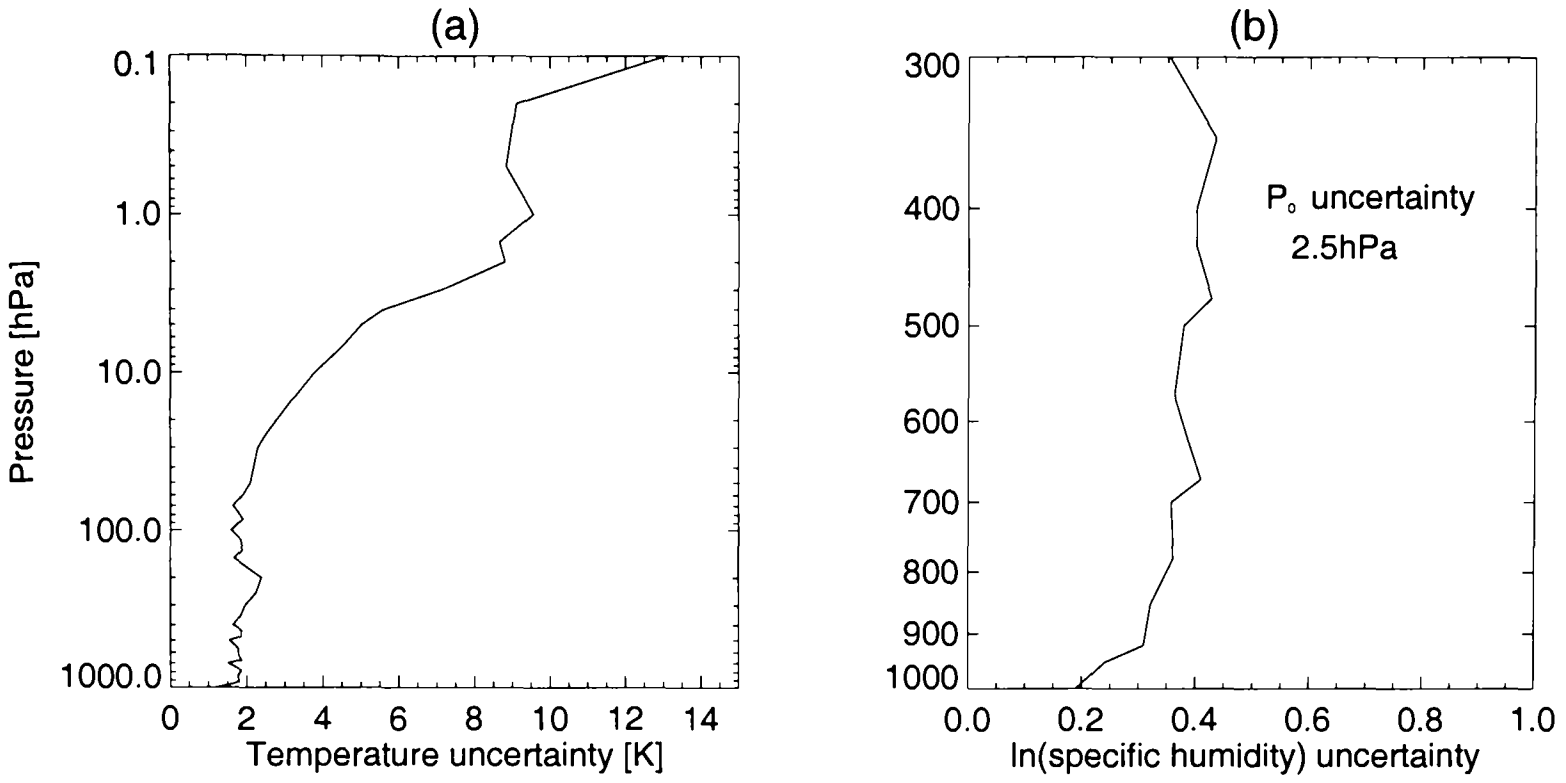


Figure 6.1: The square-root values of the principal diagonal of the UKMO *a priori* error covariance matrix (CIRA climatological errors not shown). Panels (a) and (b) show the r.m.s. errors for temperature (kelvin) and ln(specific humidity) (expressed as a fractional uncertainty), respectively, with the r.m.s error for the surface pressure element inset of panel (b).

As expected, for temperature, the lower atmosphere is reasonably well known (of the order of 1.5 kelvin) which worsens as a function of altitude. The vertical structure of this error estimate is mainly due to certain regions of the atmosphere being less dynamically active than other regions and the quality of existing observations in these regions. The actual values have been developed over many years at the UKMO and reflect the level of confidence in the UKMO unified model [Dr. B. Barwell, UKMO: private communication].

At this point, it should be stressed that these uncertainties reflect the *a priori* information obtained from the UKMO. If prior information is obtained from other sources these error estimates should be altered accordingly to reflect the information. It should also be noted that these values represent global uncertainties, i.e. latitudinal and hemispherical variations are omitted.

One method of visualising a covariance matrix is to consider its correlation matrix, which is simply the normalised covariance matrix. The correlation matrix is thus a useful tool to displaying how vertical elements are related to one another. The correlation

matrix for the modified UKMO *a priori* matrix used in the retrievals is shown in figure 6.2.

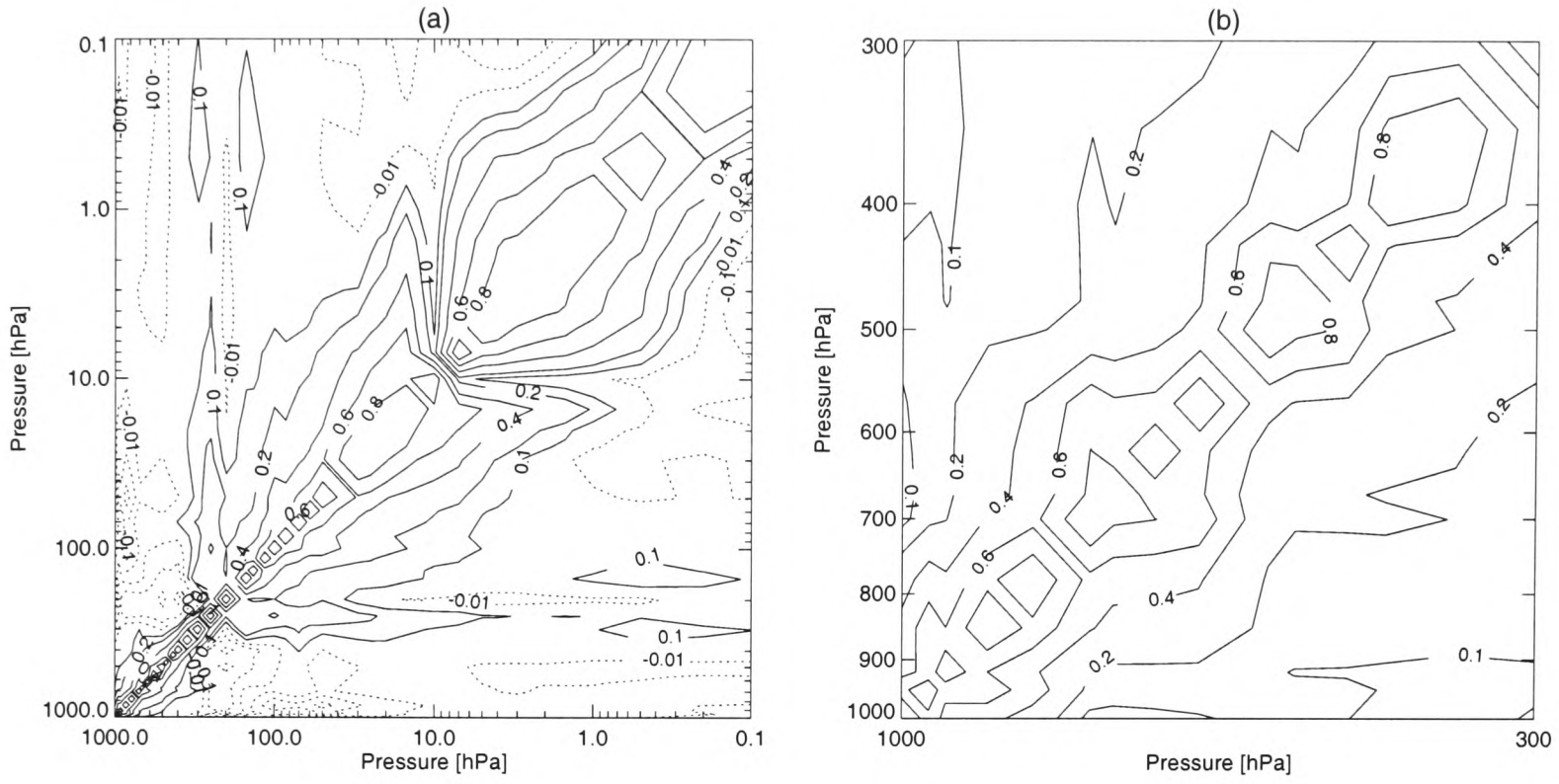


Figure 6.2: The correlation matrix calculated from the modified UKMO matrix (CIRA climatological errors not included). Panels (a) and (b) show the temperature and ln(specific humidity) inter-level correlations, respectively. The surface pressure element is uncorrelated and is not shown.

In general (for both temperature and ln(specific humidity)) it can be seen that there are strong positive vertical correlations surrounding the principal diagonal; elements further away from this diagonal, which represent the correlation between one level and the layers above/below, decrease with a correlation length. In particular, there are strong, local temperature correlations between the upper stratospheric levels. There is also a negative temperature correlation between the troposphere and stratosphere, which was found statistically by McNally (1987).

The surface pressure *a priori* error estimate is uncorrelated with both temperature and ln(specific humidity) and is not shown.

6.1.2 The measurement vector and its error covariance matrix

The measurement vector $\mathbf{y}$ contains bending angle measurements as a function of impact parameter (§2.4).

The receiver observable, the atmospheric phase delay, is low-pass filtered by the JPL to reduce noise. The cut-off frequency of the filter is tuned to pass phase variations corresponding to vertical scales of 2 to 3 km in the stratosphere and approximately 200 m in the lower troposphere (Rocken et al. 1997). From the phase observable, Doppler shifts (and subsequently bending angle profiles) for the two GPS signals are computed. Ionospheric effects are largely removed from the two bending angle profiles to form a single corrected profile (§4.2). A typical number of neutral bending angle measurements after filtering in the GPS/MET dataset used (§7.1) range from 100–200 elements, which span a vertical range of approximately 0.5–65 km.

The simulated measurements, $\mathbf{y}_n$ corresponding to the state vector $\mathbf{x}$, are computed using the forward model described in chapter 5.

In addition to the measurement error covariance matrix consisting of measurement noise estimates from the measurement system, errors from the forward modelling and forward model parameters are considered.

Measurement noise

Measurement errors are created by the hardware (and possibly the software) of the measurement system. The measurement noise estimates are taken from the work described in Luntama (1997). They include thermal noise, residual errors from the ionospheric correction (§4.2), multipath, orbit determination accuracy, and clock instabilities of low-Earth-orbit receiver and GPS satellites and ground stations (§2.4.4).

These measurement error estimates are based on phase noise levels during the measurement or estimated from other noise sources during a radio occultation event. The effect from satellite clock errors and the SA military encryption were found to be negligible by assuming the double differencing measurement technique for the occultation (§2.2.3).

The only exceptions are the error estimates caused by horizontal refractivity structure of the atmosphere and the residual error from the ionospheric correction. The former has been computed in chapter 5 and the latter is obtained from refractivity error estimates

published in Kursinski et al. (1997) and mapped to bending angle space using a forward model (Luntama 1997).

Thermal noise error estimates have been calculated using values released by ESA for the GRAS receiver (Silvestrin 1997). The ionospheric correction and the removal of clock errors by using the double differencing measurement technique will increase the thermal phase noise by factors of 3 and 2 respectively due to the summation of independent random noise variances (as opposed to dependent random noise variances), which will increase the total bending angle error.

Clock instability error estimates have been obtained from Kursinski et al. (1997). The phase noise level from clock error stays constant during an occultation. The double differencing technique can be used to reduce clock errors to an insignificant level (Hardy et al. 1993).

Multipath propagation exists when the transmitted signal is reflected from a surface near the signal propagation path. This effect distorts the transmitted signal by adding components which are attenuated in amplitude and delayed in time compared to the direct signal. The radio occultation technique is based upon phase measurement and is therefore very sensitive to multipath distortion. Because the GPS is designed for surface-based positioning applications, the beam of the transmitter antenna illuminates the whole globe. This means that during an occultation measurement the transmitted signal will always be reflecting from the Earth's surface (Luntama 1997). The magnitude of the far-field multipath propagation is proportional to the distance between the ray path and the surface and can therefore be used to define the lower limit of the radio occultation altitude range (Luntama 1997). The occultation geometry is favourable for most of the measurement: only at the end of a setting occultation (and at the beginning of a rising occultation) can multipath path effects from the Earth's surface cause significant phase errors. This multipath effect is difficult to estimate as it depends on surface reflectivity, antenna gain and atmospheric attenuation (Luntama 1997). Far-field multipath effects can also occur with sharp horizontal refractivity gradients which can can lead to multiple propagation paths. This phenomenon can occur near sharp temperature and/or water vapour gradients, usually found in the troposphere. These are transient errors and there

is no method of predicting or preventing them. Local multipath propagation occurs if any structure of the spacecraft is within the receiver antenna beam. For a realistic spacecraft any part of the structure is so near to the direct signal propagation path that the time delay of the multipath signal is too short for the receiver to filter out. This means that the received signal is a sum of the direct and multipath signals (Kursinski et al. 1997).

The derivation of the bending angle measurements as a function of impact parameter is based upon the numerical solution of the particular occultation geometry. It follows that any errors associated with GNSS satellite position data will cause an error in the derived bending angle. The Precise Orbit Determination technique used to compute the satellite position data is very accurate: the range error of a satellite in a 700 km circular orbit is expected to be less than a metre (Parkinson and Spilker 1996). The bending angle (and impact parameter) error caused by a range error of the order of one metre is insignificant (Luntama 1997). The GPS and low-Earth-orbit receiver satellite velocity errors have both been shown to contribute insignificant bending angle errors (Luntama 1997).

Error estimates from the ionospheric correction and from horizontal refractivity gradients are also included in the data used. Strictly speaking, these error sources are not from the measurement system but are direct consequences of using radio occultation to measure the Earth's atmosphere and therefore should be accounted for.

The ionospheric correction presented in §4.2 removes the dominant effect of the ionosphere on the measured phase residual. However, the correction is not complete because it assumes spherical symmetry and that the ionosphere has regular structure. The effect of these assumptions leaves a second-order residual on the signal in the neutral atmosphere after the correction. Simulations using three-dimensional ray-tracing models have shown this source of error heavily depends on the state of the ionosphere, e.g. solar activity (Kursinski et al. 1997). Although numerical representations are not able to model different atmospheric regions perfectly, it can be concluded that the correction scheme is not good enough in daytime conditions during a solar maximum (Kursinski et al. 1997). In these conditions, the ionospheric residual error is expected to limit

the radio occultation measurement in the upper stratosphere. Higher-order ionospheric corrections are possible but require knowledge of the electron density distribution in the vicinity of the ray propagation path. Unfortunately, the radio occultation technique is not able to provide this information therefore independent observations of the ionosphere are necessary (e.g. electron density). Also, existing ionospheric models do not accurately reproduce small-scale structure that is responsible for the residual error.

The measurement error estimates[4] described and used in the optimal estimation technique are shown in figure 6.3. These estimates represent normal atmospheric conditions with a relatively small multipath error (3 mm) and normal ionospheric conditions (night-time, solar maximum). Panel (a) shows the dominant error source in the lower atmosphere is clearly the errors from the assumed horizontal refractivity structure. Only at altitudes above 40 km do other sources of error start to dominate the total bending angle error curve; above this level, percentage contributions from multipath and thermal noise to the total bending angle errors increase.

These noise estimates are fully independent, i.e. their inter-level (and -quantity) covariance is zero. However, the bending angle measurements do contain a small, local correlation between successive levels due to filtering of the phase measurements. Because this correlation is small, the diagonal form used is thought to be a good approximation to the full matrix [Juha-Pekka Luntama: private communication].

A source of error that is not addressed in this work is the geometric dilution of precision (§2.2.1). Occultations are seldom diametric because the transmitter and receiver orbits are generally not co-planar, causing the ray path tangent point to drift horizontally normal to the line of sight as it descends through the atmosphere. If the motion of the occultation contains a large horizontal component, the accuracy of the 'profile', if the sequence of measurements is indeed taken to represent a vertical atmospheric profile, will be degraded. It should be noted that individual measurements in such a profile are not inaccurate (unless there is significant horizontal motion during each atmospheric phase delay observation). Horizontal motion during an occultation

[4]These estimates have been computed using a nominal bending angle profile defined using an exponential curve with a scale height of approximately 7 km [Juha-Pekka Luntama: private communication].

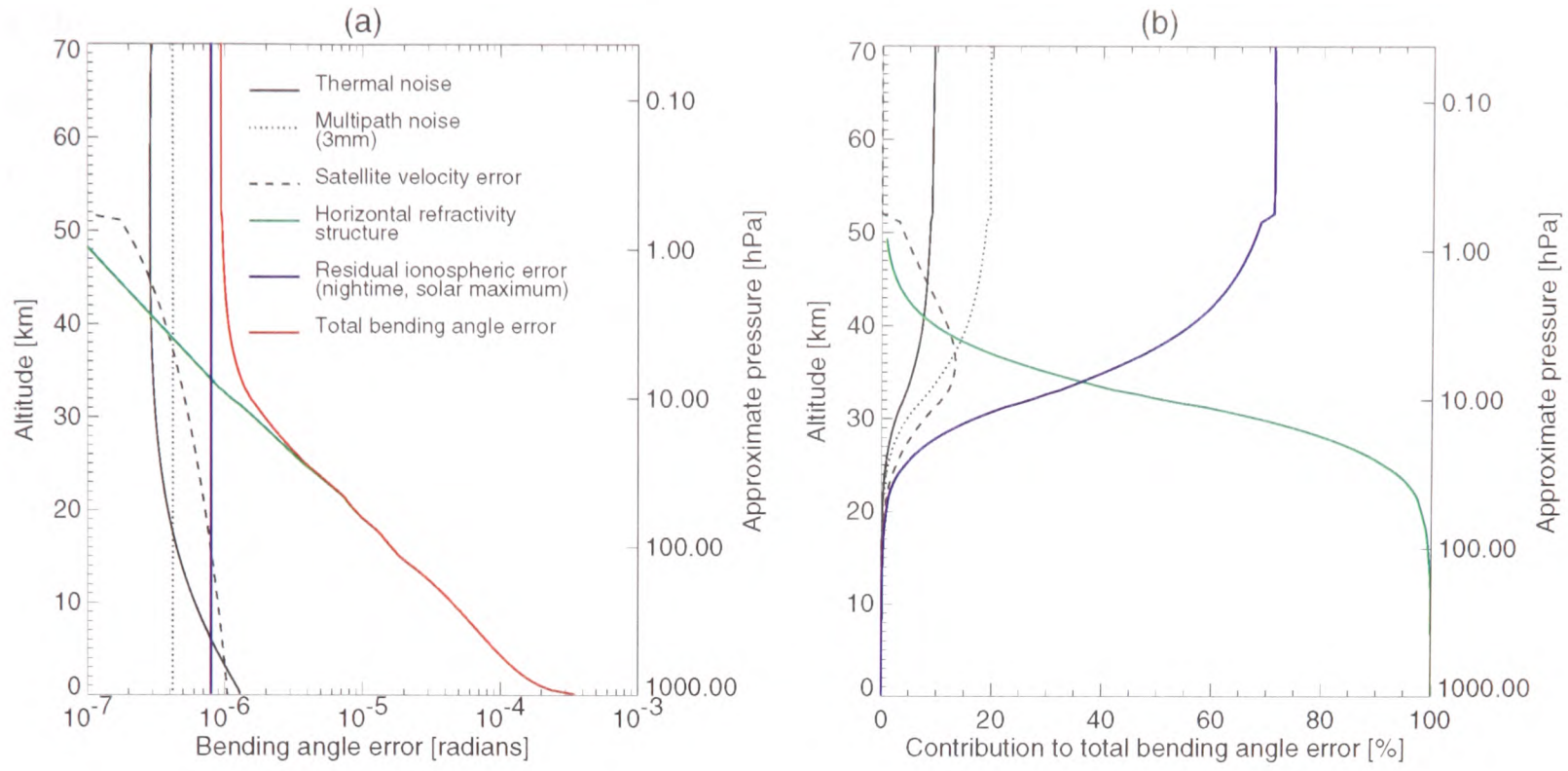


Figure 6.3: Measurement noise budget for the RO inverse method. The error estimates shown in panel (a) represent normal atmospheric conditions, i.e. small multipath error (3 mm) and normal ionospheric conditions (night-time, solar maximum), and have been computed using a nominal bending angle profile using an exponential curve with a scale height of approximately 7km. The components of the total bending angle error are thermal noise (solid black line), multipath error (dotted black line), satellite velocity errors (broken black line), horizontal refractivity structure (green line), residual ionospheric error (blue line) and total bending angle error (red line). Panel (b) shows the percentage contributions from the different error sources to the total bending angle error.

can result in one observation representing a weighted average of the atmosphere some distance away from another observation, rendering the 'profile' a weighted average of the atmosphere over perhaps thousands of kilometres. For occultations where the direction to the transmitter is within $\pm 45°$ of the receiver velocity vector, the ratio of horizontal to vertical motion of the ray path tangent point relative to the limb of the Earth ranges from 0 to 3 at the top of the atmosphere (Kursinski et al. 1997). As the ray path descends through the lower stratosphere, the vertical component of tangent point velocity slows as bending increases while the horizontal component of motion remain basically unchanged. Near the surface, the vertical descent velocity has decreased by an order of magnitude or more relative to the high altitude descent velocity (Kursinski et al. 1997).

The real observations from the GPS/MET dataset used (§7.1) have been found to be noisier than theoretical estimates (Luntama 1997). Random bending angle fluctuations

in the upper stratosphere (of the order of 10^{-5} radians) are present and an order of magnitude than can be explained to date [Juha-Pekka Luntama: private communication]. As such, a suitable error is attached to reflect the upper atmosphere measurements.

Forward modelling errors

This source of error can be further split into the numerical ability to evaluate the modelled physical system, and the physical assumptions used to model the system.

Accuracy of numerical modelling may be parameterised by the number of discrete levels used to evaluate numerical routines (particularly the Abel integral transform in the forward problem). It should be possible to minimise the error arising from numerical modelling by reducing the distance between existing levels by simple linear interpolation. A temperature profile which has 25 levels, equally spaced in ln(pressure) describing a vertical range from 1000−0.1 hPa, is used to compute a profile of refractivity as a function of geometric height, from which the forward Abel integral transform is used to compute a measurement vector. The importance of state vector interpolation is shown by considering four different cases: no interpolation (corresponding to an individual vertical level thickness of approximately 2068 metres), doubling the number of state vector levels in the same range (level thickness of approximately 1034 metres) and increasing the number of state vector levels by a factor of 10 and 50 (level thicknesses of approximately 207 and 41 metres, respectively), and comparing with an interpolated state vector with 100 inter-level levels (corresponding to a level thickness approximately 21 metres).

Figure 6.4 shows that the omission of interpolation results in differences of typically greater than 1% below 30 km increasing in the upper stratosphere to a value of approximately 6%. In general, as the number of interpolated levels increase the magnitudes of the r.m.s. differences decrease to insignificant levels; a level thickness of 41 metres leads to measurement vector differences of the order of 0.0001 %.

This source of error is especially important since NWP fields have a relatively small number of vertical levels with which to model the physical system, thus efforts should be made to minimise this source of error.

An associated error arises from representing an intrinsically high resolution problem

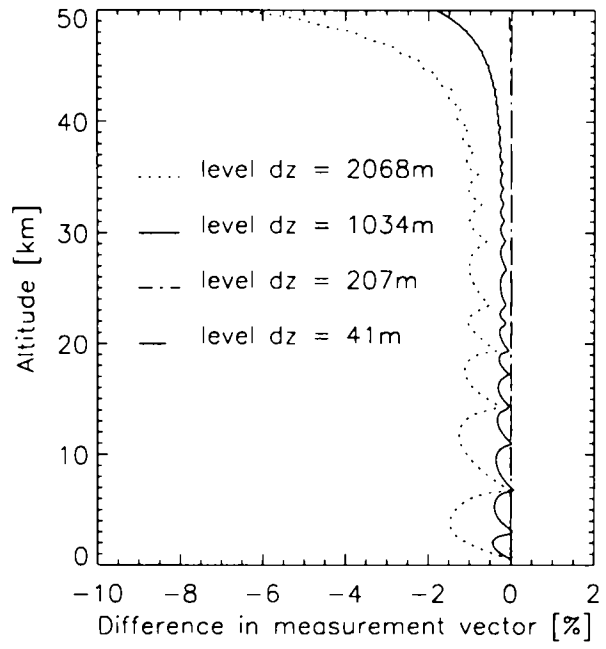


Figure 6.4: A figure to show that increasing the number of interpolated levels in between original levels reduces the error from incorrect evaluation of numerical integrals.

with a crude resolution. This type of error is called a representation error and describes the error from the inability of NWP model analyses to represent small-scale atmospheric structure, e.g. seen with GNSS radio occultation measurements. An attempt has been made to estimate this quantity by taking a profile of bending angle measurements, mapping them into refractivity space then mapping them back into bending angle space, using the inverse and forward Abel transforms, respectively. The same procedure is repeated but this time reducing the number of refractivity 'observations' by replacing them with interpolated values, assuming refractivity linearly decreased with logarithmic impact parameter. The reduced number of measurements is based upon the number and spacing of levels currently available from an NWP model, i.e. of the order of 30. The r.m.s. differences between the two forward modelled refractivity profiles are noted for an ensemble of profiles: the differences are found to be the order of 2% of the bending angle measurement in the troposphere and upper stratosphere, decreasing slightly in the middle stratosphere. This variation in the error is associated with the temperature variations in this region. This source of error is assumed to be both inter- and intra-quantity independent.

The forward model used is based on geometric optics therefore does not account for atmospheric diffraction. However, the geometric optics assumption is successful in

describing propagation characteristics above a certain diffraction limit. In the presence of sharp vertical profiles, e.g. in the vicinity of a tropical tropopause, the assumptions can lead to errors. Kursinski et al. (1997) have compared results from a geometric optics forward model with a forward model that accounts for diffraction. The temperature error in the vicinity of a sharp tropical tropopause structure is approximately 0.8 K and, except for the peak of the central error spike, are generally much less than that. The vertical extent of the error is mostly contained within ± 0.7 km (the radius of the first Fresnel zone) of the sharp structures (Kursinski et al. 1997). However, with more typical tropopause structures, errors are 0.2 K of less, with the r.m.s. error over the vertical range of the sharp structure is approximately 0.1 K. In summary, the effects of diffraction in the forward modelling are difficult to quantify; Kursinski et al. (1997) have shown in worst-case scenarios that an error does arise but in general this error is not present and is therefore not considered.

Forward model parameter errors

Other sources of error when modelling a physical system are the uncertainties associated with physical constants.

The major physical constants used in the forward model are the refractivity coefficients (c_1 & c_2, §3.1) and the local radius of curvature.

The origins of the refractivity coefficients are described in §3.1. They are experimental values which have been reported most recently by Thayer (1974). The uncertainties attached to the c_1 and c_2 are 0.02% and 1%, respectively; the latter value is relatively large due to the difficulty experienced when measuring atmospheric water vapour density. It is noted that these uncertainties do not represent the error associated with the value but the uncertainty of the measured quantity. Although these values will not be included in the overall forward modelling error covariance matrix, measurement vector differences using perturbed refractivity constants are investigated.

The local radius of curvature assumed for zero altitude is a parameter that is used to compute the bending angle observations from atmospheric phase delay, therefore any error associated with this parameter will be present in the bending angle observations.

This parameter is also used to compute geometric height levels from impact parameter levels (equation 3.2). The uncertainty of this value is assumed to be approximately 1 metre and reflects the level of accuracy with which it can be computed [Dr. J. J. Barnett: private communication].

By perturbing these constants by their uncertainties the relative changes in bending angle space can be computed. The largest source of forward model parameter error is found to be the local radius of curvature uncertainty peaking at about 0.2% near-surface which decreases quickly to a bias throughout the measurement space of magnitude 0.04% of the bending angle measurement. Both the wet and dry refractivity constants have large effects near the surface but are small compared with the error from the local radius of curvature (approximately 0.03% of the bending angle measurement); above the 10 km both errors reduce to a constant value of 0.02%.

Since the local radius of curvature is a scalar quantity, a change in one level will be positively correlated with all others. It is computed thus

$$\mathbf{S}_{\hat{\mathbf{F}}_{\mathrm{R_E}}} = \boldsymbol{\delta\varepsilon}_{\mathrm{R_E}} \boldsymbol{\delta\varepsilon}_{\mathrm{R_E}}^{\mathrm{T}} \tag{6.1}$$

where $\boldsymbol{\delta\varepsilon}_{\mathrm{R_E}}$ is a vector of bending angle errors obtained from the refractivity parameter uncertainty and $\mathbf{S}_{\hat{\mathbf{F}}_{\mathrm{R_E}}}$ is the error covariance matrix.

The total forward model parameter error covariance is computed by summing individual forward model parameter error covariance matrices (thus making different forward model parameter error independent); in this case it is just the local radius of curvature error covariance matrix.

Total measurement error covariance

The total measurement error covariance $\mathbf{S}_{\Sigma}$ is constructed by summing the covariance matrices from measurement noise, forward modelling and forward model parameters.

The diagonal terms have also been constrained using a minimum value to account

for the noisy upper stratosphere measurements: if the diagonal elements are less than this value then the minimum value is assumed.

$$\boldsymbol{\sigma}_{y_{i,i}} = \max(\boldsymbol{\sigma}_{y_{i,i}}, 10^{-6}) \tag{6.2}$$

Figure 6.5 shows how the standard deviation values of the principal diagonal from each contribution to the measurement error covariance matrix.

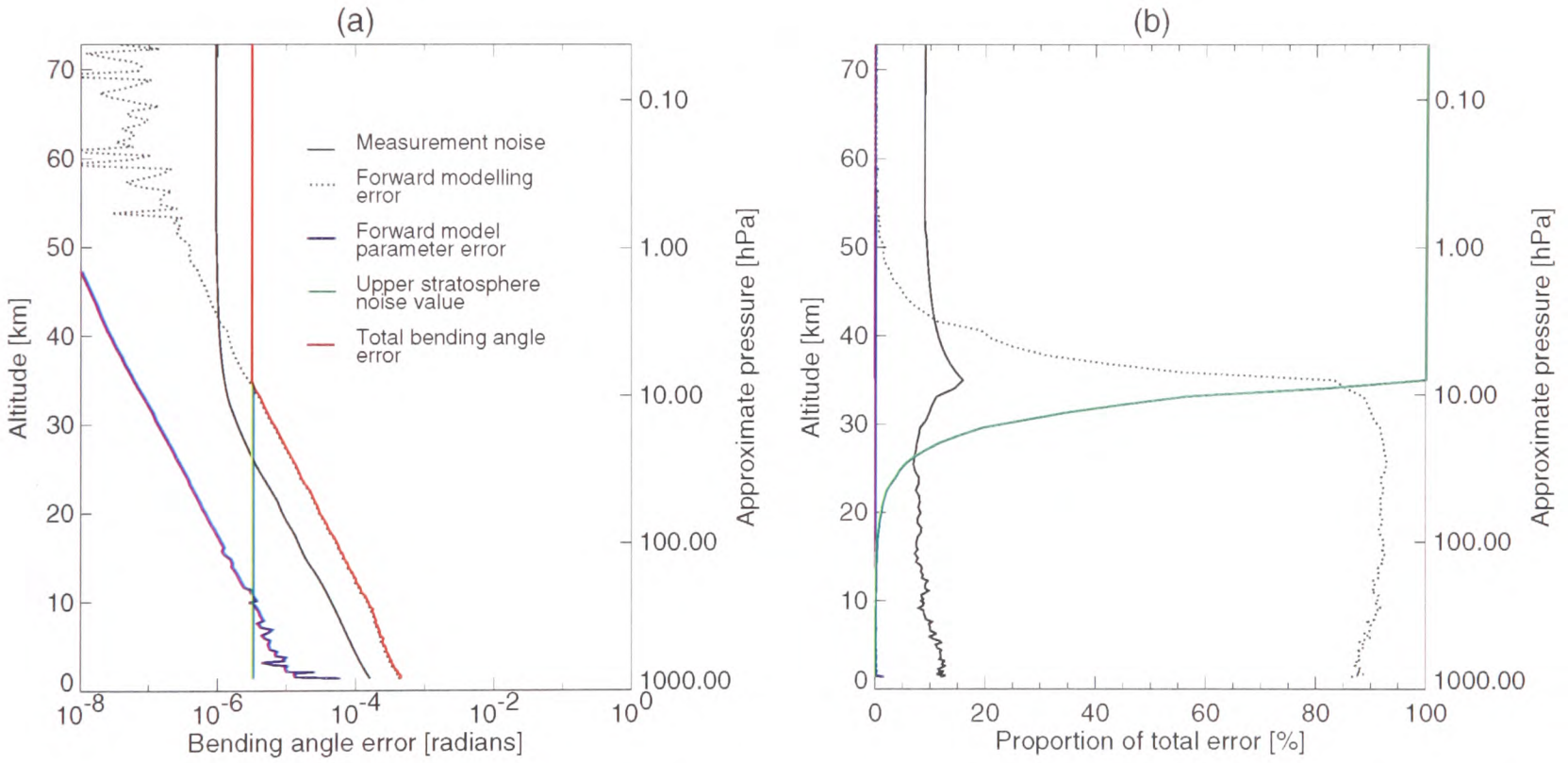


Figure 6.5: The optimal estimation inverse model total measurement error budget for a typical RO profile. Panel (a) shows typical r.m.s. values from the principal diagonal from each contribution to the total measurement error covariance matrix; and panel (b) shows the individual measurement error sources as a percentage proportion of the total error.

The dominant source of error for the majority of the vertical range considered is the forward modelling error, i.e. representation error, with the measurement noise providing a small contribution to the total error. In the upper stratosphere the total bending angle reverts to the upper stratosphere minimum noise used. At near-surface altitudes, the error from the local radius of curvature is significant.

6.2 Convergence and quality control

The method used to judge convergence relies upon values of the penalty function, i.e. if the relative change is smaller than a specified value (0.5%) then the solution is determined to have converged. This method alone is found to be a good indictator of convergence.

For the work presented in this paper, the maximum number of iterations considered is 10; if the solution has not converged (determined by the method presented above within 10 iterations) then the calculation is halted and a numerical 'flag' set.

If, after convergence has been determined, the retrieval χ^2 value is greater than the tabulated χ^2 value given the number of degrees of freedom and a set tolerance level (in this case 99.9%) then a numerical 'flag' is set.

Retrievals with flags set are omitted from any statistics, and investigated to ascertain why they have failed this quality control.

Furthermore, to ensure the solution computed at each iteration is not unphysical, the ln(specific humidity) elements are checked for super-saturation and corrected if necessary.

6.3 Characterisation of the optimal estimation inverse method

For the optimal estimate characterisation, the matrix of weighting functions $\mathbf{K}$, matrix of contribution functions $\mathbf{D}_y$ and the matrix of averaging kernels $\mathbf{A}$ are presented.

For the solution characterisation, a profile of bending angles is generated on a typical set of measurement levels from an occultation event (-39.7°S, 282.8°E) using the method outlined in §6.4. The altitude range of this occultation is 0.7−50 km.

6.3.1 Matrix of weighting functions

A weighting function is a measure of how sensitive the measurement vector is to a perturbation of a state vector parameter. If the ith element of $\mathbf{x}$ were increased by one unit, the ith column of the matrix of weighting functions would represent the corresponding change in measurement vector $\mathbf{y}$. Similarly, the ith row of the matrix

of weighting functions represents the contributions from each element in the state vector $\mathbf{x}$ to the ith element of $\mathbf{y}$. The matrix of weighting functions is denoted by $\mathbf{K}$ and may be computed thus

$$\mathbf{K} = \frac{\partial \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}})}{\partial \mathbf{x}} \tag{6.3}$$

where $\mathbf{F}$ is the forward model and $\hat{\mathbf{b}}$ represents a vector of the best estimates of the forward model parameters used to compute the measurement vector.

In practice, $\mathbf{K}$ can be computed by perturbing each element of the state vector in turn and looking at the difference between the perturbed and the unperturbed simulated measurements: the 'brute force' method.

Since the information held in this matrix is to be combined with the information from the UKMO *a priori* error covariance matrix, it was decided to compute the $\mathbf{K}$ matrix in the same variables. In this case, the variables of interest are temperature [K], ln(specific humidity) spectrum [g/kg][5] given and surface pressure.

It is important to understand the structure of the weighting functions because they represent how state space information is mapped into measurement space. Typical weighting functions are illustrated by figure 6.6. The left panel contains a typical temperature weighting function, and the right panel contains a typical ln(specific humidity) weighting function; both have been perturbed at the same state vector pressure level (850 hPa).

The weighting functions shown by figure 6.6 are not similar to those from other instruments, instead they have two peaks. A general description of these functions follows (Eyre 1994):

For the layer containing the ray, there are strong contributions caused by the changes in the vertical density gradient across the layer: a positive contribution from an increase in temperature at the lower boundary, and a negative contribution from an increase at the upper boundary.

[5]The forward operator uses specific humidity [kg/kg], and it is noted that $\frac{\partial \varepsilon}{\partial \ln(q)} = q\frac{\partial \varepsilon}{\partial q}$: if we perturb specific humidity by 1%, then this equation is equal to $100d\varepsilon$.

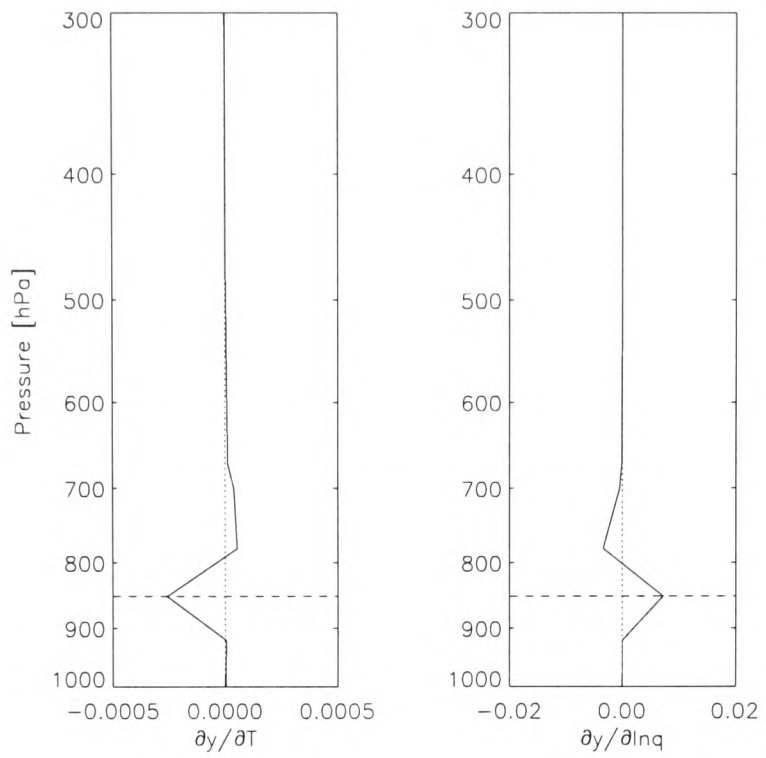


Figure 6.6: Typical radio occultation weighting functions. The left panel shows a typical temperature weighting function and the right panel shows a typical specific humidity weighting function, perturbed at the same state vector pressure level (850 hPa). The dashed lines in both plots represent y=850 hPa.

For the levels below the ray, there is a small positive contribution caused by the increase in pressure at constant height resulting from an increase in temperature below.

For specific humidity, the pattern is similar to that of temperature, but the sign is reversed in the vicinity of the ray, and the relative effect of humidity below the ray is much lower because humidity has minor effect on the hydrostatic relation.

The complete set of weighting functions for the example occultation are shown by figure 6.7. Panels (a) and (b) show the temperature and ln(specific humidity) weighting functions for the vertical range (1000−300hPa). Panel (c) shows the upper atmosphere temperature weighting functions (300−0.1hPa); the upper stratospheric levels are sensitive to changes in the state vector but are comparatively small and therefore not visible in panel (c). The measurement sensitivity to a surface pressure perturbation is shown in panel (d) which has been expressed as a percentage difference.

The measurement sensitivity to surface pressure is shown in panel (d). The magnitude of the surface pressure perturbation is 100 hPa (which corresponds to approximately 0.1% of the surface pressure). The structure of the sensitivity is related to the measurement vector level spacing.

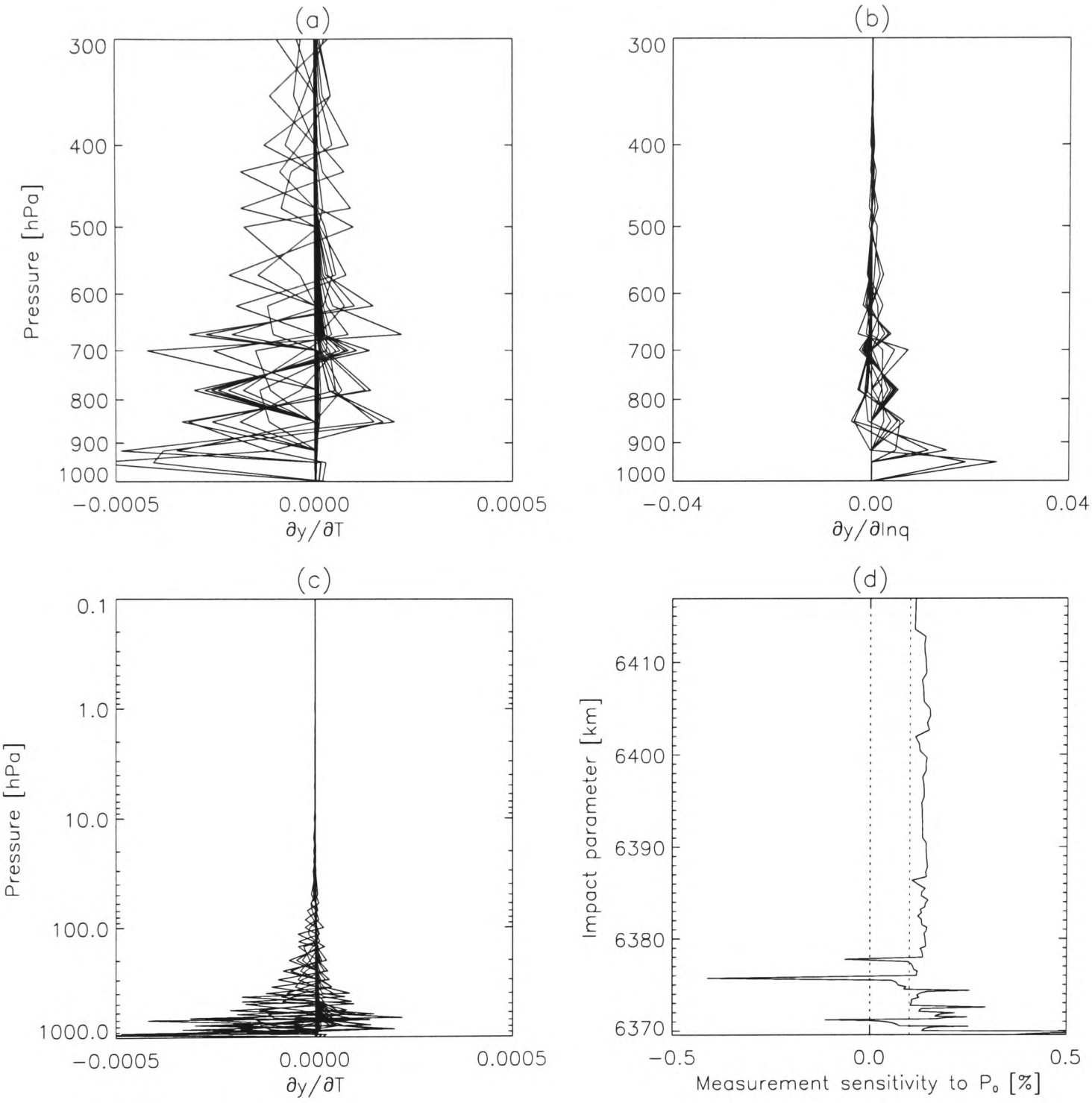


Figure 6.7: A complete set of weighting functions for an example occultation. Panel (a) and (b) show the lower atmosphere (1000–300 hPa) weighting functions for temperature and ln(specific humidity); panel (c) shows the upper atmosphere (300–0.1 hPa) weighting functions for temperature; and panel (d) shows the measurement vector sensitivity to a perturbation in surface pressure expressed as a percentage. The broken lines shown in panel (d) represent $x = 0$ and $x = 0.1$.

6.3.2 Matrix of contribution functions

The ith column of $\mathbf{D_y}$ represents the contribution to the solution due to a unit change in the ith element of $\mathbf{y}$ for the retrieval algorithm.

Figure 6.8 shows selected contribution functions for the optimal solution. Each curve is annotated with a pressure corresponding to the measurement level perturbed. Several

panels are used to illustrate the temperature contribution functions because they span a large dynamic range. The cross-quantity functions are not shown because they do not add any information to the plot.

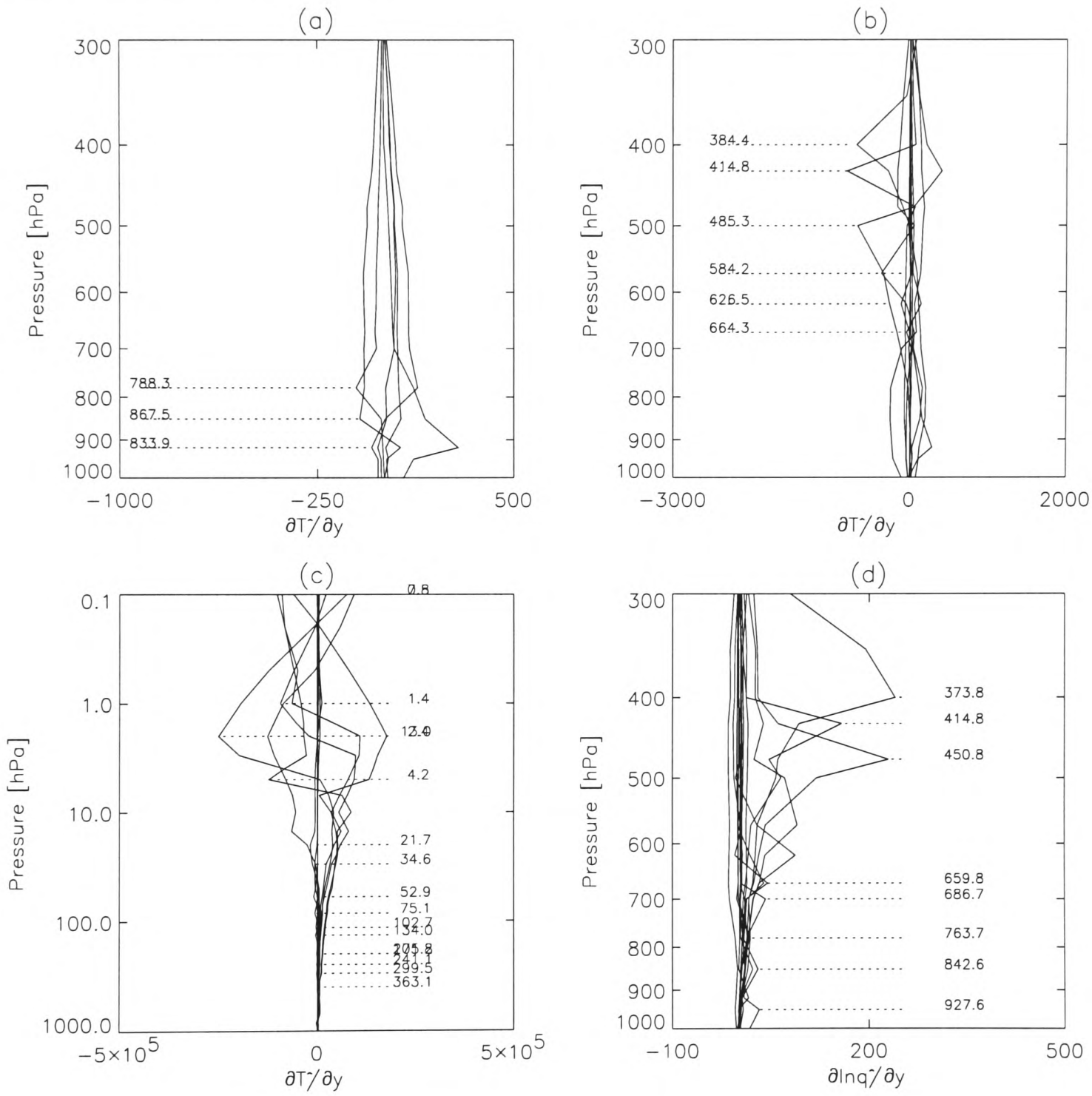


Figure 6.8: Selected columns of the matrix of contribution functions. Panels (a), (b) and (c) show the temperature contribution functions, and panel (d) shows the ln(specific humidity) contribution functions. The number corresponds to the peaks (in hPa) of the contribution functions.

In general, the peak of a function corresponds approximately to the pressure of the perturbed measurement level. In the upper atmosphere the sharpness of the functions decreases, implying that more measurements contribute to the upper atmosphere optimal solution than the lower atmosphere, reflecting the thickness of the upper atmosphere

state vector levels. Again, the shapes of the upper atmosphere contribution functions reflect the realistic statistics used to compute the optimal estimate. The general shape of the function can be explained as follows: if a measurement at pressure P is perturbed (increased), the state vector elements at that pressure P would also increase. This would mean that the state vector elements directly above and below P ($P \pm \delta P$) would decrease in order to compensate for the increase. This compensation in the state vector would cause the measurement at $P \pm \delta P$ to be reduced. This relation is similar for the specific humidity but of opposite sign (see §6.3.1).

The matrix of contribution functions also contains information about the sensitivity of the surface pressure optimal solution with measurement vector perturbations and is shown by figure 6.9.

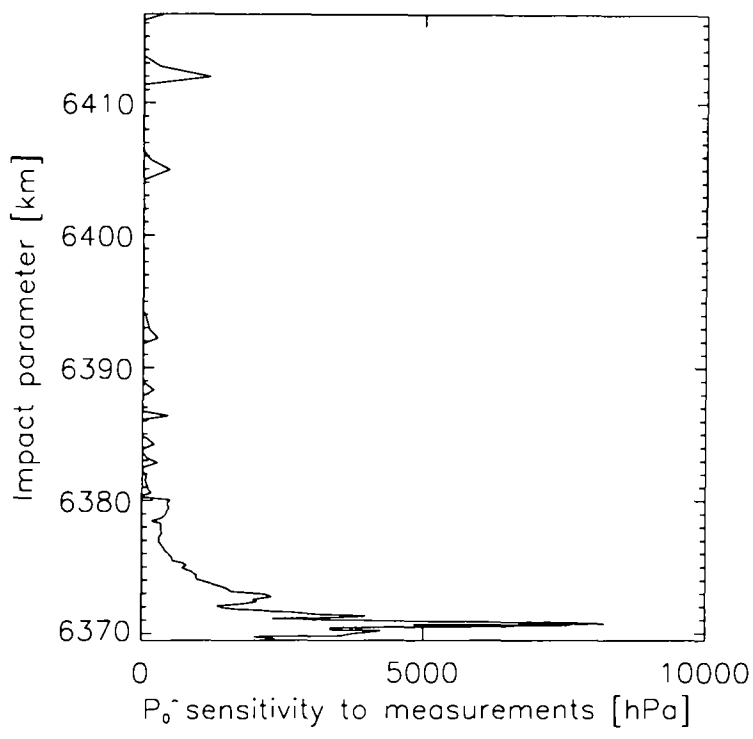


Figure 6.9: The surface pressure optimal solution sensitivity to perturbation in the measurement vector.

The surface pressure optimal solution sensitivity suggests that the greatest sensitivity lies near to the surface as expected.

6.3.3 Matrix of averaging kernels

The ith row of the matrix of averaging kernels shows how the real atmospheric state at different levels are used to construct the ith element of the retrieved state vector.

If the area underneath each kernel is computed and plotted as a function of pressure, the curve effectively represents the retrieval response to a unit shift in the state vector which can be used to gauge the retrieval performance as a function of pressure: the nearer the response is to unity the better the retrieval is.

Figure 6.10 shows the principal averaging kernels for the retrieval. Panel (a) shows the temperature averaging kernels and panel (b) the ln(specific humidity) averaging kernels with the surface pressure kernel element inset. Panels (c) and (d) show selected temperature and ln(specific humidity) kernels with the pressure at which the state vector was perturbed plotted alongside.

In general, the shape and distribution of the kernels is a direct consequence of using realistic statistics for the *a priori* and measurement errors, and the state vector used. The near-surface temperature averaging kernels are comparatively small because increasing emphasis is given to the ln(specific humidity) and surface pressure retrieval; also, the range of the measurement vector affects the level of retrieval response near-surface.

Panel (a) shows the temperature averaging kernels. The area underneath the curves (an indication of the retrieval response) remains positive throughout the vertical region described by the measurement vector. At approximately 4 hPa, the response is slightly negative, reflecting the upper limit of the particular measurement vector. Above this level the response increases due to correlations from the *a priori* error covariance matrix and from the weighting functions (equation 4.47).

Panel (b) shows the ln(specific humidity) averaging kernels. The retrieval response decreases slightly at 1000 hPa because of the measurement range. From 400−1000 hPa the response is greater than 0.5, above this range decreasing to about 0.3 at 300 hPa. The surface pressure response (0.46, in this case) is dependent on how close to the surface the measurement vector is (as shown by figure 6.14).

Panels (c) and (d) show some of the temperature and ln(specific humidity) averaging kernels in detail, with the pressure level at which the state vector was perturbed alongside each kernel. In general, the retrieval response peaks at the correct pressure.

Figure 6.11 shows the cross-quantity averaging kernels from the radio occultation

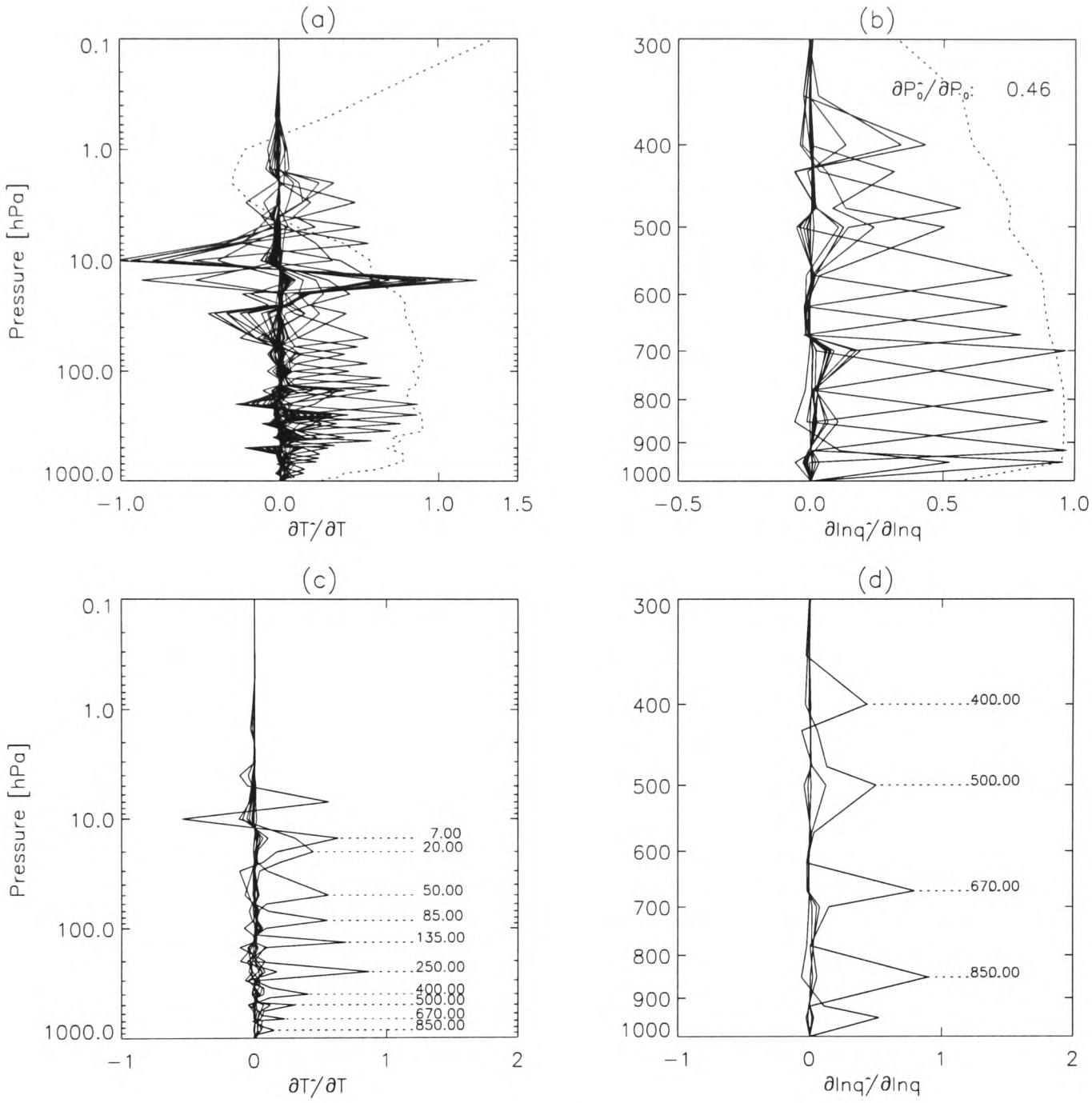


Figure 6.10: The principal averaging kernels. Panel (a) shows the temperature averaging kernels; panel (b) shows the ln(specific humidity) averaging kernels (with the surface pressure value inset); and panels (c) and (d) show selected kernels in detail. The broken lines plotted beside the principal kernels represent the area underneath the kernels.

retrieval. Since the solution is a combination of three quantities there are six cross-quantity averaging kernel sets.

These cross-quantity kernels are very complicated in structure but some understanding may be obtained from them. Panel (a) shows the sensitivity of the temperature solution to unit perturbations of the ln(specific humidity) state vector elements; panel (b) shows the sensitivity of the ln(specific humidity) solution to unit perturbations of the temperature state vector elements; panels (c) and (d) show the sensitivity of the temperature/surface pressure solution to unit perturbations in the surface

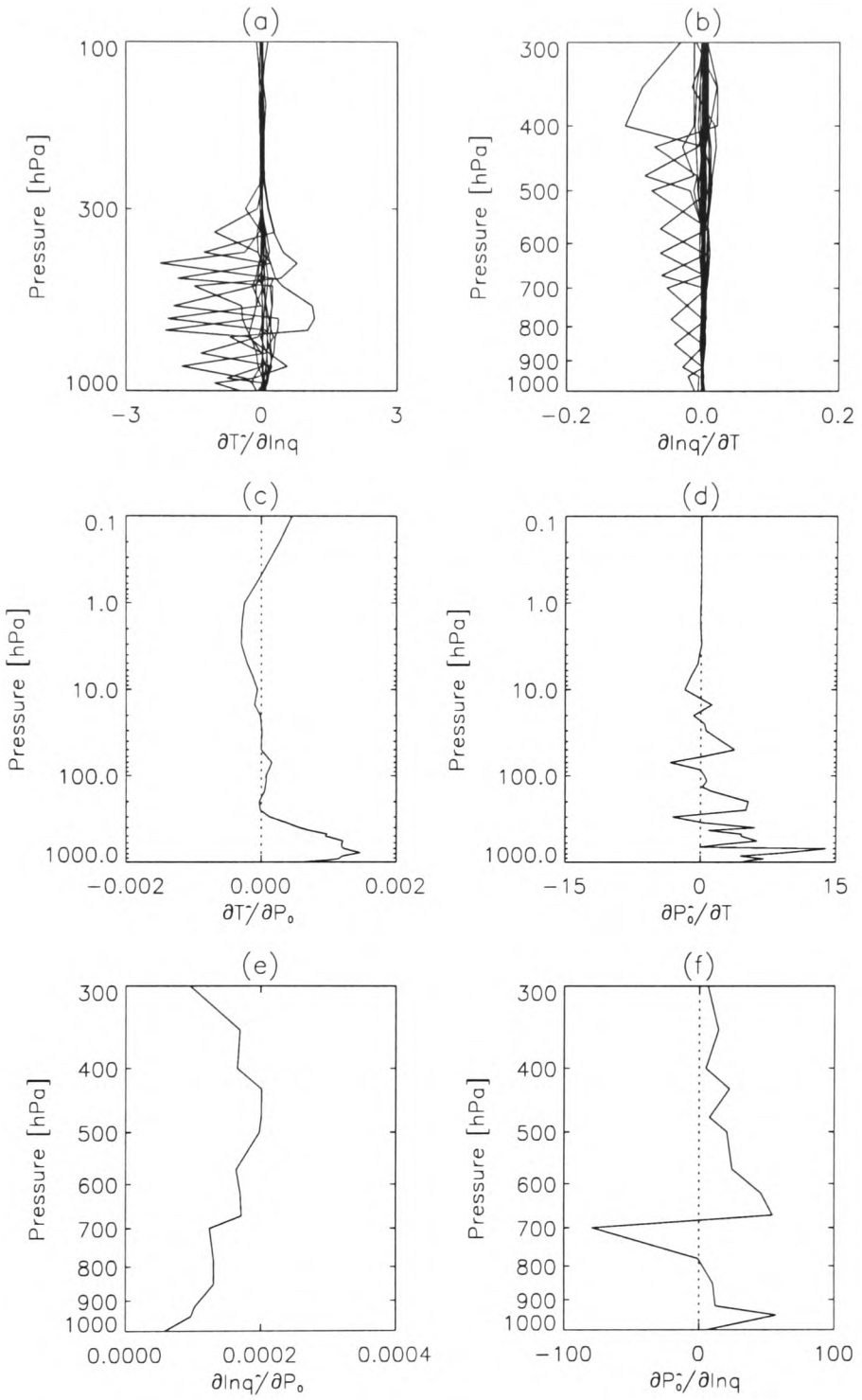


Figure 6.11: The complete set of cross-quantity averaging kernels for the radio occultation retrieval. Temperature and surface pressure are expressed in kelvin and hPa, respectively.

pressure/temperature state vector elements; and panels (e) and (f) show the sensitivity of the ln(specific humidity)/surface pressure solution to unit perturbations in surface pressure/ln(specific humidity) state vector elements.

Panel (a) shows that ln(specific humidity) perturbations do not affect the temperature solution much above the vertical range described by the ln(specific humidity) elements (i.e. 1000−300 hPa). The level of response of the temperature solution varies as a function of pressure, reflecting the realistic statistics.

Panel (b) shows that the ln(specific humidity) solution sensitivity to unit perturba-

tions in temperature increases in magnitude as a function of altitude range described by the ln(specific humidity) solution. This can be explained if the simultaneous retrieval emphasis is considered: at 300 hPa and at altitudes below, the retrieval emphasis moves from temperature to ln(specific humidity), therefore the sensitivity from temperature perturbations at low altitudes are smaller than at altitudes near 300 hPa.

Panels (c) and (e) show that the sensitivity of the temperature and ln(specific humidity) solutions from perturbing the surface pressure element is small and localised mostly in the lower atmosphere (1000−300 hPa).

Panels (d) and (f) show the sensitivity of the surface pressure solution from perturbations of temperature and ln(specific humidity) state vector elements. The sensitivity from ln(specific humidity) state vector is larger than the temperature state vector, implying greater inter-quantity sensitivity.

6.3.4 Contributory error covariance matrices

The method used to characterise the errors from the optimal method is discussed in §4.5.2. The error associated with the retrieval can be split into its constituent parts, namely error from the *a priori* error estimates, forward modelling, forward model parameters and measurement noise.

Figure 6.12 shows the principal diagonal standard deviation values from the contributory covariance matrices.

Panels (a) and (b) shows that the *a priori* provides the largest contribution to the temperature and ln(specific humidity) retrieval error above 10 hPa and 500 hPa, respectively.

Panels (c) and (d) show the measurement noise contribution to the retrieval error which is comparatively small. The contribution to the optimal temperature solution is very similar to the analytic equivalent (figure 5.20, panel (c)), indicating that the matrices of contribution functions are similar in magnitude and structure. The error contributions to the temperature, ln(specific humidity) and surface pressure retrieval errors are a direct consequence of the realistic statistics used to compute these quantities (§6.1.2).

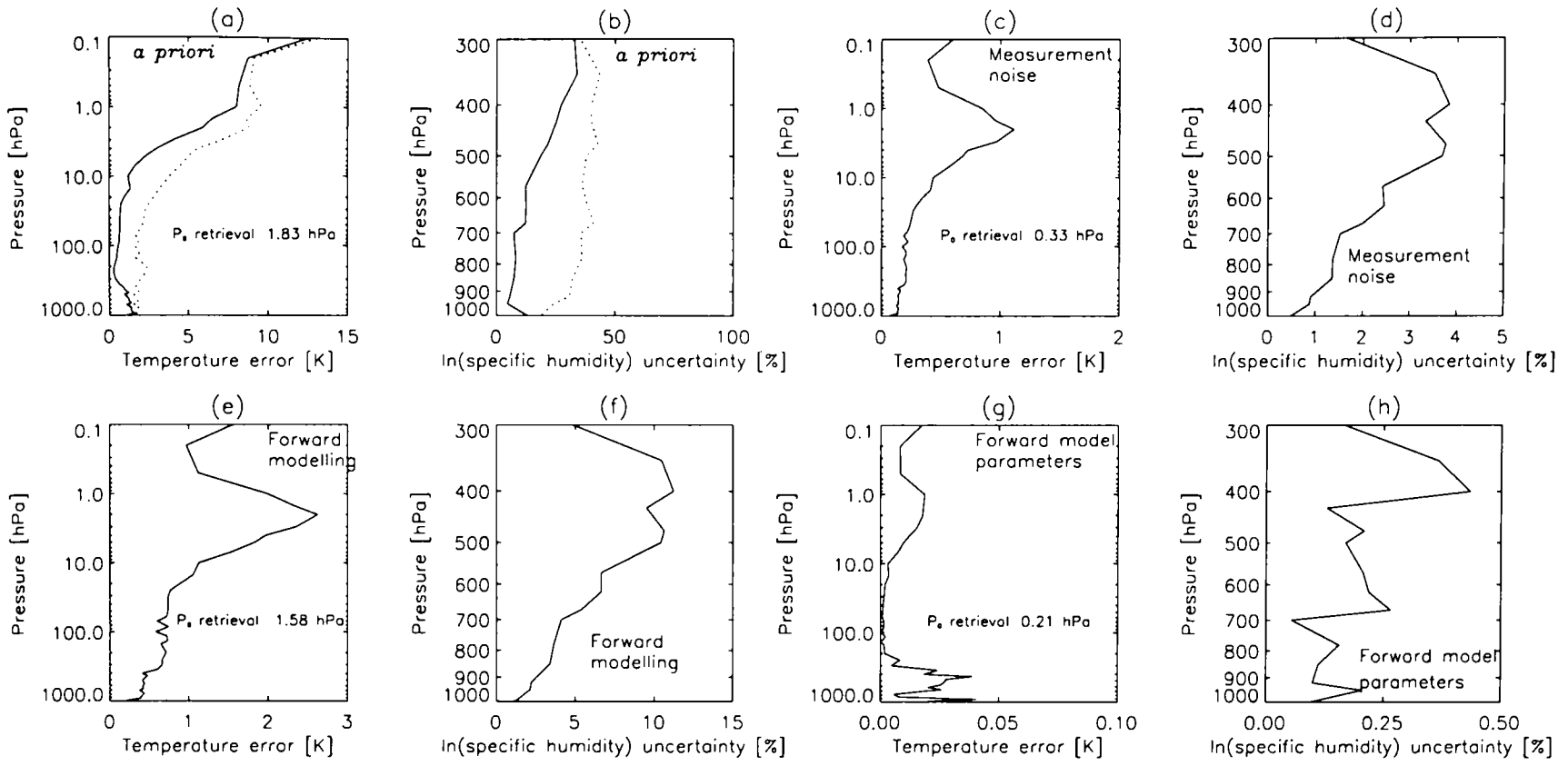


Figure 6.12: Solution error characterisation of the optimal estimation inverse method. In general, left panels show the contribution to the temperature retrieval error and the right panels show the error contributions to the ln(specific humidity) retrieval error. The surface pressure retrieval error is inset of the left panels. Panels (a) and (b) show the *a priori* contribution to the retrieval error; (c) and (d) the measurement noise contribution; (e) and (f) the forward modelling contribution; and (g) and (h) the forward model parameter contribution. The dotted lines shown in panels (a) and (b) represent the r.m.s values of the principal diagonal of the *a priori* error covariance matrix.

Panels (e) and (f) show the forward modelling contribution to the retrieval error. The representation error has been used to compute this contribution (§6.1.2). The structure of this contribution is very similar to that of the measurement noise, but has a larger associated error. In general, forward modelling represents the second largest contribution to retrieval error.

Panels (g) and (h) show the forward model parameter contribution to the retrieval error. This error represents the smallest contribution to the total retrieval error. The contribution towards the temperature solution peaks at 0.05 K near 1000 hPa due to the local radius of curvature error, and the contribution towards the ln(specific humidity) uncertainty is typically 0.2% throughout the vertical range. The contribution to the surface pressure solution represents a small fraction of the total retrieval error.

The comparison of the error characterisations between the Abelian and optimal

estimation inversion methods has to be considered carefully. First, the Abelian inversion method does not use *a priori* information to construct the solution nor does it utilise estimates of measurement noise. And second, the analytic method formalism does not allow the simultaneous retrieval of temperature and humidity (and surface pressure), therefore the specific humidity representation used in the forward model is included as a forward model parameter.

6.4 Ensemble of numerical simulations

The performance of the optimal estimation retrieval scheme has been examined using simulated data for an ensemble of profiles. For each case, a 'true' profile is established by taking one of a set of profiles of collocated UKMO model analyses (§7.1), from which to compute the 'true' measurement vector[6]. The associated background atmospheric profile is calculated by perturbing the 'true' profile thus

$$\mathbf{x}_a = \mathbf{x}_t + \sum_{i=1}^{n} \epsilon_i \lambda_i^{1/2} \mathbf{I}_i \tag{6.4}$$

where the subscript t corresponds to the 'truth', λ_i and $\mathbf{I}_i$ are the ith eigenvalues and eigenvectors of the *a priori* error covariance matrix and ϵ_i represents the ith random number drawn from a Gaussian population.

The main advantage of simulations is that the true atmospheric state is known. Another advantage of using this method to simulate atmospheric profiles is that the difference between truth and background is consistent with the statistics assumed for the *a priori* information, since the eigenvalue decomposition represents characteristic modes of the statistics used.

Measurement noise is modelled and superimposed onto the measurement vector using the method described by equation 6.4, utilising the eigenvectors and eigenvalues from the total measurement error covariance matrix (equation 6.2).

[6]The measurement vector levels are taken from associated GPS/MET impact parameter levels.

Results

Using the method described by equation 6.4, simulated measurements and realistic background profiles were produced. These were used as inputs to the inversion scheme to obtain retrieved profiles which were subsequently compared with the 'true' profiles to assess the impact of the measurements on the background information.

Five hundred profiles with random temperature, humidity and surface pressure conditions have been tested, and successful retrievals (i.e. satisfactory convergence of equation 4.42 in ten iterations or less, §6.2) were obtained in all but two retrievals. In most cases convergence is obtained within three or four iterations. In general, the χ^2 values at convergence were comparable with the number of observations, indicating that the error estimates used in the retrieval are appropriate (Marks and Rodgers 1993).

The two cases of non-convergence have been examined. They are found to be cases in which the iterated profile oscillated about the true profile. The χ^2 values associated with both these profiles have values comparable with the number of degrees of freedom and much less than would be needed to satisfy the failure criterion (§6.2). This suggests that such profiles, even though not converged (if the current convergence criterion is used), will represent good retrievals. Nevertheless, they have not been included in the statistics shown.

For each successful retrieval, the retrieval error and the background error (first-guess error) have been calculated, and the mean and r.m.s. values of these data have been computed. The square roots of the solution error covariance matrix and *a priori* error principal diagonal elements also give the r.m.s values for the retrieval errors and background errors, respectively. Attention should not be paid to the actual value of the retrieval error estimate: more emphasis should be paid to the ratio of this quantity to the assumed background errors. It is this ratio which is related to the amount of information the measurements supply to the NWP system, and may be parameterised using information content.

The information content of a particular set of measurements may be defined as the measure of the factor by which the uncertainty in the state vector is decreased by the

act of measuring it, i.e. fitting a set of measurements to an *a priori* atmospheric state (see appendix C for further details). This quantity will be used to complement the r.m.s. statistics shown.

A particular definition of information content that is useful in the case of simultaneous retrievals is the improvement vector which illustrates how optimal estimation has improved the knowledge of the background state throughout the atmospheric profile (Palmer et al. 1998).

The improvement vector $\boldsymbol{\eta}$ is given by

$$\boldsymbol{\eta}_j = 100 \times \left[1 - \left(\frac{\hat{\mathbf{S}}_{j,j}}{\mathbf{S}_{\mathbf{a}j,j}} \right)^{\frac{1}{2}} \right] \tag{6.5}$$

where j is the matrix and improvement vector element index, and all other variables are as before.

The results from the ensemble of simulated retrievals are summarised by figure 6.13. The background errors resemble those shown by figure 6.1 modified slightly by the modelled measurement noise. For the ensemble of simulated retrievals the error associated with the surface pressure optimal estimate has decreased by ~12%, indicating that radio occultation observations do contain information about the surface pressure.

Although it is clear from figure 6.13 the r.m.s. of the retrieval errors are smaller than the prior errors, as they should be, the extent of the error reduction is not clear by viewing the theoretical r.m.s. errors alone. The complimentary improvement vector for these simulated retrievals (panels (c) and (d)) provide more information about the results.

In general the temperature improvement vector suggests that optimal estimation improves considerably upon the prior knowledge of the atmospheric temperature, from the lower-troposphere to the upper stratosphere. The temperature improvement vector declines in the upper stratosphere partly because of the upper noise limit used which is comparable with the *a priori* errors in measurement space and partly because it represents the end of the measurement vectors used in the ensemble of simulations.

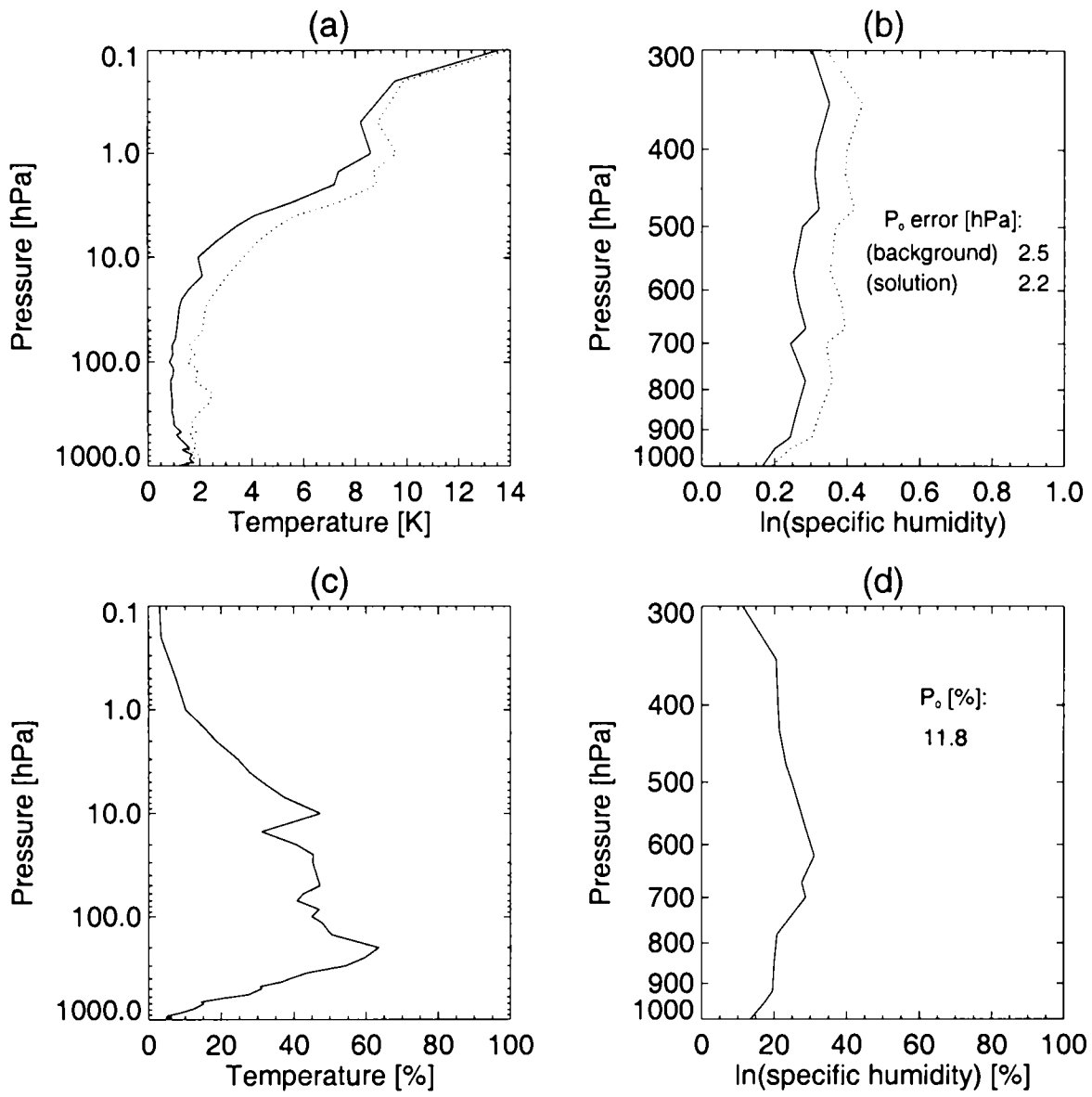


Figure 6.13: Theoretical error estimates from the ensemble of simulated retrievals. Panels (a) and (b) show the r.m.s. errors for temperature and ln(specific humidity), respectively, with the surface pressure values inset. The solid and dotted lines represent the retrieval and background errors, respectively. Panels (c) and (d) show the corresponding improvement vectors for the errors in panels (a) and (b).

The temperature improvement vector declines in the lower-troposphere where humidity becomes more significant, and both quantities are retrieved simultaneously, thus the emphasis is shifted from temperature to ln(specific humidity). It is clear from the plot that there is a gradual decline in the temperature retrieval quality as the humidity retrieval quality improves.

Both the temperature and humidity knowledge decline near the surface because the majority of occultation events presented here end typically a few kilometres above the surface.

The results shown by figure 6.13 confirm that the code and the method are suitable for the purpose of non-linear optimal estimation using RO measurements. Together, the

theoretical r.m.s. errors and the computed improvement vectors suggest that there are improvements in the prior knowledge of the atmospheric state from near-surface to the upper stratosphere. In particular, the level of surface pressure improvement provides a confirmation of the fact that the simulated RO observations improve the prior knowledge of the surface pressure.

6.5 Quality of surface pressure retrievals

It is found that the quality of the surface pressure retrieval is dependent on the vertical extent of the occultation, i.e. how near it reaches the surface. To illustrate this point, one of the simulated profiles (which spans 0.7−50 km) used for the validation of the method is used. By systematically removing observations near the surface and retrieving temperature, ln(specific humidity) and surface pressure, it can be shown how the vertical range of the occultation is important to the quality of the surface pressure retrieval. The major advantage of using one of the simulated profiles is that the truth is known, and can be compared with the retrieved value.

Panels (a), (b) and (c) show that the temperature, ln(specific humidity) and surface pressure improvement vectors get smaller in the lower atmosphere with increasing values for the lowest geometric height level, as expected. Panel (d) shows that the retrieval which includes all the observations provide enough information to retrieve accurately the true value for surface pressure; as the number of near-surface observations decreases the retrieval becomes smaller and smaller, approaching the *a priori* value.

It can be concluded from this experiment that surface pressure retrievals are only useful when the occultation reaches near zero altitude levels. This finding is important to appreciate when validating real radio occultation data.

6.6 Sensitivity of inverse model statistics

The main error estimates from the optimal retrieval have been perturbed which simulate possible improvements/problems with the measurement system or the NWP

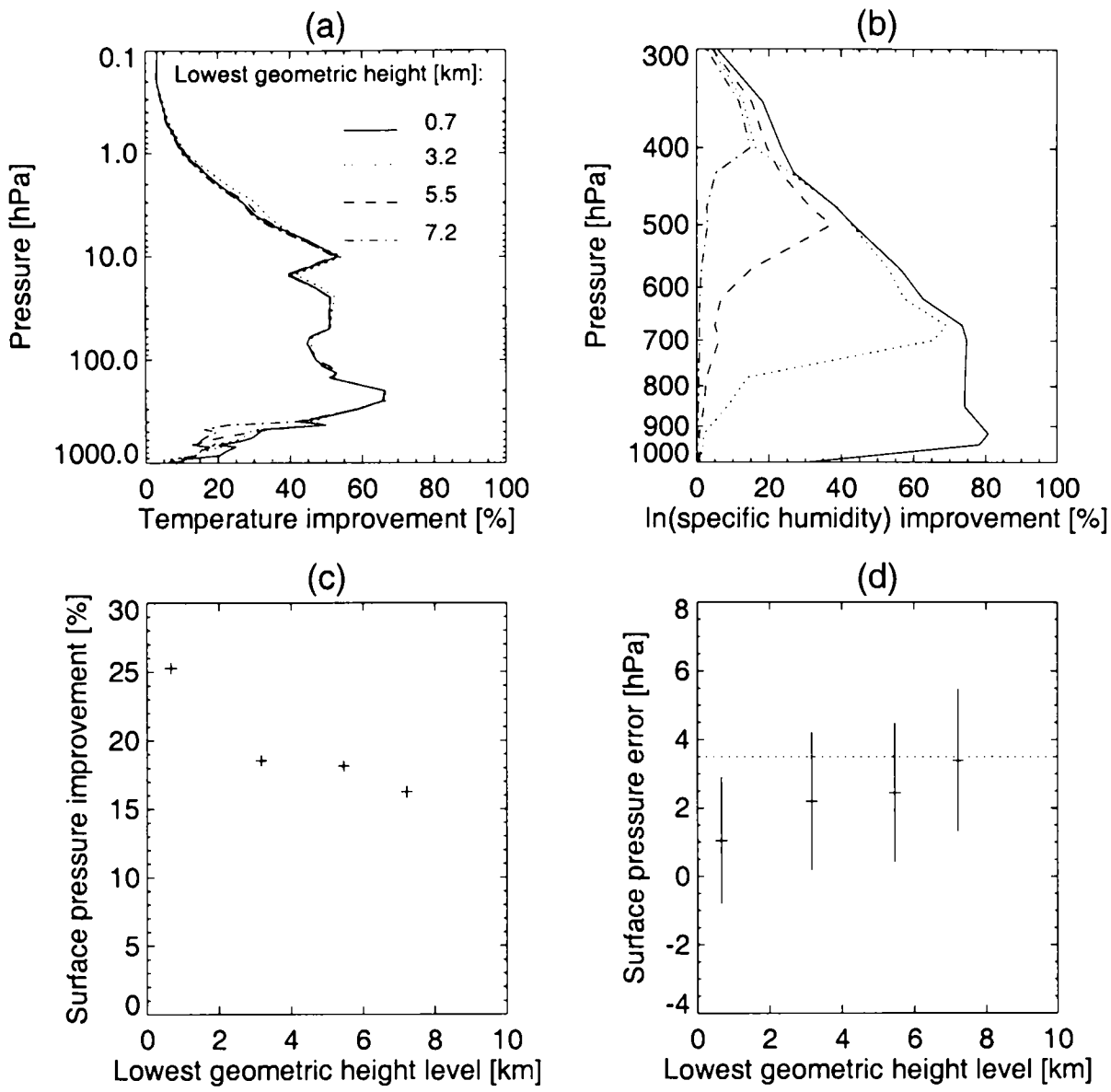


Figure 6.14: Investigating the relationship between the quality of surface pressure retrievals and the lowest geometric height level of the occultation. Panel (a), (b) and (c) show the temperature, ln(specific humidity) and surface pressure improvement vectors; and panel (d) shows the difference (with errors bars) between the truth and the retrieval. The broken line in panel (d) represents the difference between the truth and the *a priori* value.

representation of the atmosphere. The three statistics that are perturbed are the measurement noise, the forward modelling error and the forecast error since these represent the largest contributions to the solution error as shown by figure 6.12.

Perturbing the measurement noise may simulate the possible changes in measurement noise sources, e.g. improvements from high-order ionospheric corrections, or increased measurement noise, e.g. poor quality clocks aboard the GNSS receivers. This is computed by altering the standard deviation of the measurement noise provided by Luntama (1997).

Perturbing the representation error is a crude method of simulating the possibility of assimilating intrinsically high resolution GNSS radio occultation measurements with lower or higher resolution NWP models. Although it is unlikely that these measurements

will be assimilated with an NWP representation with lower resolution than the UKMO unified model (1995 version) there is a distinct possibility that a high resolution model may be used.

Perturbing the *a priori* error covariance matrix represents the effect of changing NWP knowledge of the atmosphere. Increasing these error estimates can represent the extent of the knowledge of the atmosphere in the southern hemisphere where other atmospheric information is sparse. Decreasing the error represents the NWP model having greater confidence in the model and/or greater confidence in other similar atmospheric observations used to initialise the model.

The method used to alter this error covariance matrix is to compute its correlation matrix. The new covariance matrix is constructed from the correlation matrix using altered principal diagonal values from the original covariance matrix. This method ensures that correlations present in the original matrix are retained.

Figure 6.15 shows the improvement vectors from the retrieval sensitivities outlined above. The purpose of figure 6.15 is to show the retrieval sensitivity to changes in error covariance matrix used in the retrieval. It is therefore useful to also plot the improvement vector from the unperturbed retrieval.

Panels (a) and (b) shows that doubling or halving the measurement noise has little effect on the overall improvement, reflecting the contribution from measurement noise to the total measurement error covariance matrix (figure 6.5). The changes due to ln(specific humidity) are very slight which is why the plot does not show much difference. The surface pressure improvement changes by typically by a few percent.

Panels (c) and (d) correspond to increasing or decreasing (by 50%) the representation error used to parameterise the forward modelling of GNSS radio occultation measurements from model analyses. This error is the largest source of measurement error covariance matrix for much of the vertical atmospheric considered (see figure 6.5) and as such has a large influence on the degree of improvement. The level of increase and decrease of the temperature improvement is of the order of 15% for the lower stratosphere and upper troposphere, above which this source of error does not provide a large contribution to the total measurement error covariance matrix, and below which

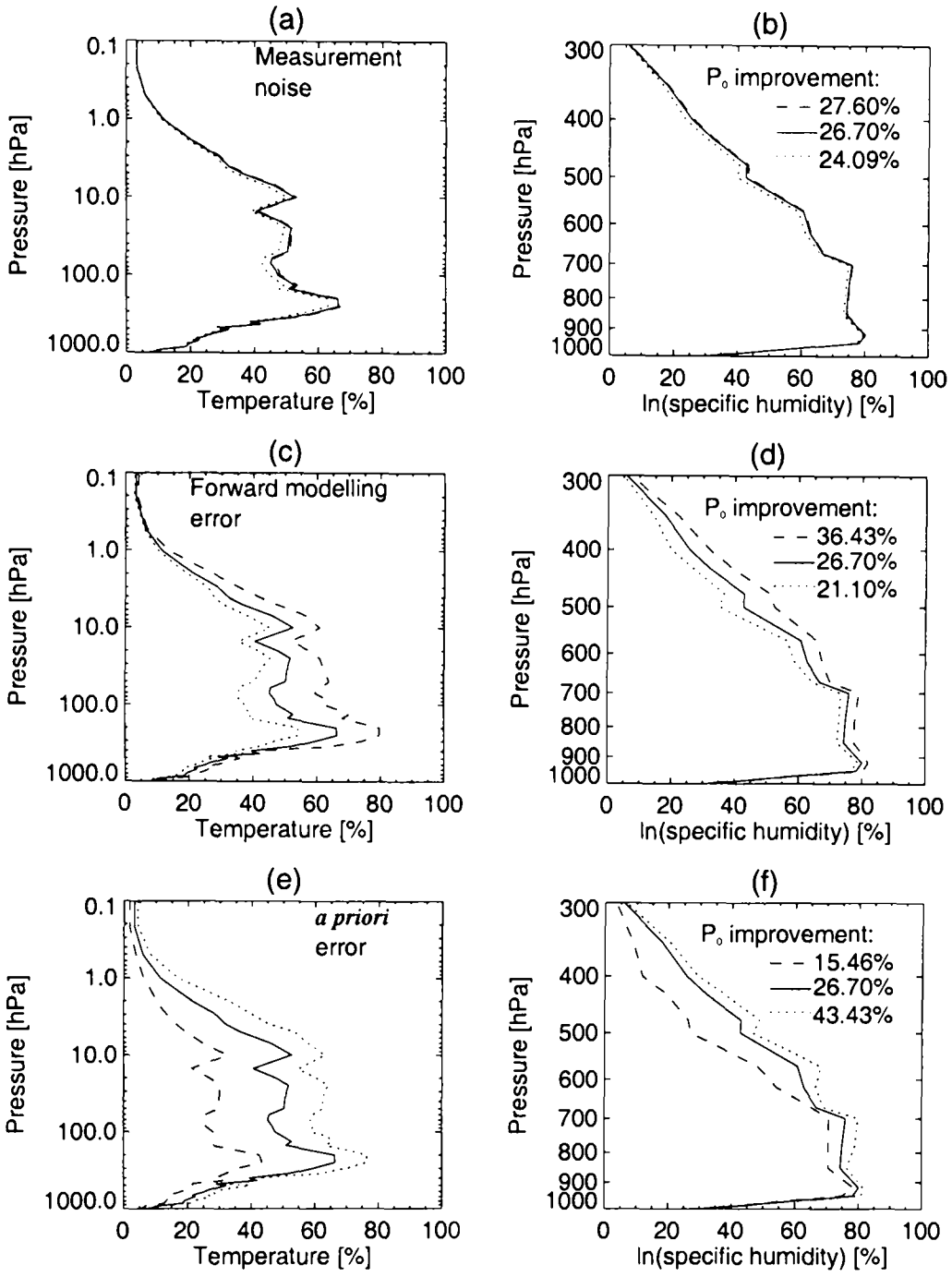


Figure 6.15: Improvement vectors from the retrieval sensitivity studies. In general, left panels show the temperature improvement vector elements and the right panels show the ln(specific humidity) improvement vector elements, with the surface pressure improvement vector element inset. For panels (a)−(d), dashed curves represent a reduction in the quantity altered, dotted curves represent an increase in the quantity altered and solid curves represent the control case; for panels (e) and (f) this dashed/dotted line convention is reversed. Panels (a) and (b) show the results from doubling/halving the measurement noise; panels (c) and (d) show the results from increasing and decreasing the representation error by ±50%; and panels (e) and (f) show the results from increasing and decreasing the principal diagonal r.m.s. errors from the *a priori* error covariance matrix by ±50%.

(the lower troposphere) the emphasis is shifted from the temperature retrieval to the ln(specific humidity) and surface pressure retrieval. The level of change of improvement for ln(specific humidity) is not quite so large, peaking at ±5%. The surface pressure improvement also varies by approximately ±5%.

Panels (e) and (f) shows that increasing or decreasing the r.m.s. errors associated with the *a priori* error covariance matrix increases or decreases the improvement of temperature, ln(specific humidity) throughout the vertical range of the observations as expected. Halving this source of error has the effect of decreasing the measurement 'weight' when the optimal estimate is computed, which corresponds to a small improvement in the background atmospheric knowledge after assimilating the measurements. Positive and negative temperature improvement is of the order of 15% throughout the range described by the measurement vector; the ln(specific humidity) improvement is of the order of ±10%; and the surface pressure improvement is of the order of ±8%.

This sensitivity study has looked at some of the extreme case scenarios in which the statistics assumed for the optimal estimation inverse method have been perturbed. It has been shown that the retrieval method is most sensitive to the background errors and forward modelling errors. The results from increasing the background error estimates are especially important to note since they represent a real possibility when dealing with any reasonable data in the southern hemisphere oceanic regions.

6.7 Chapter summary

The results shown in this chapter have successfully demonstrated a prototype optimal estimation inverse method for GNSS radio occultation observations.

The method is a non-linear, maximum likelihood technique which can accommodate most aspects of the real radio occultation problem. In particular, it is able to retrieve simultaneously temperature, water vapour and surface pressure, which is not possible using the Abelian inverse method. The optimal estimation inverse method is an iterative method but is found to be stable and to converge rapidly in most cases. Using this method a quality control scheme is implemented which allows anomalous retrieval conditions to be recognised and omitted, thus preventing gross retrieval errors.

The quality of the surface pressure retrieval is shown to depend on the vertical extent of the occultation, i.e. higher quality retrievals are attainable with occultations that reach

low altitudes. This is important to appreciate when validating surface pressure retrievals using real data.

In this thesis, a particular matrix has been used for the *a priori* error covariance. It has been shown that the retrieval accuracies, and hence the weight which should be given to the data in the subsequent model analysis, are sensitive to both *a priori* error uncertainties and values used to describe the representation error.

Chapter 7

Optimal estimation: validation

Chapter 6 dealt with the implementation of the optimal estimation inverse method, outlined procedures concerning quality control and convergence, and investigated the results from numerical simulations.

This chapter includes the second part of the overall validation: investigating how real radio occultation observations improve upon UKMO first-guess information, and comparing retrieved atmospheric profiles using real observations with correlative data from other atmospheric instruments or model analyses.

7.1 Description and preparation of the JPL GPS/MET dataset

In general, useful data from the GPS/MET receiver are found during 'prime-times' which correspond to time periods when the AS encryption (§2.2.2) of the GPS signal is disabled. Rocken et al. (1997) report four such periods since receiver launch (April, 1995) running through to February 1997.

The dataset used is from prime-times 1 (May 1995) and 2 (June, July and August 1995). There are ten days in the complete dataset used in this work: 25th April (**56**); 4th May (**53**); 5th May (**41**); 21st June (**15**); 23rd June (**52**); 27th June (**50**); 28th June (**87**); 29th June (**81**); 30th June (**71**); and 1st July (**68**), where the numbers in bold denote the number of occultation events associated with each day of the dataset. The distribution of this dataset is shown by figure 7.1.

The data format specification was written on the 27th March, 1996 by the JPL. The dataset is formatted into two levels: level 1 and level 2. The data resolution in these files

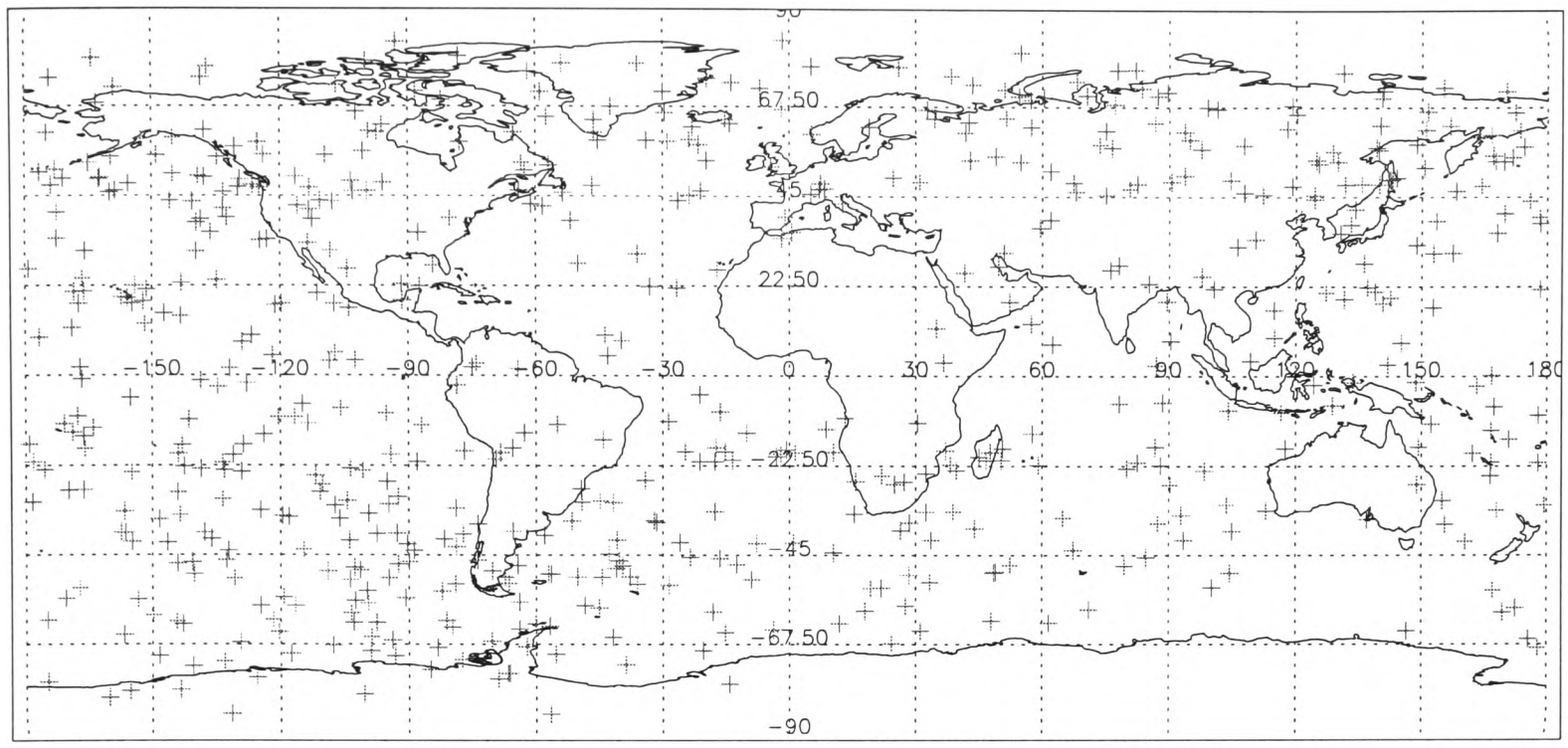


Figure 7.1: Distribution of occultation events described by the ten day JPL GPS/MET dataset. The symbols represent the geographical position of each occultation given by the average latitude and longitude.

is sufficient to ensure that round-off errors are at least an order of magnitude smaller than the anticipated precision of the measurement (JPL specification, 1995).

Level 1 data

The level 1 header includes information about the starting time of the occultation (epoch) and the local radius of curvature. The level 1 data includes information on the time since the epoch; the transmitter position and velocity vector (which are both interpolated in order to account for the transit time, approximated as the reception time less the vacuum light time between transmitter and receiver); the receiver position and velocity vector; the atmospheric phase shift[1], of both L1 and L2 signals (which are measured at a 50 Hz sampling rate as the ray path traverses the limb); the Doppler shift for both L1 and L2 signals; the bending angle and associated impact parameter[2], for both L1 and L2

[1]To reduce diffraction effects, these data are smoothed with time constants slightly less that the time required by the ray to traverse vertically the local Fresnel diameter at the limb. The phase shifts for L1 and L2 are obtained by sampling the smoothed curve at the time given by the first element of the record [from the JPL literature written to accompany the GPS/MET dataset].

[2]The bending angle and impact parameter for each wavelength are derived from the Doppler shift and satellite positions (§4.1). Spherical symmetry is assumed after locally fitting a sphere to an ellipsoid which approximates the Earth's shape at the latitude and longitude of the raypath's closest approach

signals; and the corrected values for the bending angle and impact parameter, after the ionospheric corrections (§4.2).

Level 2 data

The level 2 data are derived from level 1 information and includes details on altitude (which is defined as the height above a 36th-order mean sea level geoid); profiles of refractivity; pressure; (dry) temperature; specific humidity; assumed values of temperature and specific humidity from the ECMWF analyses, which represent ancillary data used for the retrieval of humidity and temperature, respectively; and the latitude and longitude of each tangent point.

7.1.1 Recomputing the ionospheric correction

The impetus for re-processing the bending angle data is the seemingly large number of 'corrected' bending angle profiles that did not match up with their associated L1 and L2 bending angle profiles. Furthermore, it was found that some occultation bending angle profiles were not monotonic in their structure, e.g. some profiles contained impact parameters that increased with decreasing heights. This problem has been traced back to receiver tracking software errors[3] on the GPS/MET receiver which have now been corrected [Dr. R. Kursinski: private communication]. Figure 7.2 shows an example of one such occultation during which the receiver lost signal lock. For this particular example there is also large values of noise in the upper stratosphere which render these observations 'uninvertable' if left unsmoothed, and shows where the bending angle profile becomes non-monotonic (both panels (a) and (b)). The subsequent ionosphere corrected bending angle profile (panel (c)) shows erroneous structure presumably from the L1 and L2 profiles.

---

to the ellipsoid. The impact parameter is measured with respect to the centre of the fitted sphere. The semi-major axis of the ellipsoidal approximation to the Earth is 6378.136 km and the flattening is 1/298.257 [from the JPL literature written to accompany the GPS/MET dataset].

[3]In early versions of the tracking software, if the receiver temporarily lost the radio signal the software would not terminate the occultation but continue for a few seconds using information from the internal clock until the signal was picked up again (§2.3.1)

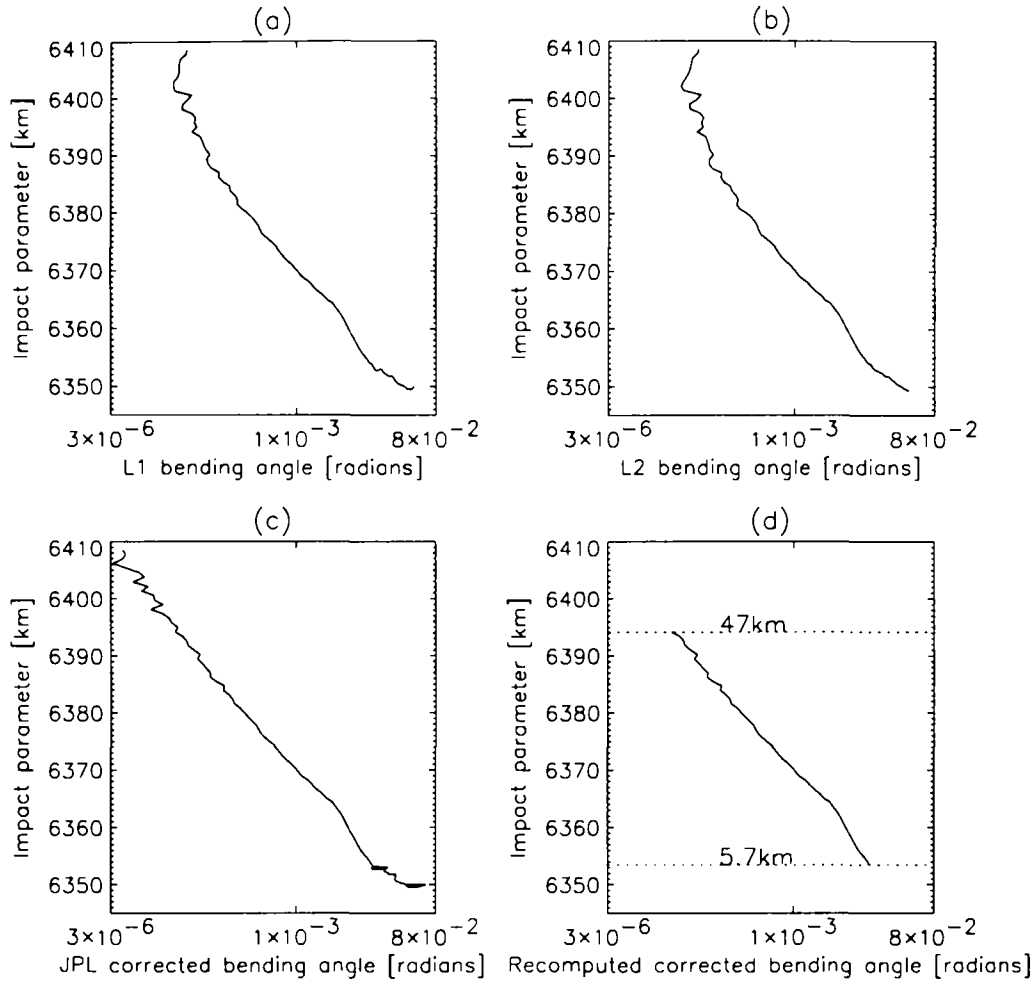


Figure 7.2: An example of an occultation (8.4°N, 145.2°E, 13:21 GMT, 1st July, 1995) which includes erroneous structure in the lower atmosphere from tracking software bugs and particularly noisy upper stratospheric observations. Panels (a) and (b) show the bending angle profiles from the L1 and L2 signals; panel (c) shows the JPL version of the ionospheric correction; and panel (d) shows the recomputed ionosphere-corrected bending angle profile. The broken lines shown in panel (d) indicate the approximate altitude range of the impact parameter values.

An attempt has been made in this work to correct for these errors by checking the L1 and L2 bending angle profiles (as a function of impact parameter) for monotonicity. Offending elements (i.e. noisy upper stratospheric observations and lower atmosphere observations) are omitted from the ionospheric correction. This form of correction is preferable to smoothing because smoothing could conceivably add erroneous structure to the profile. The ionospheric correction (outlined in §4.2) is applied to the possibly size-reduced L1 and L2 bending angle profiles. For the particular occultation shown by figure 7.2, the recomputed ionosphere-corrected profile is shown in panel (d). Although the number of measurements is reduced (in this example from 211 to 89) the resulting corrected profile is sensible and invertable. It should be noted that the example shown is

an extreme case in which a large number of observations are omitted from the correction; a typical percentage of omitted observations is 5−10%.

7.2 Computing optimal estimates from GPS/MET observations

Processing of the atmospheric phase delay from GPS/MET was carried out by the JPL to compute bending angle values as a function of impact parameter.

After the ionospheric correction was recomputed (§7.1.1), the data was used in the optimal estimation scheme (equation 4.42). The number of profiles which passed the ionospheric correction represents 97.6% of the total number of occultation events in the complete dataset.

Convergence criteria and quality control procedures (outlined in §6.2) were used to filter out profiles that did not converge or converged onto a solution with an unrealistic χ^2 value. These profiles are omitted from any subsequent statistical analyses.

Ninety seven profiles failed the defined quality control, which represents 17% and 21% of the total number of occultation events pre- and post-ionospheric correction, respectively. All the failed profiles were due to the solution converging on to unrealistically high χ^2 values, as opposed to non-convergence. The distribution of the number of failed profiles throughout the 10-day dataset is shown by figure 7.3.

Dataset days 5 and 6 show the largest quality control failure rates; in general, days 5−10 show a marked increase from days 1−4. This may be associated with receiver tracking software updates after the GPS/MET launch [Dr. R. Kursinski: private communication].

In an attempt to investigate whether a particular region is responsible for the failed profiles, the failed occultations are split further by virtue of their geographical positions. The sorting criteria for the occultations in this investigation include: land and sea; and North and South polar, tropical and mid-latitude regions.

First, occultations are split into land and sea events so that the effect of the local

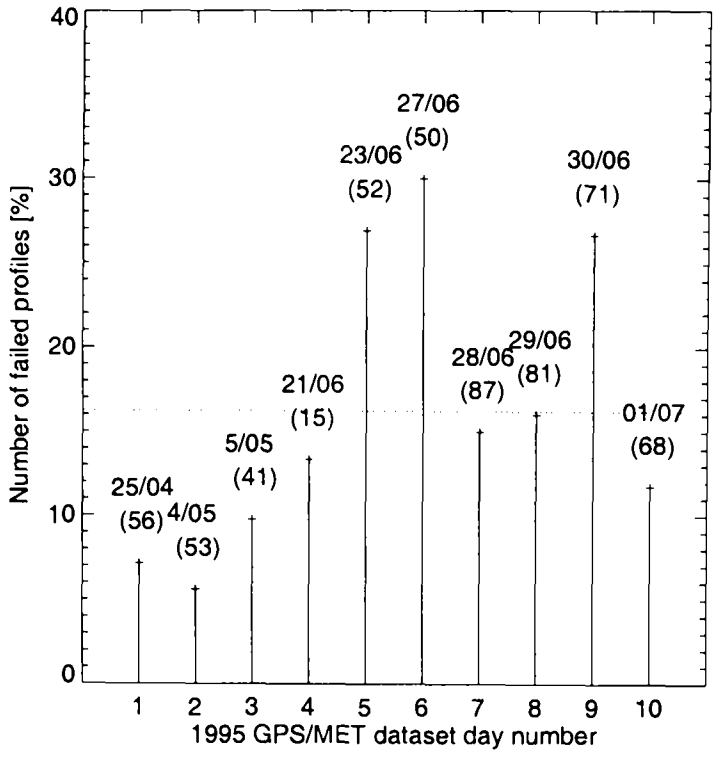


Figure 7.3: A histogram of the 1995 JPL GPS/MET dataset quality control summary. Each solid line represents the percentage (of the total number of retrievals, which is shown by the bracketed numbers) of failed profiles from each day of the dataset. The actual date of each dataset day is shown above the total number of profiles associated with each day. The broken line plotted represents the mean percentage failure rate.

radius of curvature values used can be assessed, i.e. the effect of assuming a local radius of curvature over orography. This separation is achieved by virtue of the surface elevation of the collocated UKMO NWP model analyses (used as the retrieval first-guess), i.e. if surface elevation is zero then the profile is assumed to be over the ocean. And second, occultations are split into latitude bands, i.e. North and South polar regions, mid-latitude and tropics, so that the performance of the occultation method within different regions can be investigated in detail. This separation is achieved using the mean latitude θ of each occultation. For this purpose, the North and South polar regions are defined as $\theta \geq 70°$ and $\theta \leq -70°$, respectively; the tropics region is defined as $|\theta| \leq 30°$; and the mid-latitude regions are defined between the polar region and the tropics.

Separating the occultations into their latitude bands shows the large disparity in the polar regions and the tropics. The defined tropics region has the largest quality control failure rate (32%). This large failure rate may be attributed to the large abundances of water vapour in this region and/or using first-guess estimates which are too different from the observations. From this particular separation the disparity between the North

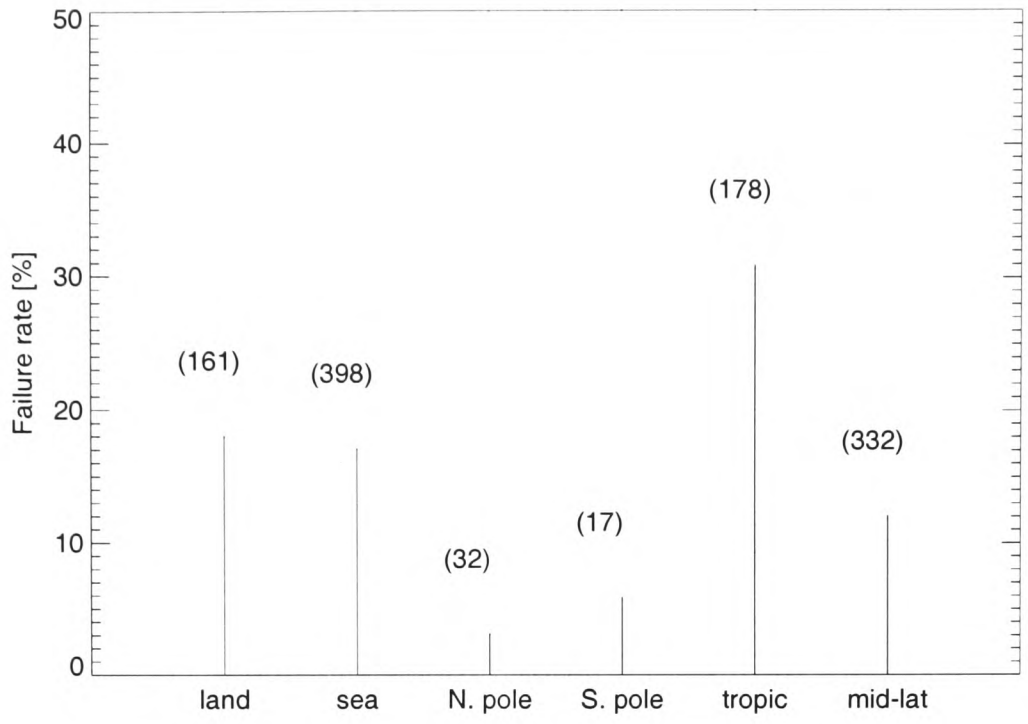


Figure 7.4: A histogram of the 1995 JPL GPS/MET dataset quality control summary based on geographical position. Each solid line represents the percentage of profiles that fail compared with the total number of profiles in that category (shown by the bracketed numbers).

and South polar occultations is also evident. The South polar region for the months described by the data is more dynamically active than the North pole; also, there are less other observations made in this polar region. No salient difference is shown for the statistics between land and sea occultations.

To investigate further why so many of the tropical occultations failed, figure 7.5 shows two tropical occultations which fail, and one which passes, the χ^2-based quality control.

Panels (a) and (b) from figure 7.5 show that the percentage difference between the observations and first-guess profile is, in general, much larger than the assumed measurement errors (§6.1.2); indeed the measurement vector at the nth iteration of the retrieval, i.e. at convergence, still falls slightly outside the measurement errors. Panel (c) shows an example occultation which does converge (in this case, within two iterations) and it can be seen that although the first-guess measurement vector is slightly larger than the measurement errors, the nth iteration measurement vector is within the errors assumed for the majority of the vertical range of the observations.

It is found that the majority of the retrievals which fail quality control in the tropics

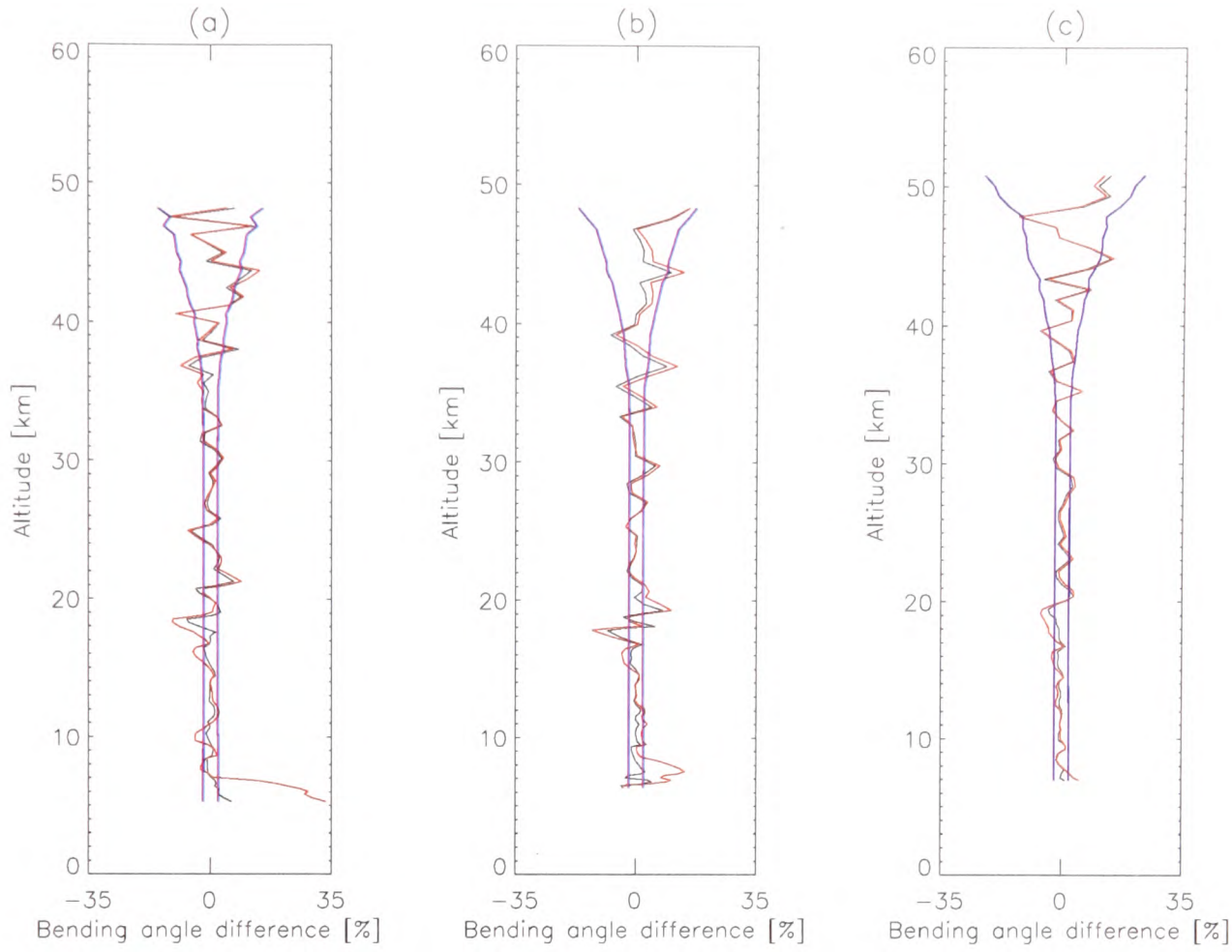


Figure 7.5: The percentage difference between the observation vector and simulated measurement vectors pre- and post-optimal estimation. Panels (a) and (b) show occultations which fail quality control. and panel (c) shows an occultation which passes quality control. Panel (a) shows an occultation at −2°N, 232°E (09:19 GMT. 25th May. 1995); panel (b) 5°N. 317°E (15:06 GMT, 23rd June, 1995); and panel (c) −14°N. 191°E (00:08 GMT. 25th May. 1995). The blue curves represent the r.m.s. values of the assumed measurement errors about zero mean. the red lines represent the differences between the first−guess and the black lines represent the differences after the nth retrieval iteration.

have large χ^2 values which derive from large first-guess−observation differences in the lower atmosphere (less than 15 km).

A possible reason for the failures is the variability of the horizontal refractivity structure errors in this region[1]. It has already been shown in the previous chapter this source of error is dominant throughout the lower atmosphere even for mid-latitudes, particularly due to horizontal water vapour variations. For this reason, the failed tropics occultations have been reprocessed using larger values for the horizontal refractivity structure component of the total measurement error covariance matrix.

Figure 7.6 shows that increasing the horizontal refractivity structure measurement

[1]Estimates of horizontal refractivity structure used were derived from a mid-latitudinal field (chapter 5), and as such only reflect typical errors in that region.

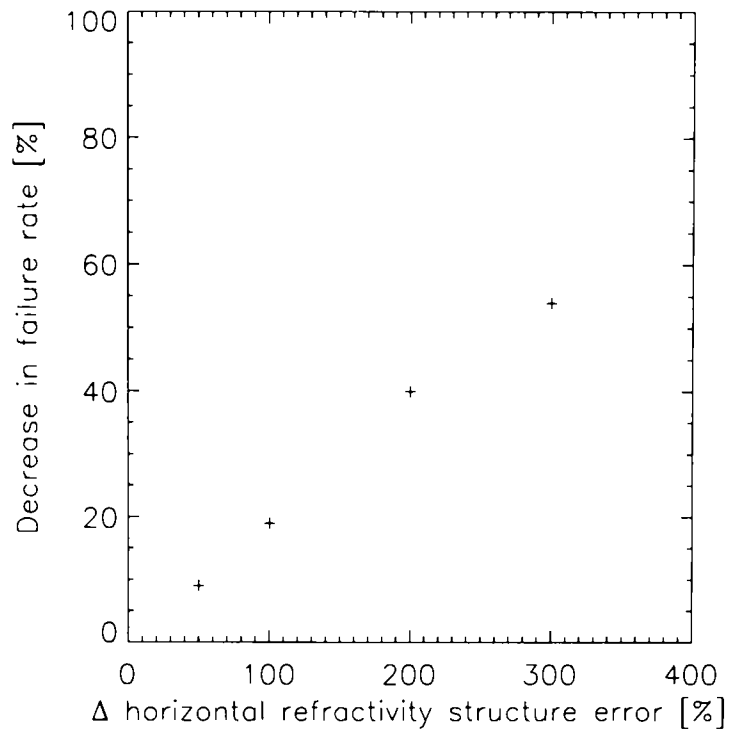


Figure 7.6: Improvement in the failure rate for tropical occultations by realistically increasing the error contribution from horizontal refractivity structure to the total measurement error estimate.

error estimates decreases the failure rate, i.e. smaller number of retrievals failing the quality control test; the improvement increases linearly over the range shown. It should be noted that results from increasing the total measurement error estimate are not surprising: increasing measurement error estimates should decrease the total χ^2 value of the retrieval. However, the horizontal refractivity structure component of the total measurement error covariance matrix represents the dominant term for altitudes less than 40 km; realistic perturbations to other measurement error sources do not alter significantly the total measurement error.

In summary, it is found for many failed tropical occultations, large observed−simulated bending angle values in the lower atmosphere (less than 15 km) are responsible for the large χ^2 values which subsequently fail the quality control. Realistic increases of the horizontal refractivity structure component of the total measurement error covariance reduce the number of failed profiles, but do not explain all failures. The main conclusion from this study of tropical occultations is that horizontal refractivity structure errors are important for retrievals using χ^2 quality control and therefore should be adjusted to account for realistic variations in refractivity along the line of sight for

different latitudes. The tropical occultations which have failed are omitted from the correlative statistics.

7.3 Description and preparation of correlative data

This section briefly describes the correlative data which is used to validate the optimal estimates computed using GPS/MET bending angle observations, and the methods used to obtain the collocation.

7.3.1 Radiosondes

The global corrected radiosonde data were originally obtained from the National Meteorological Center (NMC), via the JPL.

Radiosondes provide data from near-surface to altitudes of 20−30 km. Under ideal conditions and with careful calibration, radiosonde temperatures are accurate to within 0.5 kelvin and relative humidities are accurate to within a few tens of percent, except at very low temperatures or very high/low relative humidities (Luers and Eskridge 1995). Under operational conditions, relative humidity errors can be considerably greater than a few percent. Because of pressure errors and resulting height errors, temperature accuracy is degraded in the upper troposphere and stratosphere, with typical errors of one kelvin above 250 hPa (~10.5 km) possibly increasing to four kelvin at 10 hPa (~31 km) [Dr. J. J. Barnett: private communication].

The impetus for using radiosonde data to validate the GPS/MET radio occultation observations is that they possess high vertical resolution, resolving features with vertical scales of a few tens of metres. However, the distribution of radiosondes is sparse, especially in the southern hemisphere (figure 1.4), and the temporal coverage is not continuous with only 2 or 4 launches a day (nominally at 00:00, 06:00, 12:00 and 18:00 GMT)[5].

The nearest radiosonde to each collocation is first found by considering the nearest

[5]If a particular station only launches 2 radiosondes the nominal times will be at 00:00 and 12:00 GMT.

launch time[6]. Second, the nearest radiosonde in space is found using the mean latitude and longitude of the occultation and the geographical position of each radiosonde. For each collocation a Gaussian-type weight ζ is computed by considering the temporal and spatial difference ($t_{occ-sonde}$ and $d_{occ-sonde}$, respectively) between the occultation event and the nearest radiosonde thus

$$\zeta_d = \exp\left[-\left(\frac{d_{occ-sonde}}{c_d}\right)^2\right] \tag{7.1}$$

$$\zeta_t = \exp\left[-\left(\frac{t_{occ-sonde}}{c_t}\right)^2\right] \tag{7.2}$$

where the subscripts t and d represent time and distance, and c_d and c_t represent characteristic scales of distance and time. The values used for these parameters are 300 km and 3 hours, respectively. The former has been chosen because it represents the typical horizontal resolution of a limb-sounding observation, and the latter because 3 hours represents a large enough temporal window to guarantee 'capture' of at least some radiosonde profiles. A weight describing the collocation is given by taking the product of the temporal and spatial collocation weights. The distribution of these weights is shown by figure 7.7.

Panel (a) shows that the number of spatial collocation weights for values larger than 0.4 is, in general, smaller than the number of temporal collocation weights (panel (b)). The results shown in panels (a) and (b) imply that a reasonable number of spatial collocation will be the most difficult to achieve.

The distribution of the collocation weight, i.e. the product of the temporal and spatial collocation weights, is shown in panel (c). The number of profiles associated with each combined weight is comparable with the number of profiles of associated with each spatial weight, as expected.

These weights can be used to filter out collocated radiosondes that are too distant in time or space to expect good agreement. If a 'good' collocation is defined as 1.5 hours

[6]In general, if the nearest launch time for a particular occultation is midnight the next day and for some reason radiosonde data is unavailable the collocation procedure will use 18:00 GMT.

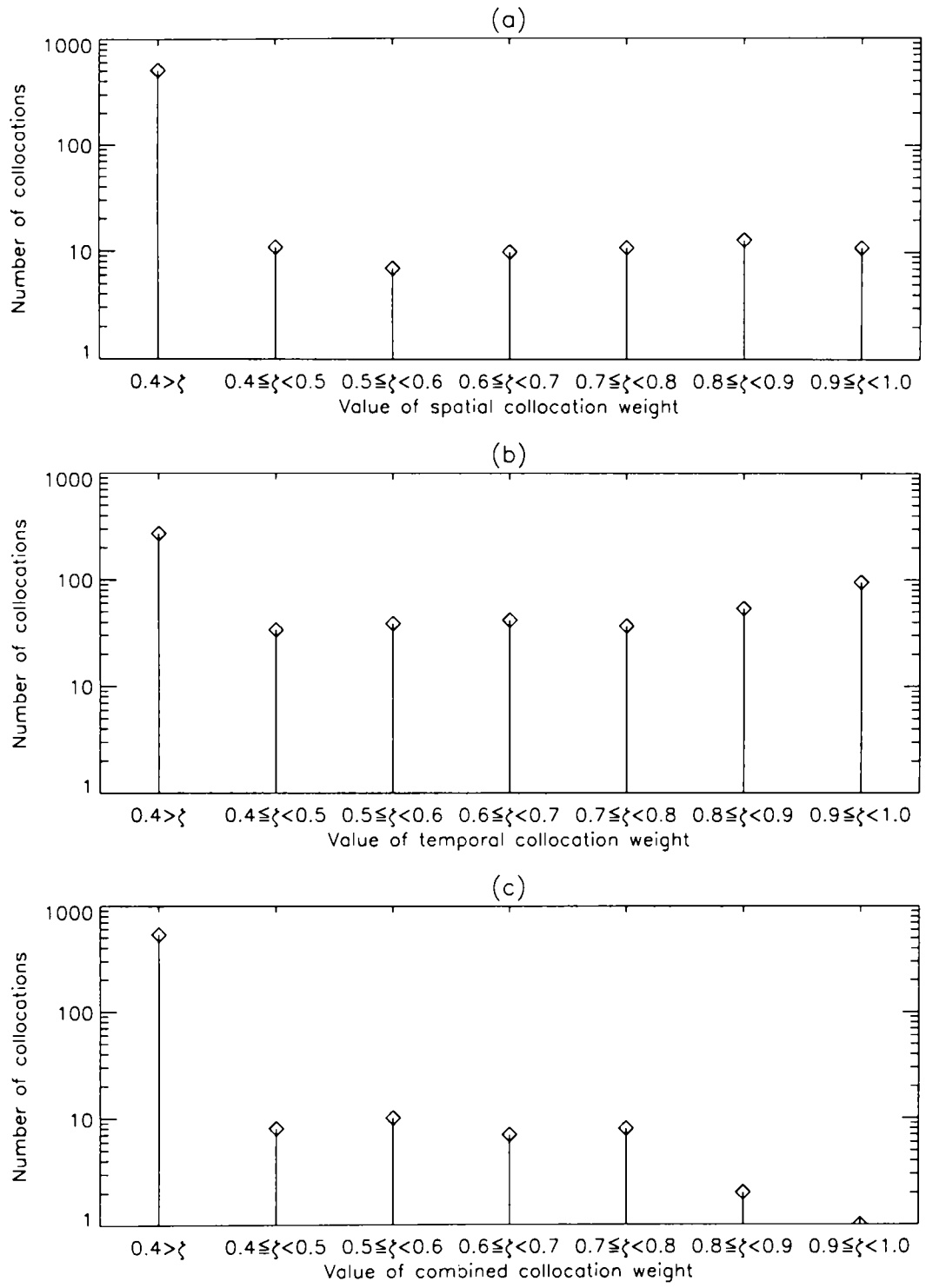


Figure 7.7: The distribution of weights given to radiosonde collocations. In particular, panel (a) shows the spatial collocation weight; panel (b) shows the temporal collocation weight; and panel (c) shows the combined collocation weight, computed using the spatial and temporal weights.

and 150 km from an occultation time (both corresponding to values of ζ equal to 0.78.), the complete collocation weight (i.e. considering time and space simultaneously) has the value of 0.6^7. The number of radiosonde collocations with values of ζ, using this definition of a reasonable collocation, is 18.

[7]For exact temporal or spatial collocations a combined collocation weight of 0.6 is equal to spatial and temporal differences of 214 km and approximately 2 hours, respectively.

7.3.2 NWP model analyses

Since closely correlative radiosonde profiles are rare, output from NWP models is also used to obtain validation statistics.

It should be noted that NWP models use atmospheric observations, e.g. radiosondes, to initialise model fields, therefore any NWP model output should not disagree too much with reasonably collocated radiosondes. Although the majority of NWP centres (e.g. the UKMO and the ECMWF) use the same set of global observations, e.g. TOVS, different operational NWP systems implement different assimilation methods and dynamical representations, and thus profiles from different NWP models should provide profiles which are independent to some extent. This point is particularly important since the first-guess for the retrieval is from collocated UKMO model data.

The NMC NWP model data used has been obtained from a set of data originally supplied as correlative data for the UARS project (Reber and Luther 1993). The NMC temperatures are derived from two analysis systems: first, tropospheric fields from 1000−100 hPa; and second stratospheric analyses from 70−0.4 hPa. The tropospheric fields from 1000−100 hPa are the 12:00 GMT gridded fields which are part of the Global Daily Assimilation System described by McPherson et al. (1979). The gridded fields are on the standard 65×65 NMC polar stereo-graphic grid oriented 80°W and 100°E. Temperature values are produced at 18 standard meteorological levels[8] from 1000−0.4 hPa for heights and temperatures. For relative humidity only the 6 lowest altitude pressure levels are used and there are no data for the southern hemisphere[9].

The collocation was computed by considering the mean latitude and longitude of each occultation. Prior to comparison with other data, the temperature values were adjusted with NMC stratospheric corrections published by Gelman et al. (1994).

The ECMWF data used has 31 vertical levels from the surface to ∼10 hPa. This data has a resolution of 0.5° and 1.0° in latitude and longitude, respectively. The collocated profiles take full account of the latitude and longitude variation throughout the

[8]18 Standard pressure levels for data (hPa): 1000., 850., 700., 500., 400., 300., 250., 200., 150., 100., 70., 50., 30., 10., 5., 2., 1., 0.4

[9]Refractivity statistics shown for this region do not include these data.

occultation. The surface pressure and surface elevation values are computed considering the mean latitude and longitude of the occultation.

7.3.3 Abelian inversion results

The Abelian inversion of the radio occultation observations represents the conventional method of retrieving profiles of refractivity, and subsequently profiles of temperature or specific humidity (method outlined in §4.3). This is a useful source of comparative data since results from this method have already been compared against independent sources of atmospheric information (Rocken et al. 1997). It should be noted that this source of comparison is not independent since both the optimal estimation and Abelian inverse methods used the same profile of bending angle observations. However, these two inverse methods should be independent enough for results from the Abelian method to be used for comparison purposes.

Because many of the bending angle observation vectors in the dataset do not reach the altitude of 50 km (either because the occultation terminated at a lower altitude or because the monotonicity test has omitted particularly noisy values in the upper stratosphere), the hydrostatic relation (used by the Abelian inversion) is initialised at an altitude comparable with the top of the ECMWF model, i.e. 30 km, thus providing a reasonable first-guess. The ancillary data used to retrieve 'wet' temperature, i.e. accounting for the presence of water vapour in the lower atmosphere, are collocated ECMWF analyses. Values of ln(specific humidity) (from partial water vapour pressure) are also retrieved using this method but are restricted to the tropics so that reasonably large values of water vapour are guaranteed (§4.3.2). Because errors may be introduced by the first-guess used for the hydrostatic relation (§5.2.2), refractivity profiles are also shown in the comparison statistics.

7.4 Validation

The statistical comparison shown in this section are based on the ten-day dataset (described in §7.1) post ionospheric correction. First, theoretical estimates of the

retrieval errors are compared with the assumed first-guess errors; second, the optimal estimates are compared with several sources of correlative data; and third, particular occultations are shown in detail with correlative data.

To validate the computed optimal estimates, several different sources of independent atmospheric information are used (outlined in §7.3). Although the optimal estimate has been computed from 1000−0.1 hPa (§6.1.1) (reflecting the vertical range of occultation observations), the validation statistics are shown from 1000−10 hPa, reflecting the vertical range described by model analyses and radiosondes.

Only optimal estimates that passed the quality control (§6.2) have been included in the statistics shown in this section.

7.4.1 Theoretical retrieval r.m.s. errors

Retrieval r.m.s. errors taken from the square-root values of the principal diagonal of the solution covariance (equation 4.51) are compared directly with the assumed *a priori* errors so that the impact of the optimal estimation retrieval method on the UKMO first-guess can be quantified. In addition, the associated improvement vector, i.e. the percentage error improvement, (§6.4) is shown for completeness.

Figure 7.8 (panel (a)) shows that the typical decrease of temperature error estimate is 1−2 kelvin in the range described by UKMO model data (i.e. surface−10 hPa). Above 10 hPa, where the first-guess represents climatology, there is a relatively large reduction over a pressure range typical of the stratopause (approximately 1.0 hPa) which decreases towards 0.1 hPa, reflecting the vertical range of a large number of occultations used in the dataset. Panel (b) shows that the largest reduction of ln(specific humidity) error is between 300−700 hPa with an approximate value of 0.15. Below 700 hPa, the improvement decreases reflecting where the majority of occultations terminated. The temperature improvement vector (panel (c)) shows the largest values between 10−300 hPa; above which the improvement slowly declines as the measurement error increases, and below which the improvement decreases as ln(specific humidity) errors become favourable. The ln(specific humidity) improvement vector (panel (d)) values

increase from 5% at 300 hPa to a typical value of 20%, decreasing from 700 hPa to a value of 3% near the surface.

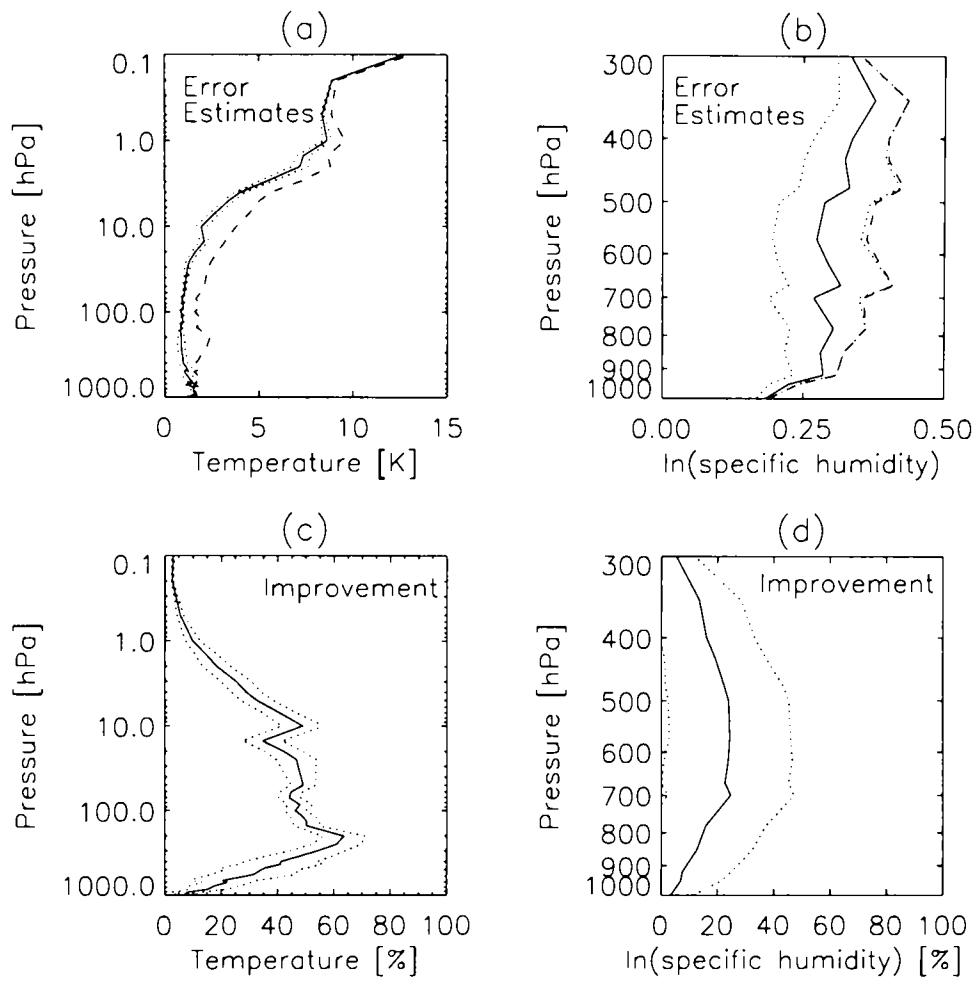


Figure 7.8: Error statistics between the optimal estimation retrieval and its first-guess. Panels (a) and (b) show the mean and mean$\pm\sigma$ values for the square-rooted principal diagonal of the solution error covariance matrix (denoted by solid and dotted lines, respectively) for temperature and ln(specific humidity), respectively; and panels (c) and (d) show the associated mean and mean$\pm\sigma$ values for the improvement vectors for (a) and (b). Dashed lines represent the square-rooted principal diagonal of the *a priori* error covariance matrix.

In general, results from figure 7.8 show that optimal estimation has a significant impact on the first-guess throughout the retrieved vertical range, with a distinct overlap region at 300 hPa where the ln(specific humidity) retrieval becomes more important than the temperature retrieval. The size of the mean$\pm\sigma$ values for the global improvement statistics indicate the large variability of the mean values, and as such it is necessary to separate the statistics for further investigation of the retrievals.

Figure 7.9 shows results separated by latitude bands. Statistics from the polar regions (panels (a)$-$(h)) show in general that the regions where there are useful reductions in temperature errors (and useful values of improvement vector) reach lower altitudes than shown by the global statistics. This is because of the small amount of water vapour

in the atmosphere, allowing the temperature component of the simultaneous retrieval to be taken down nearer to the surface. In particular, the South polar region statistics show the temperature improvement values are greater than 50% down to approximately 600 hPa; the large temperature improvement vector mean$\pm\sigma$ values for pressures larger than 300 hPa imply that the temperature improvement has a large associated variability in this region. The vertical extent of the ln(specific humidity) improvement is indicative of the cold, dry atmosphere in this region.

The statistics for the tropics region (panels (i)$-$(l)) show that the largest ln(specific humidity) error estimate reduction is between 300$-$700 hPa. The decline in ln(specific humidity) improvement below 700 hPa is a consequence of the vertical extent of occultations in this region, i.e. in general, the large abundance of water vapour at low altitudes reduces the number of radio occultation observations near the surface in this dataset. Mid-latitude statistics display similar structure to the global statistics.

The statistics for surface pressure are shown by table 7.1. As shown in the previous chapter the level of improvement in this quantity is expected to be directly related to how near to the surface the occultation reaches. Because of this the statistics are split into two sub-tables: first, showing statistics for all occultations that pass the quality control; and second, showing statistics for occultations that reach at least a pressure of 850 hPa.

In general, results from table 7.1 suggest that occultations which reach altitudes lower than 850 hPa display larger values for the mean (typically 0.65 hPa) with smaller r.m.s. values (typically 0.16 hPa). The greatest impact of the optimal estimation method is found to be in the South polar region with a reduction of 0.86 hPa in surface pressure error, and the smallest improvement is found to be the North polar region. However, the surface pressure retrieval has been shown to be dependent on the lowest retrieved geometric height level, and because such a small statistical sample is used, it is not possible to comment fully on the relevance of the comparative values. The improvement vector values are typically 29% with a standard deviation of typically 6.5%.

In general for figures 7.8$-$7.9, the reductions in temperature errors are largest near

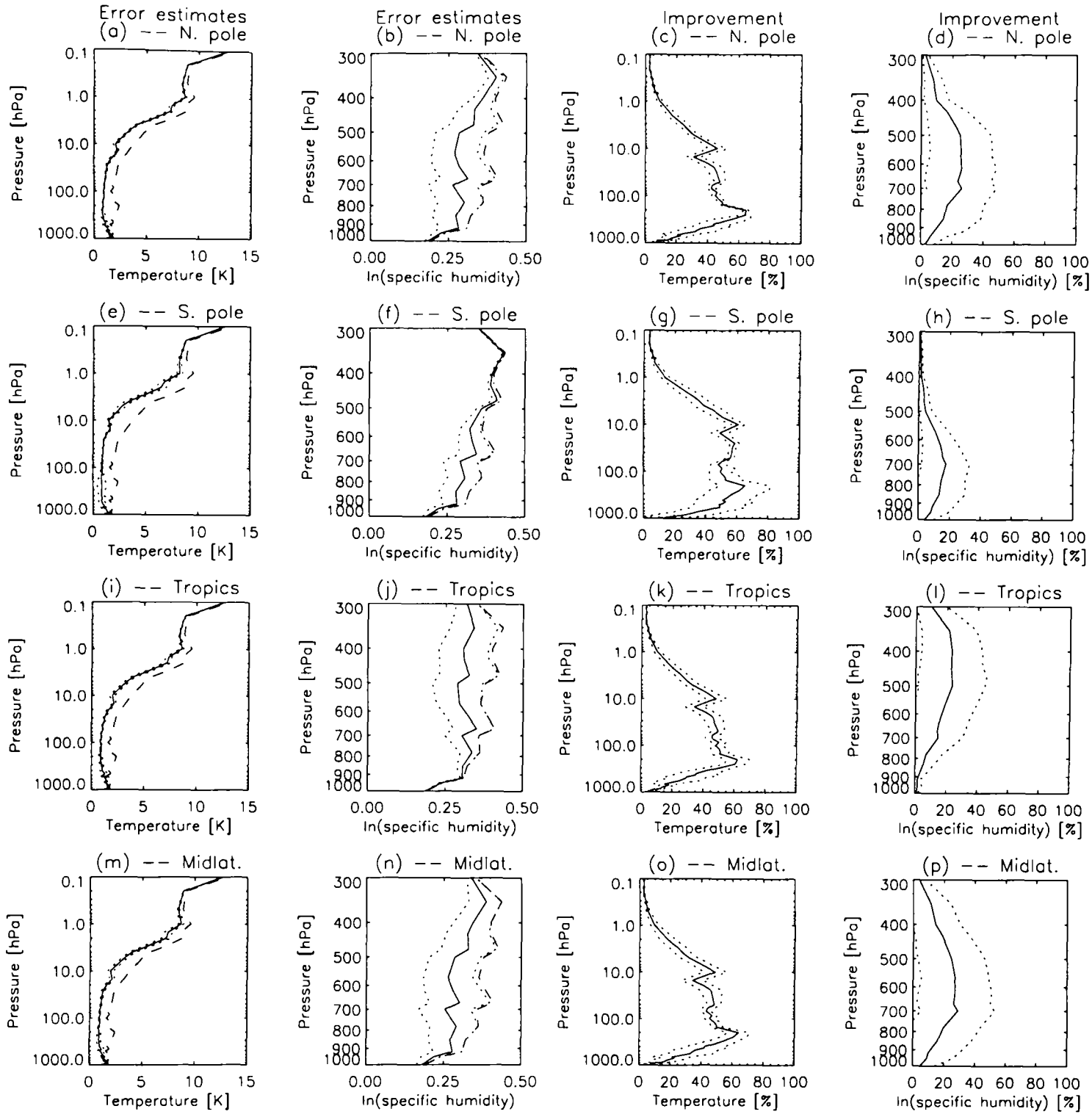


Figure 7.9: Error statistics between the optimal estimation retrieval and its first-guess shown for different latitude θ bands. The North and South polar regions are defined as $\theta \geq 70°$ and $\theta \leq -70°$, respectively; the tropics is defined as $|\theta| \leq 30°$; and the mid-latitude regions are defined to be in between the polar regions and the tropics. Left-hand side panels show the mean and mean$\pm\sigma$ values for the square-root principal diagonal of the solution error covariance matrix (denoted by solid and dotted lines, respectively) for temperature and ln(specific humidity), respectively; and right-hand side panel show the associated mean and mean$\pm\sigma$ values for the improvement vector for the left-hand side panel. Dashed lines represent the square-rooted principal diagonal of the *a priori* error covariance matrix.

tropopause pressure levels where the UKMO model is found to be biased against the ECMWF model (§7.4.5); and ln(specific humidity) error reductions are largest in the tropics (but decrease below 700 hPa due to the termination of a large number of tropical

All occultations	Reduction in error [hPa]		Improvement [%]		Number of
	mean	σ	mean	σ	profiles
Global	0.54	0.19	21.63	7.50	444
Land	0.60	0.22	23.98	9.00	129
Sea	0.52	0.16	20.65	6.56	315
N. polar region	0.53	0.14	21.12	5.67	29
S. polar region	0.92	0.25	36.65	10.46	13
Mid-latitudes	0.55	0.19	22.14	7.45	282
Tropics	0.47	0.13	18.91	5.15	120

Occultations that reach $P \geq 850$hPa	Reduction in error [hPa]		Improvement [%]		Number of
	mean	σ	mean	σ	profiles
Global	0.73	0.16	29.36	6.61	46
Land	0.67	0.16	26.94	6.90	10
Sea	0.74	0.16	29.99	6.49	36
N. polar region	0.63	0.14	25.13	6.54	3
S. polar region	0.86	0.14	34.22	6.66	4
Mid-latitudes	0.73	0.16	29.19	6.48	39
Tropics	—	—	—	—	—

Table 7.1: Mean and r.m.s. statistics of the difference between the surface pressure first-guess and retrieval errors. The upper table represents results from considering all occultations which passed quality control; the lower table represents occultations whose lowest pressure is larger or equal to 850 hPa.

occultations) where large abundances of this quantity are present, as expected. In particular, these statistics suggest that the South polar region has the largest reduction in temperature error which could reflect the atmospheric state of this region during the days of the dataset. Results from the surface pressure retrieval (for occultations which reach pressures of at least 850 hPa) show a typical improvement of 29%±6% over all regions which corresponds to error estimates decreasing by typically 0.65 hPa.

From the statistics shown in this section, the optimal estimation inverse method has a large impact on first-guess information at all latitudes.

7.4.2 Global statistics

The global statistics shown by figure 7.10 describe all the occultations compared with other correlative data. There are several points to be made from this figure.

First, the temperature comparisons (panels (a)–(d)) show, in general, better agreement with the optimal estimate than the UKMO model statistics, i.e. less bias and smaller r.m.s. values. The Abelian temperature comparison shows good agreement with the optimal estimate, except for a cold bias above 50 hPa (shown in both (a) and (c)) which derives from using the ECMWF model analyses to initialise the hydrostatic relation[10].

Although the comparison between the optimal estimate and the collocated ECMWF and NMC model analyses and radiosondes shows less bias throughout the vertical range than the first-guess comparison, the standard deviation values from the optimal estimate comparison are larger than the first-guess differences above 50 hPa.

Second, the ln(specific humidity) comparisons (panels (e)–(h)) show the optimal estimate is closer to the ECMWF analyses than the first-guess information, i.e. less biased with a smaller standard deviation. This implies that the ECMWF made better use of TOVS humidity information during this period [Dr. J. R. Eyre: private communication]. The Abelian inversion results show a negative bias between optimal estimation, UKMO model analyses and other correlative data, but is comparable with other inter-correlative statistics in this region (appendix D). The NMC and radiosonde comparison show no marked differences.

Third, the refractivity[11] comparison (panels (i)–(l)) shows excellent agreement between the Abelian inversion and the optimal estimate, with mean differences of

[10]The difference between the UKMO and ECMWF model analyses at 10 hPa may be due to a bias at the top of the UKMO unified model at the time of the occultation data set [Mr. S. Milton: private communication]. It was found that accounting for latitude and longitude variation throughout a single occultation does not change the comparison statistics significantly.

[11]Computed using equation 3.1.

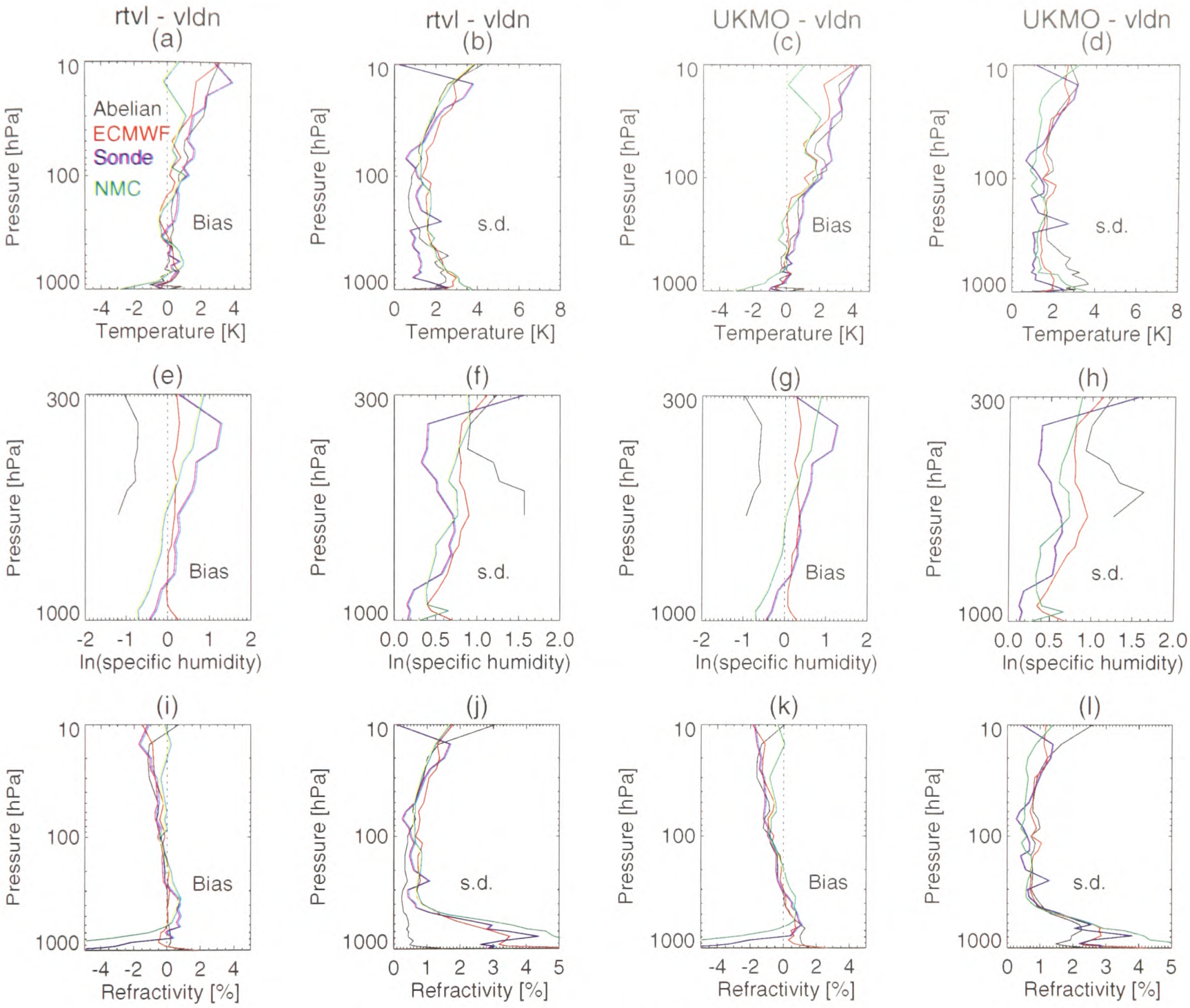


Figure 7.10: Mean bias and standard deviation global statistics between optimal estimation results and correlative ECMWF and NMC analyses, radiosondes and results from the Abelian inversion; and UKMO model information and the same correlative data. Panels (a)−(d) show the temperature comparisons; panels (e)−(h) show the ln(specific humidity) comparisons; and panels (i)−(l) show the refractivity comparisons. The header of each panel indicates the subject plotted: rtvl, vldn and UKMO represent comparison with the optimal estimation retrieval, validation (i.e. correlative data) and UKMO model data. A positive bias is where the correlative data is smaller.

typically much less than 1% (with a standard deviation of typically less than 1%) throughout the complete vertical region. The equivalent statistics for the UKMO model analyses show a marked negative bias above 300 hPa of approximately 1%, with a standard deviation of more than this amount throughout the vertical range. In general, although optimal estimate comparisons show less bias between 300−1000 hPa, strong

positive and negative biases are due to differing water vapour distributions. The largest bias values of NMC model analyses and radiosondes below 300 hPa are consistent with the temperature and ln(specific humidity) bias values in this region. The comparison between optimal estimation and the ECMWF data shows a small negative bias (0.5%) below 500 hPa, as opposed to a small positive bias in the first-guess comparison.

Overall, these global statistics show excellent agreement between the optimal estimation results and the Abelian inversion results for temperature (above levels of water vapour and below 50 hPa) and refractivity. The comparison between optimal estimation and the ECMWF ln(specific humidity) shows a smaller bias throughout the vertical region than the UKMO model analyses, which is consistent with the known differences between the UKMO and the ECMWF models in 1995. The large difference shown for the ln(specific humidity) comparison are comparable with inter-correlative data shown in appendix D.

Occultations are now separated in to their latitude bands, similar to figure 7.9, to investigate the comparison statistics further.

7.4.3 Latitude bands

In general, optimal estimate temperature comparisons in the North polar region (figure 7.11, panels (a)−(d)) show less bias than the UKMO model analyses. In particular, the Abelian inversion results show excellent agreement with a bias value of typically less than one kelvin with a standard deviation of similar magnitude between the surface and 50 hPa. The radiosonde comparison for this polar region displays a bias of typically one kelvin with a standard deviation of similar magnitude. Again, the ECMWF and the NMC model analyses and radiosondes have less bias when compared with the optimal estimate, but associated standard deviations are slightly larger.

The South polar region (figure 7.11, panels (e)−(h)) shows only results from the Abelian retrieval and collocated ECMWF analyses, reflecting the low number of radiosondes in this region[12]. In this polar region, the bias and standard deviation between

[12]NMC analysis is not shown because there are no South polar occultations within ±3 hours of midday

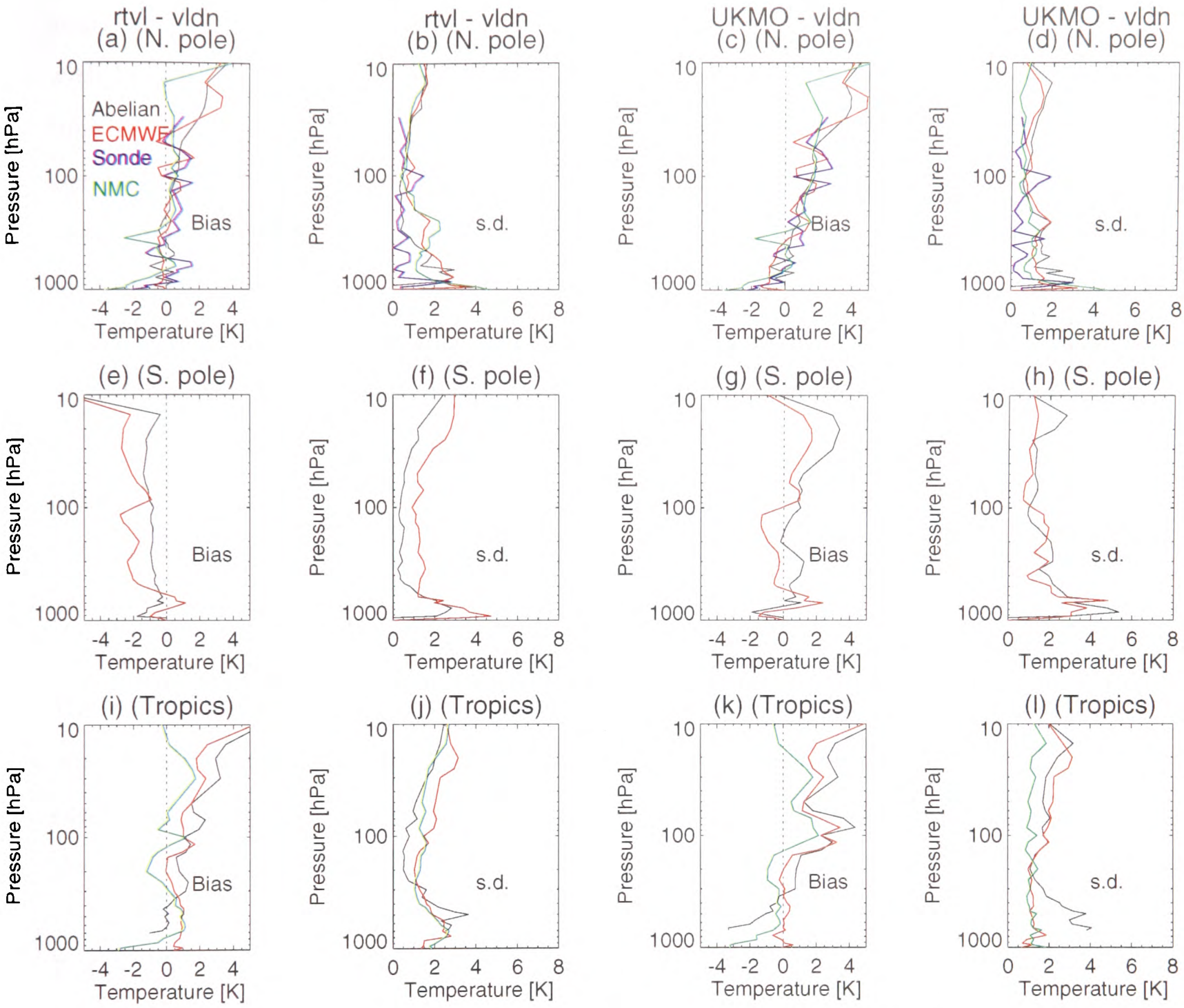


Figure 7.11: Mean bias and standard deviation temperature statistics between optimal estimation results and correlative ECMWF and NMC analyses, radiosondes and results from the Abelian inversion; and UKMO model information and the same correlative data for different latitude bands. The latitude bands used are as figure 7.9. The subject of each panel is denoted by the column and panel header. A positive bias is where the correlative data is smaller.

the Abelian retrieval and the optimal estimate is smaller than the UKMO model analyses as with the North polar region; however, bias and standard deviation values between the ECMWF analyses and optimal estimate are larger for all altitudes above 50 hPa. This implies that the UKMO and the ECMWF model analyses are similar in this geographical region at these altitudes.

The tropical region comparison (figure 7.11, panels (i)−(l)) shows that the optimal estimate is less biased when compared with the correlative data than the UKMO model

analyses, particularly at 100 hPa (typical tropopause height level); which is consistent with known differences between the UKMO and the ECMWF model analyses at the time of the dataset [Mr. S. Milton]. The warm bias shown by the Abelian inversion against the optimal estimate and other correlative data in the lower atmosphere is due to the inadequacy of the water vapour correction in regions with a high water vapour concentration (§4.3.2).

The mid-latitude statistics display very similar results to the global differences and as such are not plotted again.

In general, ln(specific humidity) comparisons in the North polar region (figure 7.12, panels (a)−(d)) show that the ECMWF correlative data exhibits less bias against the optimal estimate than the UKMO model analyses. Other correlative data do not show any marked difference.

The refractivity statistics for the defined latitude bands are shown by figure 7.13. The statistics for this quantity reflect the temperature statistics above 300 hPa and ln(specific humidity) statistics below 300 hPa. In particular, the optimal estimate shows excellent agreement (with a bias and standard deviation of typically less than 1%) with the Abelian inversion results for each latitude band shown.

7.4.4 Surface pressure validation

The first-guess for each retrieval is a UKMO model profile which is collocated to the mean latitude and longitude of each occultation in the GPS/MET dataset. The *a priori* constraint used for the retrieval allows the surface pressure solution to vary from the first-guess by typically 2.5 hPa which means if there is significant latitude and longitude variation, i.e. enough for the first-guess and truth to be different by an amount much larger than the assumed error, then the retrieval will only be able to approach the true value. Only occultations which reach pressure of at least 850 hPa are considered in table 7.2.

For this validation, the ECMWF model values are assumed to be the truth. In general, results shown by table 7.2 show that values for the mean difference between the

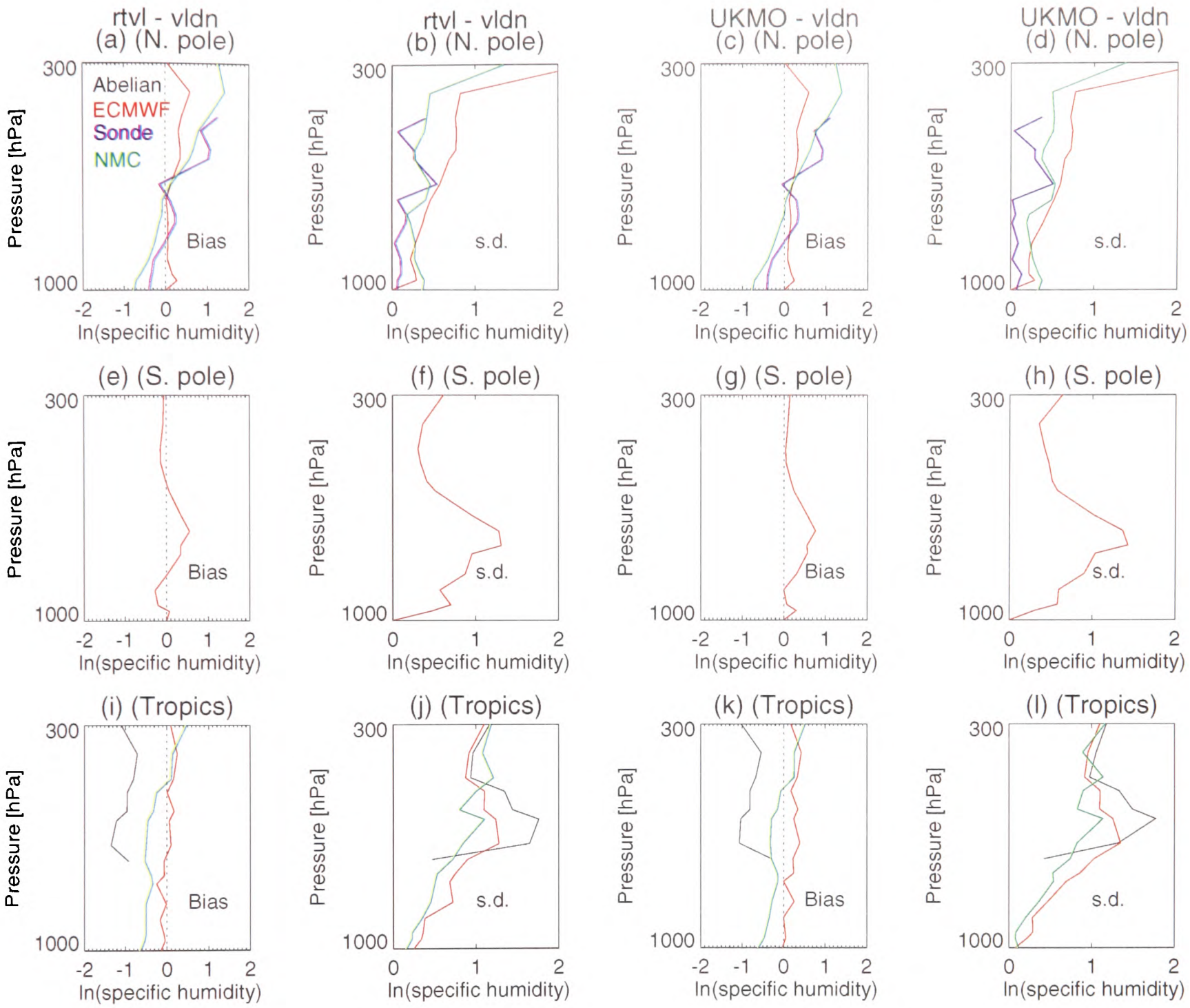


Figure 7.12: Mean bias and standard deviation ln(specific humidity) statistics between optimal estimation results and correlative ECMWF and NMC analyses, radiosondes and results from the Abelian inversion: and UKMO model information and the same correlative data for different latitude bands. The latitude bands used are as figure 7.9. The subject of each panel is denoted by the column and panel header. A positive bias is where the correlative data is smaller.

retrieval and the ECMWF collocated data are smaller than the difference between the UKMO and the ECMWF but have larger associated standard deviation values. However, the differences between optimal estimation and the ECMWF are comparable with the two model analyses (see §7.4.5).

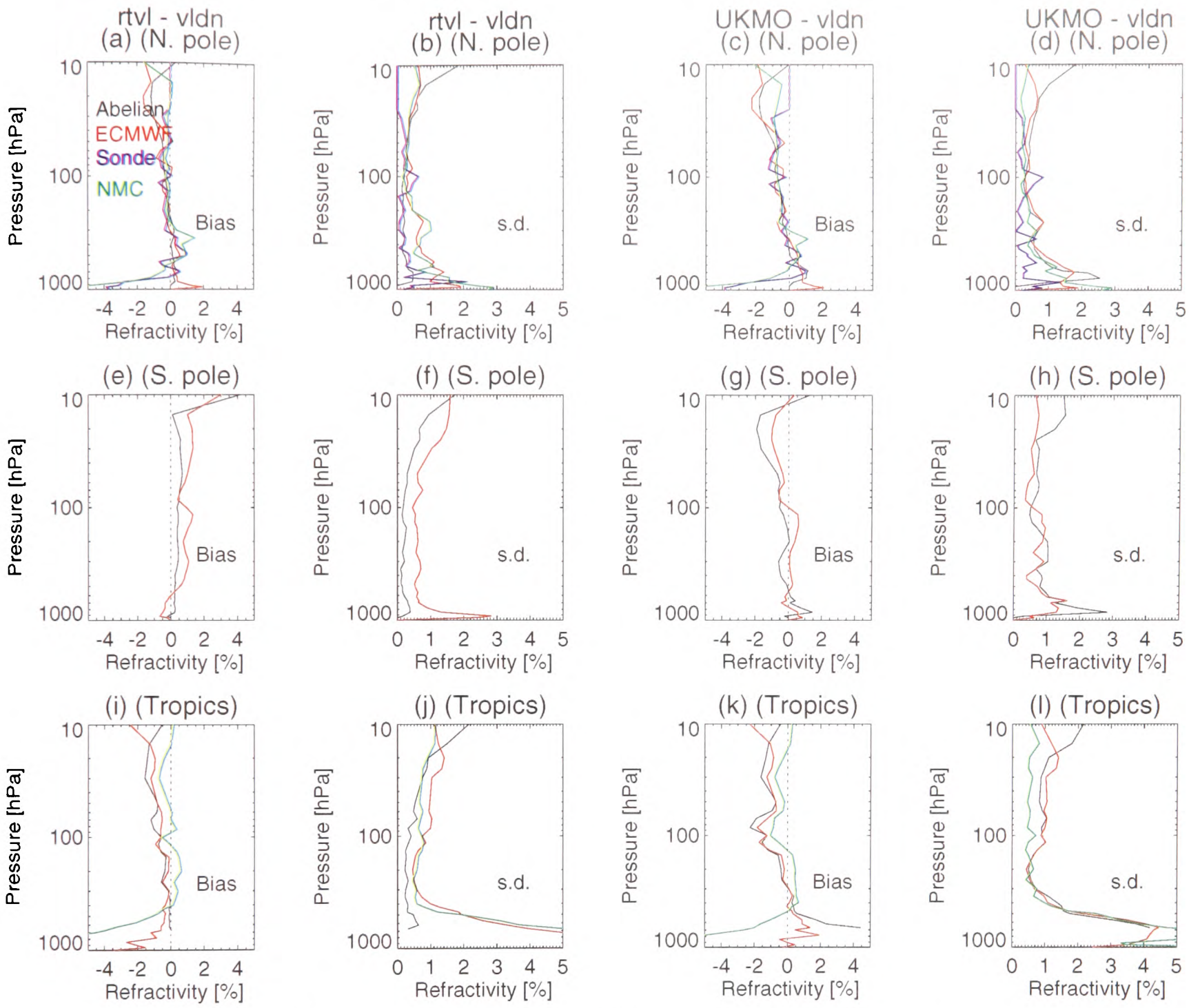


Figure 7.13: Mean bias and standard deviation refractivity statistics between optimal estimation results and correlative ECMWF and NMC analyses, radiosondes and results from the Abelian inversion; and UKMO model information and the same correlative data for different latitude bands. The latitude bands used are as figure 7.9. The subject of each panel is denoted by the column and panel header. A positive bias is where the correlative data is smaller.

7.4.5 UKMO–ECMWF model differences in 1995

As shown by figures 7.10–7.13 and table 7.2, there are biases of more than one kelvin between the UKMO and the ECMWF model analyses.

In general, the differences between the two model analyses shown by the statistics are consistent with the known biases between the models at the time of the dataset [Mr. S. Milton: private communication]. In particular, the disparity between the

Occultation that reach $P \geq 850$hPa	Retrieval – ECMWF [hPa]		UKMO – ECMWF [hPa]		Retrieval – UKMO [hPa]	
	mean	σ	mean	σ	mean	σ
Global	-0.26	5.21	-1.15	3.63	-1.41	4.24
Land	-0.06	2.99	-1.71	2.54	-1.78	1.30
Sea	-0.33	5.68	-0.99	3.86	-1.32	4.74
N. polar region	-1.25	3.43	-1.40	2.02	0.15	1.63
S. polar region	0.21	5.18	-4.64	5.23	-4.43	1.11
Mid-latitudes	-0.24	5.32	-0.98	3.25	-1.23	4.44
Tropics	—	—	—	—	—	—

Table 7.2: Mean bias and standard deviation surface pressure statistics. The number of profiles considered for each region are the same as table 7.1.

UKMO and the ECMWF model analysis largely reflects the difference in the model dynamics used in the two models rather than the differences in the assimilation systems [Mr. S. Milton: private communication].

It should be noted that these biases have been reduced in the UKMO unified model since 1995. It has been found that the biases are sensitive to the model vertical resolution. In January 1998, the UKMO increased the number of levels in the unified model from 19 to 30, with a concentration of levels about the tropopause. This has since improved both the temperature and humidity bias at the tropopause and in the stratosphere [Mr. S. Milton: private communication].

7.4.6 The importance of good temporal collocation

In general, correlative data represents information from nominal analyses times, such that any occultation should not be more than three hours from any analysis time. As mentioned in §7.3 the collocation process will always pick the nearest available analysis time; in some circumstances this may mean that the temporal collocation is greater than six hours, which will lead to increased biases and standard deviation values. Omitting

temporal collocations of more than three hours does not noticeably change the collocation statistics and are therefore not shown. This result is expected since temporal collocations of more than three hours represent a small fraction of the total collocations.

The importance of the temporal collocation is investigated by considering the retrievals between 27th–30th June[13] and comparing them with the ECMWF correlative data for the best temporal collocation and correlative data which are 6, 12 and 24 hours from the best temporal collocation.

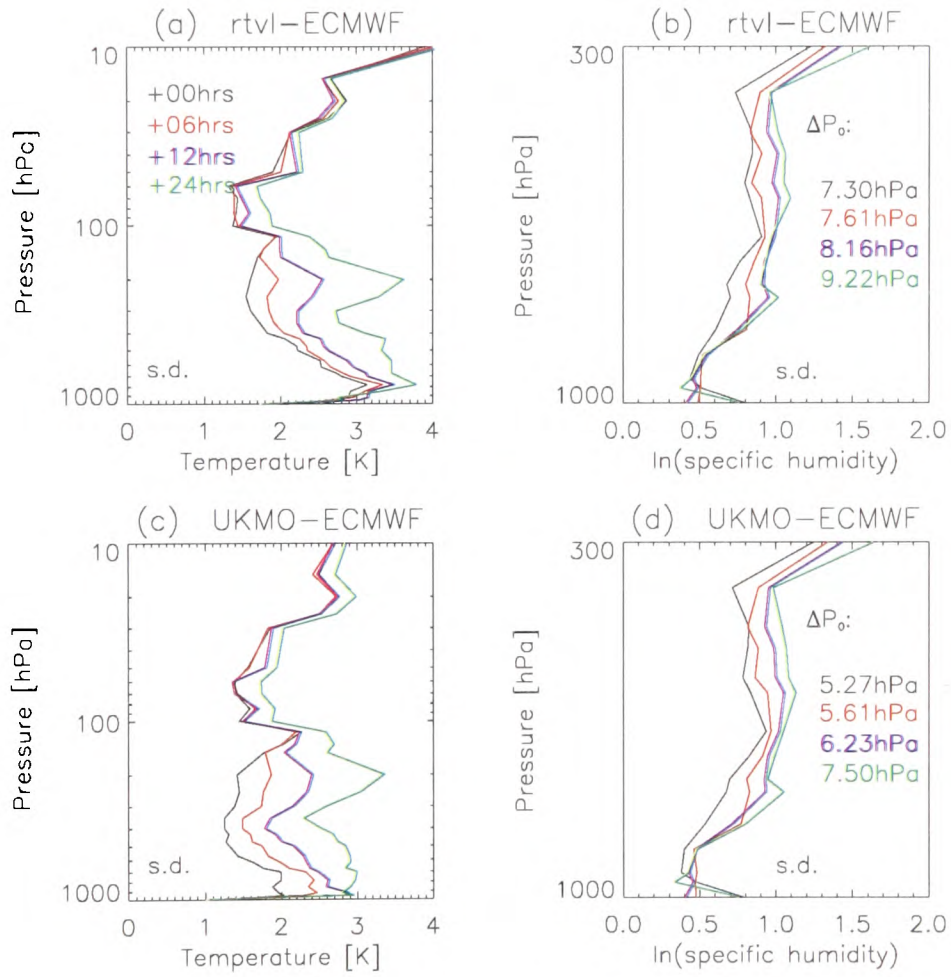


Figure 7.14: Results from investigating the importance of good temporal collocation. Panels (a) and (b) show the standard deviation values for the difference between the retrieval and the time-lagged ECMWF data (+6, +12 and +24 hours) for temperature, ln(specific humidity) and surface pressure; and panels (c) and (d) show the standard deviation values for the difference between the UKMO (first-guess) and time-lagged ECMWF data for temperature, ln(specific humidity) and surface pressure

Figure 7.14 clearly shows that decreasing the temporal collocation quality, i.e. moving further away from the best temporal collocation, increases the standard deviation values for the differences between both the retrieval and the ECMWF and the UKMO and the

[13]These days are used because they represent a complete set of successive days.

ECMWF. In particular, the larger temperature standard deviations are below 100 hPa at a pressure comparable with the tropopause.

7.4.7 Particular occultations

Two occultations are shown which exemplify the optimal estimation method and the extent to which the optimal estimate improves upon the first-guess UKMO knowledge: one at a high latitude and one at a low latitude.

Figure 7.15 shows a typical high latitude occultation (also shown by Kursinski et al. (1996)). Cold, dry conditions allow temperatures from the Abelian inverse method to be derived accurately to within 0.5 km of the surface (Kursinski et al. 1996). The collocated radiosonde agrees well with the Abelian inversion with differences of less than one kelvin for most of the troposphere (Kursinski et al. 1996). The optimal estimation appears to 'shadow' the Abelian inversion, resolving small-scale features not seen by the model analyses. In particular, panel (b) shows differences between the Abelian inversion and the optimal estimate of typically less than one kelvin with no particular bias, as opposed to a bias of more than one kelvin between the optimal estimate and model analyses.

Panels (c) and (d) shows a marked decrease in the specific humidity optimal estimate[14] relative to the UKMO model towards the ECMWF model analyses. The large percentage difference shown in panel (d) reflects the large variation between relatively small amounts of water vapour in this occultation. The magnitude of these variations is comparable with the difference between models in the North polar region (figure 7.12). Panel (c) also shows that the surface pressure retrieval decreases relative to the UKMO model value towards the ECMWF value, which is consistent with known differences between the two models in 1995.

Panels (e) and (f) show the improvement in the uncertainty of the atmospheric state relative to the UKMO model uncertainty. The temperature improvement shows no marked difference from the mean value of this quantity shown by figure 7.9; and the ln(specific humidity) improvement increases from 2% at 300 hPa to a maximum of 40%

[14]The ln(specific humidity) optimal estimates for these occultations have been converted to specific humidity because it represents a physical quantity.

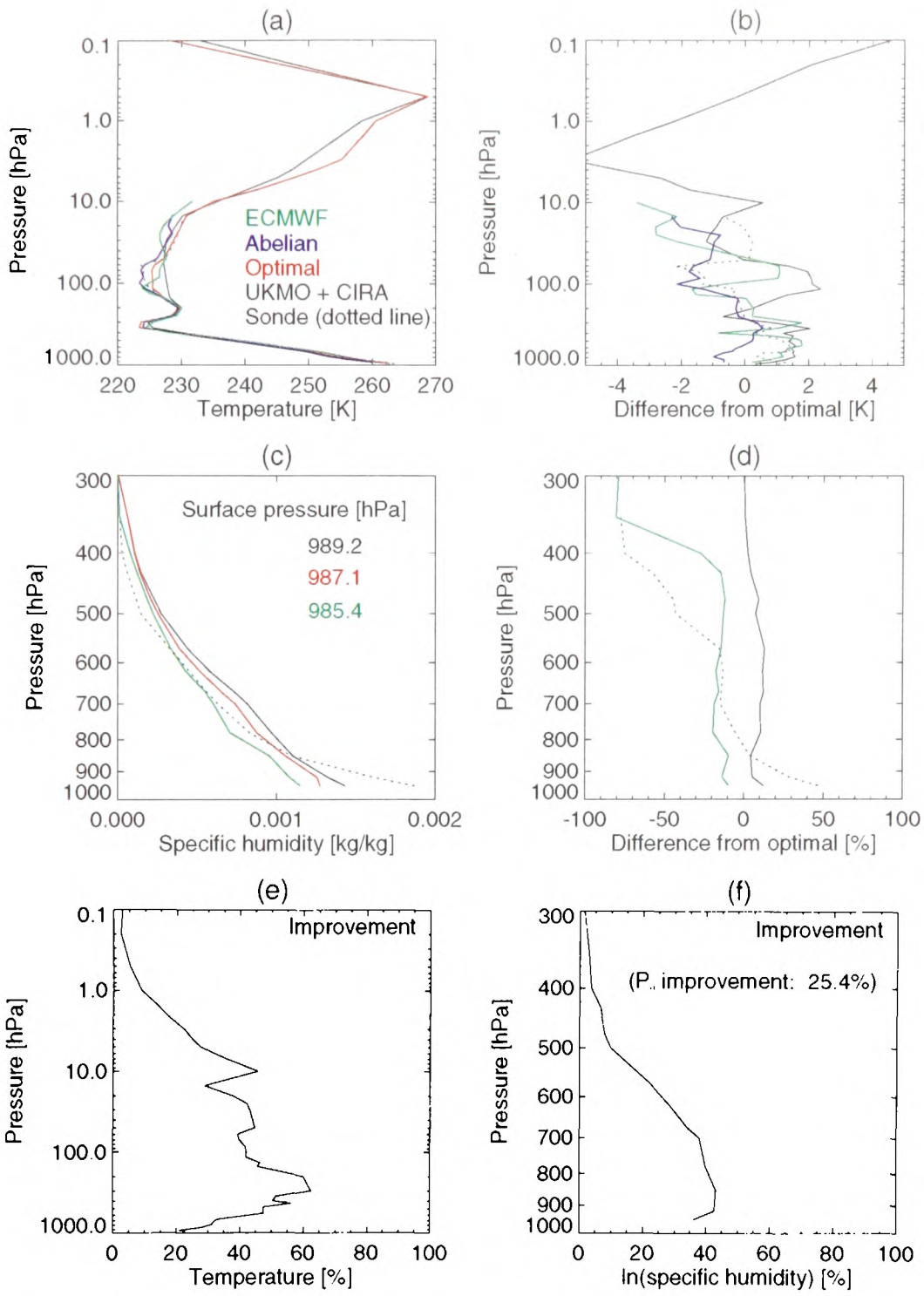


Figure 7.15: Comparison between a high latitude occultation (01:33 GMT, 5th May, 1995) and correlative data, including results from the Abelian inverse method. The occultation event is obtained at 69.1°N, 277.7°E; the radiosonde is from 68.8°N, 278.8°E at 00:00 GMT; and model analyses are from 00:00 GMT. Panel (a) shows the temperature comparison; panel (b) shows the temperature differences between the optimal estimate and the correlative data; panel (c) shows the specific humidity comparison (with surface pressure comparison inset); panel (d) shows the specific humidity differences between the optimal estimate and correlative data, expressed as a percentage; and panels (e) and (f) show the improvement vectors for temperature and ln(specific humidity), with surface pressure improvement inset.

at 900 hPa. The surface pressure retrieval has decreased the uncertainty in its value, relative to the UKMO *a priori* uncertainty, by 25%.

Figure 7.16 shows a typical low-latitude occultation. At low latitudes, the Abelian

inversion normally has problems resolving accurate temperature values much below the troposphere, even with using background estimates of water vapour. For this particular occultation, reasonable observations terminate at approximately 500 hPa. For altitudes above and just below the tropopause, the optimal estimate shows good agreement with the Abelian inversion. However, further into the troposphere the Abelian inversion develops an increasing warm bias relative to the optimal estimate and model analyses. This bias is due to the inadequacy of the low-water correction (§5.3.5) for the Abelian temperature retrieval in vertical regions of large water vapour concentration. At low altitudes the optimal estimate agrees well with model analyses with differences of less than one kelvin.

The specific humidity comparison is shown in panel (c). The specific humidity optimal estimate shows good agreement with results from the Abelian inversion and the ECMWF model analyses. In particular, the UKMO model analyses is larger than both the ECMWF model analyses and the Abelian inversion results.

As displayed by the high latitude occultation (figure 7.15), the surface pressure retrieval increases relative to the UKMO model towards the ECMWF model.

Panels (e) and (f) show a sharp overlap region at 300 hPa between the temperature and ln(specific humidity) improvement indicative of a tropical occultation retrieval. The values for the ln(specific humidity) improvement increase from 30% at 300 hPa increasing to more than 70% at 500 hPa, below which the occultation terminates and the improvement rapidly declines. The associated surface pressure improvement is 19%.

7.5 Chapter summary

There are several conclusions to be drawn from the work in this chapter, which are summarised in this section.

The first involves the reprocessing of tropical occultations. Large difference between the simulated measurements and the observations in the lower atmosphere (less than 15 km) are found to be responsible for some of the tropical occultations failing the χ^2-based quality control scheme. It is shown that increasing the error contribution for the

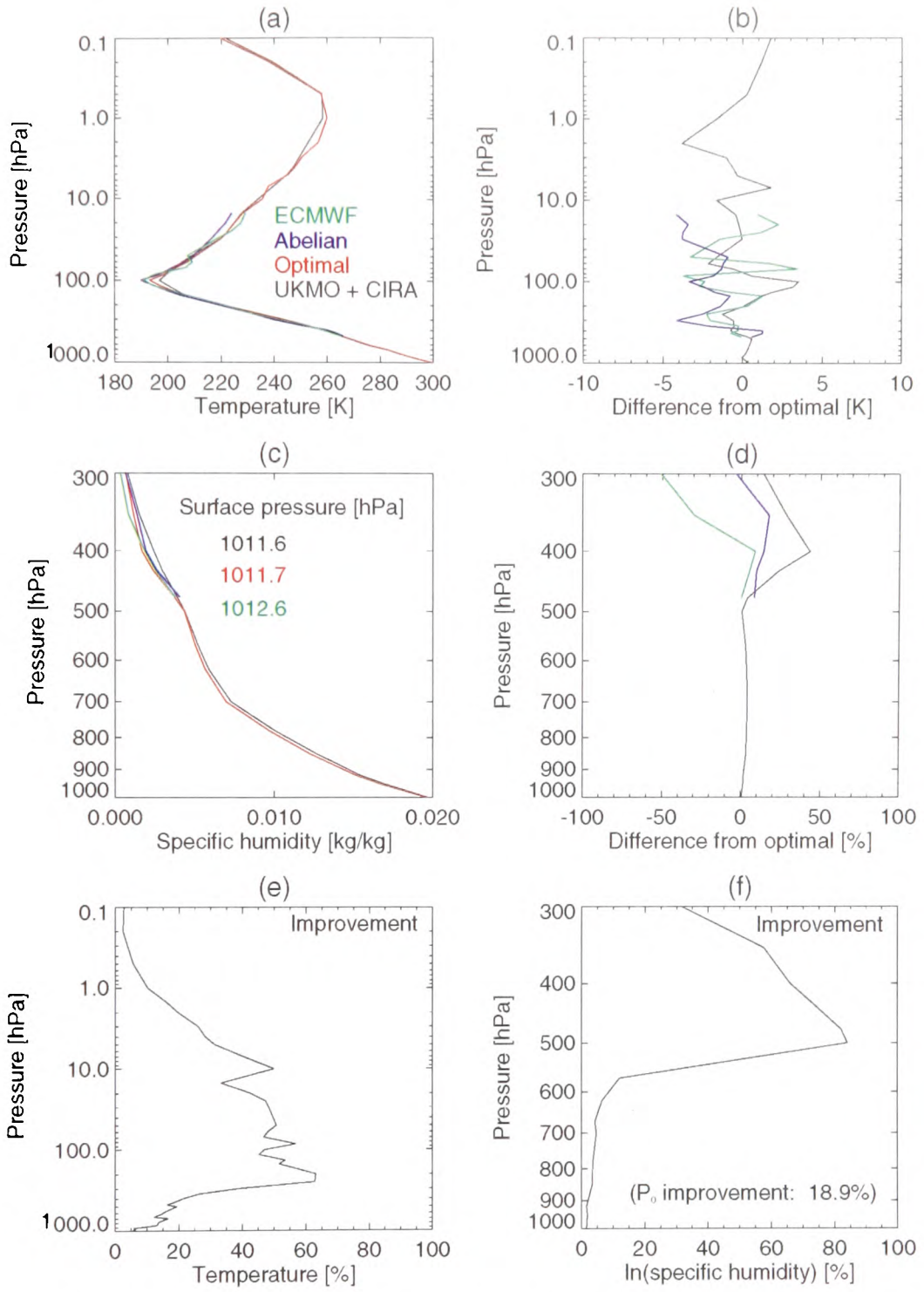


Figure 7.16: Comparison between a low latitude occultation (17:12 GMT. 29th June. 1995) and correlative data, including results from the Abelian inverse method. The occultation event is obtained at 16.9°N, 256.1°E; and the model analyses are taken from 18:00 GMT. Panels are as figure 7.15.

horizontal refractivity structure by a realistic amount causes the number of occultations which fail in this region decrease by more than 20%. This represents a failure of the quality control scheme, and highlights the global uncertainties used for the optimal estimation retrievals. It can be corrected by accounting for typical horizontal refractivity structure at different latitudes.

Second, optimal estimation using radio occultation observations provides a significant

impact compared with the UKMO first-guess knowledge of the atmosphere at all latitudes and all altitudes. In particular, high latitude occultations display large values of temperature improvement reaching pressures of more than 600 hPa, below which values of ln(specific humidity) improvement become prominent at the expense of the temperature improvement. At low latitudes, a sharp overlap region becomes apparent at 300 hPa, where the ln(specific humidity) component of the simultaneous retrieval increases. Typical values for the ln(specific humidity) improvement in this regions are larger than those shown by low latitude retrievals, and the vertical range of the improvement declines at typically 700 hPa where most tropical occultations terminated. For occultations which reach pressures larger than 850 hPa, a typical value for surface pressure improvement is 29±6%.

Third, in general between the upper troposphere and lower stratosphere the temperature optimal estimate agrees well with results from the Abelian inverse method. At high latitudes, the temperature and refractivity optimal estimates show excellent agreement with the Abelian inversion down to near-surface altitudes. In these regions, the ln(specific humidity) estimates also show a marked agreement with the ECMWF model analyses and radiosondes. At low latitudes, above and just below the tropopause, the optimal estimate compares well with the Abelian inversion. Further into the tropopause the Abelian inverse method develops a warm bias relative to model analyses and the optimal estimate due to the inadequacy of the low-water-vapour correction. At these low altitudes, the optimal estimation method shows good agreement with the model analyses for both temperature and specific humidity. Results from the low latitude comparison statistics highlight one of the main advantages of the optimal estimation method over the Abelian inverse method; that is, in the presence of large amounts of water vapour the optimal estimation inverse method is able to retrieve accurately both profiles of temperature and specific humidity. The surface pressure retrievals from occultations which reach altitudes near the surface move away from UKMO model analyses towards the ECMWF model values. In general, comparison statistics between the ECMWF and the UKMO model analyses are consistent with known differences

between the two models at the time of the dataset, providing limited justification for the retrieval of surface pressure from radio occultation observations.

Two particular occultations are shown which exemplify the optimal estimate inverse method: one at a low latitude and one at a high latitude. The optimal estimates resolve small-scale structure shown by the Abelian inverse method and collocated radiosondes, not shown by model analyses. The retrieved profiles of specific humidity and surface pressure show similar magnitude and structure to the ECMWF model values and radiosondes.

Finally, it should be noted that the results presented in this chapter represent occultation data from the first few months of the GPS/MET experiment, and may be subject to software or hardware problems. As such, a relatively small dataset has been used to demonstrate the potential of accurately retrieving simultaneously profiles of temperature, water vapour and surface pressure from radio occultation observations of the Earth's atmosphere.

Chapter 8

Conclusions and further work

Two questions were posed in the introduction to this thesis:

1. How do typical line-of-sight horizontal meteorological gradients affect the Abelian inverse method of radio occultation observations?

2. Using a non-linear optimal estimation inverse method, could radio occultation observations improve numerical weather prediction?

In an effort to answer these questions, several investigations were undertaken, the finding of which are summarised in this chapter. Furthermore, suggestions for further work are listed as an attempt to form a perspective on the research field as a whole.

8.1 Summary of conclusions

8.1.1 How do typical line-of-sight horizontal meteorological gradients affect the Abelian inverse method of radio occultation observations?

Using simple horizontal temperature gradients, temperature retrieval errors are investigated. The largest retrieval errors correspond to horizontal temperature structures which are symmetric about the vertical measurement profile, and can be explained by considering the excess refractivity contributions to the Abel integral transform. It is also shown that typical horizontal pressure gradients have a smaller effect on the Abelian retrieval than typical horizontal temperature gradients.

Applying ideas formed from simple horizontal inhomogeneities, a realistic mid-latitude atmospheric 2-d field (from an NWP model) is used to compute bending angle

profiles with and without information about local refractivity horizontal structure. The subsequent comparison using the two representations provided estimates of the bending angle errors from horizontal refractivity structure which were used for the optimal estimation inverse method. The mid-latitude temperature section is used to investigate temperature retrieval errors from typical horizontal temperature gradients; in addition, the temperature and humidity sections are considered together to investigate the effect of combining realistic horizontal humidity gradients with horizontal temperature gradients compared with the 'dry' refractivity. It is shown that humidity provides an important contribution to overall horizontal refractivity structures, which can compensate or reinforce 'dry' refractivity structures.

Regression coefficients are fitted to the temperature and refractivity retrieval errors by considering first and second-order curve fitting coefficients of the local horizontal structure. For both the retrieved temperature and refractivity errors it is shown that the second-order curve provides the largest contribution to the prediction of retrieval error. In particular, the most successful predicted values are found to be in vertical regions where horizontal variation is largest, e.g. the tropopause. However, predicted values for retrieval error only account for a small percentage of the total error. Better predictions may be possible if the atmospheric structure above the tangent point is also considered.

In summary it is shown that 'wet' horizontal refractivity gradients along the line of sight are important to consider when the Abelian inverse method is used. The inclusion of horizontal humidity gradients is especially important for altitudes below the tropopause where their contribution towards refractivity become more significant. The effect of including realistic horizontal humidity variations in addition to realistic horizontal temperature variations can have the effect of compensating for or reinforcing the temperature gradients.

8.1.2 Using a non-linear optimal estimation inverse method, could radio occultation observations improve numerical weather prediction?

An optimal estimation method is presented as an alternative inverse method with which to retrieve simultaneously profiles of temperature, specific humidity and surface pressure.

The method is a non-linear, maximum likelihood technique which can accommodate most aspects of the real radio occultation problem. In particular, it is able to retrieve simultaneously temperature, water vapour and surface pressure, which is not possible using the Abelian inverse method. The optimal estimation inverse method is an iterative method but is found to be stable and to converge rapidly in most cases. A quality control scheme is implemented concurrently which confirms sensible convergence using retrieval χ^2 values, and allows anomalous retrieval conditions to be recognised and omitted, thus preventing gross retrieval errors. The optimal estimation inverse method and quality control scheme together represent a prototype retrieval scheme for radio occultation observations.

This inverse method is validated using an ensemble of simulated retrievals. An improvement vector (the relative change in theoretical error estimates) is used throughout the work to illustrate how the vertical information content structure of radio occultation observations using optimal estimation is distributed over the simultaneously retrieved products . Simulated results show that the largest temperature improvement (using the simulated observations and measurement error estimates) are in the region 300−1 hPa range. Above this region the measurement errors are comparable with the signal and the *a priori* values are retained; below, the emphasis moves from the temperature retrieval to the ln(specific humidity) retrieval which increases to values comparable with the temperature improvement at altitudes below 300 hPa, decreasing slightly near the surface.

The quality of the surface pressure retrieval is found to depend on how near the occultation reaches the surface, i.e. the improvement to the first-guess is largest when the occultation reaches near-surface altitudes.

Using results from the simulations, it is also shown that the optimal solution is dependent on the error estimates used, i.e. the solution is sensitive to the *a priori* errors assumed and forward modelling error.

Real observations from the GPS/MET instrument are used to compute optimal estimates using the prototype retrieval scheme, which are validated using several sources of correlative data, in particular ECMWF and NMC model analyses, radiosondes and results from the Abelian inverse method.

Large differences between the simulated measurements and observations in the lower atmosphere (less than 15 km) are found to be responsible for some of the tropical occultations failing the quality control scheme. It is shown by increasing the error contribution for the horizontal refractivity structure by a realistic amount, the number of occultations which fail in this region decrease more than 10%. This point represents a failure of the quality control scheme, and highlights the global uncertainties used for the optimal estimation retrievals. It can be corrected by accounting for typical horizontal refractivity structure at different latitudes.

Optimal estimation using radio occultation observations provides a significant impact on the UKMO first-guess knowledge of the atmosphere at all latitudes and altitudes. In particular, high latitude occultations display large values of temperature improvement reach down to pressures of more than 600 hPa, below which values of ln(specific humidity) improvement become prominent at the expense of the temperature improvement. At low latitudes, a sharp overlap region becomes apparent, where the ln(specific humidity) component of the simultaneous retrieval increases. Typical values for the ln(specific humidity) improvement in this region are larger than those shown by polar occultation retrievals, and the vertical range of the improvement declines at typically 700 hPa where most tropical occultations terminated. For occultations which reach pressures larger than 850 hPa, a typical value for surface pressure improvement is 29±6%.

In general between the upper troposphere and lower stratosphere the temperature optimal estimate agrees well with results from the Abelian inverse method. At high latitudes, the temperature and refractivity optimal estimates show excellent agreement with the Abelian inversion down to near-surface altitudes. In these regions, the ln(specific

humidity) estimates also show a marked agreement with ECMWF model analyses and radiosondes. At low latitudes, above and just below the tropopause, the optimal estimate compares well with the Abelian inversion. Further into the tropopause the Abelian inverse method develops an increasing warm bias relative to model analyses and the optimal estimate, due to the low-water-vapour correction. At these low altitudes, the optimal estimation method shows good agreement with the model analyses for both temperature and specific humidity. Results from the low latitude comparison statistics highlight one of the main advantages of the optimal estimation method over the Abelian inverse method; that is, in the presence of large amounts of water vapour the optimal estimation inverse method is able to retrieve accurately both profiles of temperature and specific humidity. The surface pressure retrievals from occultations which reach altitudes near the surface move away from UKMO model analyses towards the ECMWF model values. In general, comparison statistics between the ECMWF and the UKMO model analyses are consistent with known differences between the two models at the time of the dataset, providing limited justification for the retrieval of surface pressure from radio occultation observations.

Two particular occultations are shown which exemplify the optimal estimate inverse method. The optimal estimates resolve small-scale structure shown by the Abelian inverse method and collocated radiosondes, not shown by model analyses. The retrieved profiles of specific humidity and surface pressure show similar magnitude and structure to the ECMWF model values and radiosondes.

Finally, it should be noted that the results presented in this chapter represent occultation data from the first few months of the GPS/MET experiment, and may be subject to software or hardware problems. As such, a relatively small dataset is used to demonstrate the potential of accurately retrieving simultaneously profiles of temperature, water vapour and surface pressure from radio occultation observations of the Earth's atmosphere.

8.2 Suggested topics for further work

- Re-formulation of the forward and inverse problem

  Alternative formulations of the forward and inverse problem should be investigated fully before deciding on the best approach to use. The work herein represents one such approach which uses bending angle as the observation. However, time delay of the received radio signals caused by atmospheric refraction is the true radio occultation observation. Although errors from more fundamental measurement quantities, are easier to characterise, derived quantities provide a cleaner interface to work with. Another formulation is to retrieve refractivity using the Abel transform and then use the derived quantities to compute the optimal estimate of temperature and humidity. This approach requires a smaller forward (and inverse) model but the errors are more complicated to model, e.g. refractivity errors which are derived from assumed bending angle errors will be inter-correlated between levels because of the integral nature of the Abel transform.

- Improvement of the forward model parameterisation

  The forward model used for the work herein is based upon geometric optics theory, using which it is not possible to include diffraction effects. Although diffraction effects in general are small, these effects are found to be important in the tropical lower atmosphere (Kursinski 1997).

- Optimising the forward and inverse problem

  Optimising the code to process radio occultation observations (whichever observation variable is selected) will be necessary for operational use. The largest amount of time currently used for the retrieval is for computing the matrix of weighting functions at every iteration by 'brute force', i.e. perturbing each element of the state vector individually and taking the relative change in the measurement vector. The tangent linear (TL) of the forward model represents the first derivative of model. The TL model could either be formulated explicitly or by using a source-to-source translator to generate additional code for computing the first-order derivatives

(Giering 1997). The TL model would only need to be executed once for the computation of the first derivative compared with n+1 evaluations of a forward model using the present technique, where n represents the number of state vector elements.

- Improved *a priori* and measurement error estimates

  Measurement error estimates used for the optimal estimation inverse model represent preliminary values. Measurement noise sources such as thermal noise and satellite velocity errors are well documented and can be computed accurately, but error sources such as horizontal refractivity structure are extremely variable in magnitude, depending on geographical position. An attempt is made to quantify the magnitude of typical horizontal gradients along the line of sight by using a mid-latitude field. However, it is shown that this source of error represents a large component of the total measurement error for heights below 40 km, and it is therefore important to obtain reasonable estimates. It is also shown that some tropical retrievals fail the χ^2 quality control because the horizontal refractivity structure error estimates are too small in this region. Ideally, several estimates of this error would be used depending on the latitude of the occultation.

- Impact study

  The optimal estimation work shown herein has demonstrated the potential of radio occultation observations to improve the knowledge of the atmosphere. To investigate fully the potential of GNSS radio occultation measurements an impact study needs to be performed using a large dataset of good quality observations. This would require a full assimilation of the observations into an NWP model. The 'quality' of the short, medium and long range forecasts could then be compared with and without the new source of information. Planned launches of space-borne GPS receivers capable of radio occultation are summarised by table 8.1; they should provide enough data for full assimilation.

  In addition to receiver launches shown by table 8.1, there are several proposals

Mission/receiver name	Country/ies of origin	Proposed launch date
Ørsted	Denmark	January 1999
SUNSAT	South Africa	January 1999
SAC-C	Argentina	1999
CHAMP	Germany	1999
GRAS (on METOP)	Denmark/U.K./Spain	mid 2003

Table 8.1: A summary of the planned short-term (5-years) GPS radio occultation receiver missions. This information is correct at the time of writing.

for small constellations of GPS receivers (of the order of 6) for 2001—2003, in accordance with the ESA Earth Watch programme [Mr. D. Offiler (UKMO): personal communication].

- Climate monitoring

The fundamental radio occultation observable is atmospheric phase delay, which is essentially a measure of the time difference of signal reception. This observable has the distinct advantages of being extremely stable and intrinsically self-calibrating, with no need for inter-receiver calibration.

Climate change is concerned with long-term trends in the atmosphere, with temporal scales of interest typically described over many years. As such climate monitoring requires a large time series of consistent data which is accurate enough to resolve small changes in the atmosphere.

A high density of observations is not necessary, provided that the coverage is not biased geographically or by time of day (GNSS has a big advantage for temporal coverage, since most other measurement systems are either biased to fixed synoptic times or are sun-synchronous, where GNSS for most orbits proposed is neither). A low *random* error is not essential since the averaging obtained when measuring trends will effectively average over many measurements. In addition precisely the same type of measurement is likely to be possible for several decades, i.e. for as long as the current GPS frequencies are employed.

Geopotential height measurements for constant pressure surfaces have been used in several studies to detect global warming (e.g. Wallace et al. (1993)). Leroy (1996) has shown the potential of obtaining geopotential heights of constant pressure surfaces from GPS/MET data, and has compared them with values from the global ECMWF model. Although the potential is demonstrated, the number (both spatial and temporal) of real observations is limited; if it is assumed that future GPS receivers obtain 500 occultations per day, a constellation of 16 orbiting receivers should be able to measure global average geopotential heights in the troposphere with an accuracy of 1 gpm each day.

Appendix A

Derivation of Fresnel zone radii

The radius of the first Fresnel zone (equation A.4) is derived here for purposes of further understanding the diffraction problem. Further details of this topic have been omitted here, but may be found in Hecht (1987).

To describe the problem the Huygens-Fresnel principle is used. *Every unobstructed point of a wavefront at a given instant in time, serves as a source of spherical secondary wavelets (with the same frequency as that of the primary wave). The amplitude of the optical field at any point beyond is the superposition of all these wavelets (considering their amplitudes and relative phases).*[1]. If this principle is correct it should be possible to sum all the secondary wavelets at a point P and thus obtain the unobstructed primary wave.

The primary wave may be thought of as divided up into a number of annular regions (figure A.1); the boundaries of which correspond to the intersections of the wavefront with a series of spheres centered at P of radius $r_0 + \lambda/2$, $r_0 + \lambda$, $r_0 + 3\lambda/2$, and so forth, where λ is the wavelength of the primary wave. These are called Fresnel zones[2].

Figure A.2 defines the geometry used to derive the Fresnel zone radii. The m^{th} zone is marked on the spherical wavefront by R, which tends to Q as δ tends to zero, i.e. smaller orders of Fresnel zone.

By definition, a wave that travels the path SQP will arrive out of phase by $m\lambda/2$ with a wave that traverses the path SOP. The plane described by SOQ is only a good approximation to the sphere for small values of δ, i.e. $SR \rightarrow SQ$, at which stage two Pythagorean relations $SQ = (SO^2 + OQ^2)^{1/2}$ and $QP = (OP^2 + OQ^2)^{1/2}$ can be used to

[1]But if each wavelet radiated uniformly in all directions, in addition to generating an ongoing wave, there would also be a reverse wave travelling back toward the source. No such wave is found experimentally, so an inclination factor, K is introduced to modify the radiation pattern of the secondary emitters.

[2]Also called half-period zones

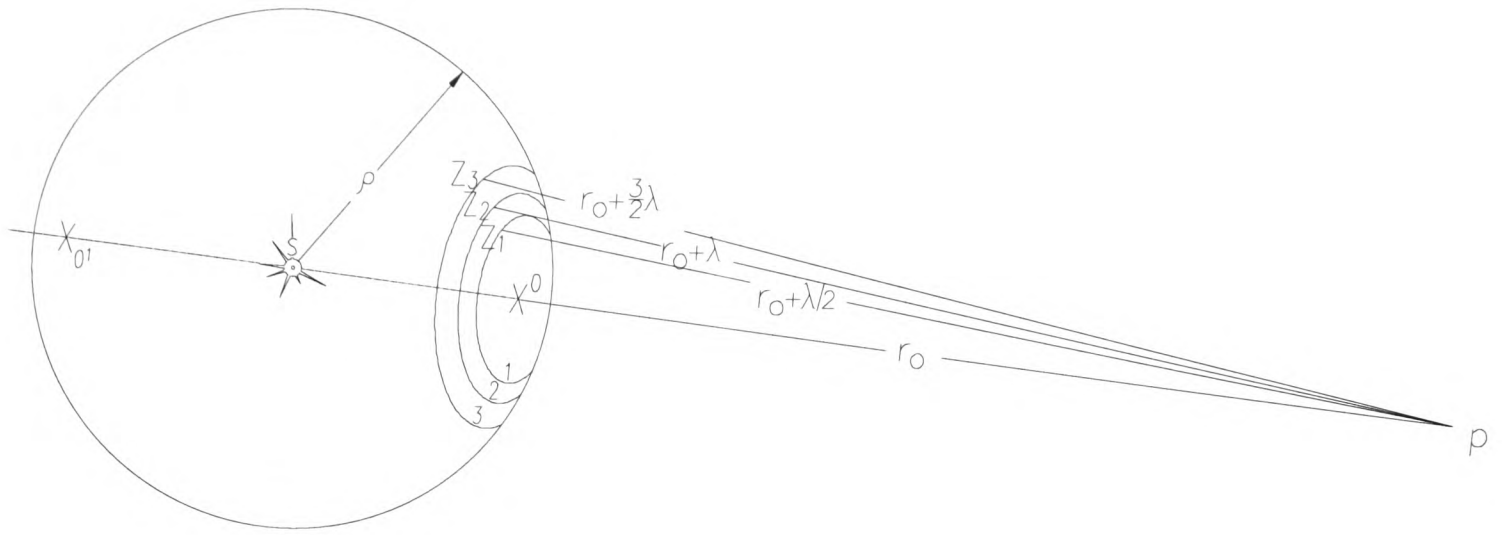


Figure A.1: Propagation of a spherical waveform used to illustrate the concept of a Fresnel zone. The spherical surface corresponds to the primary wavefront at some arbitrary time t' after it has been emitted from S at time t=0. The radius of the disturbance is shown by ρ, with consecutive Fresnel zones denoted by Z_1, Z_2 and Z_3.

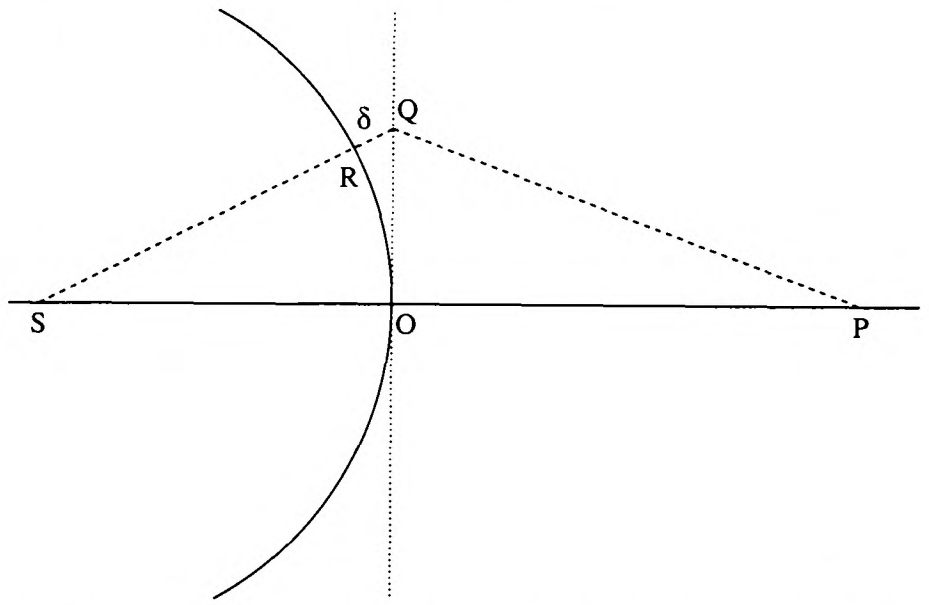


Figure A.2: A figure to show the geometry used to derived the m^{th} Fresnel zone radius.

describe accurately the geometry. If OQ is comparatively small, these expressions can be expanded using the binomial theorem, giving

$$SQ = SO + \frac{OQ}{2SO} \tag{A.1}$$

$$QP = OP + \frac{OQ}{2OP} \tag{A.2}$$

Using the fact these paths taken are $m\lambda/2$ out of phase, these relations give

$$\left(\frac{1}{SO}+\frac{1}{OP}\right)=\frac{m\lambda}{OQ^2} \tag{A.3}$$

Rearranging and substituting SO and OP with the transmitter-occultation distanceL_T and the receiver-occultation distance L_R gives

$$R_F=\left[\frac{\lambda L_T L_R}{L_T+L_R}\right]^{1/2} \tag{A.4}$$

This equation reduces to approximately $R_F=(\lambda L_R)^{1/2}$, if $L_T >> L_R$.

Appendix B

Equation for the radio refractive index of air

Over the past fifty years, there have been many measurements of the dry and moist constants of proportionality for the refractive index of air equation at radio frequencies. The latest, which is considered the most accurate, is discussed in this appendix.

The full equation quoted in Thayer (1974) is:

$$N = 77.6\frac{P_a}{T}Z_a^{-1} + 64.8\frac{P_W}{T}Z_W^{-1} + 3.776 \times 10^5\frac{P_W}{T^2}Z_W^{-1} \tag{B.1}$$

where Z_a^{-1} and Z_W^{-1} are the inverse compressibility factors (corrections for non-ideal behaviour) for dry air and water vapour. These factors are given by Owens (1967).

$$Z_a^{-1} = 1 + P_a\left[5.79 \times 10^{-7}\left(1 + \frac{0.52}{T}\right) - \frac{9.4611 \times 10^{-4}t}{T^2}\right] \tag{B.2}$$

$$Z_W^{-1} = 1 + 1650\left(\frac{P_W}{T^3}\right)\left[1 - 0.01317t + 1.75 \times 10^{-4}t^2 + 1.44 \times 10^{-6}t^3\right] \tag{B.3}$$

where t is temperature in Celsius.

In equation 3.1, the C_2 term is absorbed in c_1, with c_2 adjusted slightly to cover the difference between c_1 and C_2. The inverse compressibility factors are neglected. Neglecting the dry compressibility factor leads to errors of 0.06%, or less, in refractivity (and 0.03%, or less, for pressures above 500 hPa). Using equation 3.1, a fractional error

of 1% of (T - 359)/300 is introduced due to the simplification of two terms, and an error of less than 0.4% due to neglect of the compressibility factor (Kursinski et al. 1997).

Appendix C

Information content

The information content of a particular set of measurements can have many meanings attached to it. For the work herein, this quantity will refer to the measure of the factor by which the uncertainty in the state vector is decreased by the act of measuring it, i.e. fitting a set of measurements to an *a priori* atmospheric state.

Even with this single definition of information content there are many formulations with which this quantity may be calculated: two of which are presented and discussed below.

Different formulations of information content

The first formulation is analogous to the thermodynamical quantity of entropy. This is essentially a measure of how disordered a system is (in this case the atmospheric state). This should also by way of definition imply how much knowledge is known about the state.

Shannon and Weaver (1949) make use of this analogy to define information. If a set of n measurements whose probability density function of choice are $p_1, p_2, ..., p_n$ then a measure of the information of the atmospheric state[1] is given by

$$\mathbf{H} = -\sum_{\mathrm{i}} p_{\mathrm{i}} \log_2 p_{\mathrm{i}} \tag{C.1}$$

where $\mathbf{H}$ represents a quantity called Shannon information. The origin of the logarithmic

[1]Shannon and Weaver (1949) did not address atmospheric physics specifically but constructed a general mathematical framework from which specific cases can be derived

second base is from the general theory dealing with the transfer of 'bits' of information, i.e. 0 or 1. **H** is positive because value of p are between 0 and 1.

Using this definition of information, information content is given thus

$$\mathbf{H} = \mathbf{H}_{\text{solution}} - \mathbf{H}_{\text{a}} \tag{C.2}$$

where $\mathbf{H}_{\text{a}}$ denotes the Shannon information of the *a priori* atmospheric state and $\mathbf{H}_{\text{solution}}$ denotes the Shannon information from the optimal solution. So the information content of the measurement using this formula is the change in Shannon information when the measurements are assimilated with the background state.

If the probability density functions have Gaussian distributions with covariance **S** then the Shannon information is given by

$$\mathbf{H} = \log_2 |\mathbf{S}|^{1/2} \tag{C.3}$$

where $|\mathbf{S}|$ represents the determinant of the matrix **S**. It should be noted that the determinant of a square, symmetric matrix is equivalent to the product of its eigenvalues.

The resulting quantity (a single number) from this calculation represents how many bits of information, i.e. measurements, have been used to construct the optimal solution. In the case of a simultaneous retrieval of two of more quantities, the distribution of the measurement information is unclear using this method. In an effort to illustrate how a simultaneous retrieval improves upon the first-guess knowledge the term improvement vector is coined (Palmer et al. 1998). This approach makes use of the principal diagonal elements of the *a priori* constraint and the solution covariance, of which the square root of the elements provide an estimate of the error attached to each element of the state vector before and after the measurements have been fitted. The improvement vector $\boldsymbol{\eta}$ is given by

$$\eta_j = 100 \times \left[1 - \left(\frac{\hat{\mathbf{S}}_{j,j}}{\mathbf{S}_{\mathrm{a}j,j}} \right)^{\frac{1}{2}} \right] \tag{C.4}$$

where j is the matrix element index.

Appendix D

Cross correlative data statistics

The statistics shown in this appendix represent the mean bias and standard deviation between the sources of correlative data used to validate the optimal estimates in chapter 7. In particular, the correlative data shown are results from the Abelian inversion, ECMWF and NMC model analyses and collocated radiosondes.

If no cross correlative data exists the difference is presented to be zero.

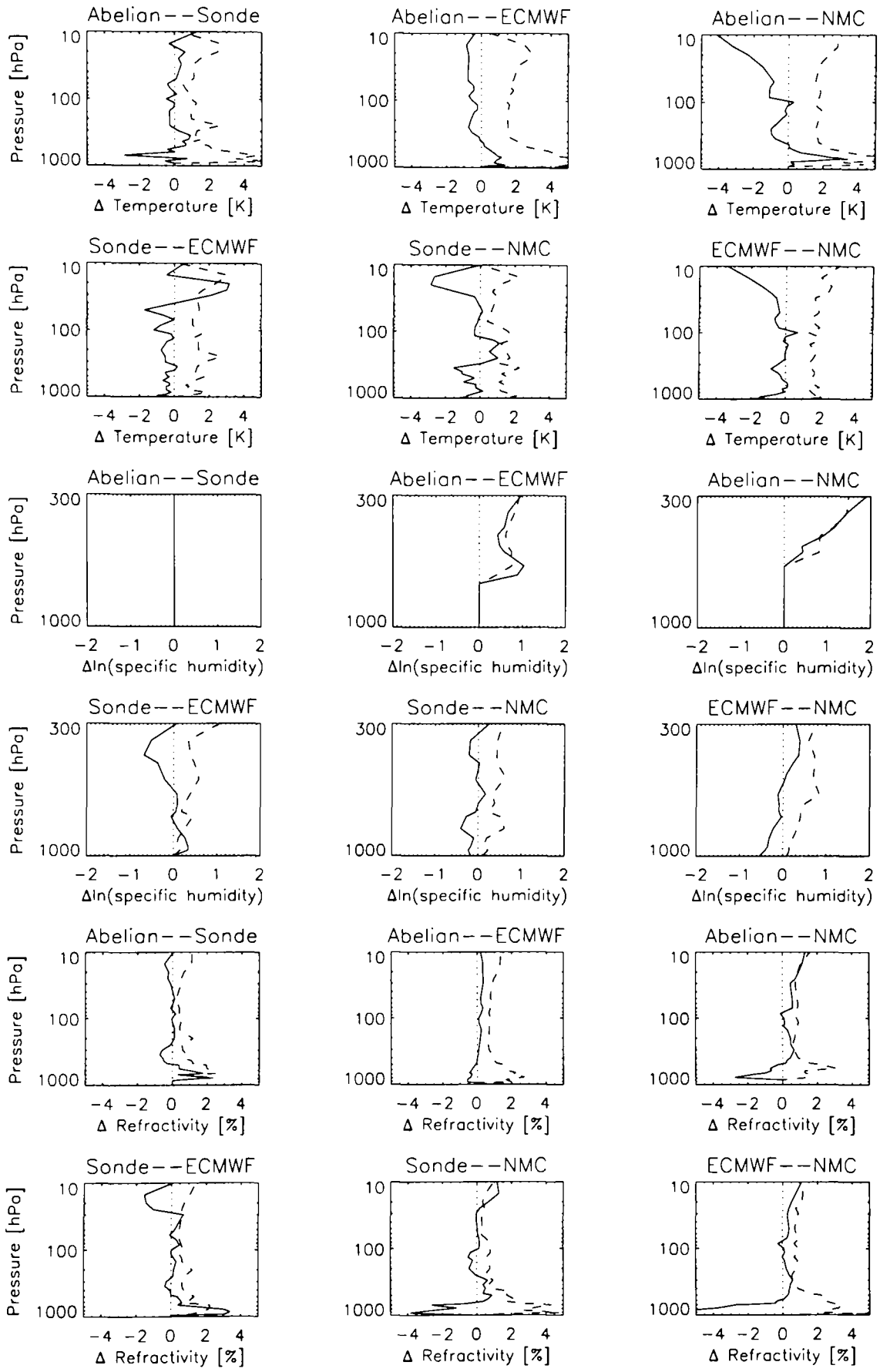


Figure D.1: Mean bias and standard deviation global statistics for the cross correlative data. The solid black and dotted lines represent the bias and standard deviation of the difference, respectively. Each panel header shows the comparison shown. A positive bias is where the first name associated with each header is larger.

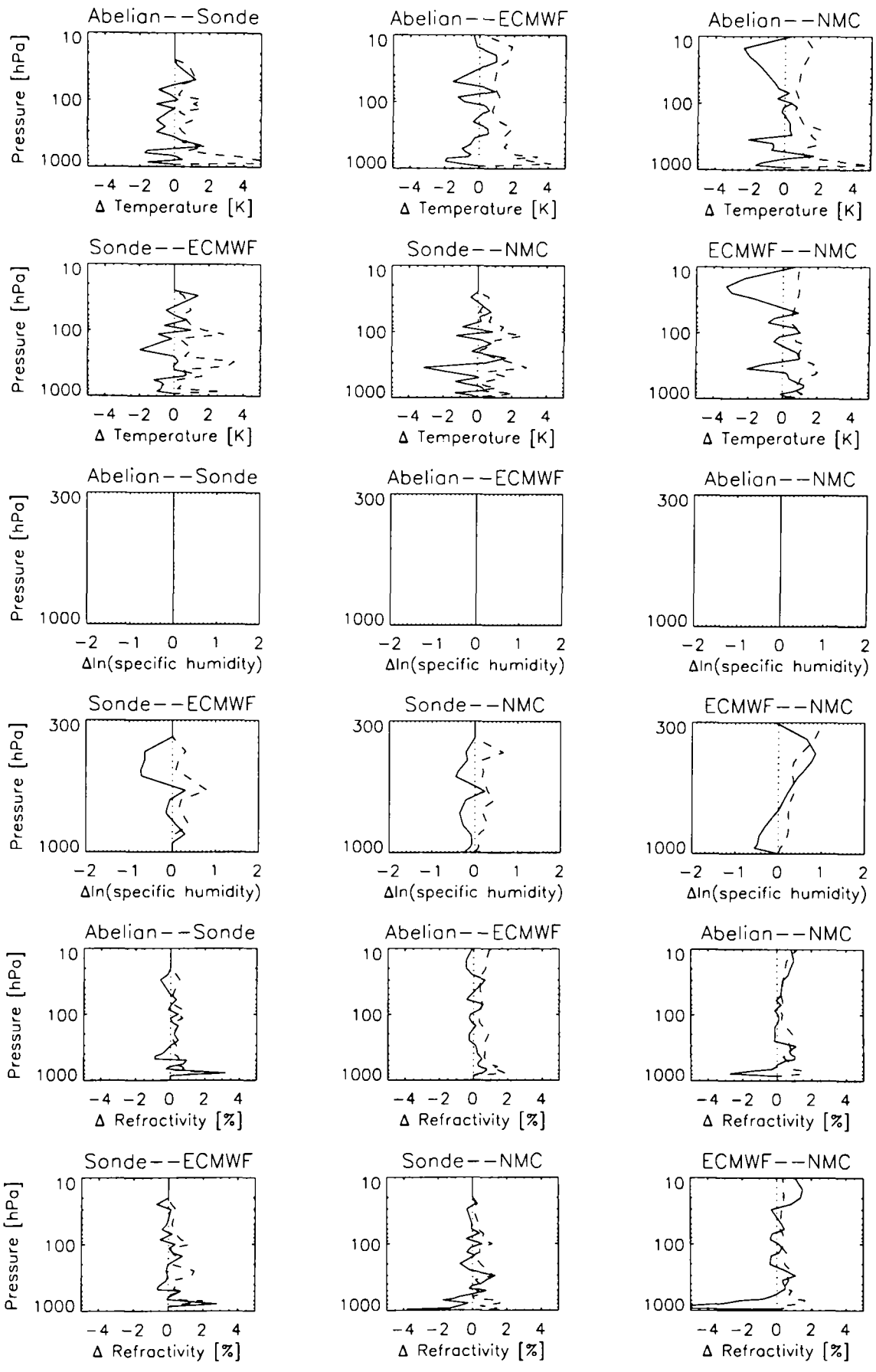


Figure D.2: Mean bias and standard deviation North polar statistics for the cross correlative data. Panels are as figure D.1.

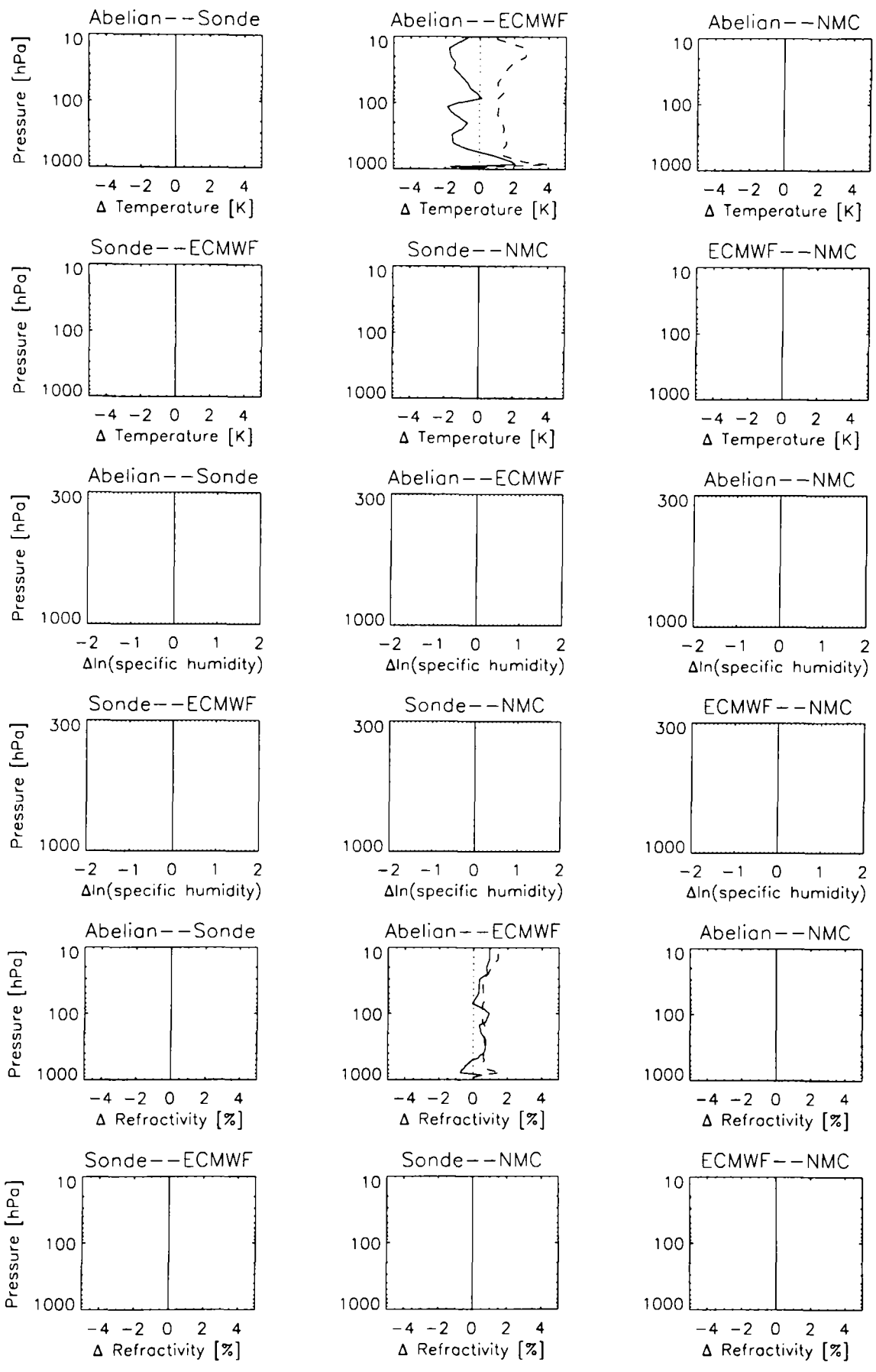


Figure D.3: Mean bias and standard deviation South polar statistics for the cross correlative data. Panels are as figure D.1.

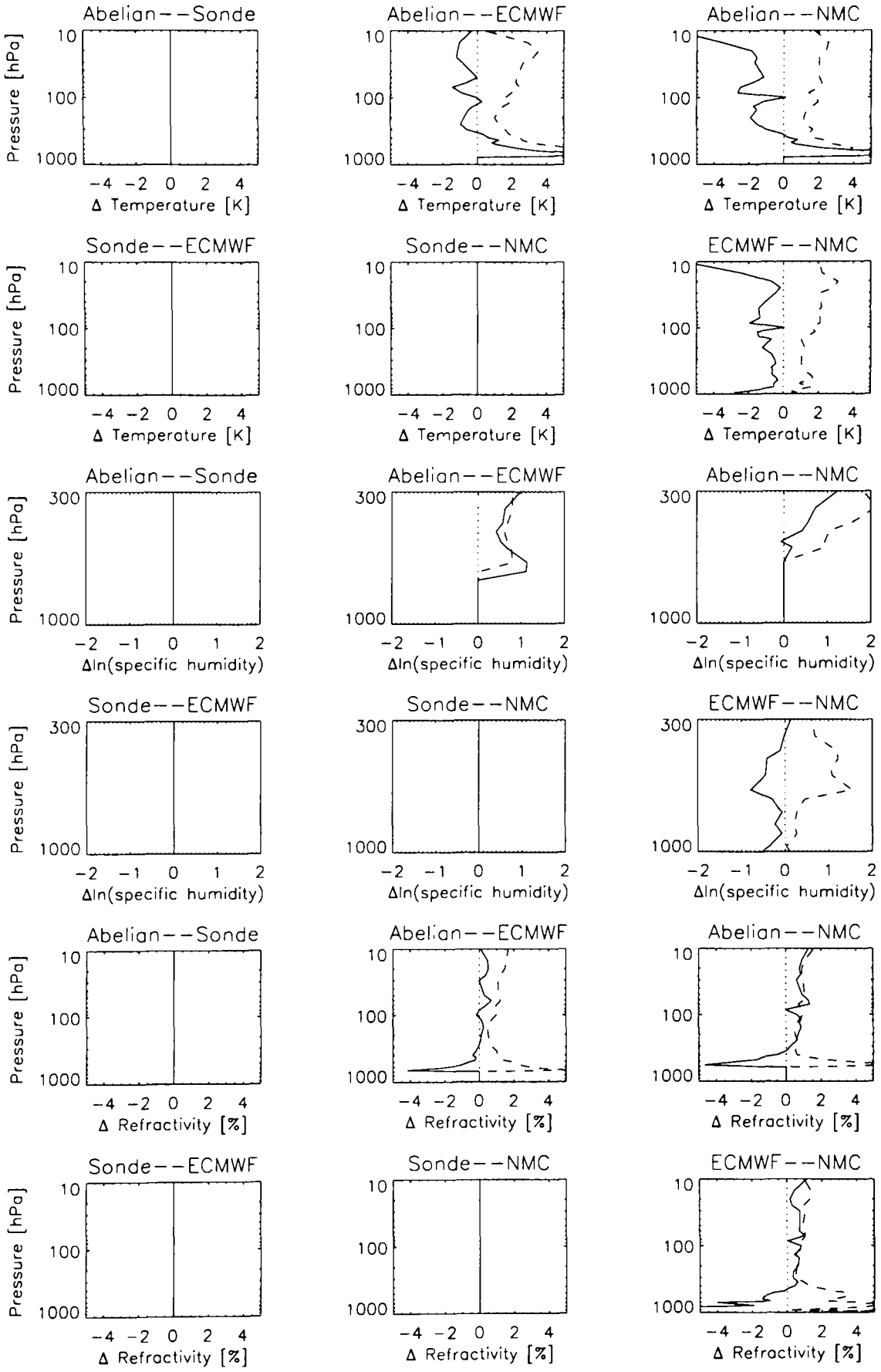


Figure D.4: Mean bias and standard deviation tropical statistics for the cross correlative data. Panels are as figure D.1.

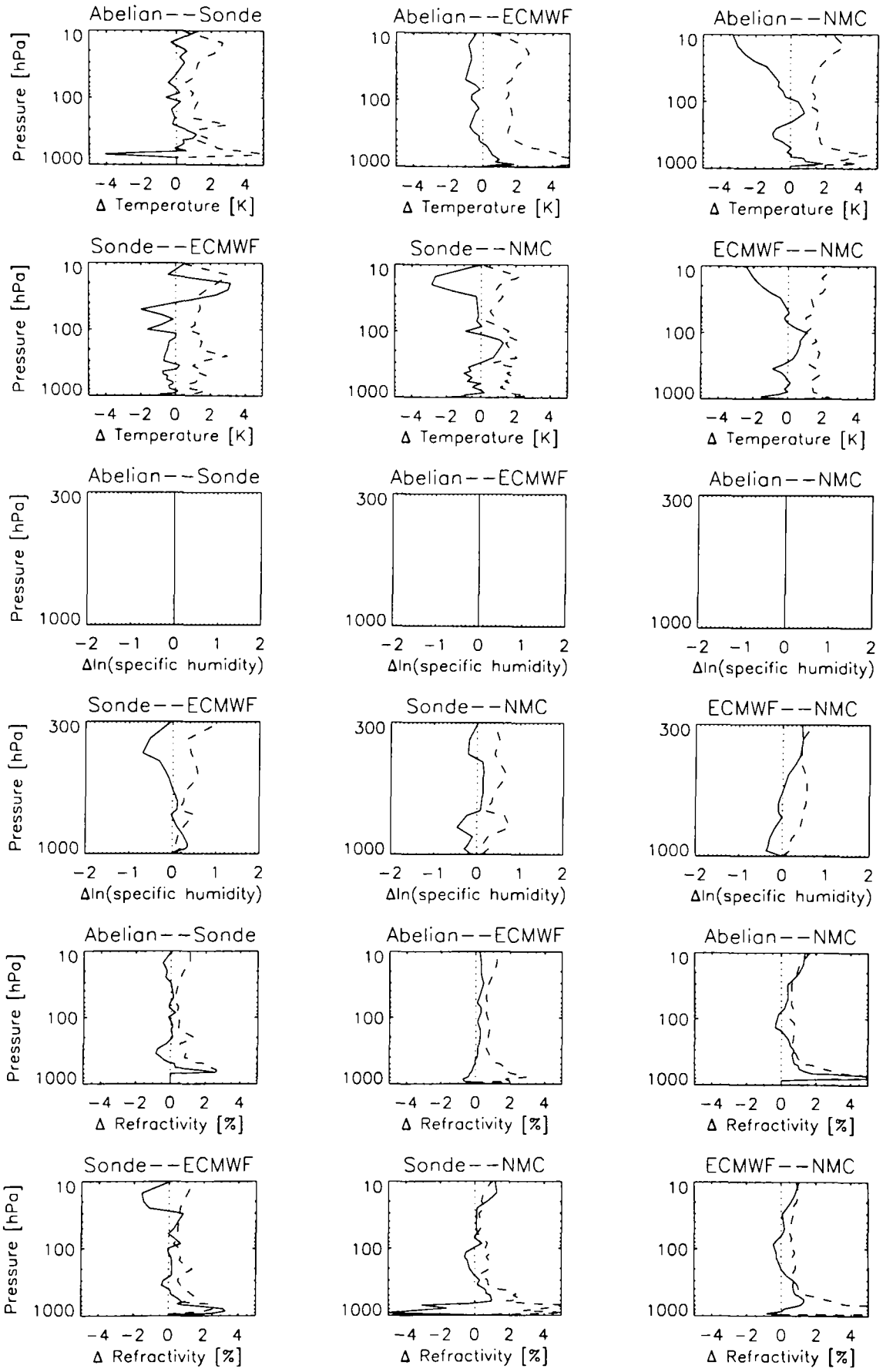


Figure D.5: Mean bias and standard deviation mid-latitudinal statistics for the cross correlative data. Panels are as figure D.1.

Appendix E

Acronyms and Notation used

This appendix has been written to provide the reader with a guide to the acronyms used in this document, and a list of notation which has been used frequently throughout.

Acronyms

AS	Anti-Spoofing signal encryption
AMS	American Meteorological Society
AMSU	Advanced Microwave Sounding Unit
ATOVS	Advanced TIROS Operational Vertical Sounder
C/A	Coarse Acquisition code
CLIMAP	CLImate and environment Monitoring with GPS Atmospheric Profiling
DD	Double difference
DGPS	Differential GPS
ECMWF	European Centre for Medium-range Weather Forecasts
EGOPS	End-to-end GNSS Occultation Performance Simulator (software developed by the University of Graz, Austria)
ESA	European Space Agency
EUMETSAT	EUropean organisation for the exploitation of METeorological SATellite
GDOP	Geometric Dilution of Precision
GLONASS	GLObal NAvigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPS/MET	GPS/METeorology experiment satellite receiver (on MicroLab-1, launched April 1995)

HIRS	High-Infrared Resolution Sounder
JPL	Jet Propulsion Laboratory
LEO	Low-Earth Orbit
LRC	Local Radius of Curvature
L1	One of two L-band frequency signals of GPS (f=1.57542 GHz; λ=0.19 m)
L2	Two of two L-band frequency signals of GPS (f=1.2272 GHz; λ=0.24 m)
MPI	Max-Planck Institute
MSU	Microwave Sounding Unit
NASA	National Aeronautical Space Administration
NAVSTAR	NAVigation System using Timing and Ranging
NCAR	National Center for Atmospheric Research
NCEP	National Center for Enviromental Prediction
NMC	National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PDF	Probability Density Function
PRN	Pseudo-Random Noise
RGPS	Relative GPS
RO	Radio Occultation
SA	Selective Availability
SATOB	Operational code for SATellite OBservations
SD	Single difference
SNR	Signal-to-Noise Ratio
SSU	Stratospheric Sounding Unit
TEMP	Operational code to radiosondes
TIROS	Television and Infra-Red Observation Satellite
TOVS	TIROS Operational Vertical Sounder
UCAR	University Corporation of Atmospheric Research
UKMO	United Kingdom Meteorological Office
UM	Unified Model (from UKMO)

VAR	VARiational assimilation (usually 1-D, 3-D or 4-D)
VCO	Voltage Controlled Oscillator
UCAR	University Corporation of Atmospheric Research
WGS-84	World Geodetic System (1984)

Notation

In general, vectors and matrices are denoted by bold lower-case letters and bold upper-case letter, respectively. Subscripts $^{-1}$ and $^{\mathrm{T}}$ denote matrix inverse and transpose, respectively. In cases where the same variable is used to denote different quantities, the usage of the variable in the text should convey the difference in meaning.

- General physics

T	Atmospheric temperature	x	Refractive radius
T^*	Virtual temperature	P or P_{a}	Atmospheric pressure
q	Specific humidity	P_0	Surface pressure
P_{w}	Partial water vapour pressure	z or h	Geometric altitude
z_0	Height above the geoid	ϕ	Geopotential altitude surface
N	Refractivity	n	Refractive index
R_{F}	Fresnel radius	R_{E}	Local radius of curvature
ρ	Atmospheric density	Φ	Gravity potential
η_{e}	Electron density distribution	f	Operating frequency
$\boldsymbol{\varepsilon}$	Bending angle	$\mathbf{a}$	Impact parameter
f_d	Doppler-shifted frequency	$\mathbf{I}$	Unit matrix
I_{i}	ith eigenvector	λ	Operating wavelength
λ_{i}	ith eigenvalue	θ	Latitude

- Forward and inverse modelling

$\hat{\mathbf{x}}$	Retrieved state vector	$\mathbf{x}_a$	*a priori* state vector
$\mathbf{y}$	Measurement vector	$\mathbf{S}_a$	*a priori* error covariance matrix
$\mathbf{S}_y$	Measurement error covariance matrix	$\mathbf{S}_\Sigma$	Total measurement error covariance matrix
$\mathbf{A}$	Matrix of averaging kernels	$\mathbf{D}_y$	Matrix of contribution functions
$\mathbf{K}_x$	Matrix of weighting functions	$\mathbf{K}_b$	Forward model parameter sensitivity matrix
$\mathbf{F}(\mathbf{x})$	Forward model	$\mathbf{f}(\mathbf{x})$	Forward function
$\boldsymbol{\eta}$	Improvement vector	$\mathbf{b}$	Forward function parameter
$\hat{\mathbf{b}}$	Forward model parameter	$\hat{\mathbf{c}}$	Inverse model parameter
$\mathbf{H}$	Shannon information	$\mathbf{H}_x$	Hessian matrix

References

Abramowitz, M. and A. Segun (1968). *Handbook of Mathematical Functions* (fifth ed.). Dover Publications, Inc.

Agency, T. D. M. (1987, September). Department of Defense World Geodetic System 1984 — its definition and relationships with local geodetic systems. Technical Report 8350.2, The Defense Mapping Agency.

Allen, C. W. (1976). *Astrophysical Quantities* (Third ed.). Athlone Press.

Backus, G. E. and J. F. Gilbert (1970). Uniqueness in the inversion of inaccurate gross earth data. *Philosophical Transactions of the Royal Society of London 266*(123), 123–192.

Bean, B. R. and E. J. Dutton (1968). *Radio Meteorology.* Dover Publications, Inc, New York.

Bengtsson, L., M. E. Gorbunov, and S. V. Sokolovskiy (1995). Space refractive tomography of the atmosphere: Modelling of direct and inverse problems. Technical report, Max-Plank Institute for Meteorology.

Born, M. and E. Wolf (1980). *Principles of Optics.* Pergamon Press.

Chahine, M. T., D. J. McCleese, P. W. Rosenkranz, and D. H. Staelin (1983). *Interaction mechanisms within the atmosphere.* American Society of Photogrammetry.

CIRA-86 (1990). Cospar International Reference Atmosphere: 1986. Part ii: Middle Atmosphere Models. *Advances in Space Research 10*(12).

Cullen, M. P. (1993). The unified forecast/climate model. *Meteorological magazine 112*, 1449–1451.

Dixon, T. H. (1991, May). An Introduction to the Global Positioning System and some Geological Applications. *Reviews of Geophysics 29*, 249–276.

ESA (1996). Atmospheric profiling mission - the nine candidate earth explorer missions. Technical report, European Space Agency. Assessment Report.

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

Eyre, J. R. (1989). Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation I: Theory and simulation for TOVS. *Q.J.R. Meteorol. Soc. 115*(489), 1001—1026.

Eyre, J. R. (1990). The information content of data from satellite sounding systems: A simulation study. *Q.J.R. Meteorol. Soc. 116*, 401—434.

Eyre, J. R. (1994, May). Assimilation of Radio Occultation Measurements into a Numerical Prediction System. Technical Report 199, ECMWF.

Fjeldbo, G. and V. Eshleman (1968). The Atmosphere of Mars Analysed by Integral Inversion of the Mariner IV Occultation Data. *Planet Space Science 16*, 1035–1059.

Fjeldbo, G. and V. Eshleman (1969). The atmosphere of Venus as studied with the mariner 5 dual-frequency occultation experiment. *Radio Science 4*, 879–897.

Fjeldbo, G. and A. Kliore (1971). The Neutral Atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments. *The Astronomical Journal 76*, 123–139.

Gelman, M. E., A. J. Miller, R. N. Nagatani, and C. Long (1994). Use of UARS data in the NOAA stratospheric monitoring program. *Adv. Space Res. 14*(9), (9)21—(9)31.

Giering, R. (1997). *Tangent linear and Adjoint Model Compiler, Users manual 1.2.* Massachusetts Institute of Technology.

Hajj, G. and L. Romans (1996, June). Ionospheric Mapping with the GPS/MET. In *Proceedings of the Institute of Navigation 52nd Annual Meeting, Cambridge, Massachusetts.*

Hajj, G. A., R. Ibanez-Meier, E. R. Kursinski, and L. J. Romans (1994). Imaging the Ionosphere with the Global Positioning System. *International Journal of Imaging System and Technology 5*, 174–184.

Hardy, K., G. Hajj, E. Kursinski, and R. Ibanez-Meier (1993, September). Accuracies of the Atmospheric Profiles Obtained from GPS Occultations. In *Proceeding of ION-GPS 93, Utah.*

Hardy, K., D. Hinson, G. Tyler, and E. Kursinski (1992, January). Atmospheric Profiles from Active Space-Based Radio measurements. In *Proceedings of the 6th Conference on Satellite Meteorology and Climatology, Atlanta.*

Hecht, E. (1987). *Optics* (2 ed.). Addison-Wesley.

Hocke, K. (1997). Inversion of GPS meteorology data. *Annales Geophysicae 15*, 443–450.

Houghton, J. (1977). *The Physics of Atmospheres.* Cambridge University Press.

Kliore, A., G. Fjeldbo, B. Seidel, D. Sweetham, T. Sesplaukis, P. Woiceshyn, and S. Rasool (1975). The Atmosphere of Io from Pioneer 10 Radio Occultation Measurements. *Icarus 24*, 407–410.

Kursinski, E., G. Hajj, W. Bertiger, S.S.Leroy, T. Meehan, L. Romans, J. Schofield, D. McCleese, W. Melbourne, C. Thornton, T. Yunck, J. Eyre, and R. Nagatani (1996, February). Initial Results of Radio Occultation Observations of Earth's Atmosphere using GPS. *Science 271*, 1107–1110.

Kursinski, E. R. (1997, March). *The GPS radio occultation concept: theoretical performance and initial results.* Ph. D. thesis, California Institute of Technology.

Kursinski, E. R., G. A. Hajj, K. R. Hardy, L. J. Romans, and J. T. Schofield (1995, September). Observing tropospheric water vapor by radio occultation using the global positioning system. *Geophys. Res. Lett. 22*(17), 2365–2368.

Kursinski, E. R., G. A. Hajj, K. R. Hardy, J. T. Schofield, and R. Linfield (1997). Observing the Earth's Atmosphere with Radio Occultation Measurements using GPS. *Journal of Geophysical Research-Atmospheres 102*(D19), 23429–23465.

Leick, A. (1990). *GPS Satellite Surveying* (first ed.). Wiley-Interscience.

Leroy, S. S. (1996, March). The Measurements of Geopotential Heights by GPS Radio Occultation. *J. Geophys. Res. 102*(D6), 6971–6986.

Lindal, G., J. Lyons, G. Sweetham, V. Eshleman, D.P.Hinson, and G. Tyler (1987). The Atmosphere of Uranus: Results of Radio Occultation Measurements with Voyager 2. *Geophys. Res. Lett. 92*(A13), 14987–15001.

Lindal, G., G. Sweetham, and V. Eshleman (1985). The Atmosphere of Saturn: An Analysis of the Voyager Radio Occultation Measurements. *The Astronomical Journal 90*(6), 1136–1146.

Lindal, G. F. (1992, March). The Atmosphere of Neptune: An Analysis of Radio Occultation Data Acquired with Voyager 2. *The Astronomical Journal 103*(3), 967–982.

Luers, J. K. and R. E. Eskridge (1995). Temperature corrections for the VIZ and Vaisala radiosondes. *J. Appl. Met. 34*, 1241–1253.

Luntama, J.-P. (1997). Atmospheric Profiling with Radio Occultation. Master's thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology, Finland.

Marks, C. J. and C. D. Rodgers (1993, August). A Retrieval Method for Atmospheric Composition from Limb Emission Measurements. *J. Geophys. Res. 98*(D8), 14939 — 14953.

McNally, A. P. (1987, August). *Inverse methods for retrieving tropospheric temperature profiles from satellite sounding data.* D. Phil. thesis, Oxford University.

McPherson, R. D., K. H. Bergman, R. E. Kistler, G. E. Rasch, and D. S. Gordon (1979). The NMC Operational Global Data Assimilation System. *Mon. Wea. Rev. 107*, 1445—1461.

Melbourne, W. G., E. S. Davis, C. B. Duncan, G. A. Hajj, K. R. Hardy, E. R. Kursinski, T. K. Meehan, L. E. Young, and T. P. Yunck (1994, April). The

Application of Spacebourne GPS to Atmospheric Limb Sounding and Global Change Monitoring. Technical report, Jet Propulsion Laboratory.

Mortensen, M. D. and P. Høeg (1998). Inversion of GPS occultation measurements using fresnel diffraction theory. *Geophys. Res. Lett. 25*(13), 2441.

Oort, A. H. and E. M. Rasmusson (1971, September). *Atmospheric Circulation Statistics.* U.S. Department of Commerce/National Oceanic and Atmospheric Administration.

Owens, J. S. (1967). Optical refractive index of air: dependence on pressure, temperature and composition. *Appl. Opt. 6*, 51—58.

Palmer, P. I. (1997, September). Scientists exploit the Global Positioning System. *Physics World, Institute of Physics 10*(9), 31—33.

Palmer, P. I., J. J. Barnett, and J. R. Eyre (1998, May). 1DVAR analysis of GPS radio occultation measurements. In *9th Conference on Satellite Meteorology and Oceanography*, Volume 1, pp. 34—37. AMS, EUMETSAT, SMF.

Parkinson, B. W. and J. J. Spilker (Eds.) (1996). *Orbit Determination*, Volume 2 of *Global Positioning System: Theory and Application.* American Institute of Aeronautics.

Phinney, R. and D. Anderson (1968). On the Radio Occultation Method for Studying Planetary Atmospheres. *J. Geophys. Res. 73*, 1819–1827.

Rangaswamy, S. (1976, August). Recovery of Atmospheric Parameters from the APOLLO/SOYUZ-ATS-F Radio Occultation Data. *Geophys. Res. Lett. 3*(8), 483–486.

Reber, C. A. and M. J. Luther (1993). The Upper Atmosphere Research Satellite (UARS) mission. *J. Geophys. Res. 98*, 10643—10647.

Rocken, C., R. Anthes, M. Exner, D. Hunt, S. Sokolovskiy, R. Ware, M. Gorbunov, W. Shreiner, D. Feng, B. Herman, Y.-H. Kuo, and X. Zou (1997, December). Analysis and validation of GPS/MET data in the neutral atmosphere. *J. Geophys. Res. 102*(D25), 29849—29866.

Rodgers, C. D. (1976). Retrieval of Atmospheric Temperature and Composition from Remote Measurements of Thermal Radiation. *Reviews of Geophysics and Space Science 14*, 609–624.

Rodgers, C. D. (1990). Characterization and Error Analysis of Profiles Retrieved from Remote Sounding Measurements. *Q.J.R. Meteorol. Soc. 95*, 5587–5595.

Rodgers, C. D. (1998). Inverse methods: theory and practise. (to be published), website: http://www.atm.ox.ac.uk/user/rodgers.

Shannon, C. E. and W. Weaver (1949). *The Mathematical Theory of Communication* (8th ed.). University of Ilinois Press.

Silvestrin, P. (1997). Preliminary requirement specifications of the GNSS receiver for atmospheric sounding (GRAS). Technical report, ESA ESTEC.

Smith, E. K. and S. Weintraub (1953). The constants in the equation for atmospheric refractive index at radio frequencies. *Proc. IRE 41*, 1035 – 1037.

Syndergaard, S. (1998). Modeling the impact of the earth's oblateness on the retrieval of temperature and pressure profiles from limb sounding. *J. Atmos. Solar-Terr. Phys. 60*(2), 171—180.

Thayer, G. D. (1974, October). An improved equation for the radio refractive index of air. *Radio Science 9*(10), 803–807.

U.S.-Air-Force (1961). *The Handbook of Geophysics* (Revised ed.). MacMillan.

Vorob'ev, V. V. and T. G. Krasil'nikova (1994, April). Estimation of the Accuracy of the Atmospheric Refractive Index Recovery from Doppler Shift Measurements at Frequencies Used in the NAVSTAR System. *Physics of the Atmosphere and Ocean 29*(5), 602–609. Russian Edition: September -October, 1993.

Wallace, J. M., Y. Zhang, and K.-H. Lau (1993). Structure and seasonality of interannual and interdecadal variability of the geopotential height and temperature fields in the northern hemisphere troposphere. *J. Climate 6*, 2063—2082.

Ware, R., M. Exner, D. Feng, M. Gorbunov, K. Hardy, W. Melbourne, C. Rocken, W. S. S. Sokolovskiy, F. Solheim, X. Zou, R. Anthes, S. Businger, and K. Trenberth

(1996). GPS Sounding of the Atmosphere from Low-Earth Orbit: Preliminary Results. *Bulletin of the American Meteorological Society 77*, 19–40.