Detection of Stratospheric Gravity Waves Using HIRDLS Data

Corwin Wright

Atmospheric, Oceanic and Planetary Physics

Trinity College, University of Oxford

Hilary 2010

Abstract

Detection of Stratospheric Gravity Waves Using HIRDLS Data

Corwin James Wright, Trinity College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2010

Temperature measurements from the HIRDLS instrument on NASA's Aura satellite are analysed for the purposes of detecting and studying internal gravity waves in the terrestrial stratosphere. A detailed description of the methodology used to obtain these data is given, including details of the instrument correction processes used to compensate for errors introduced by a blockage in the instrument optics. A short précis of the relevant theoretical considerations related to atmospheric gravity waves is then outlined.

The thesis then discusses the use of the Stockwell (time-frequency) Transform for the detection of gravity waves in HIRDLS data, together with a detailed analysis of the limitations of this method, and the results obtained from this analysis are analysed by comparison to other instruments and climatology. It is concluded that the Stockwell Transform is an appropriate method for the analysis of the HIRDLS dataset, and that the results obtained are robust.

We apply these results to analyse stratospheric gravity wave activity during the 2005/06 Arctic sudden stratospheric warming. By comparing the magnitude and form of the gravity wave results to local wind data obtained from ECMWF operational analyses, we conclude that a heavily deformed stratopause observed during this period by other instruments was most probably associated with wind-based filtering of the gravity wave spectrum during this period.

Acknowledgements

Firstly, I would like to thank my supervisor, **John Barnett**, for all his help and advice over the course of my DPhil. I would also like to thank **Anu Dudhia** for stepping in to check my thesis chapters, and **Scott Osprey** for many (many!) useful conversations about gravity wave dynamics.

Mention must of course go to my doubtless long-suffering series of officemates - **Luis Millán Valle**, **Rosalind West**, **Reno Choi** and **Jaemin Lee** - who have had to put up with my constant stream of physics and programming questions over the years. Especial thanks are also due to **Jo** (-anna, -sephine, -ella...? We'll work it out one of these days) **Barstow**, **Benjamin Grandey** and (again) **Luis Millán Valle** for proofreading chapters of my thesis. Any remaining mistakes are, of course, entirely their fault...

Elsewhere in AOPP, I would like to thank **Jamie Banks**, **Lucy Bricheno**, **Ramin Lolachi** (and his hair), **Cécile Merlet**, **Andy Sayer**, **Roland Young** and the whole **AOPP Coffee Club** for many useful hours of entertainment and distraction, and far far too many biscuits. I'd also like to thank **Sarah and Monika** for all their logistical support, both in terms of keeping the department running and in dealing with all the paperwork a DPhil appears to involve.

Outside AOPP, both at Trinity College and elsewhere, especial thanks for keeping me sane (or at least distracted) go to **Katherine Atkinson**, **Mette Bundvad** (who has taught me far more about the Book of Ecclesiastes than I *ever* wanted to know), **Sunita Darbe** (especially for the free holidays in Boston and Paris...), **Elizabeth Davidson**, **Sarah de Haas**, **Jen Lorden**, **Puck Rombach** (and the fluorescent yellow, orange and pink socks), **Charlie Turner**, **Jana Uehlecke** (who I suspect has seen at least two-thirds of the graphs in this thesis over our many breakfasts) and **Seth Wilson**.

And, finally, I would like to thank **my parents** for all their help and support in getting this far.

Contents

Chapter 1

Introduction

This thesis describes work undertaken using the HIRDLS instrument to detect gravity waves in the terrestrial stratosphere.

First, the design and operations of the HIRDLS instrument will be discussed (chapter 2), including a description of the theory behind the retrieval method used to obtain atmospheric temperature data from HIRDLS radiance measurements and the corrections applied to accommodate for post-launch hardware problems. The physical mechanisms underlying atmospheric gravity waves and our motivations for studying them, together with a brief description of their sources and a short precis of previous similar studies, will then be discussed in chapter 3, followed by a detailed description of the methods used to detect such waves in the HIRDLS dataset in chapters 4 and 5. Validation of the dataset and methodology will then be discussed in chapter 6, followed by a case study in chapter 7 to show the power of this dataset. Finally, chapter 8 will draw together the conclusions from this work and suggest future directions for the research.

Several versions of the HIRDLS dataset have been released as the Kapton correction process has been improved, and work on this continues. For consistency, this thesis uses the publically-released V004 of the HIRDLS dataset, known internally within the HIRDLS project as v2.04.19. The precise corrections used to produce this version are discussed in section 2.5.

Chapter 2

HIRDLS

The **HI**gh-**R**esolution **D**ynamics **L**imb **S**ounder, or HIRDLS, is an instrument on NASA's Aura platform, part of the 'A-Train' afternoon satellite constellation. Designed by a team led by groups at the University of Oxford in the UK and the National Center for Atmospheric Research (NCAR) in the US, it is a 21-channel filter radiometer for measuring the thermal IR emission from the Earth's limb (Gille et al., 2003; Hepplewhite et al., 2005).

2.1 Mission

2.1.1 Introduction

Many previous experiments, using a range of approaches including balloon and rocket sondes, aircraft-mounted direct measurement, ground-based methods such as RADAR and LIDAR, and space-based remote sensing, have enabled characterisation of the stratosphere at a range of horizontal and vertical scales, with significant success (Andrews, 2000; Gille et al., 2003). However, unobserved phenomena almost certainly still exist which require new data to explain. Of especial interest are a range of processes which occur on small length- and timescales but which are widely distributed across the globe, and hence require detection by an instrument capable of scanning over a broad spatial area (Gille et al., 2003). Accordingly, there is a clear need for a satellite-based remote sensing experiment with high spatial resolution which can scan for these localised phenomena over the whole globe, a niche HIRDLS is designed to fill.

The principle scientific aims of the HIRDLS mission are to elucidate the key elements of

chemistry and dynamics in the upper troposphere, stratosphere and mesosphere via high-resolution monitoring of temperature, trace chemicals and geopotential height gradients. From these data, it is hoped that information on processes involving the roles of planetary and gravity waves in transporting and mixing chemicals can be obtained (Gille et al., 2003; Schoeberl et al., 2006).

2.1.2 Infrared Remote Sensing of the Atmosphere

A remote-sensing experiment is one which obtains information on a process or object without coming into direct contact with it. In the case of satellite-based terrestrial remote sensing, this applies specifically to the measurement of the Earth's surface and/or atmosphere from orbiting spacecraft.

Two key types of remote sensing exist in this context. The first, active remote sensing, uses a radiation source to illuminate the target and measures the intensity of the returned signal. Examples of this include LIDAR and microwave radar systems. Such systems are often based on land and aircraft, and are more rarely used aboard satellites. The second type is passive remote sensing, which uses natural radiation sources where the radiance is generated by geophysical processes or transmitted light from the Sun or stars.

The terrestrial atmosphere is largely transparent to optical wavelengths. This makes optical-wavelength instruments excellent for tasks which involve looking through the atmosphere, whether making astronomical observations from the Earth's surface or recording surface data from orbit. However, this same property makes the optical regime a poor choice for instruments intended to observe the atmosphere itself. Accordingly, measurements for this purpose must be made in a different part of the electromagnetic spectrum.

A wide range of potential spectral regions exist in which observations can be made, each of which contain many spectroscopic peaks due to the actions of chemicals and other atmospheric processes of interest. However, most of these peaks are due predominantly to the action of a single chemical, for example water at sub-mm wavelengths and oxygen in the UV. Two main options present themselves if we wish to observe a significant range of chemistry: the microwave/mm-wave region and the infra-red. The Aura satellite supports instruments designed to observe in both these wavelength regions: HIRDLS and the Microwave Limb Sounder (MLS); the former in the the infrared and the latter in the microwave.

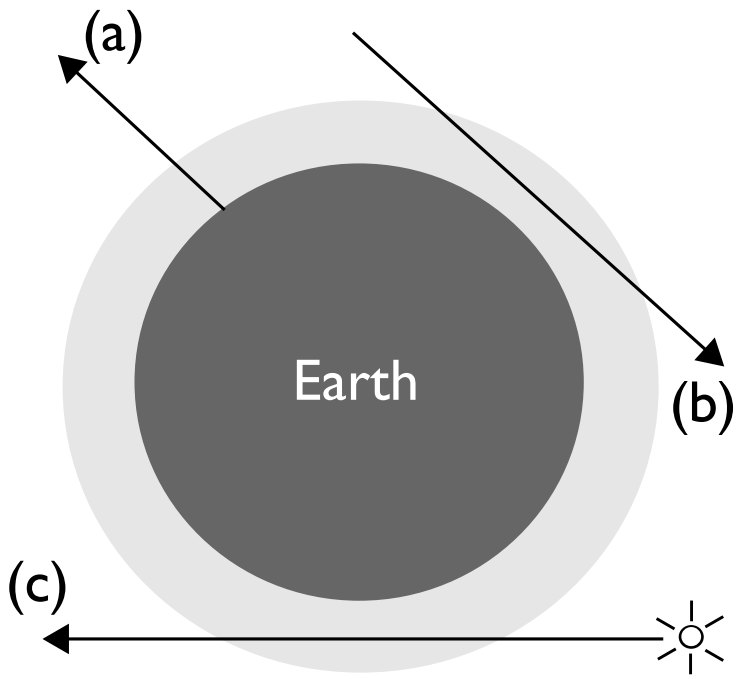


Figure 2.1: *Diagram showing the three main different types of satellite remote sensing: (a) nadir sounding; (b) limb sounding and (c) occultation sounding. In each case, the line illustrates the path through the atmosphere, with the satellite at the end of the arrow. After Andrews (2000).*

2.1.3 Satellite Remote Sensing Geometries

Satellite-based instruments designed to measure the terrestrial atmosphere can adopt any one of several potential viewing geometries, depending on their intended objective (figure 2.1). Each of these geometries has key advantages and disadvantages.

2.1.3.1 Nadir Viewing

A nadir-viewing instrument (figure 2.1 a) observes vertically below or at a slight angle to the path of the satellite as it travel along its orbital track. Such instruments can either observe a fixed path below the satellite or scan horizontally across the path of the instrument, but cannot scan vertically through the atmosphere. Measurements can usually be obtained from all points along the orbital track, allowing a regular and predictable spatial and temporal pattern of measurements to be obtained.

Nadir sounding provides significant advantages in terms of horizontal resolution: as the instrument cannot scan vertically, the sensor time can be devoted to obtaining higher-resolution data in the horizontal plane of the scan. However, this increased horizontal resolution comes at the cost of two disadvantages. The first is the extremely limited vertical resolution. The second

is the warm background signal from the earth: as the instrument is observing not just the atmosphere but the ground beneath it, this signal needs to be removed or compensated for before the atmospheric data can be properly analysed. This is not necessarily trivial due to the variability of this ground signal caused by, for example, topography and urban features.

2.1.3.2 Limb Viewing

A limb-viewing instrument is one which, rather than looking vertically downwards through the atmosphere as with a nadir sounder, instead looks through the atmosphere at a tangent to the Earth (figure 2.1 b). This means that the optical path of the instrument passes entirely through the atmosphere rather than being directed towards the surface as with nadir-sounding instruments, eliminating the terrestrial signal and replacing it with the cold background of space.

The radiation measured by limb sounding is that emitted by the atmosphere, rather than the sum of the atmospheric and terrestrial signals as in the case of nadir sounding. Since the radius of the Earth is significantly greater than the depth of the atmosphere, the horizontal line of sight through the atmosphere seen by a limb sounder is significantly longer than the vertical path observed by a nadir sounder, allowing a larger mass of emitting material to contribute to the measured radiance and hence allowing the detection of emission from low gas concentrations (Elachi and Van Zyl, 2006).

As with a nadir sounder, limb sounders scan in a regular sequence determined primarily by the orbit of the satellite, giving a regular sequence of evenly-spaced measurements against the cold background of space. A key advantage of limb sounding, however, is that additional information can be obtained by adjusting the atmospheric depth of the point of the optical path closest to the Earth, known as the tangent point. The disadvantages of this arrangement are that limb sounding requires a precise knowledge of the viewing geometry, that horizontal resolution is limited, and that it cannot usually be used below the tropopause.

2.1.3.3 Occultation Viewing

An occultation-viewing instrument (figure 2.1 c) is a special case of a limb-sounding instrument. The optical path of the instrument is directed through the atmosphere towards a bright light source, for example the Sun in the case of stellar occultation instruments such as the SOFIE

instrument on the AIM satellite or satellites from the GPS constellation in the COSMIC mission (both discussed further in chapter 6). By measuring the diffraction and/or attenuation of the signal through the atmosphere, a significantly higher signal-to-noise ratio can be obtained over normal limb sounding (Russell III, 1980).

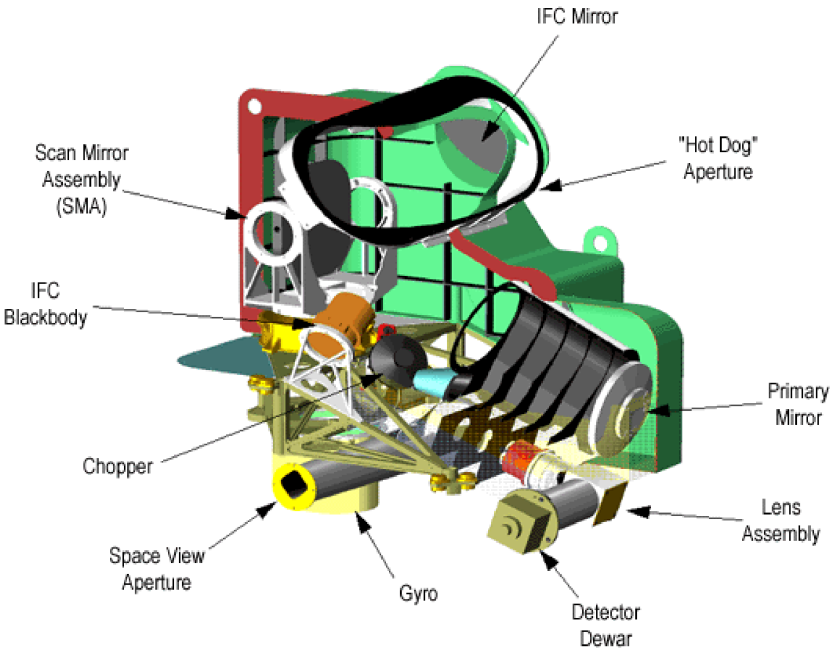
However, this method is fundamentally limited in that it is reliant upon the input signal, and hence only a limited number of occultations can be performed in any given day. For example, the SOFIE instrument only measures around 30 high-latitude solar occultations a day (Gordley et al., 2009), and, whilst the COSMIC constellation measures a significant number of profiles each day due to the high number of GPS satellites, these profiles are distributed irregularly in both time and space rather than along a single scan track as with the other two classes of instrument (Barnett et al., 2008).

The key scientific objectives of the HIRDLS instrument, as discussed above, are largely based upon a requirement for very high vertical resolution combined with good horizontal and temporal coverage. The vertical resolution requirements necessarily exclude the use of a nadir sounder for these purposes, whilst the spatial and temporal requirements make occultation viewing unsuitable. This combination of requirements hence makes a limb sounding design optimal for achieving the desired scientific objectives.

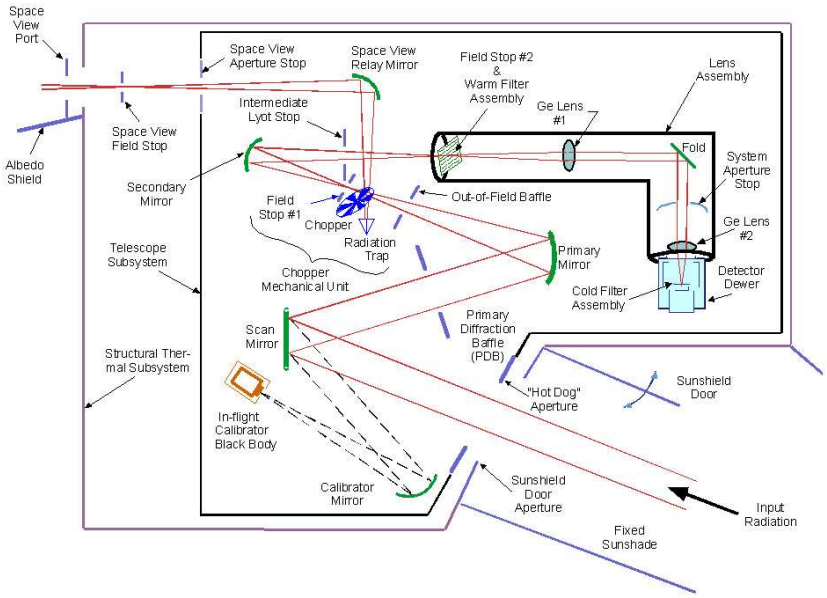
2.2 The HIRDLS Optical System

HIRDLS is a scanning spectroradiometer which images the atmospheric limb onto a focal plane detector array of 21 individual detectors in the mid-infrared, designed to measure with a high degree of spatial resolution of around 1–1.2 km vertically by approximately 500 km horizontally (Barnett et al., 2003), and with a horizontal field of view perpendicular to the line of sight of a few km. Each of the detectors is designed to have a different frequency passband, through a combination of filtering at the detector and blocking in the optical system (Moorhouse, 2003). Channels are distributed over a range of wavelengths from 6.12–17.76 μm, corresponding to the radiance peaks of important atmospheric components (figure 2.3).

Figure 2.2(a) shows the basic internal layout of HIRDLS, and figure 2.2(b) an optical schematic. Key features to note are the aperture and scan mirror, both to be discussed further in section 2.4.



(a) Physical schematic of HIRDLS



(b) Optical schematic of HIRDLS

Figure 2.2: *HIRDLS schematics*

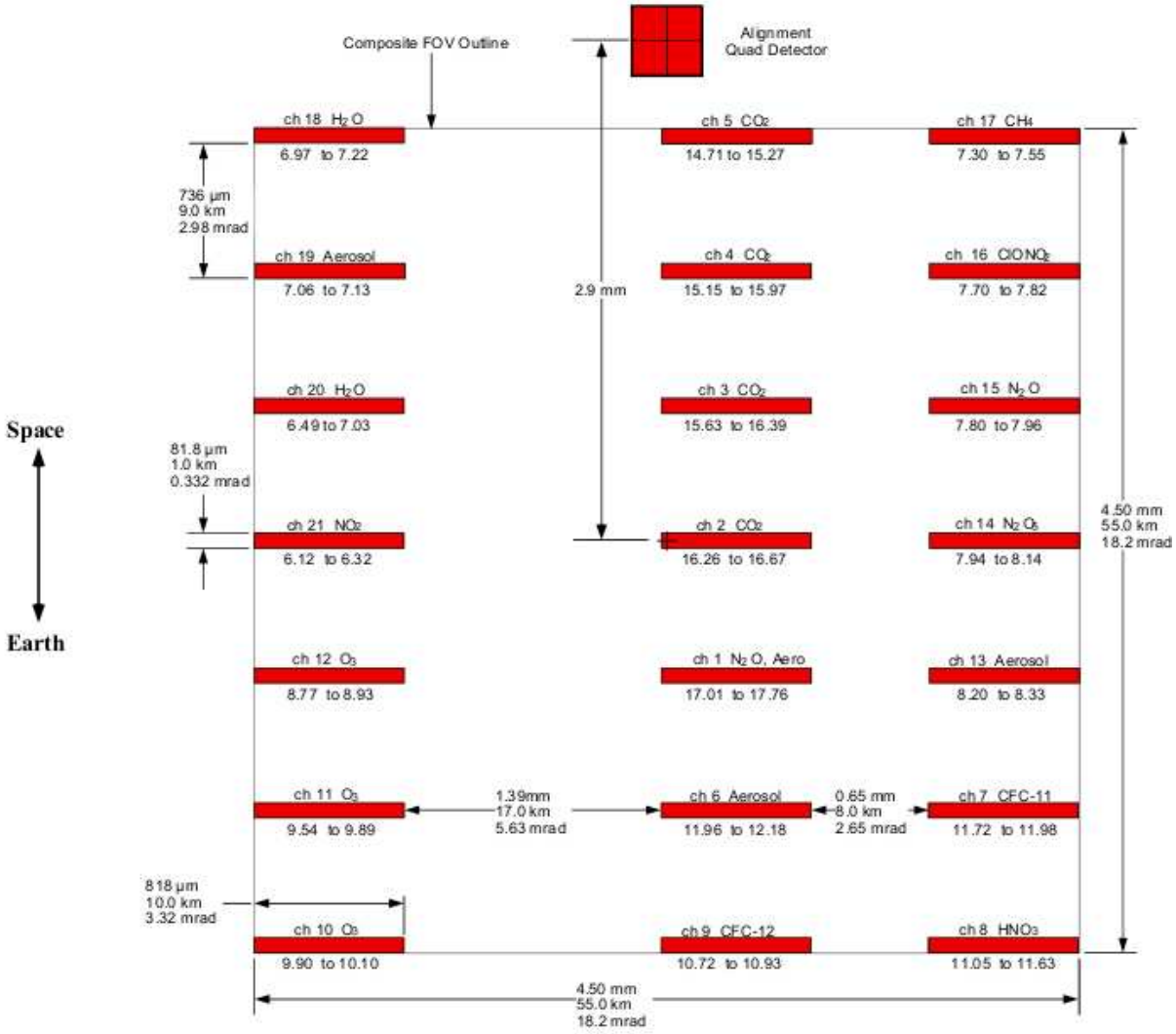


Figure 2.3: *Diagram showing the distribution of channels on the HIRDLS detector plate. Distances on the diagram in kilometers refer to distances at the tangent point. The quad-detector at the top of the diagram is used for alignment purposes*

Incident radiation passively detected from the atmospheric limb enters the instrument through the main aperture and is reflected off the 2-axis scan mirror onto the primary mirror, which focuses the beam onto a rotary chopper that chops the limb view with a view of space from the side of the instrument. The chopped signal is insensitive to the temperature of any preceding component in the optical train, enhancing signal precision.

The beam is then recollected onto the secondary mirror and directed through a series of band-defining warm multi-layer dielectric interference filters. It is then passed through two lenses and focused into a detector dewar held at 62 K, passing through an aperture stop to reduce any out of field signal which may have reached this point though the system. Finally, the image is focused onto the main detector plate, where 21 independent interference filters, each with a bandpass slightly larger than that at the warm-filtering stage, lead to 21 HgCdTe detectors. This double-filtering arrangement serves to reduce ghosting from refractive optical components in the system, and also improves out-of-band signal rejection (Barnett et al., 2003; Gille et al., 2003; Moorhouse, 2003).

Each detector is nominally designed to be the equivalent of 1 km high by 10 km wide when projected onto the limb, in order to give high vertical resolution while still receiving a strong signal from the atmosphere, but is in practice slightly larger than this in the vertical. The scanning mode utilised by HIRDLS for the majority of the mission (internally known as Scan Table 23, used from the 1st of May 2006) performs 27 pairs of vertical up and down scans lasting around 15.5 seconds ($\sim$7.75s each way), followed by a 1 – 2 second space view before the next scanning sequence, taking measurements at 120 vertical levels; the scanning modes used before this utilised a similar scanning sequence but with a slightly altered spaceward limit on the scans (Barnett and Gille, 2006a).

2.3 Limb Sounding and Temperature Retrieval

2.3.1 Limb Sounding and Radiative Transfer

Figure 2.4 shows the basic principle behind the limb sounding process. As the optical path of the instrument is directed through the atmosphere at a given angle, it passes through multiple atmospheric layers of varying density. Atmospheric density decreases approximately exponen-

tially with height (Andrews, 2000), and so a density maximum exists at the point in the optical path closest to the planetary surface. This point, known as the tangent point, contributes the bulk of the radiance signal measured. By shifting the optical path, for example by means of internal mirrors within the instrument, we can adjust the height of the tangent point and hence the height of our atmospheric measurements. This allows us to scan through the atmosphere vertically and collect measurements at a number of points. Figure 2.4 illustrates this for three tangent point heights $z_{tan,a}$, $z_{tan,b}$ and $z_{tan,c}$.

Assuming an optimally functioning instrument, the radiance measured consists of the sum of the radiance emitted throughout the entire optical path of the instrument. If we assume local thermodynamic equilibrium (LTE) and negligible scattering, then we can state that the radiation emitted from any given point is given by the Planck function

$$B(\nu, T) = \frac{2hc^3\nu^3}{\exp(hc\nu/k_BT) - 1}, \tag{2.3.1}$$

where h is Planck's constant, c the speed of light, k_B Boltzmann's constant, T the temperature of the emitter, and ν the wavenumber of the signal. If we also define the transmittance between a source at position x_s and the instrument detector at position x_d as

$$\tau(\nu, x_s, x_d) = \exp\left(-\int_{x_s}^{x_d} \Sigma_i k_i(\nu, x)\, \rho_i(x)\, dx\right), \tag{2.3.2}$$

where the summation is over i absorbers of monochromatic absorption coefficient $k_i(\nu, x_{obs})$ and number density $\rho_i(x)$, then we can state that the radiance at wavenumber ν measured by the instrument after a distance x is

$$L(\nu, x_d) = L(\nu, x_s)\,\tau(\nu, x_s, x_d) + \int_{x_s}^{x_d} B(\nu, T(x))\,\frac{\delta\tau(\nu, x, x_d)}{\delta x} dx, \tag{2.3.3}$$

where $\delta\tau$ is the small change in transmittance and δx is the small change in position, and therefore that

$$L(\nu, x_d) = L(\nu, x_s)\,\tau(\nu, x_s, x_d) + \int_{x_s}^{x_d} B(\nu, T(x))\,\tau(\nu, x_s, x_d)\,\Sigma_i k_i(\nu, x)\,\rho_i(x)\,dx \tag{2.3.4}$$

For infrared emissions detected via limb sounding, the background emission term $L(\nu, x_s)$

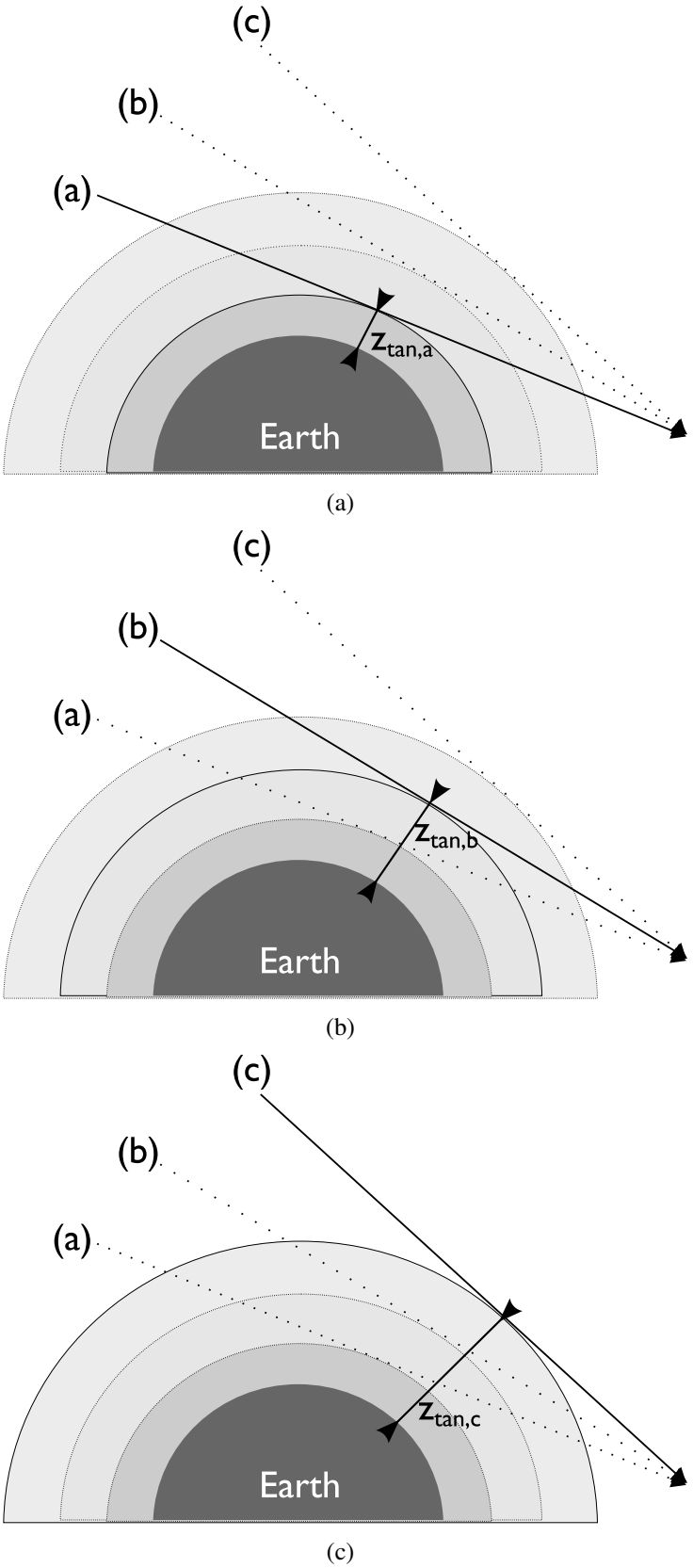


Figure 2.4: *Illustration of how the tangent point of a limb-sounding instrument varies as the optical path is shifted in angle.*

is neglible due to the cold background of space, and hence the detected radiance in the line of sight of the instrument

$$L(\nu, x_d) = \int_0^\infty B(\nu, T(x))\,\tau(\nu)\,(\Sigma_i k_i(\nu, x)\,\rho_i(x))\,dx, \tag{2.3.5}$$

where we have replaced our integral from the source to the detector with one along the (infinite) line of sight, and simplified the notation used for the transmittance $\tau(\nu, 0, \infty)$ to just $\tau(\nu)$.

It can be shown geometrically that the majority of the optical path through the atmosphere will lie at or very close to the tangent height z_{tan}. Additionally, as discussed above, the density term in equation 2.3.5 combined with the exponential density-dependence of the atmosphere with height means that the dominant part of the signal comes from the tangent point. Hence, we assume that the whole line of sight is at the same temperature as the tangent point, and that our measurement comes from a infinitely narrow field of view, allowing us to state that

$$\begin{aligned} L(\nu, z_{tan}) &\simeq B(\nu, T(z_{tan})) \int_{-\infty}^{z_{tan}} \tau(\nu)\,\Sigma_i k_i(\nu, x)\,\rho_i(x)\,dx \\ &\equiv B(\nu, T(z_{tan}))\,\epsilon(\nu, z_{tan}), \end{aligned} \tag{2.3.6}$$

where $\epsilon(\nu, z_{tan})$ is the emissivity of the path, $\epsilon = 1 - \tau$. This tells us that, if we know the distribution of the emitting gas, we can compute the temperature of the gas by inverting the Planck function $B(\nu, T(z_{tan}))$. If we examine the radiation emitted from gases which have well-known distributions through the atmosphere, such as CO_2 or O_2 which are nearly uniformly mixed, then, provided no other atmospheric constituents overlap the measurement band in frequency, the measured radiances can tell us the temperature of the emitting gas. Conversely, if we know the temperature of the emitting gas, then we can obtain $\epsilon(\nu, z_{tan})$, and hence derive the amount of absorber present, that is to say the concentration of the gas (Houghton et al., 1984).

Obviously, our assumption that the whole line of sight is at the height of the tangent point introduces some inaccuracy to our measurement. In practice, a range of heights will be sampled at any given tangent height, governed by a weighting function $\kappa(z)$ which weights the contributions to $L(\nu)$ from different altitudes z. Moorhouse (2003) shows that this weighting function

Figure 2.5: *Calculated HIRDLS weighting functions for channel 2 for the field of view as measured before launch, reproduced from Moorhouse (2003).*

can be expressed as a shape of the form

$$\kappa(\nu, z) \propto \frac{1}{\sqrt{z - z_{tan}}} \exp\left(-\frac{z - z_{tan}}{H}\right) \tag{2.3.7}$$

where z is the height in the atmosphere, derived by considering the relationship $(R + z_{tan})^2 + x^2 = (R + z)^2$, for radius of the Earth R, and H is the atmospheric scale height. Figure 2.5 shows, as an example, the resulting weighting functions for the CO_2-sounding HIRDLS channel 2 for a range of tangent heights. As this figure shows, measurements at a lower altitude will have a significant contribution from vertically higher levels, whilst vertically higher measurements experience this problem to a significantly lesser degree. Consequently, as our measurements reach lower altitude levels, they become much less finely resolved in height, providing a lower bound to the tangent heights measurable by the instrument.

As the instrument scans vertically, the satellite also moves along its orbital path, causing a shift in the geolocation of the measurements as a function of height. We can make an estimate of this distance for HIRDLS. Aura completes 14.25 orbits a day, with an orbital period of around 6060 seconds, and performs a vertical scan every 8 seconds. Accordingly, in the time taken to perform a complete vertical scan, Aura will have travelled

$$\left(\frac{2\pi \times R_E \times 14.25}{24 \times 60 \times 60} \times 8\right) \text{m}, \tag{2.3.8}$$

where R_E is the radius of the Earth, giving a distance travelled per scan of 54 km. A full

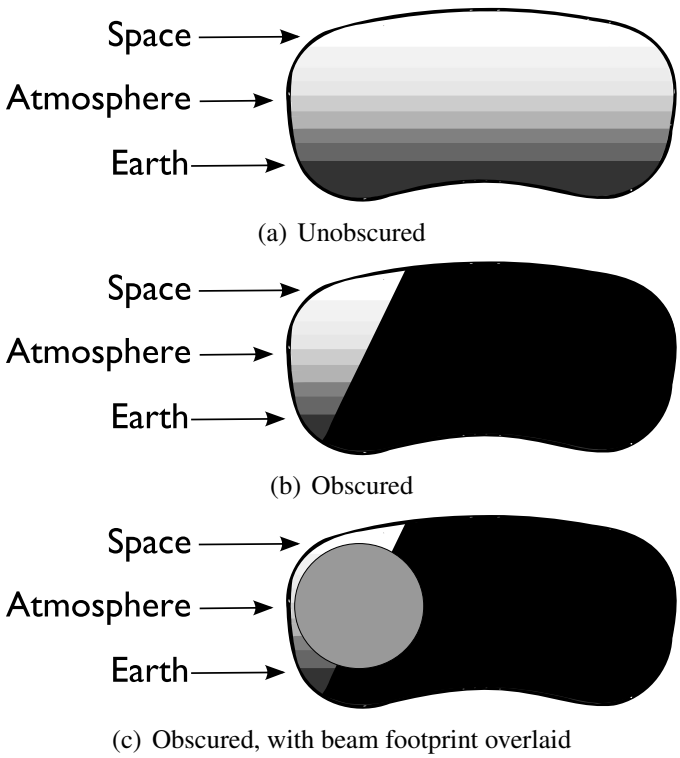
vertical scan runs from the surface to a height of ∼121 km and consists of 121 levels, and hence the horizontal distance along-track between individual height levels is approximately 0.5 km. Over the height range used in this thesis to estimate gravity wave signals, 15 km—60 km, this equates to a range over the scan period of 20 km, or a variation of ±10 km from the mean point of the measurement.

2.3.2 Retrieval of Temperature from Radiance

The method used to retrieve temperatures from radiance in the HIRDLS data is beyond the scope of this thesis, but follows the method described by Lambert et al. (1999), with appropriate modifications due to the issues described in sections 2.4 and 2.5. Briefly, an *a posteriori* retrieval algorithm is used, based on a radiative transfer forward model with an initial guess derived from GEOS-5 assimilated meterological data and an *a priori* value derived from the CIRA86 atmospheric climatology. This is described further by Gille et al. (2008) and (Khosravi et al., 2009). The averaging kernels for the temperature results are narrowly peaked, with an FWHM of ∼1 km, and have "almost no smoothing by the a priori" (Khosravi et al., 2009).

2.4 Post-Launch Hardware Problems

Aura was launched on the 15th of July 2004 into an ascending-node 705 km sun-synchronous polar orbit at 98° inclination, crossing the equator at 13:45±15min local solar time, with an intended design lifetime of five years (Schoeberl et al., 2006). Following launch and 24 days of outgassing, the instrument coolers were activated during the evening of the 9th of August, and reached the designed operating temperature of 62K early on the 10th of August. At this point, the sunshield door was opened, and initial radiance measurements were taken. Several problems immediately became apparent: a significantly higher and more uniform level of measured radiance than would be expected was observed at all viewing angles other than a small region at extreme azimuthal angles, the in-flight calibration black body view was completely blocked (Gille et al., 2008), and more drag was encountered when moving the internal mirrors than had been anticipated pre-launch, with the elevation mechanism encountering around 20% more damping than in ground-based measurements (Barnett et al., 2005; Barnett and Gille, 2006a).



(a) Unobscured

(b) Obscured

(c) Obscured, with beam footprint overlaid

Figure 2.6: *Artist's impression of the impeded field of view of the HIRDLS instrument.*

From these data, together with tests on the ground-based engineering model and additional purpose-built rigs, it was concluded that a blockage was obscuring a significant proportion of the aperture and touching part of the scan mirror (Barnett et al., 2005). Numerous attempts were made to dislodge or shift the blockage by means of door and mirror movements, but were unsuccessful.

The blockage was later ascertained to most probably be due to a sheet of Kapton thermal blanketing material, used to line the optical cavity to ensure cleanliness, coming loose during launch. The most probable cause of this was that one of the two layers of this film could have separated during the depressurisation caused by the launch, with the inner layer being dragged up towards the aperture, snagging and tearing on the back of the scan mirror as it did so. The force of the launch rocket could then have pulled it over the scan mirror, giving the observed results (Gille et al., 2008).

This blockage, covering around 80%–90% of the instrument's aperture (figure 2.6), results in a significant loss of signal entering the detector system, and also introduces several errors into the results obtained (section 2.5.1). With the scanning mirror at its most extreme azimuth of around 47° from the velocity vector and the sunscreen door fully retracted, a partially clear

view of the external scene can be obtained, and it is believed that good atmospheric profile results can be measured (Barnett et al., 2005; Barnett and Gille, 2006b; Schoeberl et al., 2006), at the cost of a loss of coverage of the Antarctic continent and a reduction in the longitudinal resolution of measurements (Barnett and Gille, 2006a) (see figure 6.1(e) for an impression of the resulting data coverage). Due to the viewing angle being at such a large horizontal angle from the orbital track of the satellite, the blockage also precludes the intended simultaneous coincident measurements with other A-Train and Aura instruments, in particular the Microwave Limb Sounder (Gille et al., 2008).

Aside from this key issue, the instrument operated as well as or better than expected from launch until March 2008, allowing it to achieve a significant fraction of the mission objectives originally set (Gille et al., 2005; Hepplewhite et al., 2005). However, in March 2008, the instrument chopper stopped rotating, blocking the atmospheric signal completely from the detectors. Attempts continue to restart the chopper, but have as yet been unsuccessful. Accordingly, HIRDLS data are currently only available for the period from January 2005 to March 2008.

2.5 Correcting for the Obscuration

Due to the physical nature of the obscuration, the loss of spatial coverage due to the Kapton blockage cannot be resolved. However, extensive work to compensate for the radiometric errors introduced by the Kapton blockage has been done at the University of Colorado at Boulder (UCB) and at Oxford, allowing recovery of the data obtained for the reduced region of coverage. This section closely follows Gille et al. (2008) and Eden et al. (2005), and the work is described in further detail by those sources.

2.5.1 Description

The blockage emits a strong radiance signature, which is the most probable cause of the unexpectedly high radiances measured by the instrument in initial tests. This signal is a combination of the emission from the material itself and some backscattered signal from inside the instrument. Additionally, at the high azimuthal angle where the view is least obstructed, the optical

beams from the detector plate are significantly out-of-focus, resulting in a very wide beam footprint. As shown in figure 2.6(c)[1], this means that the region of the aperture which remains open is significantly smaller than the footprint of the optical beam from the detector plane. Hence, even with the azimuthal angle set at this extreme position, any measured signal will include some component due to the blockage rather than the intended atmospheric measurement. This problem is particularly severe at the bottom of the aperture, where the obstruction is greatest. The blockage also prevents viewing of the onboard calibration black body.

As the beam scans vertically, the signal varies in part due to the blockage and in part due to the atmosphere which we intend to measure. At the top of the scan, only space is observed through and around the blockage, and consequently the measurement is a combination of the blockage and a negligible signal from space, whilst at the bottom the signal consists of the blockage and the desired atmospheric signal. Due to the shape of the remaining opening, as the instrument scans vertically upwards the signal will decrease as less of the beam is obstructed by the blockage, and vice-versa. An additional oscillation overlies this signature due to movement of the Kapton film: as the scan mirror is moved vertically, oscillations are created in the film due to vibration.

With extended study, this behaviour has become well-understood, and consequently methods have been developed to compensate for the problems introduced. The aim of these methods is to correct the signals in order to restore the radiances to as near as possible to what would be observed by an equivalent unobscured instrument.

2.5.2 The Kapton Correction

2.5.2.1 Calibration

As initially designed, measured radiances would have been calibrated from the space view at the top end of each limb scan, allowing the computation of a radiometric offset for each channel. The long-term gain to the signal over the instrument's lifetime would then have been calculated via a comparison to the in-flight calibration (IFC) black body, with a scan via the IFC mirror once per minute. However, due to the blockage, neither of these approaches remain possible.

[1] Directional references in the text to the aperture and blockage should be taken to refer to this diagram unless otherwise stated

During ground calibration of the instrument, data suggested that the space view signal could be estimated to a high degree of accuracy from the temperatures of the reflecting surfaces in the instrument optical path (i.e. the primary, secondary and space view mirrors and the back of the optical chopper blade). As this is a minimal source of error, this is how the offset is now computed. However, the blockage of the in-flight calibration black body means that the long-term gain to this signal cannot be easily measured. Instead, it is necessary to resort to a combination of reliance on prelaunch measurements and comparison of the measured radiances to long-term climatological means to check for any deviation from this. No significant deviation has been observed to date.

2.5.2.2 Pitch Manoeuvres

In order to study the effects of the blockage without the additional complications introduced by the atmospheric signal, several pitch manoeuvres have been carried out by the Aura spacecraft. In these manoeuvres, the whole spacecraft is pitched down by 5.25°, raising the HIRDLS line-of-sight (which is directed out of the rear of the spacecraft) above the top of the atmosphere for all optical channels. In this configuration, only the signal from the blockage is measured, allowing better characterisation of its effects. This manoeuvre prevents the collection of atmospheric data by all four instruments on Aura (HIRDLS, MLS, TES and OMI) for the period of the pitch, and hence is only done sparingly.

The success of this method relies on the assumption that the behaviour of the Kapton blockage is not affected significantly by the change in pitch angle. Studies at an intermediate pitch angle of 2.62° suggest that this is a fair assumption.

2.5.2.3 Deoscillation

The first element in the actual correction of the data is to remove the oscillatory signal discussed above. As the scan mirror moves vertically through elevation angle θ, contact between the mirror and the blockage causes vibrations in the blockage, with a sharp spectral peak near 1.9 Hz (as compared to a vertical scan time $\sim$15 s).

Data from pitch manoeuvres suggest that this oscillation has a typical peak-to-peak amplitude in the upper part of the scan $\sim 12 \times 10^{-4}\,\mathrm{Wm^2sr}$, which compares to a noise level

$\sim 1 \times 10^{-4}\,\mathrm{Wm^2sr}$. These peak amplitudes are excited at certain elevation angles ('trigger points'), which are presumably where the contact between mirror and blockage occurs. The oscillations behave differently depending on the scan direction due to the way in which they are triggered by this contact, and hence the directions are compensated for separately. This may have ramifications on the final corrected product, but also provides reassurance that any wave signals we may detect in the data are not introduced due to this process: an oscillation introduced due to the corrections process will appear differently in the scans in each direction, and hence not in adjacent profiles.

To model these oscillations, the radiometric data from the pitch manoeuvres were analysed via singular vector decomposition (SVD), giving 16 empirical orthogonal functions (EOFs) spanning the range of elevation angles. The oscillations observed in the space view at the top of the scan determine the amplitudes of these EOFs, which are then subtracted from the data[2]. This removes the oscillation from the signal, with no residual effects observed in the radiances or the final retrieved products.

2.5.2.4 Radiance Correction

The next step in the correction is to remove the enhanced radiance levels produced by the blockage. As discussed above, the blockage emits radiance into the optical path of the instrument, as it emits and reflects radiation. The magnitude of this emission varies with orbital position, and increases as scans move downwards towards the Earth due to the blockage obscuring a larger portion of the beam. This is the largest of the problems affecting the dataset, with consequent ramifications for the instrument's accuracy.

The current correction method is based upon data from a single pitch manouevre, that of the 6th of July 2006. Data from up and down scans were analysed separately to compute a mean effect, and departures from this mean analysed once again via SVD to produce EOFs governing the deviations from this mean. Data from the top of the scan are used to scale the amplitudes of the EOFs, and the resultant estimate of the additional radiance due to the blockage is removed. This is illustrated in panels (a) and (b) of figure 4 of Gille et al. (2008).

The current approach has been shown to work well for channels with large signal-to-noise

[2]This inherently assumes that the oscillations are an additive radiative term and do not modulate the incoming radiation.

ratios (SNRs), namely temperature T and ozone O_3, but does not compensate sufficiently for the retrieval of results from channels with lower SNR. Consequently, work continues to improve upon this method, in order to allow the retrieval of additional chemical species from the collected data. Residual errors due to this step are $\sim 10 \times 10^{-4}$ Wm^{2}sr, i.e. ten times as large as the noise level.

2.5.2.5 Open Area Fraction Correction

After the radiances have been corrected for the influence of the blockage, they then need to be scaled to compensate for the reduced area visible through the gap. By comparing the measured signal at the viewing angle of 47° and at an azimuthal angle ϕ where only the blockage is visible, an estimate for the open area fraction D as a function of elevation angle θ has been computed (panel (c) of figure4 of Gille et al. (2008)). The radiances obtained from the radiance correction step previously are then scaled by this, such that a corrected radiance profile

$$R_c = \frac{R - R_s}{D}, \tag{2.5.1}$$

is obtained, for measured radiance R, radiance correction R_s, and corrected radiance R_c.

2.5.2.6 Final Radiance Adjustment

Following the implementation of the above correction steps, the resulting radiances were found to have the anticipated form, with features such as the tropical tropopause in the right location, but with magnitudes reduced from those expected based on an analysis of the GEOS-5.01 dataset produced by the Goddard Modelling and Assimilation Office (GMAO). To correct for this, a final adjustment is made to the radiance values, based on this data. Measured radiances for channels 2–5 (the CO_2 channels used to measure temperature – see figure 2.3) are compared to the radiances that would be expected based on the GMAO data. This comparison is based upon the data for the 18th of May 2006.

Work was undertaken to examine the latitudinal effect of this difference, but the required correction was found to be the same at all latitudes, and in both ascending (northbound) and descending (southbound) orbits, for this day. Consequently, a scaling factor is applied uniformly to the entire mission. However, there could still be some residual latitudinal variation in the

resulting corrected data.

The size of the correction varies with height, and only data above 9 km in altitude are scaled, with the upper limit of the correction varying between 50 and 65 km depending on channel, at which point the scaling then smoothly varies towards unity over an additional 15 km upwards. The value varies with height and latitude, from approximately 1.1 at low altitudes to 0.8 at the top of the adjustment range. With time, it is hoped that the other corrections can be improved sufficiently that the required scaling factor will approach unity at all levels, and this step will no longer be necessary. Consequently, this step acts as a diagnostic of the correction algorithms used.

2.5.3 Accuracy of Correction

Barnett and Gille (2008), together with several recent papers (e.g. Kinnison et al., 2008; Nardi et al., 2008) and multiple internal documents (e.g. Waymark, 2009; Millán Valle, 2007), discuss the validation of the HIRDLS temperature product against a range of other data sources, including sondes, ground- and air-based LIDAR, ACE-FTS, MLS, ECMWF operational analyses, MIPAS and COSMIC.

Analyses of the dataset version used in this thesis, V004, suggest that the most recent data corrections, as described above, give temperature results which closely agree with those of other instruments, including the resolution of fine structure to high vertical resolutions. Temperatures in V004 typically agree with ECMWF analyses to within 1 K from 400 hPa to 1 hPa and with sample radiosondes to within 0.5 K from $\sim$300 hPa to above 10 hPa (Barnett and Gille, 2008).

Analyses (Gille et al., 2008) of a previous version of the dataset, V003, showed a small positive bias of around 1–2 K from 200 hPa to 10 hPa, above which HIRDLS temperatures are then slightly lower than other sources. Random errors range from $\sim$0.5 K at 20 km altitude to $\sim$1 K at 6 km, close to predicted values.

Chapter 3

Gravity Waves

Gravity waves are a type of wave in which fluid pressure, density, temperature and velocity fluctuate together, and arise in stratified fluids as a result of the interaction between buoyancy and gravity. As we would expect, these waves arise in the atmosphere just as in any other fluid, and can propagate over considerable distances without a corresponding transport of mass.

3.1 Motivation

Atmospheric gravity waves appear on a large range of scales, from localised internal gravity waves to planetary-scale Rossby waves (Andrews, 2000)[1], and contribute significantly to the behaviour of the atmosphere, making up "an essential part of the dynamics of the atmosphere on all meteorological scales" (Nappo, 2002). They strongly affect the circulation of the middle atmosphere by transporting and redistributing momentum, and are believed to largely determine the large-scale structure of the middle atmosphere (Fritts, 1984), accounting for the majority of temperature and velocity variances at small- and mesoscales (Fritts and Luo, 1992).

In a fluid bounded above and below, such as the ocean, gravity waves propagate primarily horizontally, as any waves which propagate vertically are reflected from the upper and lower surfaces to create standing waves. However, in a fluid with no significant upper boundary, such as the atmosphere, gravity waves can propagate freely in the vertical direction, transporting energy and momentum to higher levels. For such waves, phase is a function of height (Holton, 1979), and amplitude increases with height as $\exp(z/2H)$ (for altitude z and scale height H)

[1] In this thesis, the term 'gravity wave' shall be taken to exclude planetary-scale waves unless otherwise specified.

due to the decreasing atmospheric density (Lindzen, 1973; Herzog et al., 2001). Vertically-propagating gravity waves carry a flux of horizontal momentum, transferring it away from low altitudes and returning it to the mean flow at altitudes and locations far removed from the region of wave generation, via processes such as dissipation, saturation, wave-breaking and critical-level absorption (Lane et al., 2001). Inclusion of parameterisations of these processes into numerical weather prediction and climate models have significantly improved their predictive capability (Karoly et al., 1996). Accordingly, an improved understanding of gravity wave processes is important to our understanding of the climate system as a whole.

Historically, the earliest global climate models (GCMs) neglected the contribution of gravity waves to large-scale atmospheric flow. However, studies such as those of Jones and Houghton (1972) and Lindzen (1973) suggested that such waves (and, more specifically, their breaking) might provide a source of thermal forcing necessary to explain the observed structure of the upper and middle atmosphere. This would serve to slow down the modelled zonal winds (Meriweather and Gerrard, 2004) to closer to their measured values, and hence allow significant improvements in model accuracy to be achieved by treating gravity waves with a parameterisation. Physically simplistic models developed from the early 1980s onwards have confirmed that this does indeed help to improve models, but when applied to full-scale GCMs, results have tended to be variable at best. Accordingly, significant additional observations are required to help parameterise gravity waves, in order to further enhance these models by producing a global climatology of gravity waves and their generation mechanisms (Meriweather and Gerrard, 2004; Fritts and Alexander, 2003; Hamilton, 1999).

As a result, the study of gravity waves has become increasingly important, with extensive studies undertaken to elucidate their global distribution and their underlying physics. Such waves are believed to arise from a wide variety of sources, including but not limited to topography (Bacmeister et al., 1994; Carslaw et al., 1998; Leutbacher and Volkert, 2000b, etc; see also section 3.3.1), convection (Alexander et al., 2004; Clark et al., 1986, section 3.3.2.1), and wind shear (Fritts and Alexander, 2003). Once generated, the propagation of gravity waves is complicated by the nature of the middle atmosphere: winds varying in space and time cause refraction, reflection, ducting and focusing; instability processes generate turbulence which can either enhance or disrupt the wave; and interaction with other middle atmospheric phenomena

can exchange energy between these processes and the wave (Fritts and Alexander, 2003).

3.2 The Fluid Dynamics of Gravity Waves

As discussed above, gravity waves are a fluid-dynamical phenomenon arising from the interaction between gravity and buoyancy. Consequently, their behaviour is governed by the fundamental equations of fluid dynamics, and we can make some simple predictions about their basic form.

Two key types of gravity wave exist in the atmosphere: internal gravity waves and inertia-gravity waves. In this context, internal waves are those waves with short enough length scales (tens to hundreds of kilometres in the horizontal) that the rotation of the Earth can be neglected, and inertia waves those with long enough length scales to be affected by Coriolis effects.

3.2.1 Internal Gravity Waves

It is useful to derive the expected behaviour of internal gravity waves. The following derivation borrows heavily from those of Houghton (1977), Holton (1979), and Lynch and Cassano (2006). We start with the Navier-Stokes equation, an expression of Newton's second law in fluid-dynamical terms:

$$\frac{D\mathbf{u}}{Dt} + \frac{1}{\rho}\nabla p + g\hat{\mathbf{k}} = \frac{\mu}{\rho}\nabla^2\mathbf{u} \tag{3.2.1}$$

where ρ denotes the density of the fluid, $\mathbf{u}$ the velocity, p the pressure, g the acceleration due to gravity in the z-direction, μ the dynamic viscosity, and $\hat{\mathbf{k}}$ is a unit vector in the z-direction. The term $D/Dt \equiv \partial/\partial t + \mathbf{u}\,.\nabla$ is known as the 'advective' or 'material' derivative, and represents the time-rate of change following the motion of a 'fluid blob', as opposed to $\partial/\partial t$, the rate of change with respect to time at a fixed point (Andrews, 2000).

To proceed further, we neglect frictional effects and make the Boussinesq approximation, whereby we assume that variations in density only have an effect when coupled with gravity, that is to say that such variations have no effect on the horizontal force balance. We also assume that flow is two-dimensional in the x–z plane, such that the flow speed in the y-direction v is zero, i.e. that the velocity of the fluid can be represented as $\mathbf{u} = (u, 0, w)$, and that the temporal and spatial scales of the gravity waves are such that their Rossby number is low enough for

terms arising due to Coriolis forces to be comparatively small. For these assumptions, equation 3.2.1 simplifies to the set of equations (Lynch and Cassano, 2006)

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0 \tag{3.2.2a}$$

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + w\frac{\partial w}{\partial z} + \frac{1}{\rho}\frac{\partial p}{\partial z} = 0 \tag{3.2.2b}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0. \tag{3.2.2c}$$

We further assume that the flow is adiabatic, allowing us to express the potential temperature θ in the form

$$\frac{\partial \theta}{\partial t} + u\frac{\partial \theta}{\partial x} + w\frac{\partial \theta}{\partial z} = 0. \tag{3.2.3}$$

We now apply perturbation analysis to our expressions, whereby the variables in equations 3.2.2 are described as being the sum of an underlying background flow state and a small perturbation to that flow, i.e. for a given variable Ψ, $\Psi = \bar{\Psi} + \Psi'$, where $\bar{\Psi}$ is the background and Ψ' the perturbation. If we assume that the background vertical flow $\bar{w}$ is zero and that pressure and potential temperature vary with height as $p = \bar{p}(z) + p'$ and $\theta = \bar{\theta}(z) + \theta'$ respectively, then we can describe the terms in equations 3.2.2 and 3.2.3 as

$$\begin{aligned} u &= \bar{u} + u' \\ w &= w' \\ p &= \bar{p}\,(z) + p' \\ \theta &= \bar{\theta}\,(z) + \theta' \\ \rho &= \bar{\rho} + \rho' \end{aligned} \tag{3.2.4}$$

If we further assume that hydrostatic equilibrium $d\bar{p}/dz + \bar{\rho}g = 0$ applies then, by substituting these perturbation forms into equations 3.2.2 and 3.2.3, we can rearrange and eliminate small terms,i.e. those which are quadratic in Ψ' and apply the approximation

$$\frac{1}{\bar{\rho} + \rho'} = \frac{1}{\bar{\rho}}\left(1 + \frac{\rho'}{\bar{\rho}}\right)^{-1} \approx \frac{1}{\bar{\rho}}\left(1 - \frac{\rho'}{\bar{\rho}}\right) \tag{3.2.5}$$

to obtain

$$\frac{\partial u'}{\partial t} + \bar{u}\frac{\partial u'}{\partial x} + \frac{1}{\bar{\rho}}\frac{\partial p'}{\partial x} = 0 \tag{3.2.6a}$$

$$\frac{\partial w'}{\partial t} + \bar{u}\frac{\partial w'}{\partial t} - \frac{\theta'}{\bar{\theta}}g + \frac{1}{\bar{\rho}}\frac{\partial p'}{\partial z} = 0 \tag{3.2.6b}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0 \tag{3.2.6c}$$

$$\frac{\partial \theta'}{\partial t} + \bar{u}\frac{\partial \theta'}{\partial x} + w'\frac{\partial \theta'}{\partial z} = 0. \tag{3.2.6d}$$

If we already know the underlying state $(\bar{u}, \bar{w}, \bar{p}, \bar{\theta})$, equations 3.2.6 become four equations in four unknowns (u', w', p', θ'), which can be reduced (Lynch and Cassano, 2006) to

$$\left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right)\left(\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2}\right) + N^2\frac{\partial^2 w'}{\partial x^2} = 0, \tag{3.2.7}$$

i.e. a single equation in a single unknown w', where N represents the Brunt-Väisälä or buoyancy frequency, $N^2 = -(g/\rho)\partial\rho/\partial z$. By assuming a wavelike solution to equation 3.2.7 of the form $w' = w_0 \cos(kx + ly + mz - \omega t)$, we hence derive the dispersion relationship for gravity waves,

$$(\omega - \bar{u}k)^2\left(k^2 + m^2\right) - N^2k^2 = 0 \tag{3.2.8}$$

where w_0 is the wave amplitude, ω the frequency of the wave, and k and m the horizontal and vertical components of the wavevector $\mathbf{k}$. If we now neglect the background flow $\bar{u}$, the dispersion relation becomes

$$\omega = \pm\frac{Nk}{|\mathbf{k}|} \tag{3.2.9}$$

which can be substituted into the general expression for phase speed in direction α, $c_\alpha = \omega/\mathbf{k}_\alpha$ to obtain the result

$$c \equiv (c_x, c_y, c_z) = \left(\frac{\omega}{k}, 0, \frac{\omega}{m}\right) = \left(\frac{\pm N}{|\mathbf{k}|}, 0, \frac{\pm Nk}{m\,|\mathbf{k}|}\right) \tag{3.2.10}$$

and similarly into the general expression for group velocity $c_{g,\alpha} = \partial\omega/\partial\mathbf{k}_\alpha$ to obtain

$$c_g = \left(\frac{\pm Nm^2}{|\mathbf{k}|^3}, 0, \frac{\mp Nkm}{|\mathbf{k}|^3}\right). \tag{3.2.11}$$

It can be shown by considering all possible combinations of phase and group velocity that, when $\bar{u}$ is zero (Lynch and Cassano, 2006) the signs of c_x and k are always identical, that the signs of c_z and m are always the same, and hence that the wave will propagate in the direction of the wavenumber vector $\mathbf{k} = (k, 0, m)$.

We can further obtain a prediction of the internal structure of a gravity wave. In doing so, we no longer utilise the Boussinesq approximation. We first need to substitute our wavelike solution into equation 3.2.6c, giving us

$$\frac{\partial u'}{\partial x} + \frac{\partial}{\partial z}\left[w_0 \cos\left(kx + mz - \omega t\right)\right] = 0,$$
$$\therefore u' = -w_0 \frac{m}{k} \cos\left(\phi\right), \tag{3.2.12}$$

where phase $\phi \equiv (kx + mz - \omega t)$. Similar expressions in terms of ϕ can be derived for p' and θ', namely

$$p' = -\rho w_0 \omega \frac{m}{k^2} \cos\left(\phi\right), \tag{3.2.13}$$
$$\theta' = \frac{w_0}{\omega} \frac{d\bar{\theta}}{dz} \sin\left(\phi\right), \tag{3.2.14}$$

which can then be applied to an example wave.

Take a wave propagating westward and vertically downward, that is to say one with negative k and m, at a fixed time τ. As all our key variables are defined in terms of ϕ, we can consider lines of constant phase C in terms of co-ordinates (x, z) as

$$\phi = kx + mz - \omega\tau = C, \quad \therefore z = \frac{-k}{m}x + C'. \tag{3.2.15}$$

The relationships in u', w', p' and θ' arising from these are illustrated in figure 3.1

From equation 3.2.15 and the dispersion relation 3.2.9, we can also show that the angle ζ between the wavefronts of constant phase ϕ and the vertical z can be described as

$$\cos\left(\zeta\right) = \frac{k}{\sqrt{k^2 + m^2}} = \frac{k}{|\mathbf{k}|}$$
$$\therefore \omega = N \cos\left(\zeta\right). \tag{3.2.16}$$

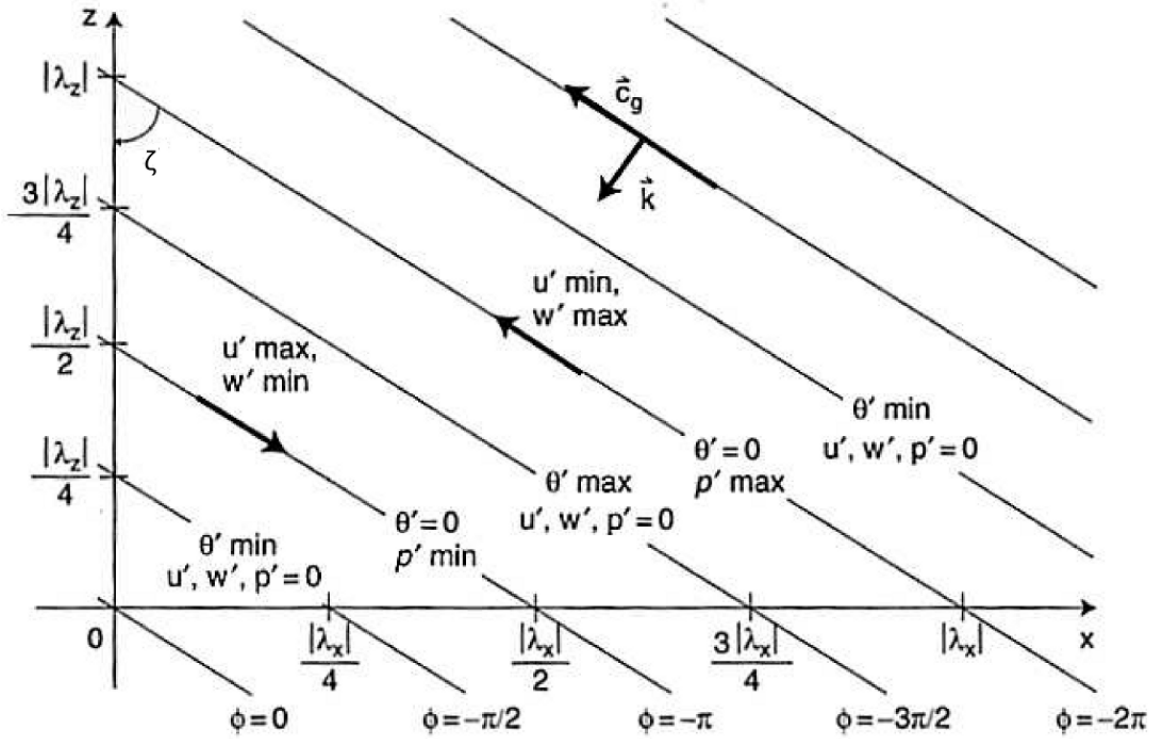


Figure 3.1: *Illustration of the phase relations between u', w', p' and θ'. The diagonal lines represent constant phase ϕ.*

This expression, first shown by Rayleigh (1883), tells us that the wave frequency ω can be related to just the buoyancy frequency N and the angle ζ. This also tells us that the frequency of our gravity waves is constrained to the range $0 \leq \omega \leq N$.

Internal gravity waves amplify vertically as $e^{z/2H}$ due to stratification of mass in the atmosphere (Salby, 1996), where H is the scale height of the atmosphere ($\sim$7 km), and range in period from a minimum of $2\pi/N$, which equates to a fraction of an hour in the atmosphere, to longer periods on the order of a day for waves with a near-vertical wavevector (Staquet and Sommeria, 2002). In the latter case, the Coriolis force becomes significant enough to be a important influence on the gravity waves, and the wave is better described as an *inertia*-gravity wave. This thesis focuses upon the detection of internal gravity waves, and accordingly the derivation of the form of inertia-gravity waves will be neglected.

3.2.2 Momentum Flux

As gravity waves propagate vertically through the atmosphere, they carry with them a pseudo-momentum flux (referred to hereafter as simply momentum flux), i.e. a vertical flux of hori-

zontal momentum. Upon reaching the upper atmosphere, wave breaking processes lead to this momentum being transferred, effectively transporting it from the troposphere to higher altitude regions. This transfer of momentum allows wave motions to communicate information from one region of the atmosphere to another (Salby, 1996). Due to the low densities of the upper stratosphere and mesosphere, momentum transferred vertically from the troposphere can hence have a significant effect on atmospheric mean flow. As gravity waves conserve momentum flux if they propagate without dissipation through a background atmosphere that varies only in the vertical (Fritts and Alexander, 2003), significant momentum can reach high altitudes in this way when vertically-propagating waves break.

As it represents a vertical flux of horizontal momentum, momentum flux in the atmosphere is a vector quantity, i.e. it has both a magnitude and a direction. We can describe a vertical flux of horizontal pseudomomentum (Fritts and Alexander, 2003) as

$$(F_{px}, F_{py}) = \bar{\rho} c_{gz} \frac{E}{\hat{\omega}} (k, l) \tag{3.2.17}$$

which, assuming that the waves propagate with minimal dissipation through a background atmosphere which varies only in the vertical, can also be expressed as

$$(F_{px}, F_{py}) = \bar{\rho} \left(1 - \frac{f^2}{\hat{\omega}^2}\right) \left(\overline{u'w'}, \overline{v'w'}\right). \tag{3.2.18}$$

and will be referred to hereafter as simply momentum flux. Here, $c_{g,z}$ is the wave group velocity in the vertical direction, E is the total gravity wave energy, k and l are the wavenumbers of the wave in the x and y directions respectively, (u', v', w') is the vector of perturbations to the underlying wind, $\overline{\rho}$ is the background atmospheric density, $f = 2\Omega \sin(\phi)$ is the Coriolis parameter (where Ω is the rotation rate of the Earth and ϕ the latitude), $\hat{\omega} = \omega - k\bar{u} - l\bar{v}$ is the wave intrinsic frequency, i.e. the frequency of the wave in a frame of reference moving with the background wind $(\bar{u}, \bar{v})$, F_{px} and F_{py} represent the momentum flux in the x and y directions, and the overline over the wind perturbations indicates the time-average over a full wave-period.

Due to wave filtering processes in the lower and middle atmosphere (see chapter 7), the wavevector is usually directed opposite to the underlying wind fields at upper stratospheric and mesospheric heights (Osprey, 2002). Accordingly, breaking waves usually act to reduce the

magnitude of the mean flow at these altitudes.

Since HIRDLS takes near-instantaneous sequential 1D profiles of atmospheric parameters as it travels along an orbital track, rather than collecting data as time series at a fixed location or measuring a 3D field, HIRDLS temperature data cannot be used to determine the direction of measured momentum fluxes. However, an estimate can still be made of the magnitude of the momentum flux based upon measured wave properties. We calculate this as follows, closely following the derivation of Ern et al. (2004).

We start by explicitly writing down the time-average using the real parts of the wind perturbations in equation 3.2.18 to obtain

$$F_{px} = \bar{\rho}\left(1 - \frac{f^2}{\hat{\omega}^2}\right)\frac{1}{\tau}\int_0^\tau \Re\{u'\}\Re\{w'\}dt, \tag{3.2.19}$$

where τ is the time-period of the wave, $\omega = 2\pi/\tau$. If we then take wavelike solutions to the properties that vary in a gravity wave, of the form

$$\left(u', v', w', \frac{\Theta'}{\bar{\Theta}}, \frac{p'}{\bar{p}}, \frac{\rho'}{\bar{\rho}}\right) = \Re\left\{\left(\tilde{u}, \tilde{v}, \tilde{w}, \tilde{\Theta}, \tilde{p}, \tilde{\rho}\right)\exp\left[i\left(kx + ly + mz - \omega t\right) + \frac{z}{2H}\right]\right\} \tag{3.2.20}$$

where z is the altitude, we can substitute the solutions for u' and w' to give

$$F_{px} = \frac{\bar{\rho}}{2}\left(1 - \frac{f^2}{\hat{\omega}^2}\right)\Re\{\tilde{u}\tilde{w}^* e^{z/H}\}, \tag{3.2.21}$$

where w^* is the complex conjugate of w. We then use the polarisation relations

$$-i\hat{\omega}\tilde{\Theta} + \frac{N^2}{g}\tilde{w} = 0 \tag{3.2.22a}$$

$$\tilde{p} = \left(\frac{\hat{\omega}^2 - f^2}{\hat{\omega} + ifl}\right)\tilde{u} = \left(\frac{\hat{\omega}^2 - f^2}{\hat{\omega}l - ifk}\right)\tilde{v} \tag{3.2.22b}$$

and the expression

$$\tilde{w} = \frac{-\hat{\omega}}{N^2 - \hat{\omega}^2}\left(m + i\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)\right)\tilde{p}, \tag{3.2.23}$$

all three of which are derived by Fritts and Alexander (2003), to obtain the results

$$\begin{aligned} F_{px} = &\left[-1 + \frac{\hat{\omega}^2}{N^2}\right] \\ &\times \left[1 + \frac{1}{m^2}\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)^2\right]^{-1} \\ &\times \left[1 + \frac{fl}{mk\hat{\omega}}\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)\right] \\ &\times \frac{1}{2}\bar{\rho}\frac{k}{m}\left(\frac{g}{N}\right)^2 \left|\tilde{\Theta}\right|^2 \exp\left(z/H\right) \end{aligned} \tag{3.2.24}$$

and

$$\begin{aligned} F_{py} = &\left[-1 + \frac{\hat{\omega}^2}{N^2}\right] \\ &\times \left[1 + \frac{1}{m^2}\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)^2\right]^{-1} \\ &\times \left[1 + \frac{fk}{ml\hat{\omega}}\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)\right] \\ &\times \frac{1}{2}\bar{\rho}\frac{l}{m}\left(\frac{g}{N}\right)^2 \left|\tilde{\Theta}\right|^2 \exp\left(z/H\right), \end{aligned} \tag{3.2.25}$$

where c_s is the speed of sound and g the acceleration in the z-direction due to gravity. Allowing for both the x and y directions in our coordinate system, the total vertical flux of horizontal momentum due to our gravity waves

$$F_{ph} = \sqrt{F_{px}^2 + F_{py}^2} \tag{3.2.26}$$

and the horizontal wavenumber $k_h^2 = k^2 + l^2$. We can hence express F_{ph} as

$$\begin{aligned} F_{ph} = &\left[1 - \frac{\hat{\omega}^2}{N^2}\right] \\ &\times \left[1 + \frac{1}{m^2}\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)^2\right]^{-1} \\ &\times \left[1 + \left(\frac{f}{m\hat{\omega}}\right)^2\left(\frac{1}{2H} - \frac{g}{c_s^2}\right)^2\right]^{\frac{1}{2}} \\ &\times \frac{1}{2}\bar{\rho}\frac{k_h}{m}\left(\frac{g}{N}\right)^2 \left|\tilde{\Theta}\right|^2 \exp\left(z/H\right). \end{aligned} \tag{3.2.27}$$

This expression is in terms of the potential temperature perturbation $|\tilde{\Theta}|$, which we earlier

(equation 3.2.20) defined as Θ'/Θ. We can replace this with the real temperature perturbation due to the gravity wave amplitude T' via the relationship

$$|\tilde{\Theta}| = \frac{\Theta'}{\bar{\Theta}} = \frac{T'}{\bar{T}}. \tag{3.2.28}$$

We still need to make some approximations to this to be in terms of quantities measurable by HIRDLS. In particular, the wave intrinsic frequency $\hat{\omega}$ is unknown. If we assume that our waves lie in the range $N \gg \hat{\omega} \gg f$, known as the midfrequency approximation, then we can state that

$$\hat{\omega}^2 = \frac{N^2 k_h^2}{m^2}. \tag{3.2.29}$$

This approximation allows us to significantly simplify our expression for F_{ph}. As the first three terms in equation 3.2.27 define the deviation from the midfrequency approximation, this assumption allows us to replace their product by 1, giving us an expression for the horizontal flux of vertical momentum in terms of quantities we can measure with HIRDLS:

$$F_{ph} = \frac{1}{2}\bar{\rho}\frac{k_h}{m}\left(\frac{g}{N}\right)^2\left(\frac{T'}{\bar{T}}\right)^2 \exp\left(z/H\right). \tag{3.2.30}$$

This describes a force per unit area due to the wave, giving us a momentum flux in units of Pascals. The term $\exp(z/H)$ compensates for the altitude dependence of $\bar{\rho}$, which decays exponentially with height. It should be noted that this expression describes merely the absolute value of the along-track component of the full momentum flux, and gives only a constraint on the full vector formulation. The use of the midfrequency approximation has been shown by Ern et al. (2004) to account for around a 10% difference to measured values in observations from the CRISTA instrument.

It is important to emphasise is only an estimate of the magnitude of the momentum flux along the vector path of the measurements made by the instrument; it is not possible with the single track of HIRDLS data to obtain an estimate of both components of the horizontal momentum vector.

3.2.2.1 Validity of the Midfrequency Approximation

The use of the midfrequency approximation for observations from HIRDLS can be justifed for most latitudes by considering the full form of the dispersion relation (Fritts and Alexander, 2003),

$$\hat{\omega}^2 = \frac{N^2\left(k^2+l^2\right)+f^2\left(m^2+1/4H^2\right)}{k^2+l^2+m^2+1/4H^2}. \tag{3.2.31}$$

For sampled horizontal wavelengths $\sim$100s of km, vertical wavelengths $\sim$10 km, and typical stratospheric values for scale height ($\sim$7 km) and the Brunt-Väisälä frequency ($\sim$0.02 rad s^{-1}), it can be shown that the intrinsic wave frequency $\hat{\omega}$ is typically $\sim 10 \times 10^{-3}$ rad s^{-1}, falling comfortably between N ($\sim 10 \times 10^{-2}$ rad s^{-1}) and f ($\lesssim 1.5 \times 10^{-4}$ rad s^{-1}).

Figure 3.2 illustrates this for three different wavelength combinations, contrasted with the limits upon the regime. The red line indicates the Brunt-Väisälä frequency N, the green line the Coriolis parameter f, and the three blue lines the value of $\hat{\omega}$ for three different wave cases. The solid blue case is that of a wave in the typical range of our studies, with a vertical wavelength λ_z = 12 km and horizontal wavelength λ_h= 700 km – these values represent approximately the mean wavelengths derived from our ST wave analysis (chapter 4). The other three lines represent more extreme cases: the dotted line represents a wave at the lower limits in size detectable by HIRDLS (λ_z = 2 km, λ_h= 200 km), the dashed line a wave flattened in the vertical but long in the horizontal (λ_z = 2 km, λ_h= 2000 km), and the alternately dashed and dotted line a wave long in the vertical and short in the horizontal (λ_z = 40 km, λ_h= 200 km). In all of these cases except for the long flattened wave, the predicted intrinsic frequencies of the waves are comfortably in the midfrequency range; the flattened case is unlikely to be observed in our results as both the extremely short vertical and extremely long horizontal wavelength both lie more than three standard deviations from the typical observed waves (chapter 4) and are also unlikely to be observed by the scan pattern of HIRDLS.

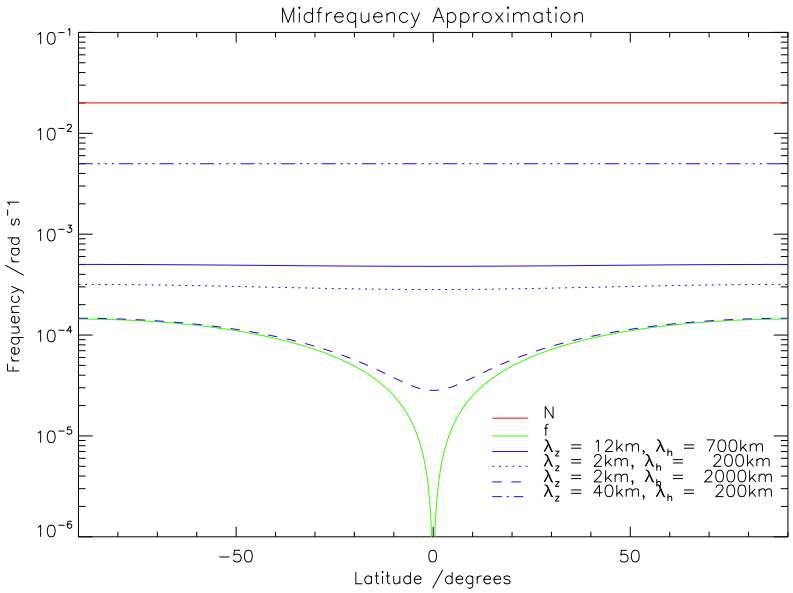


Figure 3.2: *Figure illustrating that the studied waves fall in the midfrequency range between N and f, and hence can be described mathematically using the midfrequency approximation. See text for further details.*

3.3 Sources

3.3.1 Mountain Waves

A special case of gravity waves in the atmosphere are those which arise due to flow over orography, which are called mountain waves. These waves arise frequently in a stable atmosphere, and have a wide range of generation mechanisms. As with the general case of gravity waves in a clear atmosphere, we can use the basic assumptions of fluid dynamics to obtain an idea of their physical behaviour. This derivation again follows that of Lynch and Cassano (2006).

Firstly, consider the mathematically simplified case of an infinite series of identical sinusoidal ranges expressed in terms of a perturbation to a zero-elevation mean orography $\bar{h}$,

$$h(x) \equiv \bar{h} + h' = \bar{h} + h_0 \cos(kx)\,, \tag{3.3.1}$$

where k represents the horizontal wavenumber of the series of ridges and h_0 the height of each ridge. Since the wind cannot flow through the surface, the surface flow must be parallel to our

orography, and hence the perturbation vertical velocity at zero height

$$\begin{aligned} w'(x,0) = \frac{Dh}{Dt} &= \frac{Dh'}{Dt} \\ &= \bar{u}\frac{\partial h'}{\partial x} \\ &= -\bar{u}kh_0 \sin(kx) \end{aligned} \tag{3.3.2}$$

If we now take a stationary wave, such that $\partial/\partial t = 0$, then we can simplify equation 3.2.7, and integrate twice with respect to x to obtain

$$\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2} + \frac{N^2}{\bar{u}^2} w' = 0. \tag{3.3.3}$$

By substituting a wavelike solution $w' = w_0 \sin(kx + mz)$ into this expression[2], we obtain the dispersion relation

$$m = \pm\sqrt{\frac{N^2}{\bar{u}^2} - k^2} \tag{3.3.4}$$

which, when applied to the wavelike solution gives us the perturbation vertical velocity in two dimensions

$$w'(x,z) = -\bar{u}kh_0 \sin(kx + mz). \tag{3.3.5}$$

This expression suffices for low wind speeds, $|\bar{u}| < N/|k|$. For higher wind speeds however, forcing for the wave occurs at a frequency greater than N, and equation 3.3.4 tells us that the atmosphere cannot support waves of this frequency. In this situation, as equation 3.3.2 still applies, our solution must instead be wavelike in the horizontal direction but take a different form in the z-direction, namely the evanescent form

$$w' = w_0 e^{-z\sqrt{k^2 - \frac{N^2}{\bar{u}^2}}} \sin(kx). \tag{3.3.6}$$

Now, instead of our infinite series of peaks, consider an isolated ridge $h'(x)$, represented as a sum of sinusoidal functions with varying heights h_n and wavenumbers $k_n \propto \frac{1}{\lambda_n}$. In this case, the flow due to each individual sinusoidal term is as derived above (equation 3.3.5 or equation 3.3.6 as appropriate), and hence by summing each individual term we can construct the overall

[2] a sinusoidal, as opposed to cosinusoidal, solution is used here in order to be consistent with our expression for the perturbation vertical velocity, equation 3.3.2

flow over this single peak. For a wide ridge, this will be dominated by terms with large λ_n giving a wavelike solution vertically, whilst for a narrow ridge small λ_n terms will dominate to give an evanescent solution.

These mountain-like waves also exist in other fluids, for example due to rough orography on the ocean bed, and can be shown experimentally under laboratory conditions, using a denser fluid and consequently a smaller orography, provided we consider them on short enough timescales and large enough length scales to prevent standing waves forming due to reflection from the boundaries to form standing waves. Aguilar and Sutherland (2006) describe experimental work undertaken to study mountain waves arising due to topography in a salt-stratified solution, with peaks comparable in relative size to the Rocky Mountains. Figure 4 of Aguilar and Sutherland (2006) shows results from this experiment plotted as a function of space and time for sinusoidal and rectangular hills, demonstrating clear wavelike signatures arising from flow over orography. These results are broadly equivalent to waves we can observe with HIRDLS.

Fritts and Alexander (2003) provide a brief précis of the current state of knowledge regarding mountain waves. 2D responses to 2D topography are already well understood in terms of their dynamics, and the effects of 3D topography have also been investigated in many studies, both numerical and observational, giving good estimates of their length scales ($\sim$10s to 100s of km in the horizontal) and amplitudes (ranging from small to breaking). However, simulations fail to capture fully the effects of small-scale topography. Experimental studies with aircraft and radar have measured momentum fluxes $\sim$0.01–0.5 Pa in the upper troposphere/lower stratosphere (UTLS) region, with the bulk of this momentum flux associated with waves on smaller horizontal scales.

3.3.2 Other Gravity Wave Sources

The mountain waves discussed above are believed to be one of the main sources of atmospheric gravity waves (Guest et al., 2000), especially at northern midlatitudes where significant landmass exists. However, several sources of non-orographic waves are also believed to contribute significantly to the atmospheric momentum budget (Lane et al., 2001). This section will briefly discuss several other such sources.

3.3.2.1 Deep Convection

In the tropics, deep convection is believed to contribute significantly to the gravity wave spectrum, accounting for up to 25% of the forcing required to drive the quasi-biennial oscillation (Piani et al., 2000; Pfister et al., 1993). Lane et al. (2001) suggests three possible mechanisms by which such convective waves arise in deep tropical cloud, which can be briefly summarised as follows:

- Vertical forcing by variations in diabatic heating within the cloud
- Cloud updraft developing in regions of strong vertical wind shear, causing the cloud to partially block the flow and driving the wind over the top of the cloud, in a broadly equivalent manner to topography acting as a barrier to form mountain waves
- Individual updrafts within a convective system striking the tropopause and suddenly decelerating, causing the particles to oscillate about their level of neutral buoyancy.

Lane et al. (2001) argue that the third of these is the most likely, whilst Clark et al. (1986) instead suggest that the second is the dominant source.

3.3.2.2 Frontogenesis and Jets

Another probable source is frontogenesis. Reeder and Griffiths (1996) discuss a two-dimensional modelling study of wave generation from dry weather fronts, with the model producing waves with horizontal wavelengths of $\sim$500–1000 km and vertical wavelengths $\sim$3–10 km. Ray-tracing studies of the stratospheric gravity waves generated in their model suggest they originate from the frontal region, leading the authors to argue that cross-front circulation may be viewed as an unsteady source of gravity waves. This view is also suggested by the results of a sonde study of waves over Macquarie Island in the south-west Pacific (Guest et al., 2000). Wind and temperature measurements were used to study the seasonal and synoptic cycle over the island, with a total of 47 sonde flights undertaken over the course of 1994. A seasonal cycle of gravity wave activity was observed, with a minimum in winter when the background flow was strongly westerly and vertical shear strongly positive. A common synoptic pattern was observed, with a strong upper-level jet and associated surface front upstream of the island on sixteen of the twenty-one days on which gravity waves were detected, which was absent on fifteen of the twenty-one days on which gravity waves were not detected. This synoptic pattern

had a similar seasonal cycle to that of the observed gravity wave activity. Ray-tracing studies of the gravity waves indicated that the observed waves were most probably generated within this jet-front system.

3.3.2.3 Geostrophic Adjustment

Fritts and Luo (1992) and Luo and Fritts (1993) discuss 2D and 3D models of gravity wave excitation via geostrophic adjustment. Tropospheric sources such as those discussed above primarily produce motions with high intrinsic frequencies, but observations at different altitudes show that the gravity wave spectrum is also made up of motions at or near the inertial frequency, of which geostrophic adjustment could be a significant source. Their studies simulated a Gaussian jet; at positions away from the centre of the jet, motions were observed with wave characteristics, with a typical vertical wavelength $\sim$6–8 km imposed by the jet structure.

3.3.2.4 Convective Boundary Layer Processes

Kuettner et al. (1987) discuss the generation of gravity waves via convective processes at the planetary boundary layer, suggesting that a deep layer in the troposphere is filled with internal waves whenever there is both convective activity in the planetary boundary later and moderate-to-strong wind shear. The work described by this study and by Clark et al. (1986) conclude that this wave-generation process is most probably due to the obstacle effect encountered by the convective currents: as upper winds become aligned at a large angle to the axes of cloud-streets in the upper boundary layer, the wind is directed over them, leading to the launch of vertical-propagating gravity waves in order to conserve their momentum in a similar way to the launch of mountain waves. A similar effect was also observed by Pfister et al. (1993)

3.4 Previous Observational Studies

A wide variety of methods have been used to analyse gravity waves in the middle atmosphere, including radiosondes, rocket soundings, LIDAR, RADAR, and aircraft measurements as well as satellite results. Fritts and Alexander (2003) discuss results from these studies in significantly greater depth; the key results described in this review shall briefly be discussed here.

3.4.1 Radiosondes, Rocket Soundings and Aircraft

Radiosondes, rocket soundings and aircraft measurements inherently have poor temporal and spatial coverage, as they require the instrument to be physically located within the gravity wave feature. However, this same property makes them highly accurate in their measurements, as the remote-sensing limitations discussed in chapter 2 do not need to be considered.

Radiosonde measurements have helped to improve understanding of gravity wave processes significantly at their operational altitudes, typically up to $\sim$25 km. Temporal coverage has generally been poor, however – a typical radiosonde campaign will launch at most a few balloons per day. Windspeed measurements from sondes worldwide have shown rotation in the horizontal winds consistent with the presence of gravity waves, with "abundant evidence ... that the perturbations in these vertical profiles are dominated by low $\hat{\omega}$ inertia-gravity waves with frequencies $\sim 1 - 3f$" (Fritts and Alexander, 2003; Vincent and Alexander, 2000). Results have shown seasonal maxima in local winter in northern and southern midlatitudes (Yoshiki and Sato, 2000), with a spring maximum at Antarctic latitudes (Allen and Vincent, 1995). The tropics around Malaysia and the Indian Ocean have shown a slightly different cycle peaking in the period December to February, consistent with seasonal patterns of deep tropical convection (Vincent and Alexander, 2000). There is also radiosonde evidence of gravity wave behaviour being strongly correlated with the quasi-biennial oscillation, with an excess of eastward-propagating waves observed during periods of strong westward winds in the lower stratosphere with eastward shear above (Sato and Dunkerton, 1997).

Rocket soundings during the 1970s and 1980s at a variety of sites in the Northern Hemisphere and tropics, meanwhile, have been used to infer seasonal and latitudinal variations in gravity wave activity to a resolution $\sim$1 km up to a height of 60 km. These rocket measurements similarly suggest a seasonal cycle of gravity wave activity at high latitudes, peaking in winter. In the tropics this seasonal cycle is less pronounced, with a preference again observed for eastward-propagating waves.

Finally, aircraft studies have failed to make a significant contribution to our understanding of gravity wave variability, due to their limited duration and coverage. However, individually-targeted aircraft campaigns have made a significant contribution to the local dynamics of localised gravity wave sources, providing much-needed quantitative constraints.

3.4.2 LIDAR and RADAR

LIDAR and RADAR measurements provide an different means of observation, operating as a remote sensing process directed from the surface upwards.

Due to the limited availability of the equipment required, LIDAR studies have been focused on a small number of locations, and only a few such sites have operated for long enough periods of time to provide seasonal variability data. What few studies there have been suggest, similarly to radiosondes and rocket soundings, that the seasonal cycle at high latitudes is characterised by a peak in winter and a minimum in summer, although this is largely made up of observations in the monthly mean rather than more temporally-localised measurements (Fritts and Alexander, 2003).

RADAR results, meanwhile, provide the greatest range of vertical coverage of the techniques discussed so far, allowing wind-perturbation measurements to be made throughout the stratosphere, mesosphere and lower thermosphere. The temporal coverage of studies to date has been limited, based on a small number of campaigns at a limited number of sites, but the results of some of the longer-term campaigns, all of which have been at midlatitudes, are illustrative, once again showing a seasonal asymmetry biased towards winter. Specific RADAR campaigns have also produced detailed results associated with both mountain and convective waves (Fritts and Alexander, 2003).

3.4.3 Space-based and Satellite Measurements

We finally consider satellite measurements. These provide coverage on a global scale, but are limited in terms of their spatial and temporal resolution; very few satellites have sufficiently fine spatial resolution to accurately resolve the relatively small spatial and short temporal scales at which internal gravity waves exist.

A key distinction for our measurements, as discussed above in section 2.1.3, is the scanning geometry used. Limb sounders will tend to have a high vertical resolution, with instruments such as HIRDLS and the COSMIC GPS constellation having vertical resolutions at or approaching 1 km. However, as a result of their scanning geometry, these instruments will also tend to have extremely poor horizontal resolution: to take the same two examples, HIRDLS has a profile spacing of around 80–100 km along its orbital track, whilst the profiles obtained from

the COSMIC constellation are irregularly spaced depending on occultation opportunities and consequently have an extremely poor horizontal resolution (Liou et al., 2007). The weighting functions along the line of sight of the instrument further contribute to this horizontal resolution issue. Nadir sounders, meanwhile, will tend to have the reverse of these characteristics, namely a high horizontal but poor vertical resolution. This is further discussed in section 3.5.1. Future planned missions, such as the proposed PREMIER mission (Ungermann et al., 2009), may be designed to produce three-dimensional retrievals which will allow high-enough resolution in both directions to fully analyse gravity waves, but such data are not currently available.

To date, limb sounding measurements have provided the most useful results on gravity waves. Early studies such as those undertaken with LIMS produced broad climatologies of large-scale inertia-gravity waves, with maps of the observed temperature variances showing peaks in the lower stratosphere and in the winter jet at higher altitudes, and additional evidence for gravity wave variances due to the semiannual oscillation (Fetzer and Gille, 1994). More recent studies such as analyses with UARS-MLS and GPS (both CHAMP and COSMIC) have significantly contributed to middle-atmospheric gravity wave climatologies, with both datasets once again showing a stronger gravity wave signal in winter than summer and with low measured variances at the equator (Wu and Waters, 1997). Longitudinal variations in the UARS-MLS results also provided evidence suggesting deep convection as a source of waves observed in summertime (McLandress et al., 2000).

Another significant limb sounding experiment was the CRISTA mission, carried by the Space Shuttle during 1994. This observed wave variances over major mountains ranges such as the Andes and the ranges of central Asia, and showed clear evidence that mountain waves were a significant source of gravity wave variances in the stratosphere, according well with model predictions (Eckermann and Preusse, 1999).

More recently, nadir-sounding instruments such as AIRS have enabled high-resolution 2D horizontal retrievals of gravity waves. Gravity wave studies using such instruments are at an early stage, as instruments have only recently reached the required resolutions, but several studies have been produced using them, such as Alexander and Barnet (2007), which gives a climatology of gravity waves for a month over Patagonia, and Alexander and Teitelbaum (2007), in which a large-scale mountain wave over the Antarctic peninsula is analysed in detail. Eck-

ermann et al. (2006) also show a mountain-wave event over Scandinavia to be observable in results from the AMSU microwave sounder.

3.5 Measurement Limitations

Each of the detection methods discussed above inherently involves filtering the observed gravity wave spectrum through the capabilities of the detector used. Consequently, we must bear in mind that the gravity wave results produced and observed by our instrumental measurements provide only an approximation to the real gravity wave activity in the atmosphere. This section will briefly discuss three such issues.

3.5.1 The Observational Filter

Alexander and Barnet (2007) discuss the observational filtering effect. Figure 3 of this paper illustrates the effects of resolution and weighting-function sizes for a range of limb, nadir, and sublimb instruments. The coloured background illustrates the temperature perturbations arising due to gravity waves generated by a convective cloud model (where the white region in bottom centre is the cloud itself), and the various ovals overlaid indicate the weighting-function footprints of eight instruments. The positions of the ovals on the figure are arbitrary; it is their sizes which indicate their filtering effects. Note that UARS-MLS is a limb-sounder by design, but the results shown on this figure are for sublimb gravity wave measurements taken from it. HIRDLS is shown towards the lower right-hand side: the oval is narrow in the vertical to illustrate the high vertical resolution of the instrument, but is elongated in the horizontal direction due to the wide spacing of the horizontal profiles and wide horizontal weighting functions. Note also that the AMSU and AIRS instruments provide two-dimensional data in the horizontal (with a 50 km and 12.5 km horizontal footprint respectively), allowing them to resolve waves in the axis projected out of the page.

3.5.2 Source Intermittency and Group Velocity

Figure 2 of Alexander and Barnet (2007), illustrates the effects of source intermittency and variable group velocity on gravity wave observations. For two waves, one propagating rapidly

in the vertical and other more slowly, emitted at the same time from an intermittent gravity wave source, the slower-propagating wave is significantly more likely to be passing through the detection area of the measurement instrument. This is due to it spending a significantly longer period of time traversing the atmosphere before breaking, and consequently being much more likely to be observed. However, the faster-moving wave packets could in practice make a significantly larger contribution to the atmospheric momentum budget, whether due to larger numbers of them being emitted or due to their carrying more momentum.

3.5.3 Averaging

Finally, temporal and spatial averaging of gravity wave results observed by instruments may serve to skew our results. Consider an intermittent gravity wave source, such as a mountain range which only generates waves for a given wind direction or convectively-generated waves. This source will only emit some of the time, and the form and magnitude of the wave will also vary. Other atmospheric processes, such as wind, through which the wave propagates will also vary temporally, further distorting the observed perturbations and fluxes as a function of time. Consequently, any study which involves the averaging of gravity wave results over a period of time will inherently reduce the quality of the product produced. In order to provide good spatial coverage, it will sometimes be necessary in this thesis to average results over a longer or shorter time period: this will inherently reduce the temporal accuracy of the results considered.

Chapter 4

Gravity Wave Detection Using the Stockwell Transform, Part I: Method

4.1 Overview

Alexander et al. (2008) describe in brief the technique used to detect gravity waves in this thesis. This process will be outlined in significantly greater detail below, but can be summarised as follows. Initially, the height region of interest is selected, and zero-padded at either end to avoid wave-detection overruns (section 4.5.1). We next remove the mean background from the data, together with any large-scale planetary wave signal present, and S-Transform the resultant profiles (section 4.2). The profiles are then analysed in pairs, taking the covariance spectra of each adjacent profile-pair (section 4.6), and the physical parameters of our gravity waves (temperature perturbations, horizontal and vertical wavelengths, and momentum fluxes) calculated from the resulting data.

4.2 The S-Transform

The Stockwell Transform (ST) is an extension of the idea of the continuous wavelet transform based on a moving scalable Gaussian window, which provides a joint time-frequency representation of the spectral components of a time series (Stockwell et al., 1996). The ST has been used for gravity wave analysis in several studies, including Stockwell (1999), Stockwell and Lowe (2001), Wang et al. (2006), and Alexander et al. (2008).

4.2.1 Derivation of the S-Transform

It is helpful to briefly derive the form of the S-Transform, in order to fully understand its application. This derivation closely follows that outlined by Stockwell (1999). To avoid confusion between spatial and temporal frequencies, we shall derive this for a time-varying signal $h(t)$, with the understanding that the result derived applies equally to the analysis of a signal varying instead with altitude, $f(z)$.

We first consider the Fourier Transform $H(f)$ of a time series $h(t)$, defined as

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp(-2\pi i f t)\, dt. \tag{4.2.1}$$

The results of this transform do not localise the temporal position of the input signal in any way, and consequently application of this expression to a time-varying signal $h(t)$ will return only the time-averaged frequency distribution present in the signal. For many purposes, this is entirely suitable, but in our case we are interested not just in the spatial frequencies present in a given temperature profile, but also in the altitude of the signals. To rephrase this in terms of our temporally-varying signal, we are interested not just in the temporal frequencies present in the signal, but when they happen. Accordingly, we would like to adapt our analysis to also return this information.

One approach to provide this temporal information is to "window" our signal with a windowing function $g(t)$. We can then instead consider the spectrum

$$I(f) = \int_{-\infty}^{\infty} h(t)\, g(t) \exp(-2\pi i f t)\, dt, \tag{4.2.2}$$

which will, for an appropriately chosen windowing function $g(t)$, return the frequencies present only within the temporal region defined by the window.

To derive the S-Transform, we define our windowing function as the normalised Gaussian

$$g(t, \tau, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(\tau - t)^2}{2\sigma^2}\right), \tag{4.2.3}$$

which is a function of time t, temporal translation τ, and window width σ. This then allows us

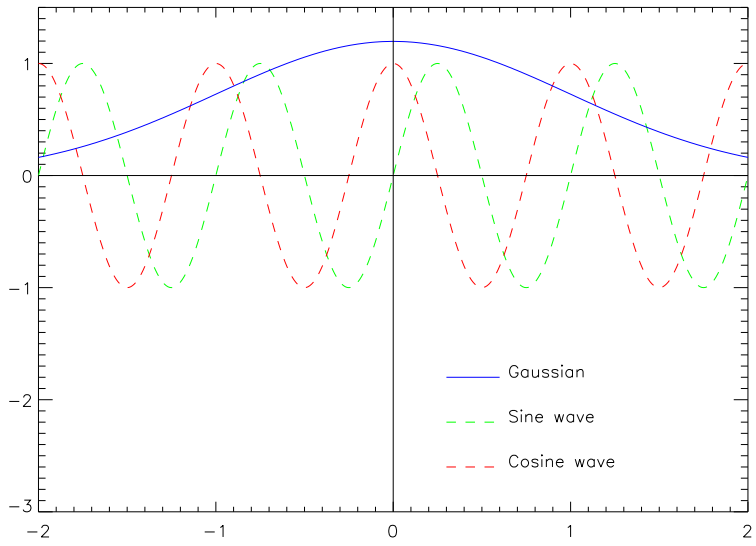


Figure 4.1: *Illustration of the relative widths of the Gaussian wavepacket $g(t, \tau = 0)$ (blue) used in the S-Transform and sine (green) and cosine (red) waves of equal frequency f. The Gaussian curve has been scaled by a factor of three to aid visual interpretation; since the windowing term of the S-Transform equation (4.2.3) is not related to the amplitude of the input waves, this does not affect the results.*

to define our frequency spectrum in terms of the three variables τ, σ and t as

$$S^{\star}(\tau, f, \sigma) = \int_{-\infty}^{\infty} h(t) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(\tau - t)^2}{2\sigma^2}\right) \exp(-2\pi i f t)\, dt. \tag{4.2.4}$$

A normalised Gaussian is chosen for this purpose because it is "the most compact [windowing function] in time and frequency" (Stockwell, 1999). Figure 4.1 shows the width of this Gaussian as compared to sine and cosine waves of equal frequency f, illustrating the relative contributions of adjacent cycles of an ideal wave which contribute to the transform at a given frequency.

This expression is a function of three independent variables, and is consequently impractical as a tool for analysis (Stockwell, 1999). We can resolve this by adding the additional simplifying constraint that the window width σ is proportional to the period $(1/f)$ of the signal, giving us the S-Transform equation

$$S(\tau, f) = \frac{|f|}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(t) \exp\left(\frac{-(\tau - t)^2 f^2}{2}\right) \exp(-2\pi i f t)\, dt. \tag{4.2.5}$$

This expression returns a two-dimensional result for a one-dimensional input signal, the first dimension being the frequencies present in the signal and the second being the time at which said frequencies are present. Stockwell et al. (1996) gives several examples of such results.

4.2.1.1 Wavelet Formulation

A possibly more intuitive way of considering equation 4.2.5 is to rearrange into the form

$$S(\tau, f) = exp\,(2\pi i f \tau)\, W(\tau, d). \tag{4.2.6}$$

In this case, $W(\tau, d)$ represents the continuous wavelet transform of a mother wavelet $w(t, f)$:

$$W(\tau, d) = \int_{-\infty}^{\infty} h(t) w(t - \tau, d) dt \tag{4.2.7}$$

$$w(t, f) = \frac{|f|}{\sqrt{2\pi}} \exp^{\frac{-t^2 f^2}{2}} exp\,(-2\pi i f t)\,. \tag{4.2.8}$$

It should however be noted that the expression in equation 4.2.7 does not have zero mean, and hence does not strictly meet the definition of a continuous wavelet transform as it does not form an admissible wavelet (Stockwell et al., 1996).

4.2.2 The Discrete S-Transform

The above derivation assumes that the input signal is a smoothly-varying function of time. However, in practice, this is not the case: our input signal consists of the results of physical measurements, which exist over a finite range and are discrete. Accordingly, we cannot directly use the continuous analytic S-Transform equation above, but must instead use a discrete form. We again follow Stockwell (1999) for this derivation.

We first re-write equation 4.2.5 as a convolution between the windowing function $g\,(t)$ and a function $p\,(\tau, f)$. From the definition of convolution,

$$f\,(x) * g\,(x) \equiv \int_{-\infty}^{\infty} f\,(x)\, g\,(x - \chi)\, d\chi, \tag{4.2.9}$$

we can rearrange 4.2.5 as

$$S\,(\tau, f) = \int_{-\infty}^{\infty} p\,(t, f)\, g\,(\tau - t, f)\, dt = p\,(\tau, f) * g\,(\tau, f)\,, \tag{4.2.10}$$

where

$$p\,(\tau, f) = h\,(\tau)\ \exp\,(-2\pi i f \tau) \tag{4.2.11}$$

and

$$g(\tau, f) = \frac{|f|}{\sqrt{2\pi}} \exp\left(\frac{-\tau^2 f^2}{2}\right). \tag{4.2.12}$$

We next take the Fourier transform of the right-hand side of equation 4.2.10, which we define as $B(\alpha, f)$. By application of the convolution theorem of the Fourier transform, we hence obtain

$$B(\alpha, f) = P(\alpha, f)\, G(\alpha, f), \tag{4.2.13}$$

where P and G are the Fourier transforms of p and g respectively. This can then be rearranged as

$$B(\alpha, f) = H(\alpha + f) \exp\left(\frac{-2\pi^2\alpha^2}{f^2}\right) \tag{4.2.14}$$

and inverse-transformed to give us the ST in the form

$$S(\tau, f) = \int_{-\infty}^{\infty} H(\alpha + f) \exp\left(\frac{-2\pi^2\alpha^2}{f^2}\right) \exp(2\pi i\alpha\tau)\, d\alpha. \tag{4.2.15}$$

We next proceed to derive the discrete analogue of this expression. The discrete Fourier transform is given by

$$H\left(\frac{n}{N\Xi}\right) = \frac{1}{N}\sum_{k=0}^{N-1} h[k\Xi] \exp\left(\frac{-2\pi i n k}{N}\right), \tag{4.2.16}$$

where $h[kt]$ denotes a discrete time series corresponding to $h(t)$ with $k = 0, 1, ..., N-1$ and Ξ the time-interval between steps, and $n = 0, 1, ..., N-1$. By analogy, we can derive the discrete form of equation 4.2.15. To do this, we discretize τ, f and α as $j\Xi$, $n/N\Xi$ and $m/N\Xi$ respectively, where j, m and $n = 0, 1, ..., N-1$. We then sum over the direction of integration and rearrange, to obtain our discrete analogue

$$S\left(j\Xi, \frac{n}{N\Xi}\right) = \sum_{m=0}^{N-1} H\left[\frac{m+n}{N\Xi}\right] \exp\left(\frac{-2\pi^2 m^2}{n^2}\right) \exp\left(\frac{-2\pi i m j}{N}\right). \tag{4.2.17}$$

This discrete form of the S-Transform is the expression implemented in the S-Transform routine used for the work described in this thesis[1].

[1]The IDL source code used for the S-Transform process in this thesis was sourced from Robert Stockwell, developer of the transform (Stockwell, 1999), and can be obtained from `http://www.cora.nwra.com/~stockwel/` (checked March 2010)

4.2.3 Localised Cross-Spectral Analysis

The process outlined above allows us to compute the vertical wavelength of our signal as a function of height. We can obtain additional information on the gravity wave spectrum present by considering the cospectrum of two adjacent profiles, via the utilisation of localised cross-spectral S-Transform analysis ('the cross-ST') (Stockwell, 1999).

To do this, we analyse the S-Transformed profiles in adjacent pairs. For each transformed profile $Q(z)$, we compute the cospectrum $C_{i,i+1} = Q_i Q_{i+1}^*$, i.e. the product of this and the adjacent profile. From equation 4.2.15, we can consider the effect of a change in wave phase across the gap between the profiles.

Consider a sinusoidal wave signal

$$h(t) = \exp(2\pi i \omega t), \tag{4.2.18}$$

with Fourier transform

$$H(\alpha) = \frac{1}{2}\delta(\alpha - \omega) \tag{4.2.19}$$

and hence S-Transform

$$S_h(\tau, f) = \frac{1}{2}\exp\left(\frac{-(f-\omega)^2}{(f/2\pi)^2}\right)\exp(2\pi i(f-\omega)\tau), \tag{4.2.20}$$

and a second wave identical to 4.2.18 but with a constant phase shift ϕ

$$g(t) = h(t)\,\exp(i\phi) = \exp(2\pi i\omega t + i\phi), \tag{4.2.21}$$

and S-Transform

$$S_g(\tau, f) = \frac{1}{2}\exp\left(\frac{-(f-\omega)^2}{(f/2\pi)^2}\right)\exp(2\pi i(f-\omega)\tau)\,\mathrm{e}^{i\phi} = S_h(\tau, f)\,\mathrm{e}^{i\phi}. \tag{4.2.22}$$

If we multiply S_h with the complex conjugate of S_g, S_g^*, we can show that

$$S_h(\tau, f)\,.\,S_g^*(\tau, f) = \frac{1}{4}\exp\left(\frac{-2(f-\omega)^2}{(f/2\pi)^2}\right)\mathrm{e}^{i\phi} \tag{4.2.23}$$

and hence that the phase change ϕ manifests itself directly in the cospectrum output. This is true for the general series (Stockwell, 1999), and can also be applied to the discrete form of the ST equation. Consequently, we can state that

$$C_{i,i+1} = Q_i Q_{i+1}^* = \hat{T}_i \hat{T}_{i+1}^* \, \mathrm{e}^{i\Delta\phi_{i,i+1}}, \tag{4.2.24}$$

where $\hat{T}_i$ and $\hat{T}_{i+1}$ represent the temperature perturbations present in each profile and exp $(i\Delta\phi_{i,i+1})$ the change in phase between the two profiles, computable as

$$\Delta\phi_{i,i+1} = \arctan\left(\frac{\Im\left(C_{i,i+1}\right)}{\Re\left(C_{i,i+1}\right)}\right). \tag{4.2.25}$$

By considering this change in phase in combination with the geolocation information provided by the satellite, specifically the along-track distance between profiles Δx, we can hence obtain a measurement for the horizontal wavenumber $k_h = \Delta\phi/\Delta x$ of the observed waves.

We also compute the covariance spectrum

$$|C_{i,i+1}| = \left|Q_i Q_{i+1}^*\right| = \left|\hat{T}_i \hat{T}_{i+1}^* \, \mathrm{e}^{i\Delta\phi_{i,i+1}}\right| \tag{4.2.26}$$

and, as a result of this, the covarying amplitude

$$\hat{T}_{i,i+1} = \sqrt{|C_{i,i+1}|} \tag{4.2.27}$$

This provides a measure of the temperature perturbation present at each frequency and altitude across the profiles, representing the perturbation in temperature due to an observed gravity wave.

4.3 A Note on Notation

During this and subsequent chapters, we will often need to specify the parameters of artificial waves created for analysis. To avoid repetition, it will be useful to define a notation for these waves.

The key parameters to be described for each wave are:

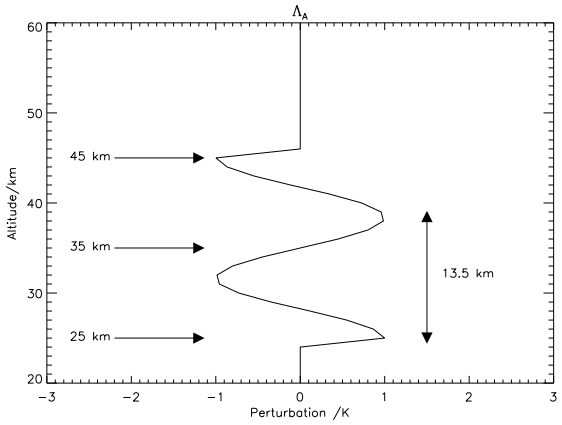


Figure 4.2: *An artificial wave, described in our notation as* $\Lambda_A\,[13.5, (25, 35, 45)]$

- vertical wavelength λ_z
- the altitude at which the wave begins h_B
- the altitude at which the wave ends h_T
- the altitude at which the wave is centred[2], i.e. the magnitude is zero, h_C
- the amplitude of the wave, A

In each case, we are using a sinusoidal as opposed to cosinusoidal wave. All altitudes referred to are defined vertically upwards from the surface. Waves will hence be denoted as

$$\Lambda_i = [\lambda_z, (h_B, h_C, h_T)]\,,$$

where λ_z, h_B, h_C, and h_T are specified in km, and i is used to refer to the specified wave. For example, the identifier

$$\Lambda_A = [13.5, (25, 35, 45)]$$

describes a wave Λ_A with a vertical wavelength of 13.5 km, defined over the altitude range 25 km– 45 km and centred on an altitude of 35 km. Figure 4.2 illustrates this wave.

All waves described in this chapter, unless otherwise specified, will be generated on a 1 km altitude scale, intended to approximate the vertical resolution of HIRDLS data. An amplitude of 1 K will be assumed unless otherwise specified.

[2]note that, while it often will be, this is not necessarily the mean of h_T and h_B, but may take other values when we wish to shift the wave in phase

4.4 ST Examples on Test Cases

Figure 4.3 illustrates the magnitudes of the complex results of S-Transform analysis applied to artificial wave signals. The first subfigure shows $\Lambda_1 = [15, (25, 30, 35)]$, and the subfigure immediately below this the results of the S-Transform as applied to this profile: the resulting distribution is clearly peaked near this height and wavelength (illustrated by the orange lines, which cross at the maximum; the slight inaccuracy of this fit will be discussed further in section 5.2). The amplitudes will be discussed further in section 5.3.

The second column shows another artificial wave, $\Lambda_2 = [5, (40, 45, 50)]$. Again, the input height and wavelength are clearly resolved in the output spectrum, as indicated by the cyan lines. The resulting fit is significantly stronger than that found for the previous wave, as illustrated by the tighter spatially peaked region at the input wavelength and height. This is because of the larger number of complete wave cycles present in this test case, allowing the transform to better compute the optimal fit (see section 5.2.2).

Finally, the third column shows a composite artificial profile composed of these two waves. For comparison to Λ_1 and Λ_2, their maxima are indicated on the ST result plot by the coloured horizontal and vertical lines. Due to the more complicated input signal, significant additional noise is observed in the spectrum; the higher-frequency wave at high altitude is still clearly resolved at the the right wavelength and height, but the longer-wavelength, lower-altitude wave signal is now smeared out significantly across the S-Transform spectrum, to the point where it cannot be fairly said to have been accurately resolved. However, this fit can be improved significantly by zero-padding the signal.

4.5 Data Preparation

4.5.1 Height Extraction and Zero-Padding

To analyse our data using the S-Transform, we need to perform some preliminary preparation on the data. Initially, we extract the height range we wish to study, which we define, based on the parameters of HIRDLS studies as discussed in chapter 2, as the height range from 20 to 60 km above the surface.

We next zero-pad the data at top and bottom. This is because the S-Transform process, in the

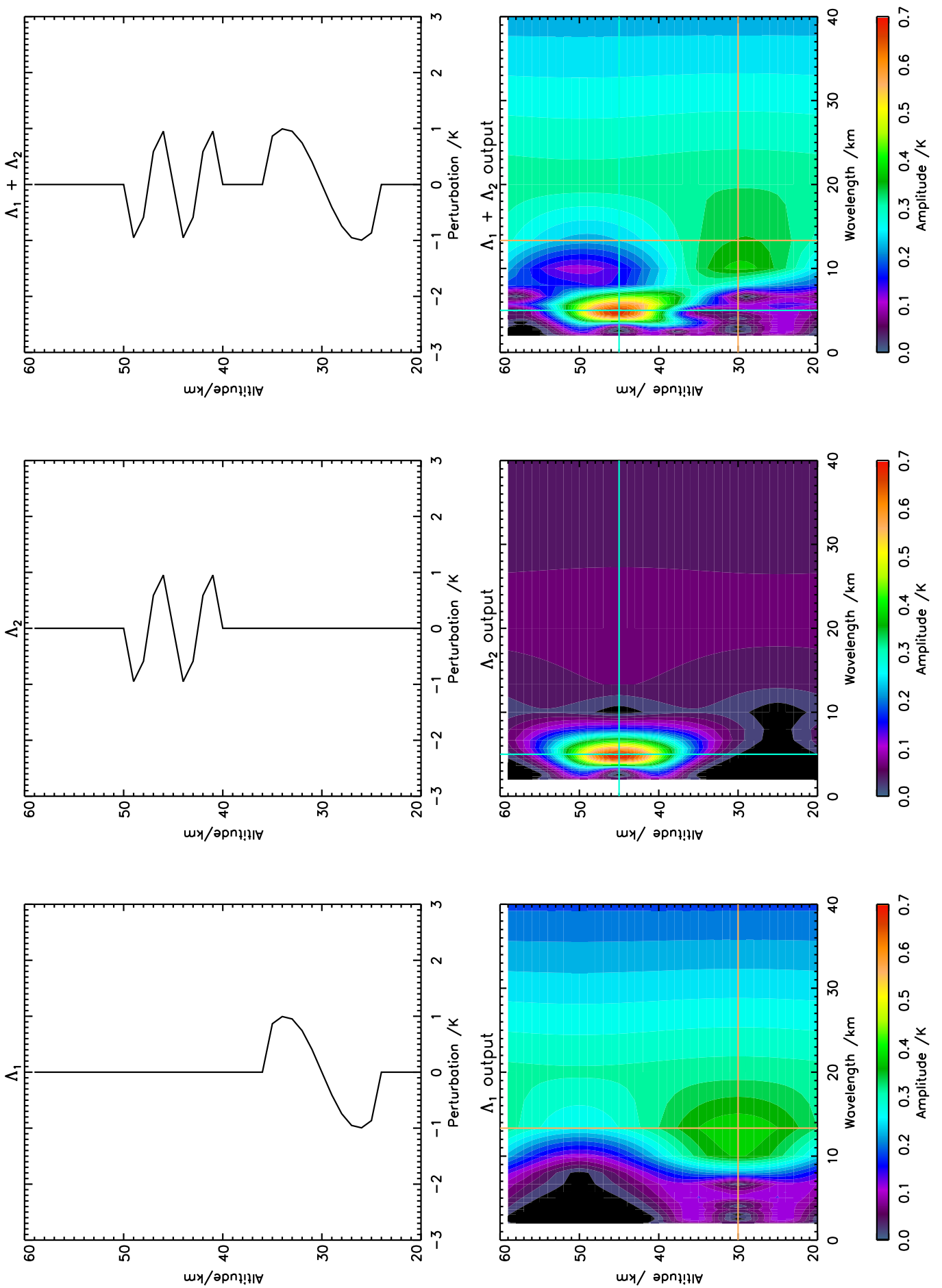


Figure 4.3: *Illustration of the S-Transform method as applied to artificial wave signatures. Upper row shows input signals, and lower row the S-Transforms of these signals as a function of wavelength and altitude. The overplotted solid lines indicate the locations of peaks in the S-Transform output.*

same way as a Fourier Transform, wraps the input data around from one end to the other as part of the analysis, and hence spurious unphysical waves may be detected in the wraparound region. Accordingly, by zero-padding the data, we improve the quality of the wave fit and hence of our results. These padding regions are removed after the dataset has been S-Transformed (section 4.6), and before any subsequent analysis.

The effects of this are demonstrated in figure 4.4, which illustrates the results of the S-Transform as applied to Λ_1 and Λ_2 with 20km of zero-padding at top and bottom. If we compare this to figure 4.3, discussed above, we observe that the quality of the fit in each case is significantly sharper, with the peaks better resolved in wavelength and height and with much less noise surrounding the peak-wave regions. The weaker peak (lower altitude) of the combined signal, in particular, is now much better resolved.

4.5.2 Removal of Background Planetary Wave Signal

Our next task is to remove the background from our signal. This is necessary because, over the vertical range of our measurements, the temperature structure of the atmosphere varies significantly, and this variation may manifest itself in the Transform results as a spurious wave signal.

Figure 4.5 illustrates the results of the Transform as applied to a reference daytime temperature profile, with the artificial wave Λ_2 (as described above) inserted, increasing in amplitude from left to right. In all three cases, the majority of detected power is in longer-wavelength modes, due to the background atmospheric temperature variation being much greater in amplitude than the artificial wave added. This results in the added wave at low amplitudes being largely indistinguishable from noise. Accordingly, it is important to remove this background.

For our purposes, we shall define this background as the mean and the first three planetary wave modes, that is to say those waves which fit around the Earth at a given latitude between one and three times. Section 5.5 discusses the selection of this criterion in further detail.

To calculate this background, we need to analyse the whole Earth on a regular scale, rather than on a profile-by-profile basis. To do this, we take each day's data and then bin it onto a regular grid on a scale of 2.5° latitude by 24° longitude. This resolution gives us 72 bins in the meridional direction and 15 bins in the longitudinal direction, and is chosen due to the inter-orbit

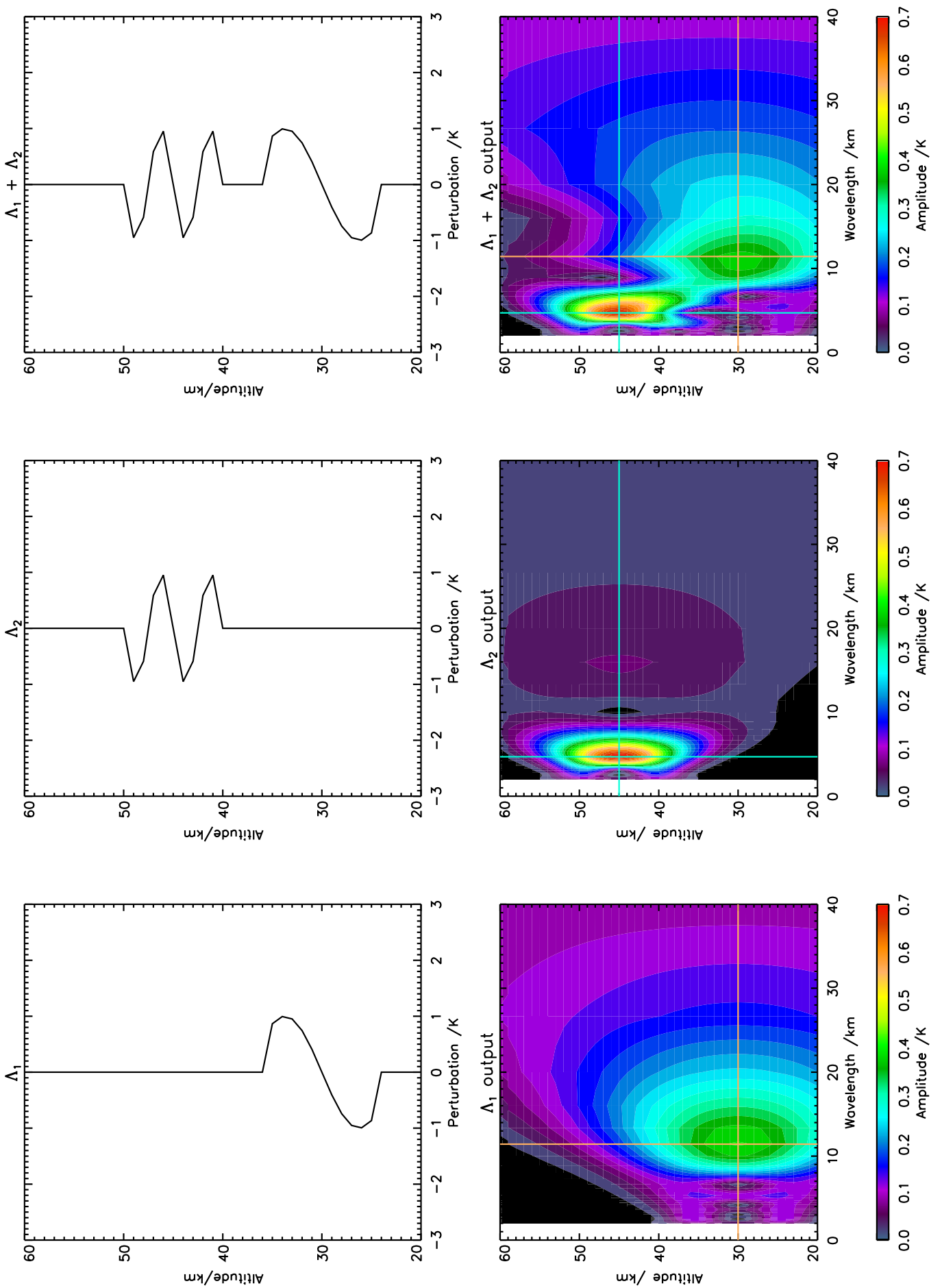


Figure 4.4: *Illustration of the S-Transform method as applied to artificial wave signatures, with twenty levels (20 km) of zero-padding added to the top and bottom of the signals (not shown). Input wave signals are otherwise identical to those of figure 4.3.*

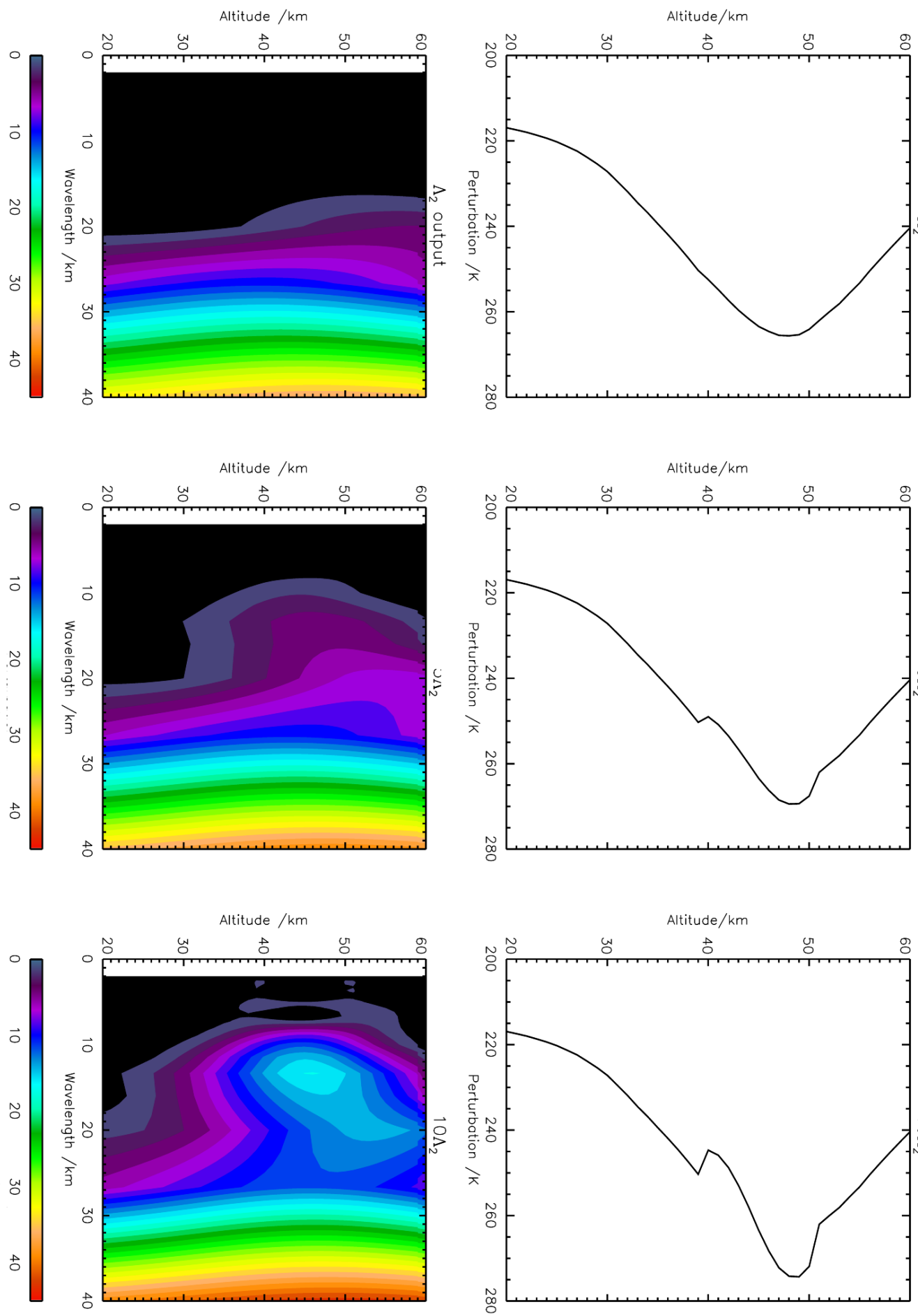


Figure 4.5: *S-Transform as applied to unsmoothed atmospheric profile data with an artificial wave of form Λ_2 added, of amplitude 1 K in column A, 5 K in column B, and 10 K in column B.*

spacing of Aura's orbital track: this grid scale is the finest which will still give full coverage over the Earth for each full day of data. To allow for any missing data, we then interpolate over any gaps in this grid in the meridional direction, i.e. for a given latitude we interpolate over the longitudinal-distributed values. Each vertical level in the data is analysed independently.

We next individually analyse each zonal series of values. For each such line, we analyse for planetary-scale waves using the Fourier transform. To do this, we first duplicate the series onto the beginning and end of itself: that is to say, for a series $[x_0, x_1, \ldots, x_{15}]$, we produce a new series $[x_0, x_1, \ldots, x_{15}, x_0, x_1, \ldots, x_{15}, x_0, x_1, \ldots, x_{15}]$. The purpose of this step is to ensure that the Fourier series properly extracts any waves not properly centred in the window due to zero-padding in the Fast Fourier algorithm; as the Earth is spherical, our wrapped-around results remain physical. We then apply the Fast Fourier Transform to this series, and extract those portions of the the signal corresponding to the background mean (the 0^{th} element of the transform result), and to waves which fit once, twice and three times around the Earth (the 3^{rd}, 6^{th} and 9^{th} elements respectively, due to our duplication of the series). The results returned by the FFT encode information about the amplitude, frequency and phase of the signal and, accordingly, we can compute and then simply subtract the magnitude of each planetary wave from each individual point in the original signal. This leaves us with just the small perturbations to these larger-scale signatures. Figure 4.6 illustrates sample planetary wave data obtained from this analysis, which will be discussed further in section 5.5.

At the equator, the mode-3 planetary wave corresponds to a zonal wavelength of approximately 13 000 km. This falls away rapidly as we approach the poles: at the southern limit to HIRDLS data at 62.5S this has fallen to approximately 3000 km, and at the northern limit of $\sim$80N to approximately 1200 km; even at high latitude this is still significantly greater than our profile spacing. Accordingly, this removal of the large-scale background should not directly affect the results due to our internal and inertia-gravity waves, which exist on much smaller length scales, $\sim$100s of km.

4.6 Gravity Wave analysis

We now compute the spectrum of gravity waves present in the data. Firstly, we S-Transform the data on a profile-by-profile basis. We then remove any wavelengths detected in the analysis

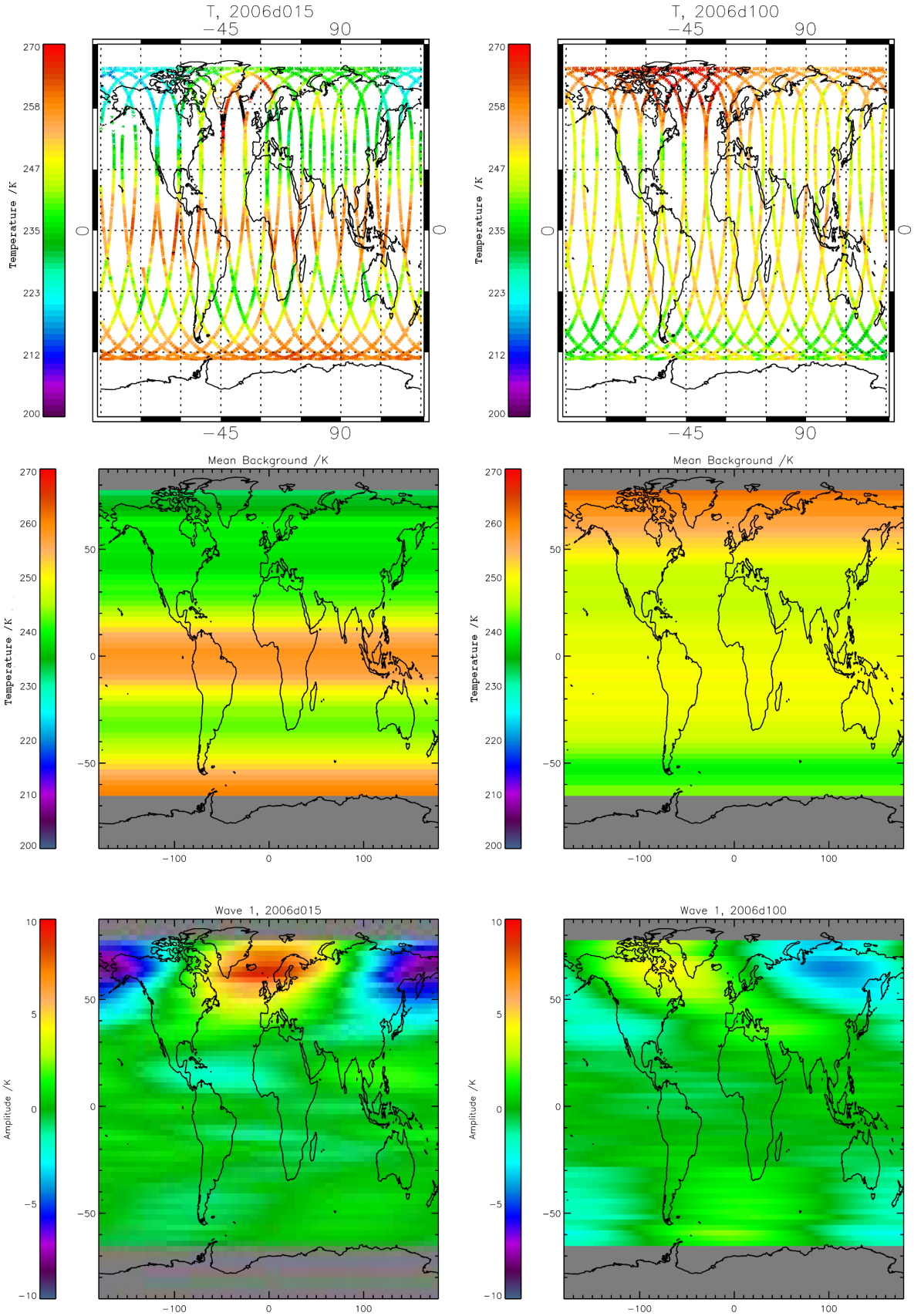


Figure 4.6: *Example planetary wave results at the 100 hPa (∼16 km) level, derived from the FFT analysis. The original data is shown at the top, planetary waves mode 1–3 in the middle three rows, and the resulting temperature perturbations (after S-Transform analysis) at bottom. Strong planetary wave signals in the lower modes are observed in the Arctic in northern-hemisphere winter but not summer, as would be expected (see section 5.5), and amplitudes are generally low (∼ a few K) otherwise. (continues)*

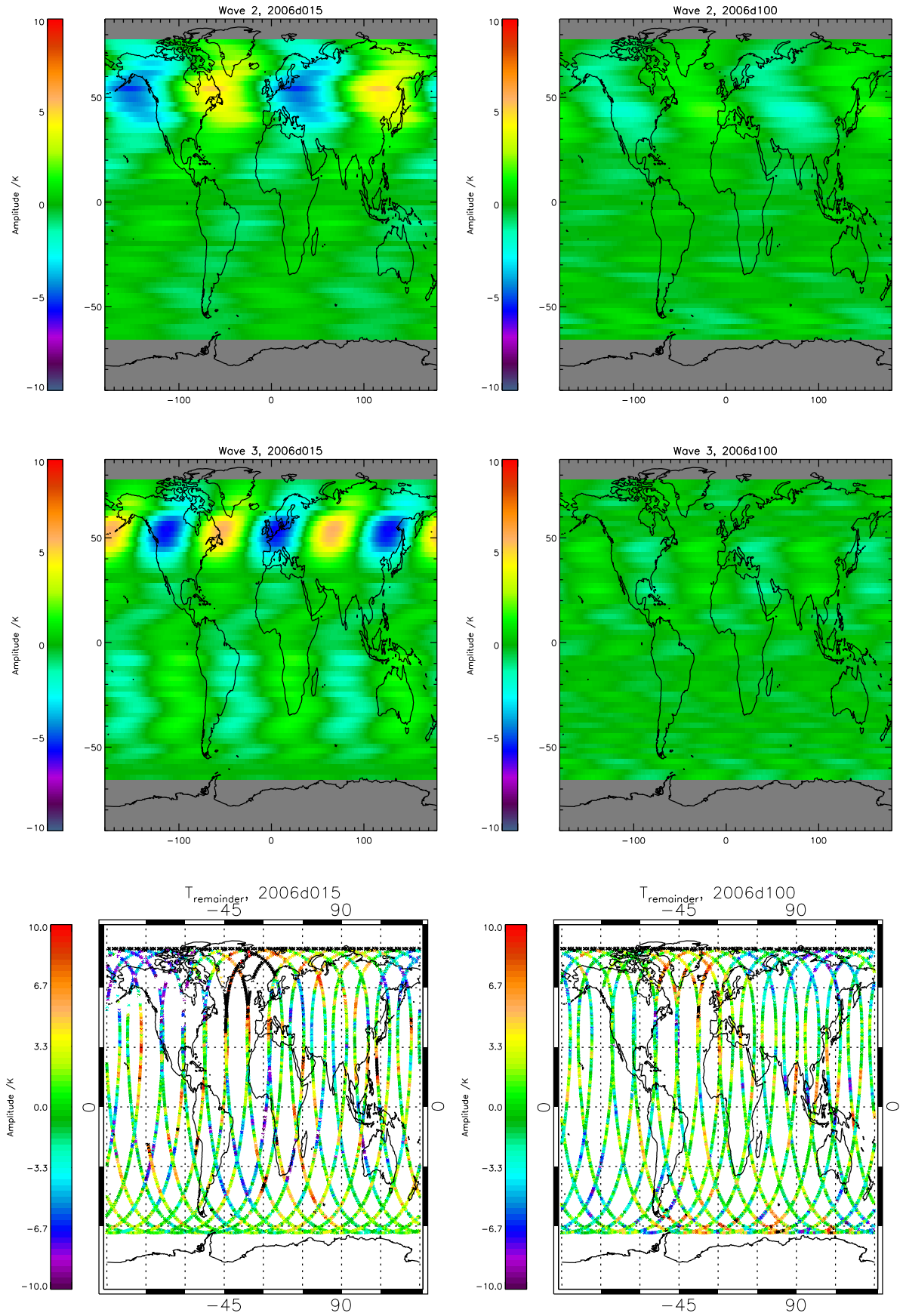


Figure 4.6: *(continued) Example planetary wave results at the 100 hPa (∼16 km) level, derived from the FFT analysis. The original data is shown at the top (as perturbations from the global daily mean), planetary waves mode 1–3 in the middle three rows, and the resulting temperature perturbations (after S-Transform analysis) at bottom. Strong planetary wave signals in the lower modes are observed in the Arctic in northern-hemisphere winter but not summer, as would be expected (see section 5.5), and amplitudes are generally low (∼ a few K) otherwise.*

with a vertical wavelength greater than 16 km. This focuses our results on internal gravity waves with short vertical wavelengths, which are minimally affected by the zero-padding of our data — see section 5.2.2 below for further discussion of this. We also remove any data which lies outside the height range we are interested in studying, including our earlier zero-padding.

We then take the cospectrum of each pair of profiles, as described in section 4.2.3, and analyse each height level individually. For each height level, we find the peak in the covariance spectrum, giving us the dominant vertical wavelength present across both profiles. Once we have this value, we then calculate the covarying temperature amplitude $\hat{T}_{i,i+1}$ and horizontal wavenumber $k_h = \Delta\phi/\Delta x$, where Δx is the along-track separation distance between the adjacent profiles, computed from the HIRDLS geolocation data. We also compute the momentum flux present

$$M_{i,i+1} = \frac{\rho}{2}\frac{\lambda_z}{2\pi}k_h\left(\frac{g}{N}\right)^2\left(\frac{T'}{\bar{T}}\right)^2, \tag{4.6.1}$$

as derived in section 3.2.2. Note that the terms used here are slightly different from those used in section 3.2.2 as this expression is in terms of the quantities measured by HIRDLS rather than theoretical considerations: the m of equation 3.2.30 is replaced by $\lambda_z/2\pi$, and the F_{ph} by $M_{i,i+1}$ to emphasise that this is a momentum flux measured across two profiles rather than the absolute value which would be obtained from a full three-dimensional knowledge of atmospheric conditions.

Figure 4.7 illustrates the cospectrum analysis for a series of artificial profile-pairs, based on the waves

$$\begin{aligned}\Lambda_3 &= [4, (25, 32.500, 40)]\\ \Lambda_4 &= [4, (25, 32.500, 40)]\\ \Lambda_5 &= [4, (25, 33.125, 40)]\\ \Lambda_6 &= [12, (25, 32.500, 40)]\,,\end{aligned}$$

where Λ_4 has an amplitude of -1 K, equivalent to inverting the phase of the wave. The change in centre-altitude of Λ_5 is equivalent to a $\pi/2$ phase shift applied to Λ_3. Short wavelengths are used to provide sufficient wave cycles across the wave packet for the Transform to resolve the signal well (see section 5.2).

The first case (top row) is that of two identical wave signals, both represented by Λ_3. The three plots shown on this row illustrate, from left to right, the input waves (offset by ± 2 K), which in this case are identical, the covarying amplitude $\hat{T}_{i,i+1}$ of these signals, and the phase change between these waves $\Delta\phi$. In this case, the resolved peak amplitude is 0.96 K, approximately equal to the input signal 1 K (see section 5.3), and is resolved at the correct height and wavelength. The third plot, $\Delta\phi$, is much noisier, but the value at the height and wavelength of the peak resolved signal is 0.08 cycles, approximately equal to zero and hence in agreement with our expectations.

The second case is two waves in antiphase. In this case, the covarying amplitude is once again resolved at the correct amplitude, height and wavelength. The measured phase difference between the two waves is 0.42 cycles: the correct value for waves in antiphase should be 0.5 cycles, but given the resolution limits applied to the input signal, this is a reasonable approximation to the input signal.

In the third case, the two waves provided are identical in wavelength and amplitude, but differ in phase by $\pi/2$, that is to say by $1/4$ of a wave cycle. Once again, the results obtained at peak covarying amplitude agree with the input.

Finally, the fourth case shows two waves chosen to be significantly different in wavelength, but over the same altitude range. As the middle figure shows, the resulting covarying amplitudes are ill-resolved, with only a weak peak value measured.

4.6.1 Examples Using An Ideal Wave

Figure 4.8 illustrates fully the use of our S-Transform analysis to obtain a profile of peak temperature perturbations as a function of height. The top panel illustrates four identical waves of the form Λ_3, with the wave progressing by 1/4 of a wavecycle between each profile. The second row then shows the cospectra of the S-Transforms of each adjacent profile pair.

The resulting temperature perturbation profiles are then shown at the bottom of the page. For each cospectrum, the peak temperature amplitude at each height level is extracted, and plotted as a function of height. The grey highlighted region illustrates the region of full overlap between the two adjacent waves, the coloured bars at right the full extent of the two adjacent waves, colour-coded to match the profiles at top, and the dashed line the input wave amplitude.

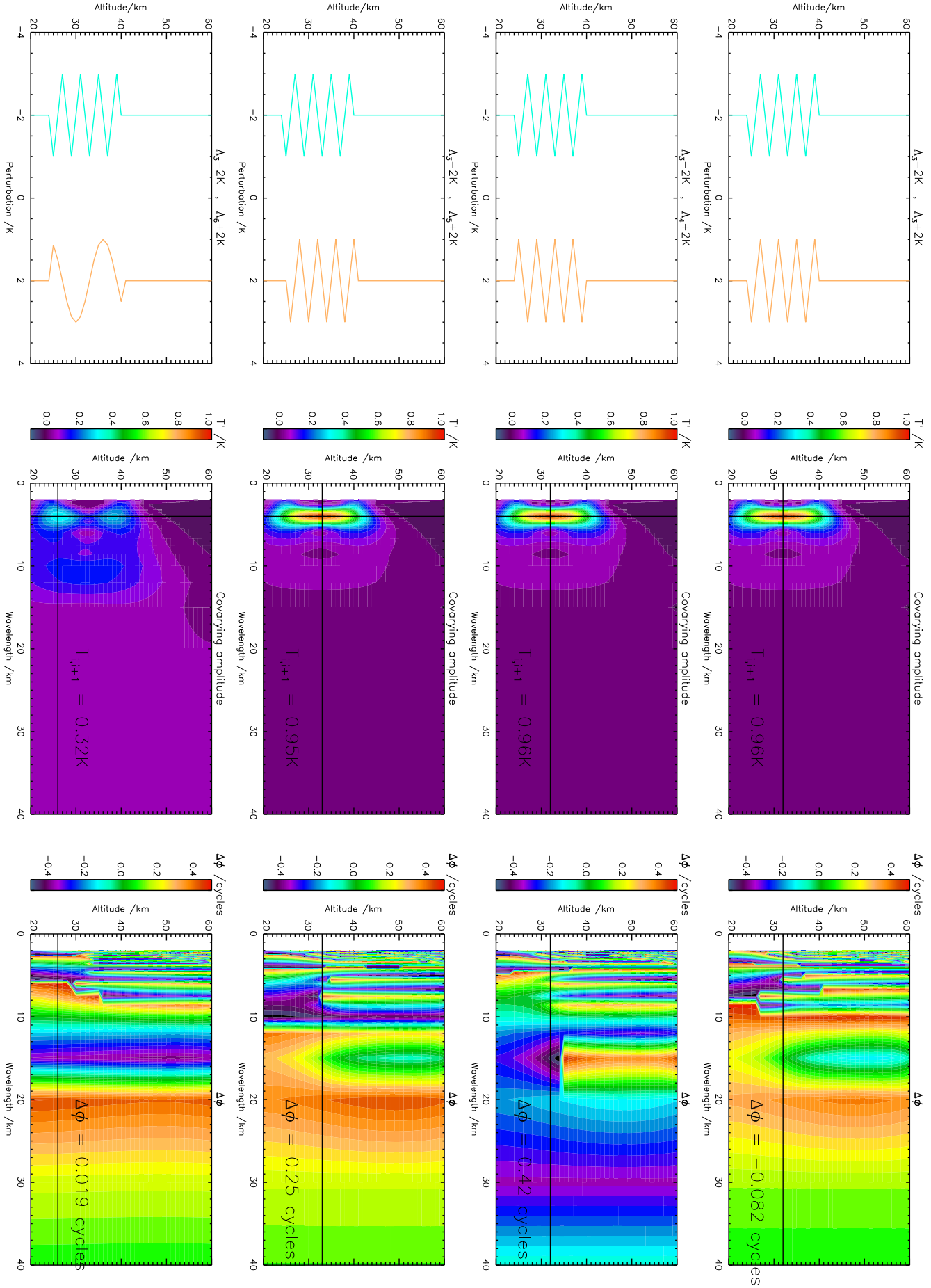


Figure 4.7: *Illustration of the application of the cross S-Transform to our data. See text for further details.*

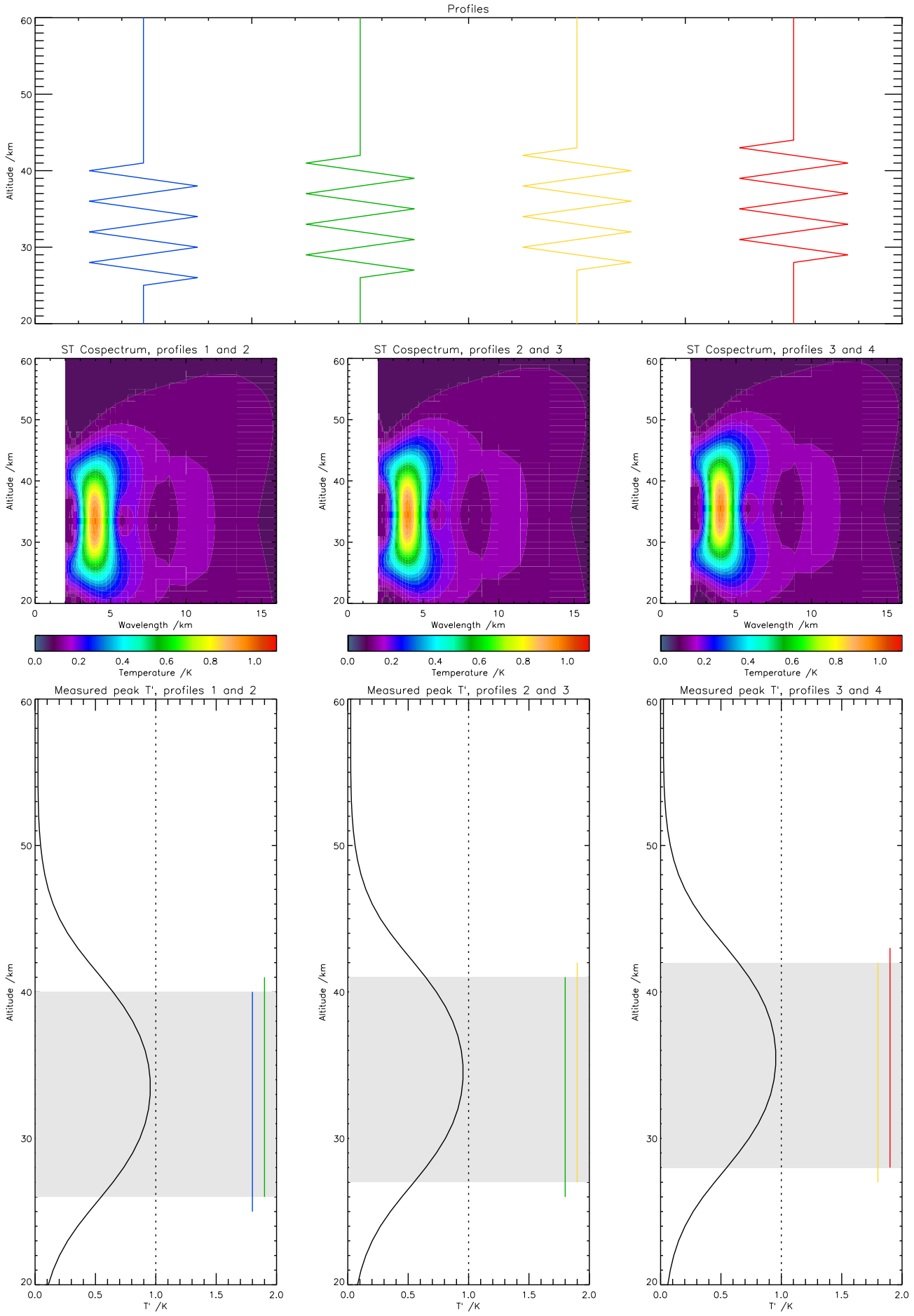


Figure 4.8: *Sample analysis of the extraction of a temperature perturbation profile from an ideal wave.*

The results show a peaked region in the temperature perturbation profile, with the measured perturbation value approaching closest to the input wave amplitude at the centre of the region. There is still a small discrepancy between the input and output temperature amplitudes even at this peak; this will be discussed further in section 5.3.

Figure 4.9 illustrates an identical analysis, with the final product being the vertical wavelength of the dominant temperature feature at each level, and figure 4.10 the absolute phase differences between the profiles; the latter figure also includes the ST-derived phase difference cospectrum of each profile pair. The results for wavelength are excellent, with a very close fit to the input wavelength throughout the region where the waves exist; the measured wavelength peak only deviates from this value at large vertical distances from the wave packet. The phase results are similar, with the phase difference between the profiles being computed near-exactly throughout the overlap region.

4.7 Real Examples

Unlike the ideal wave cases illustrated above, we expect more ambiguous results from real wave cases. The real atmosphere has multiple additional confounding factors, such as noise and waves which may be deformed, overlapping or simply significantly undersampled. Accordingly, it is instructive to consider some real cases.

Figure 4.11 illustrates the analysis process as applied to a strong wave event observed at 59S, 51W on 2006d121, and figures 4.12 and 4.13 illustrate respectively a weak wave event (at 63S, 72W) and a region with no significant observed gravity wave activity (71N, 97E) on the same day.

We shall first explain the layout of each figure, and then proceed to compare the results obtained. For each figure, subfigure A illustrates a curtain plot along the scan track of HIRDLS, showing the measured atmospheric temperature as a function of height in the vertical axis and distance along the scan track in the horizontal axis. Subfigure B shows the same data, but with the mean background and first three planetary wave modes removed, leaving just the small-scale temperature perturbations. Subfigure C then shows the same perturbation profiles used to generate subfigure B, with the individual profiles offset to allow them to be individually distinguishable. Two of the profiles shown in subfigure C are shown as solid lines while the rest

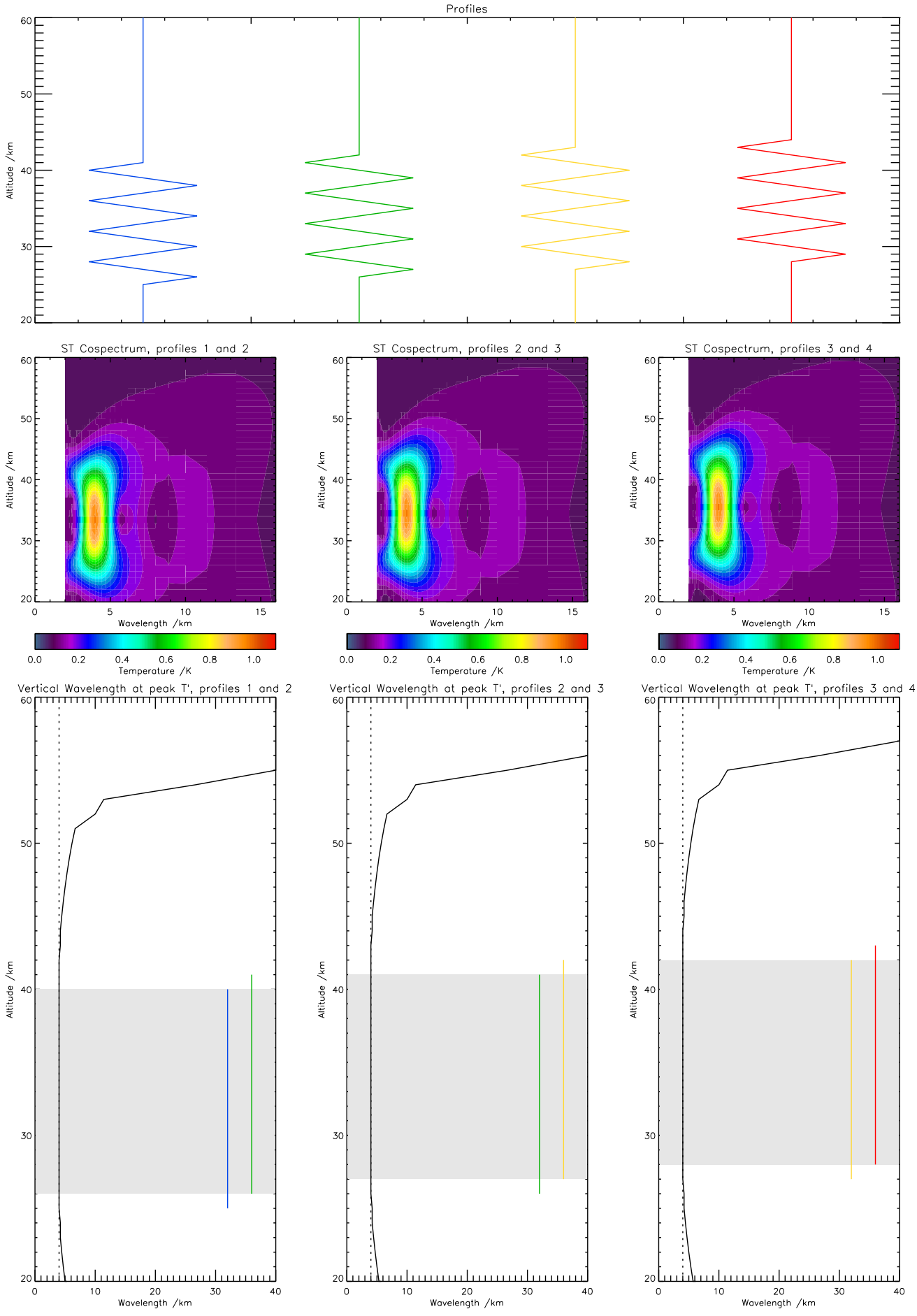


Figure 4.9: *Sample analysis of the extraction of a vertical wavelength profile from an ideal wave.*

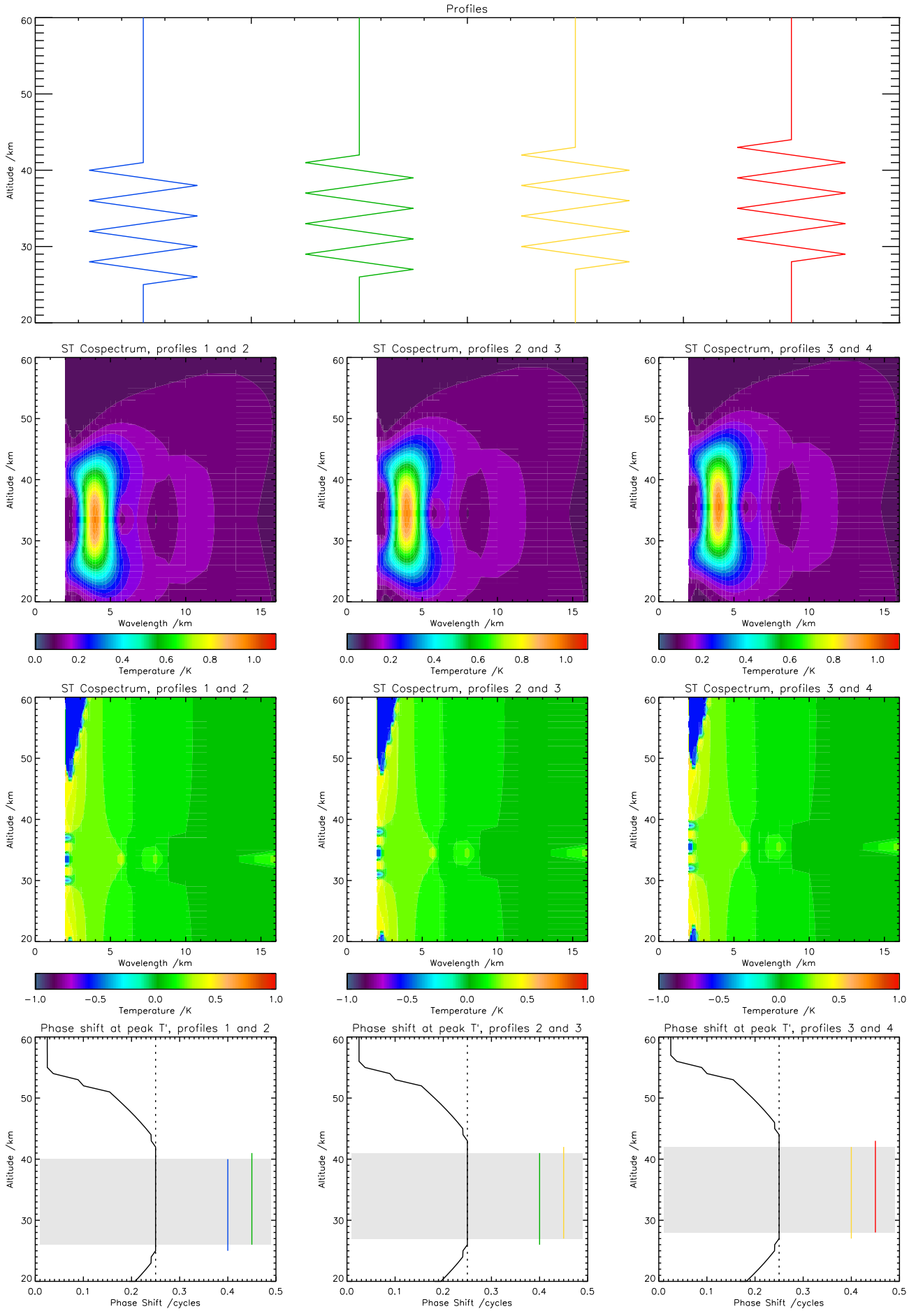


Figure 4.10: *Sample analysis of the extraction of a wave phase profile from an ideal wave.*

are shown as dashed lines: the two solid-line profiles are the ones which the remainder of the figure will proceed to show the analysis of using the S-Transform, and are also shown as vertical dotted lines in subfigures A and B.

We then move on to the S-Transform analyses of the data. Subfigures D and E show the S-Transform amplitude spectra of the two selected profiles, and subfigure F the cospectrum of the profiles. From this, we compute subfigures G and H, which respectively represent the peak measured temperature perturbations and the vertical wavelengths at this cospectral peak for each height level, computed as discussed in section 4.6.

Consider first the strong wave case, figure 4.11. In this figure, a major wave event is clearly visible even in the unfiltered temperature data (subfigure A), centred in the middle of the curtain plot. With the background removed (subfigure B), a strong gravity wave is observed with a significant temperature amplitude of 10s of K. This wave is also clearly visible in the individual profiles (subfigure C), with the profiles overlapping despite their offset. Planetary wave amplitudes are low at this time of year in this location, leaving the variance to be explained by smaller-scale gravity waves.

The S-Transforms of the two selected profiles (subfigures D and E) both show a significant peak of around 10–15 K in a vertical wave mode of around 16 km vertical wavelength, close to the cutoff limit applied to our data, consistent with the data shown in subfigures A and B. When considered together, their cospectrum (subfigure F) leaves a 9.5 K peak in this mode; when we select the mode with the highest temperature amplitude at each height, the resulting profile (subfigure G) shows gravity wave variances peaking at around 48 km in altitude and 9.5 K. Allowing for a proportion of variance to be carried by other-wavelength modes, this seem a good estimate for this wave.

We shall next consider a substantially weaker wave event, to illustrate that the ST analysis process also resolves these. Figure 4.12 illustrates such an event; this feature is faintly visible from around 40–55 km in altitude at the centre of subfigure B as a series of slightly-slanted horizontal lines in temperature perturbation, with an amplitude of around 10 K, and shows up fairly well in the individually-selected profiles. The S-Transformed profiles show a clear feature of nearly 6 K at a wavelength of 8 km, and this feature is faithfully reproduced in the cospectrum and resulting peak temperature and wavelength profiles.

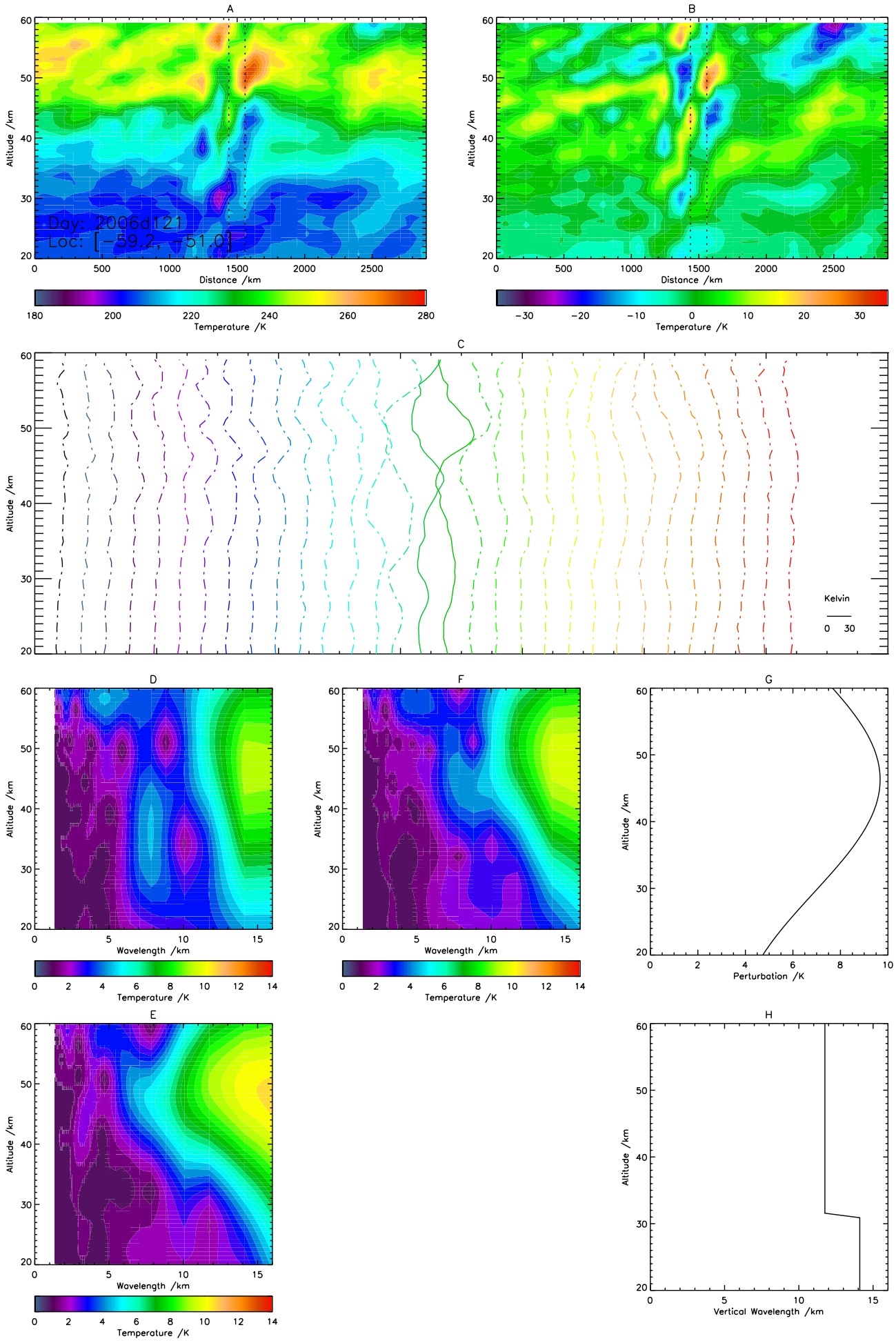


Figure 4.11: *Sample analysis of a strong gravity wave.*

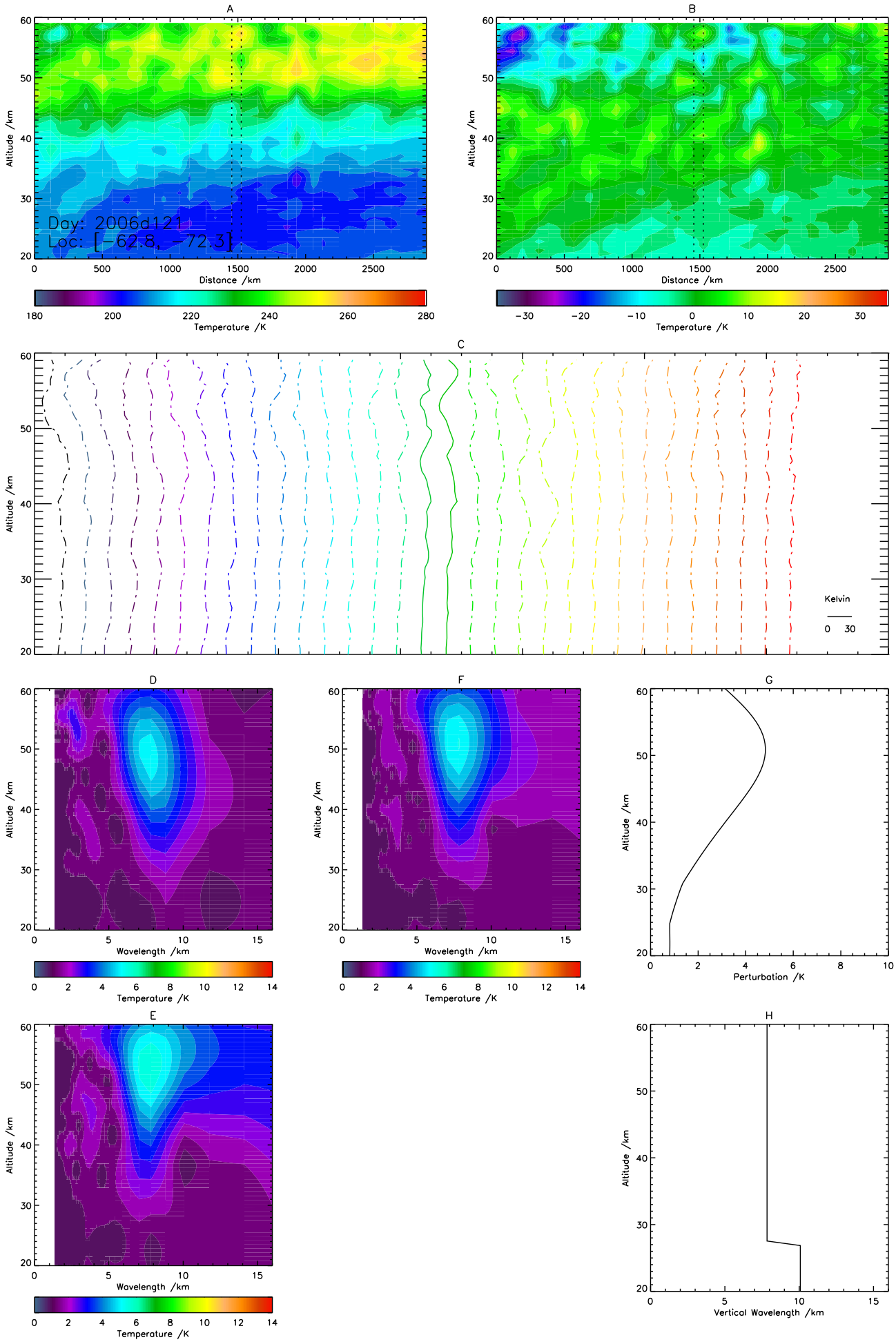


Figure 4.12: *Sample analysis of a weak gravity wave.*

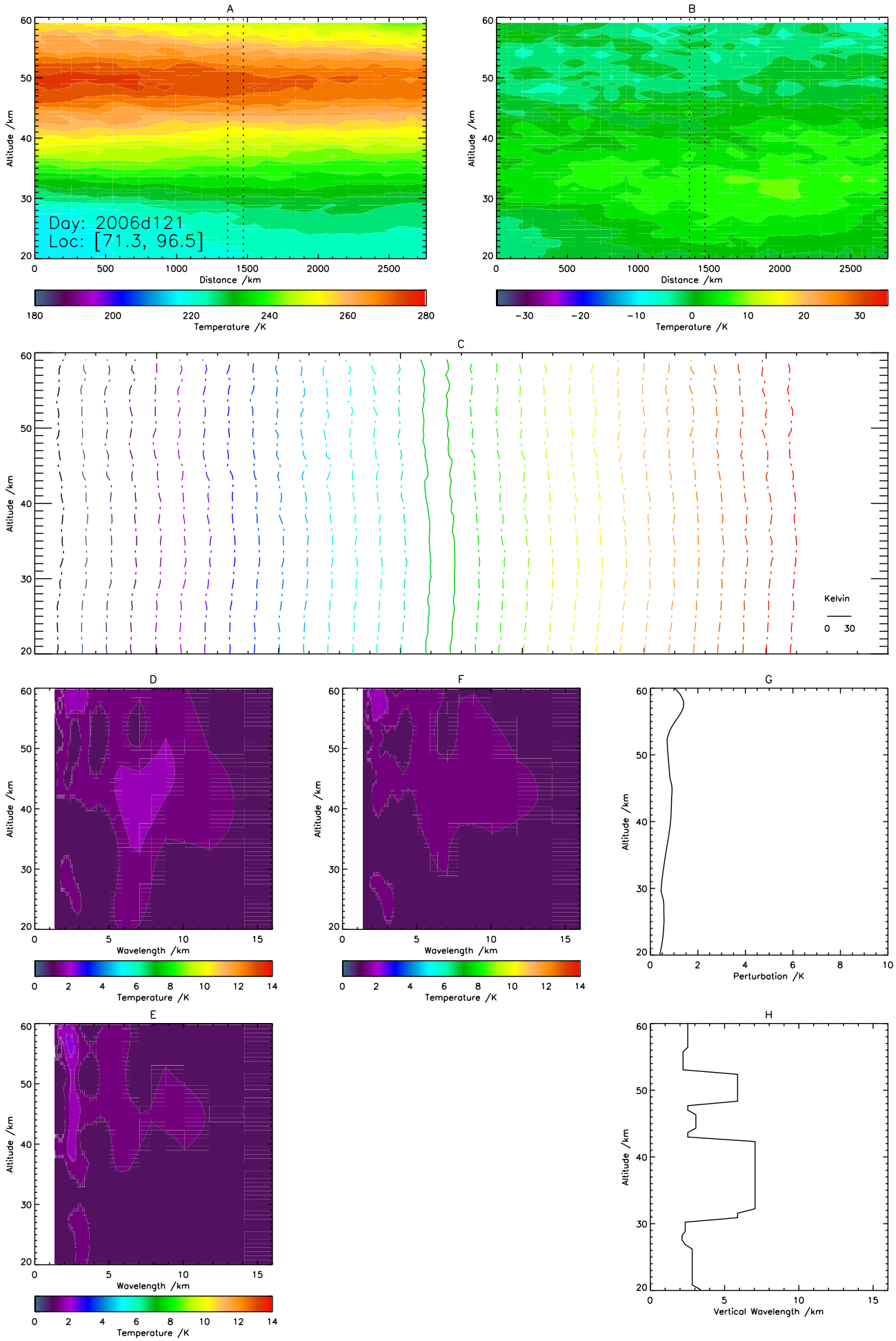


Figure 4.13: *Sample analysis of region with no significant gravity wave activity.*

Finally, we consider a case with no significant observable gravity wave activity, figure 4.13. In this figure, we see only noise in the S-Transform amplitude spectra and in the cospectrum, and peak temperature perturbations across the entire height range are less than 1 K. A range of vertical wavelengths are observed, but these can be discounted as noise due to the extremely low temperature amplitudes.

Chapter 5

Gravity Wave Detection Using the Stockwell Transform, Part II: Performance Assessment

5.1 Overview

The previous chapter discussed the theory behind gravity wave detection using the Stockwell Transform and the details of this method as applied to HIRDLS data. This chapter will discuss the limitations of this method, focusing on the motivation behind the selection of the various fixed parameters chosen and on the effects of random noise upon the gravity wave signals detected.

5.2 Frequency Filtering

As discussed in the previous chapter, we apply a cutoff to our S-Transform resolved data at 16 km. This section discusses the reasons for the selection of this criterion. It then continues further to investigate the effects of varying the number of analysed wave cycles in a given wave packet upon the measured wavelengths and amplitudes of cross-ST analysed data.

5.2.1 Resolution Effects

Figure 5.1 illustrates three sample waves,

$$\begin{aligned}\Lambda_7 &= [10, (30, 40, 50)] \\ \Lambda_8 &= [20, (30, 40, 50)] \\ \Lambda_9 &= [30, (30, 40, 50)]\end{aligned}$$

The S-Transform of each of these waves is also illustrated, and the peak vertical wavelength as detected by the S-Transform analysis is indicated on each plot. As we increase the input wavelength from left to right, the detected wavelength becomes significantly less accurate: at 10 km, the detected peak is exactly where it would be expected, but for the 20 km wave, we return a peak at 16 km, and the 30 km wave is even less accurately resolved, with a returned wavelength of 20 km.

Figure 5.2 analyses this dependence more systematically. Each green cross on the plot indicates the S-Transform derived peak wavelength of a wave $\Lambda_i = [\lambda_i, (30, 40, 50)]$ with the vertical wavelength of the wave λ_i varying in the x-direction in 100 m steps. The red line indicates the 1:1 line of measured wavelength equalling input wavelength. The results of this analysis can be divided into three distinct regimes.

In the first regime, at very low input wavelength, the analysis fails completely to resolve the waves, returning entirely spurious wavelength data. This is because of insufficient data coverage: with only one level every kilometre, any wave of wavelength less than 2 km cannot be resolved at all, and ones slightly longer than this will be poorly-resolved at best. Consequently, aliasing effects are strongly observed in the output. This is a limitation of our analysis technique – we cannot eliminate the possibility that aliased short waves are detected by our analysis methods as longer wavelength structures in the range we wish to analyse.

The second regime, covering approximately the range from 2–20 km, is made up of waves which are initially well-resolved, but with the accuracy falling off as the vertical wavelength increases, due to the fixed height range providing a smaller number of wave cycles to fit to (see section 5.2.2).

Finally, the third regime, at high vertical input wavelengths, completely fails to resolve our

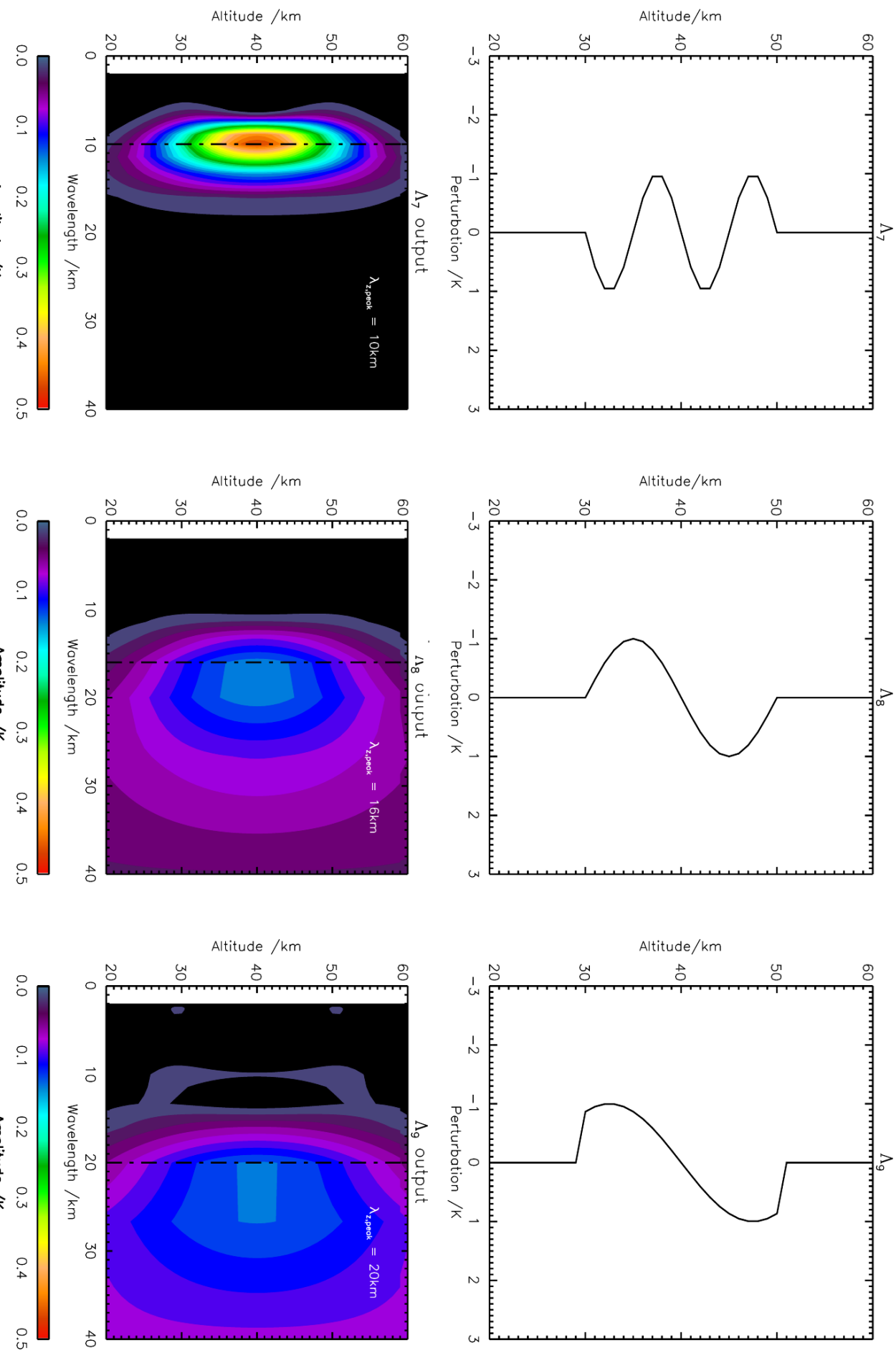


Figure 5.1: *S-Transform output for three sample waves of increasing vertical wavelength, showing the increasing inaccuracy of fit produced for a fixed height range covered by the wave packet.*

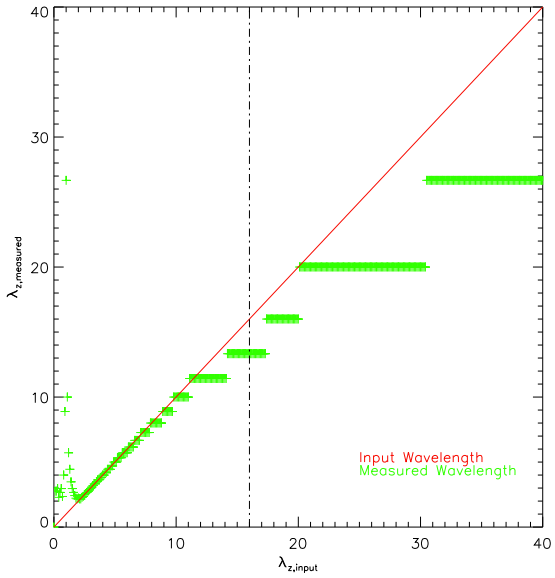


Figure 5.2: *Vertical wavelength measured for a series of waves* $\Lambda_i = [\lambda_i, (30, 40, 50)]$. *The vertical dotted-and-dashed line indicates the 16 km cutoff applied to our data, the red line a 1:1 correspondence between the input and output wavelengths, and the green crosses values computed via our ST method.*

input waves, returning highly inaccurate values for wavelength for all inputs. This is due to the wavelength of the wave being significantly greater than the height range covered by the wave packet: there is insufficient information present in the input profile to correctly identify the wavelength under consideration.

5.2.2 Wavecycle effects

Figure 5.3 shows an equivalent analysis to figure 5.2, varying both vertical wavelength λ_z and the height range ('packet size') covered by the wave $h_T - h_B$ on 100 m scales, for window top height h_T and bottom height h_B. In each case the wave is centred on an altitude of 40 km. The coloured contours show the percentage difference between the input and detected wavelengths, specifically $\lambda_{z,input} - \lambda_{z,measured}$. The region where the input vertical wavelength is less than 2 km is indicated by the vertical dashed line, and the 16 km cutoff by the vertical alternately dotted-and-dashed line.

In the upper plot, the full analysed range of packet sizes and wavelengths is shown, with the horizontal solid line indicating the packet size $h_T - h_B$ used for figure 5.2. This shows a substantial error in the detected wavelengths for waves with long wavelengths defined only over short windows (lower right), with the S-Transform analysis significantly underestimating

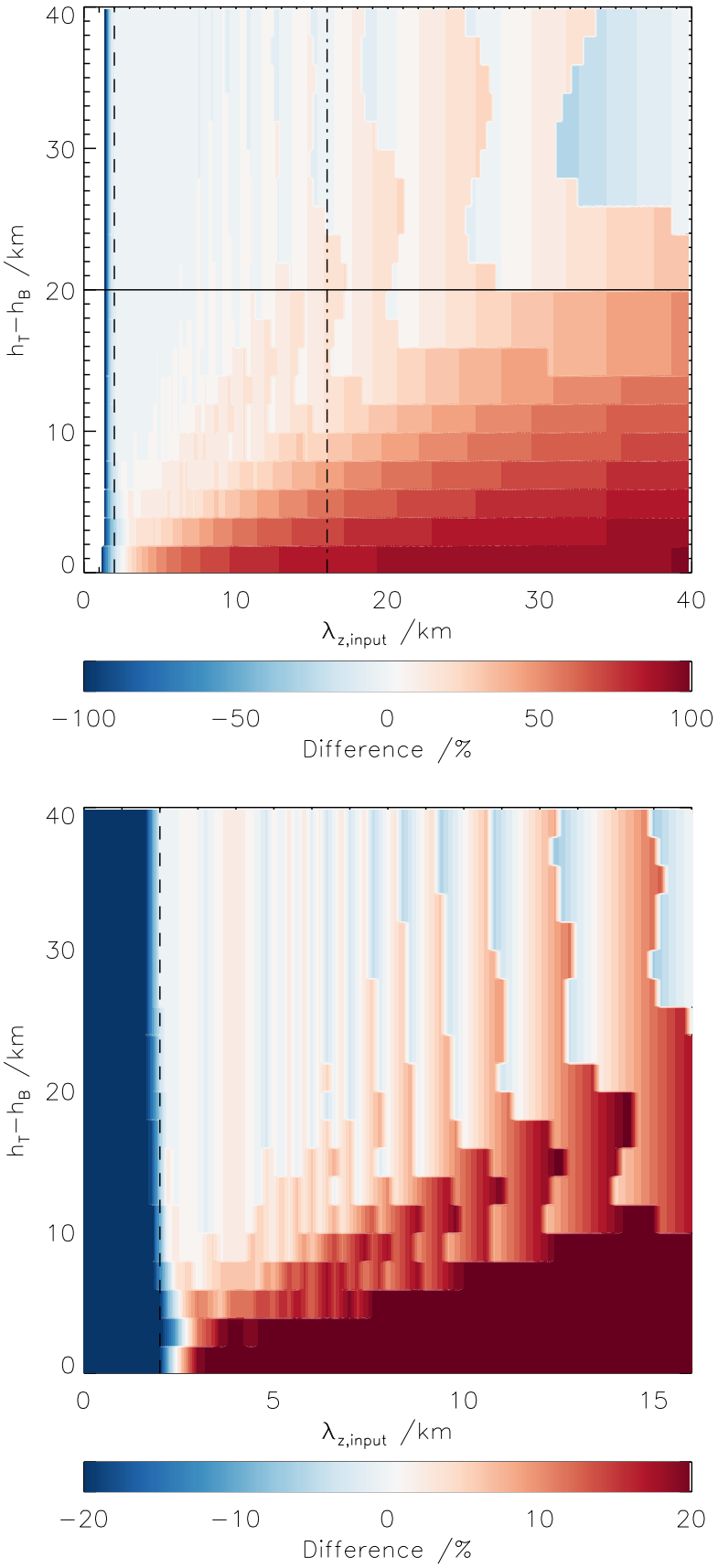


Figure 5.3: *Differences between input and measured vertical wavelengths arising from S-Transform analysis as applied to ideal waves of varying height ranges and vertical wavelengths. The vertical dashed line indicates the 2 km vertical wavelength, the vertical alternate dashed-and-dotted line the 16 km vertical wavelength cutoff, and the horizontal solid line the packet size used for section 5.2.1.*

the vertical wavelength present. At the bottom of the plot, this error reaches 100%. In this region, only a small fraction of a wave cycle is present in the profile being analysed, which leads to this marked discrepancy; for packet sizes comparable to or greater than the wavelength of the signal, the measured wavelength is typically relatively close to the input. An inherent 'blockiness' is also observed in the results, particularly in the lower-right region of the figure. These blocks, typically $\sim$2 km across in both directions, are due to the 1 km resolution of the input data.

The lower plot shows the same data on a 5$\times$ amplified scale, focusing on the portion of the input wavelength spectrum below our wavelength cutoff of 16 km. Once again, the bottom region of this plot shows a substantial error, saturating the colour scale used; this is again due to the high ratio between the input wavelength and the packet size. This plot also once again illustrates that wavelengths less than $\sim$2 km on our 1 km grid cannot be resolved due to the resolution of our data, with large differences for these waves saturating the scales - since waves this short are not Nyquist sampled, they cannot be directly observed. The rest of the plot shows generally small errors, typically less than our vertical resolution of 1 km.

Figure 5.4 reproduces the results of figure 5.3, but analysed instead in terms of the number of complete wave cycles represented by the wave packet; that is to say, for an input wave of vertical wavelength 5 km, the range from zero cycles to ten cycles would represent a height range varying from 0 km to 50 km covered by the repeating wave pattern. Discrepancies are expressed as a percentage of the input wavelength. All waves are centred on an altitude of 40 km, with λ_z analysed in 100 m steps and the number of wave cycles in 0.1 cycle steps.

The top plot once again represents the full analysed range of wavelength and cycles, with the lower plot showing the analysed wavelength range up to 16 km on a 5$\times$ amplified scale. The overlaid black curve divides the plot into two distinct regions: above and to the right of the line, less than one complete wave cycle can exist in our 40 km height range, whilst below and to the left of the line at least one full wave cycle is present in our analysis[1]. Similarly, the dotted horizontal line separates waves with less than one full wave cycle, below the line, from those with at least a full wave cycle, above the line.

The results of this figure tell a similar story to those of figure 5.3, but with several important distinctions. Firstly, it becomes much more apparent that those waves for which S-Transform

[1] Accordingly, all the data shown in figure 5.3 is represented by the region below this curve.

analysis significantly overestimates the vertical wavelengths are indeed those waves for which we have much less than one wave cycle present: all measured discrepancies above the two-wavecycles line are less than 20% of a wavelength, whilst discrepancies at high vertical wavelengths below the line peak significantly above this, saturating the short colour scale on the lower plot.

Secondly, we again see a region of strongly overestimated wavelengths in the low-λ_z region to the left of the plot. This confirms that, for waves of vertical wavelength similar to or less than the Nyquist frequency of the input data, we incorrectly compute the result.

Finally, in the region above the black curve, varying the number of wave cycles present has no effect upon the measured wavelength. This is intuitive: since the waves are always centred on the 40 km altitude level, all analyses with more cycles than fit into the range for a given wavelength will present the same apparent wave pattern to the S-Transform analysis routine, and confirms that our analysis routine is functioning as expected.

5.3 Wave Amplitude

Along with vertical wavelength, the key property which our S-Transform analysis determines is the measured temperature perturbation due to the gravity waves observed, i.e. the amplitude of the waves. This section will discuss this in further detail.

5.3.1 Variation of Amplitude for 1D input

Figure 5.5 illustrates S-Transform output for three waves, each of the form Λ_1 but differing in amplitude, from 1 K at the left through 5 K in the centre to 10 K at right. The first key feature to observe from these three plots is that, for appropriately chosen colour scales, the S-Transform output from the three signals is identical: each output plot has the same peaks and troughs, with the wave peak at the same height and vertical wavelength. Accordingly, provided there are no other confounding factors, the S-Transform of an individual profile should return the appropriate wavelength for a wave of any given amplitude. This is as we would expect: in the absence of noise, the only factor which should affect our results as we change the amplitude of the signal is numerical accuracy, which is not a significant consideration at the precision scales

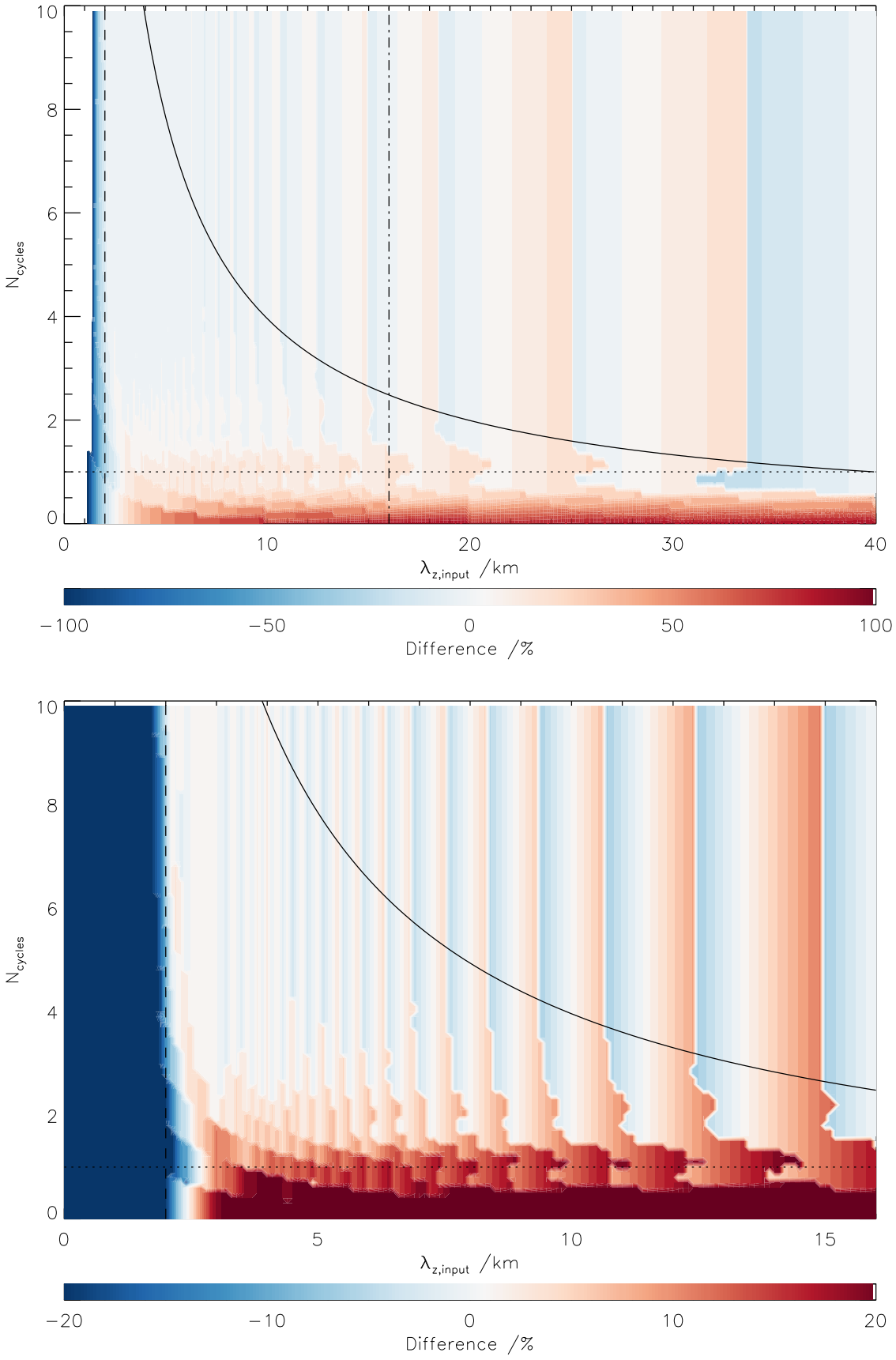


Figure 5.4: *Figure illustrating the errors in S-Transform analysis as applied to ideal waves of varying vertical wavelengths, expressed in terms of number of complete wave cycles present in the analysis range.*

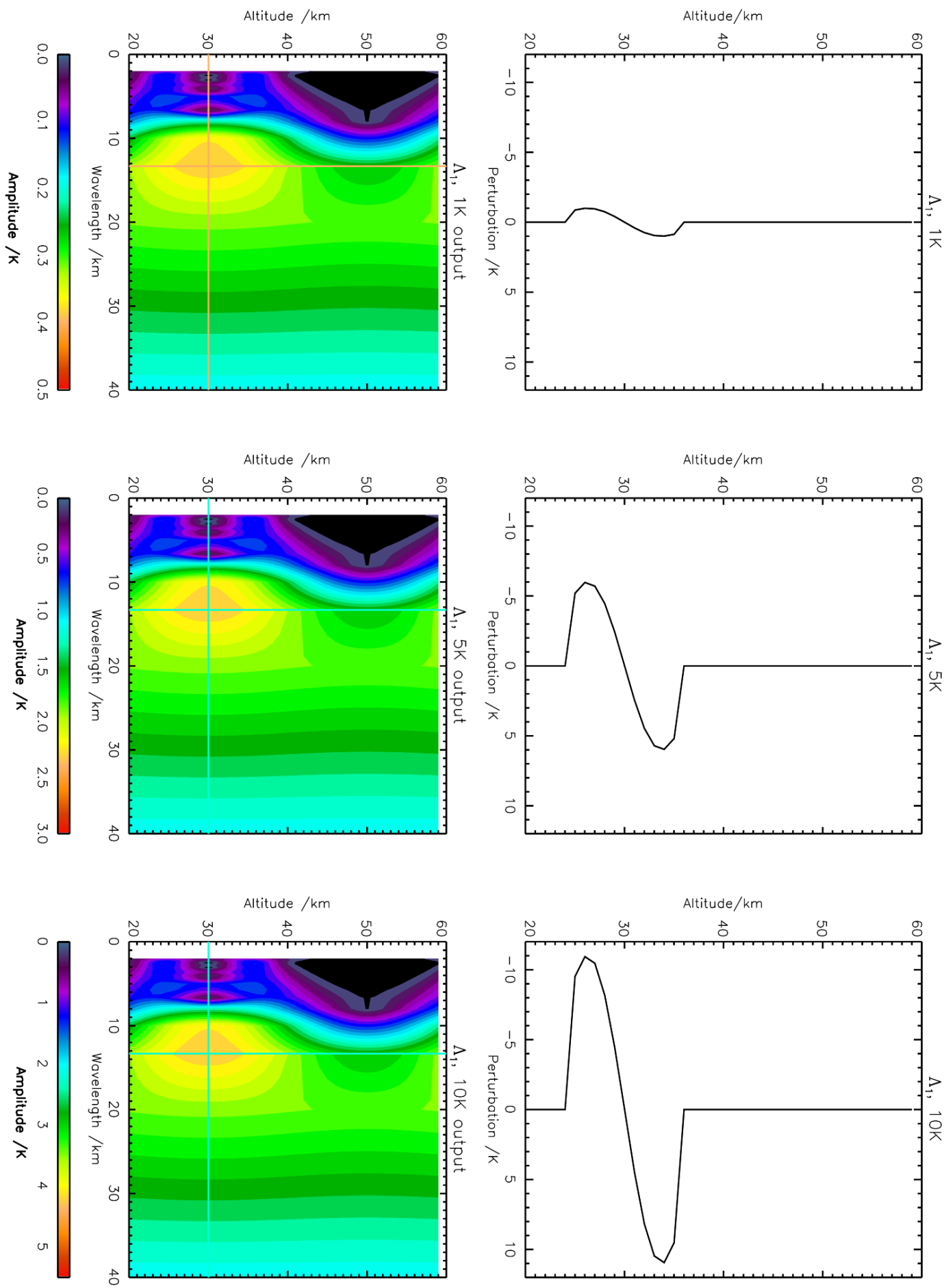


Figure 5.5: *Effects of varying the amplitude of a wave Δ_1 on the S-Transform output for a single profile. Note the different colour scales used for each plot.*

we are considering.

However, this figure also highlights another potential error source in our analysis. The colour scales for each plot illustrate the computed amplitude of the wave in each profile. Crucially, these values are significantly lower than the input amplitudes of the waves. This difference is a function of the wavelength and height range over which the wave is detected, and represents the wavelength fitting issue discussed in section 5.2.

Figure 5.6 considers this further. We discuss first the top panel, showing the effect of varying the number of complete wave cycles present and the wave amplitude, for a constant vertical wavelength of 5 km. The dotted line towards the bottom of the graph indicates the one-wavecycle limit and the solid line towards the top the limit at which the wave cycles exceed the height of the observational window. From this graph we see that only the number of wavecycles present has an effect upon the percentage difference between the input and measured temperature amplitude of the wave, regardless of the input temperature amplitude. This agrees with the data presented in figure 5.5, where varying the amplitude of the wave shifted the returned temperature amplitude of the wave but did not affect the form of our wavelength-height plots. Another key feature to observe from this graph is that in no region is the difference $T'_{input} - T'_{output}$ positive; that is to say, varying the number of wave cycles or the input temperature amplitude only results in an underestimate of the input temperature perturbation. This is because, in the absence of noise, the amplitude of the S-Transform output peak is inherently limited to the peak value of the signal, and inaccuracies in the fit can hence only lead to values lower than this peak.

The lower panel of figure 5.6, meanwhile, considers the effects of varying the number of cycles and the input vertical wavelength for a wave of fixed temperature amplitude 1 K. The axes and overplotted lines of this figure are identical to those of the lower panel of figure 5.4, with the colour contours representing the percentage change in measured temperature amplitude. This plot shows that the input vertical wavelength and number of wave cycles present, unlike the temperature amplitude, have a strong effect upon the measured temperature amplitude for a single S-Transformed temperature profile. The results show a very similar distribution to those for the discrepancy in vertical wavelength discussed in section 5.2.2, with a region of large discrepancy below the one complete wavecycle line and no variation above the full-window-

size curve. At wavelengths less than 2 km differences range up to 100% between the input and measured wavecycles; this is consistent with the wave not being resolved at all due to the 1 km resolution of the input data. Aside from these two previously-described regions, however, the difference between the input and measured temperature perturbations typically amount to only a few percent.

5.3.2 Cospectrum Analysis

According to our definition in chapter 4 of the covarying temperature amplitude as

$$\hat{T}_{i,i+1} = \left| \hat{T}_i \hat{T}^*_{i+1} \, \mathrm{e}^{i\Delta\phi_{i,i+1}} \right|, \tag{5.3.1}$$

we expect the returned peak temperature amplitude of a pair of perfect waves to be equal to the square root of the product of their amplitudes. This section will investigate our data to confirm that this relationship exists, as a check upon our implementation of the method.

Figure 5.7 illustrates the effects of varying wave amplitude upon cross-ST analysis, in the same format as figure 4.7. The first row illustrates two identical waves of amplitude 1 K. As we would expect, the resulting cospectrum returns an amplitude of approximately this value, with a minimal phase difference between the profiles, and with the correct vertical wavelength. The second and third rows analyse similar identical wave-pairs, with amplitudes of 5 K and 10 K respectively, once again returning the correct amplitude, vertical wavelength and phase. These three wave-pairs are intentionally chosen to be equivalent to those discussed in section 5.3.1 above.

The final row shows the case of two waves of the same wavelength and height range, but differing amplitudes, 1 K and 10 K. Once again, the correct vertical wavelength and a near-zero phase difference are measured for the waves, and the amplitude calculated for this wave-pair is 3.1 K, approximately equal to $\sqrt{1^2 + 3^2}$ as expected.

Figure 5.8 analyses the relationship between identical-amplitude profile pairs further, showing the measured covarying temperature amplitude as a function of input wave amplitude, for a fixed wave Λ_3, with the 1:1 line overplotted. As the figure shows, the relationship between input amplitude and measured amplitude is constant, with the measured amplitude representing 96% of the input for all amplitudes. This precise ratio is a function of the input waveform rather

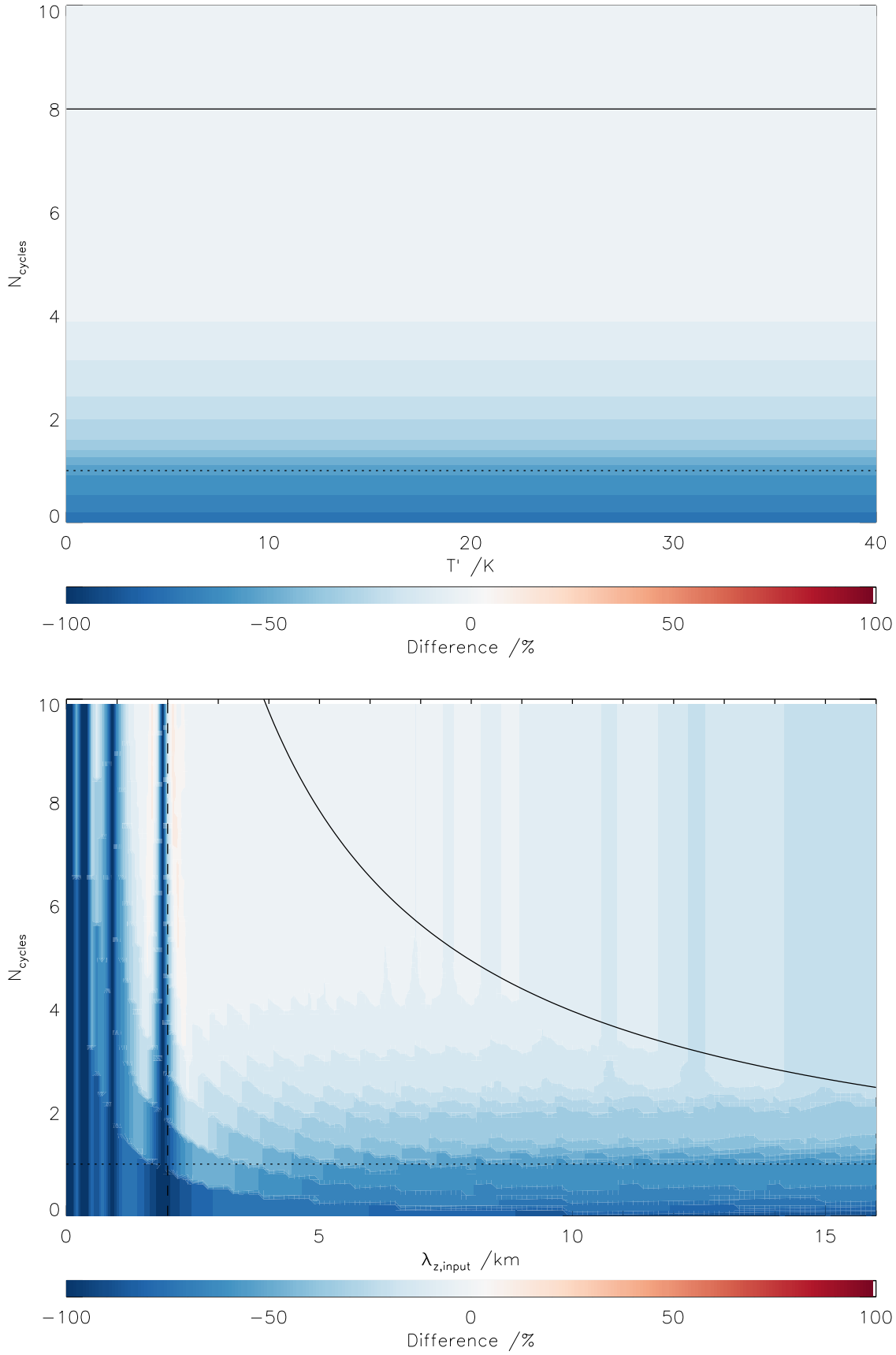


Figure 5.6: *Effects of varying the number of complete wave cycles in an S-Transformed profile on the measured temperature amplitude, as a function of (top) the wave amplitude T' at constant λ_z and (bottom) input vertical wavelength λ_z at constant T'*

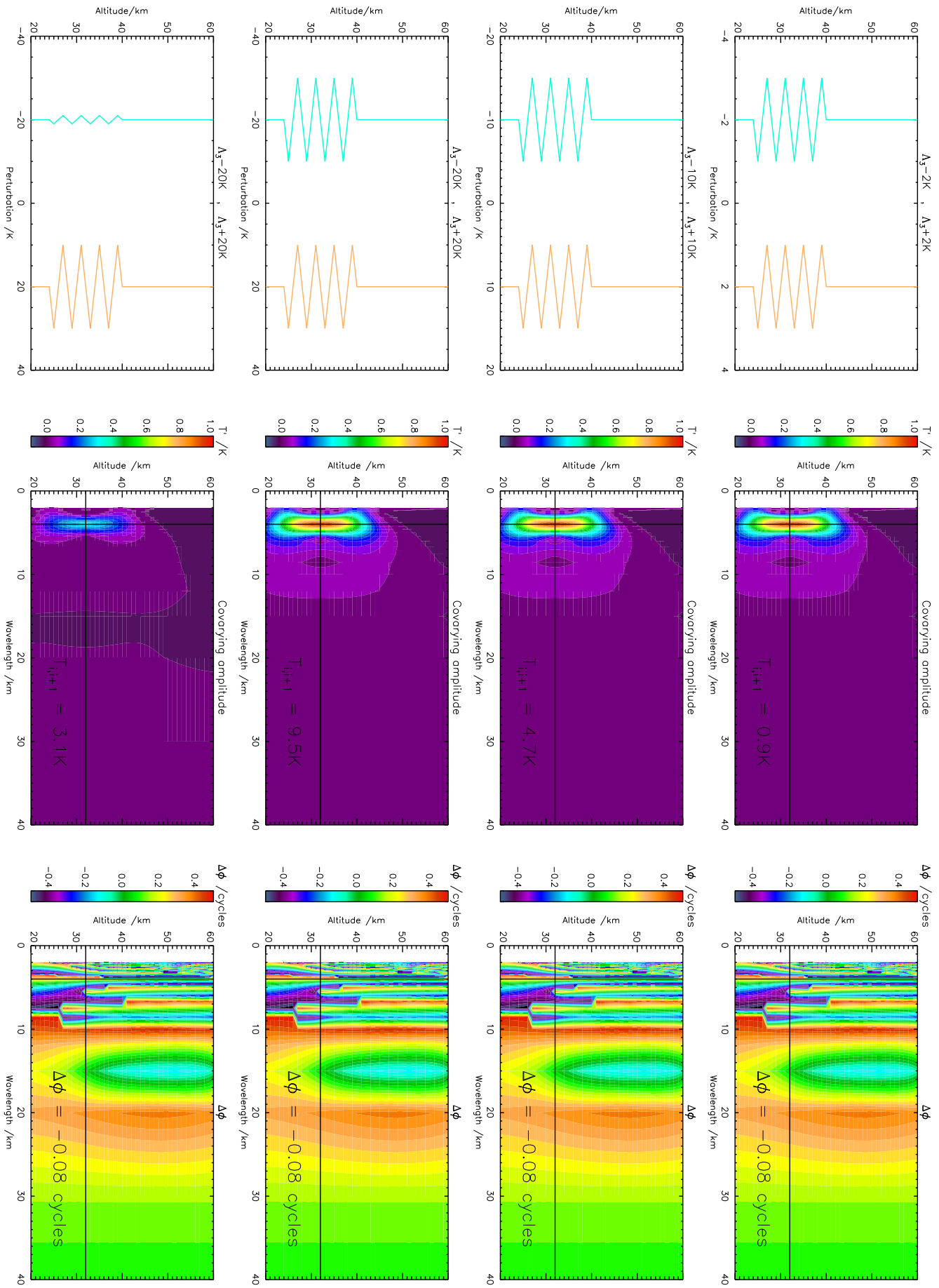


Figure 5.7: *Effect of varying temperature input upon cross-ST analysis.*

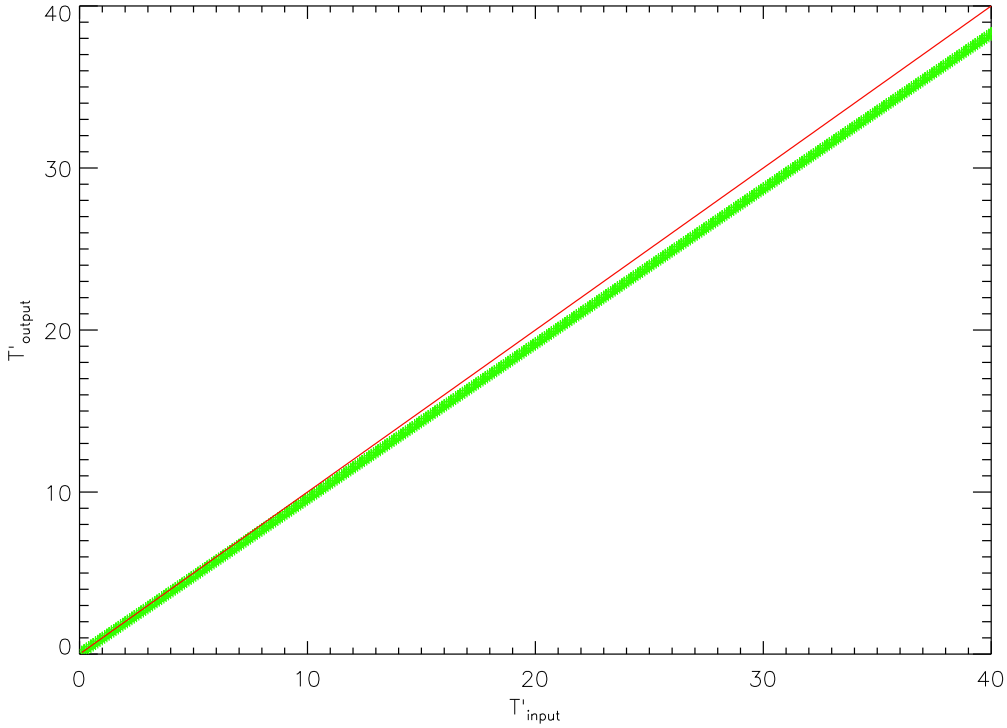


Figure 5.8: *Illustration of the constant relationship between input and output temperature amplitude for identical waves analysed using the cross-ST.*

than the amplitude, and hence represents imperfections of the quality of the wave fit rather than in the temperature perturbation calculation. These results are consistent with the result shown in the top panel of figure 5.6.

Figure 5.9 considers the case of the temperature perturbations computed from the cross-ST of two waves Λ_3 identical in form but differing in their amplitude. The first column shows the returned temperature perturbations as a function of the two input wave amplitudes, the second the difference between the returned value and the square root of the product of the amplitudes, and the third the percentage difference between the returned values and the arithmetic mean of the input amplitudes. In each case, the upper figure shows the full range of values, and the lower figure a slice through the upper figure where the amplitude of wave 1 is 10 K, as indicated by the dashed line. The diagonal line on each of the upper plots indicates the 1:1 line where the input temperature amplitude of each wave is equal. Both directions are analysed in steps of 0.1 K.

Considering initially the first column, we see that the measured value along the 1:1 correspondence line is near-precisely the input amplitude, with only a very small difference at high

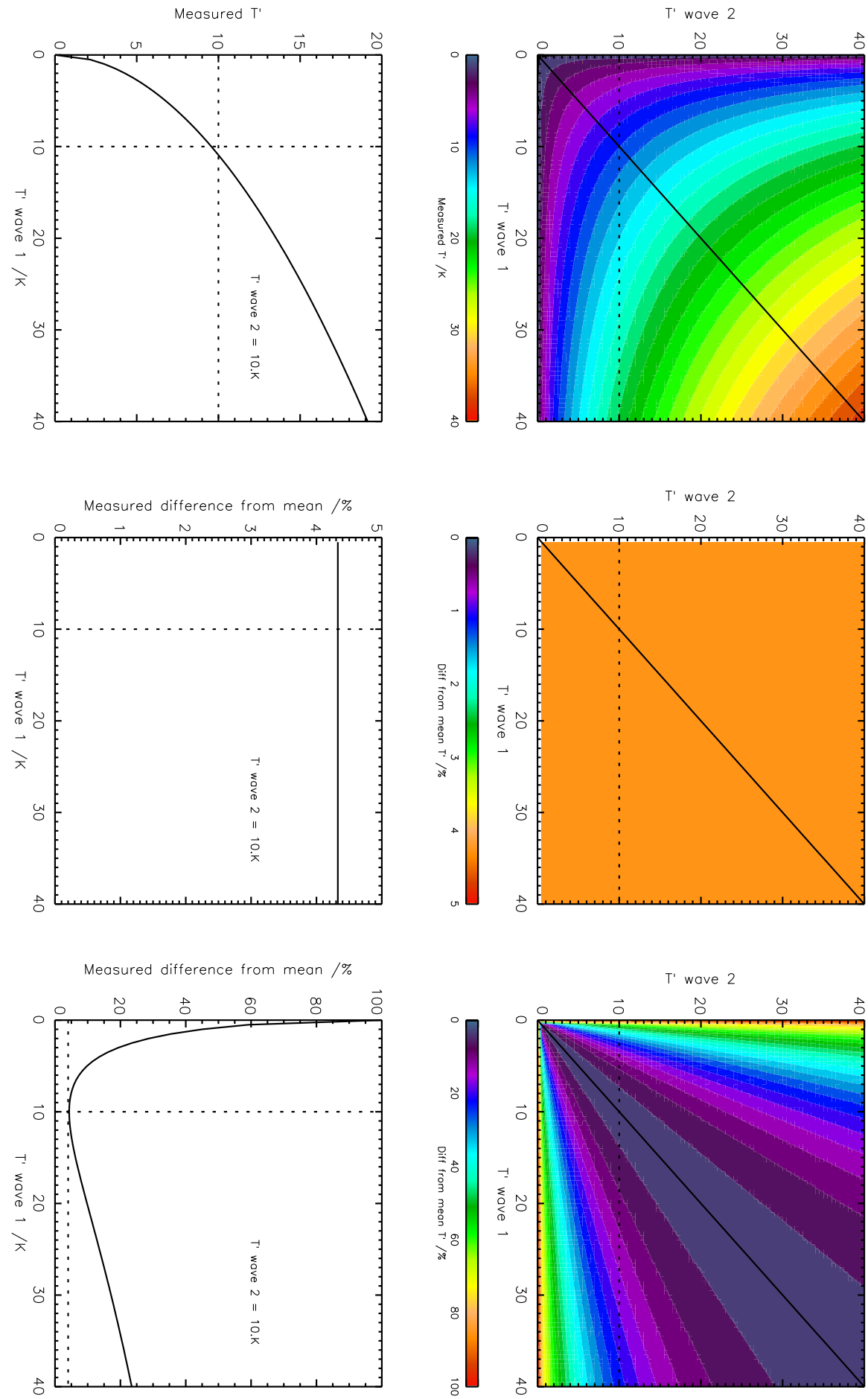


Figure 5.9: *Measured wavelengths from the cospectra of pairs of S-Transformed profiles of identical form but differing amplitudes. Left-hand column shows returned temperature amplitudes as a function of the input wave amplitudes, the middle column the difference between the returned value and the square root of the product of the amplitudes, and the right column the percentage difference between the returned result and the arithmetic mean of the inputs. In each case, the upper plot shows the full space covered by the analysis in two dimensions, and the lower plot a slice through the upper plot at the location of the dashed line.*

amplitudes. This is consistent with the previously shown results and with the result we would expect from equation 5.3.1. The lower plot shows a representative slice through the upper plot at T_1'=10 K; as we deviate from the 1:1 correspondence point, indicated by the crossover of the horizontal and vertical lines, we see the characteristic shape of the curve $y = \sqrt{x}$.

The second column confirms this relationship, showing the difference between the computed temperature perturbation and the square root of the product of the input temperature perturbations. Throughout the entire analysis region, this value is exactly 4%, which as discussed above is the deviation from the expected result we expect for the waveform used.

Finally, the third column shows the percentage difference between the computed temperature perturbation and the arithmetic mean of the input perturbations. The horizontal line in the lower plot represents the 4% diffeence value, which is the minimum we can attain with this waveform. This plot illustrates the difference between what the S-Transform analysis returns and a more naïve interpretation of the results as the mean of the values of the two adjacent profiles.

5.4 Phase

5.4.1 Effects of Phase Variation on a Continuous Wave

Figure 5.10 illustrates the returned phase difference between two adjacent waves offset by a known input phase difference, measured as a proportion of a complete wave cycle, for an input waveform $\Lambda = (20, (20, 40, 60))$. The red line indicates the 1:1 correspondence line of input and output phase, with the green crosses indicating the returned values. As can be seen, for this simple wave, the analysis routine reproduces the input phase difference near-perfectly, with slight offsets observed based on the 1 km resolution of the input data. There is a steplike form to these results: this is due to small shifts in phase representing a shift of less than one full level in our input data.

Figure 5.11 shows the measured phase difference for pairs of input profiles of the form $\Lambda_i = (i, (20, 40, 60))$, with the upper plot indicating the returned phase difference and the lower plot the difference between this and the input phase. Only the phase difference range from 0 to 0.5 cycles is shown; the phase difference range from 0.5 to 1 is identical except for the reversal of sign due to the symmetry of the input waves, so this diagram represents the shift in the absolute

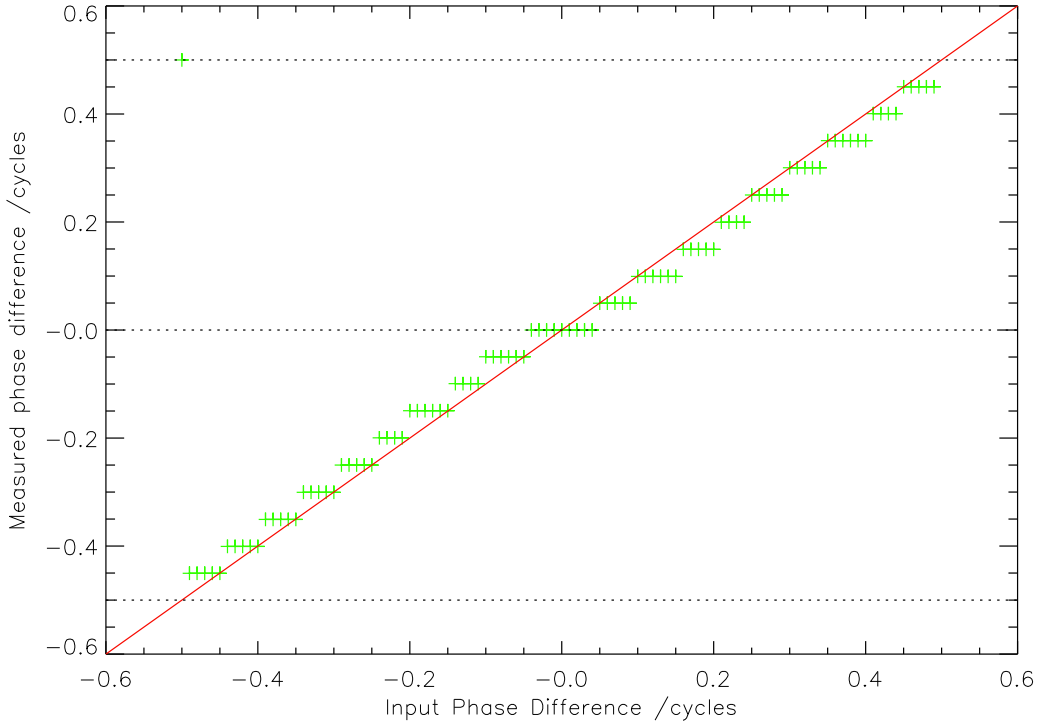


Figure 5.10: *Measured phases for a pair of identical waves* $\Lambda = (20, (20, 40, 60))$ *with a known phase offset, as a function of that phase offset.*

phase difference across the full range. The data is analysed on a 100 m scale in wavelength and on a 0.01 wavecycle scale in phase, and the horizontal dashed line at the bottom of each plot marks a wavelength of 2 km.

Considering first the lower plot, we see that, except for very short wavelengths, the measured phase differences are less than a tenth of a wave cycle throughout. The region of high discrepancy at the bottom of the plot can be explained by comparison to the upper plot, where it can be seen that the measured phase difference in this region is zero; the curve between the zero phase difference region at the bottom and left of the diagram accordingly corresponds to the region where we would expect the phase difference to amount to a shift of less than one full level on our 1 km altitude scale. Similarly, a pattern of steadily-varying regions with sudden jumps is recorded across both plots, consistent with the shift being stepwise due to the defined altitude levels.

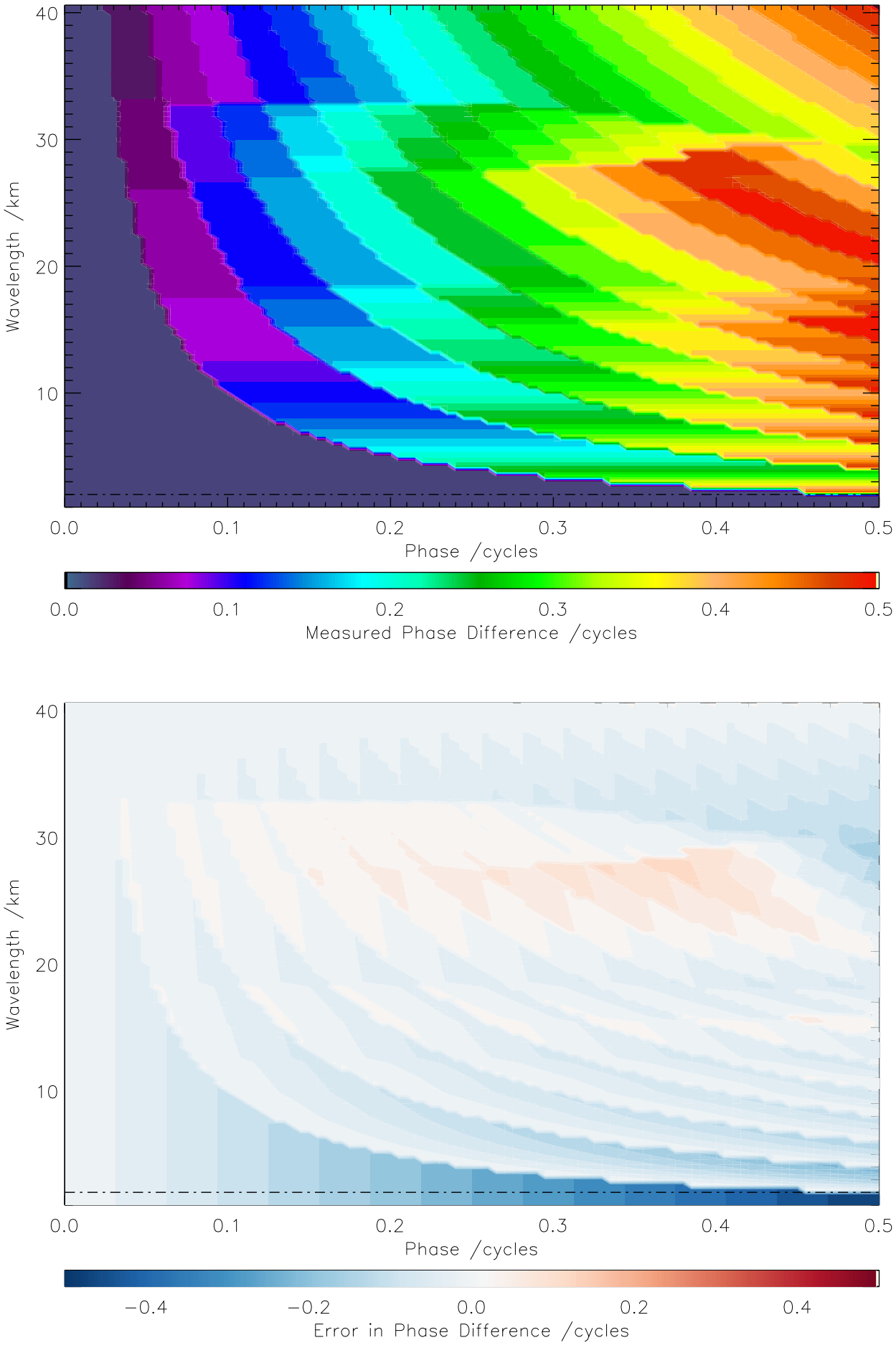


Figure 5.11: *Measured phase differences for a pair of perfect continuous waves, as a function of phase shift and wavelength*

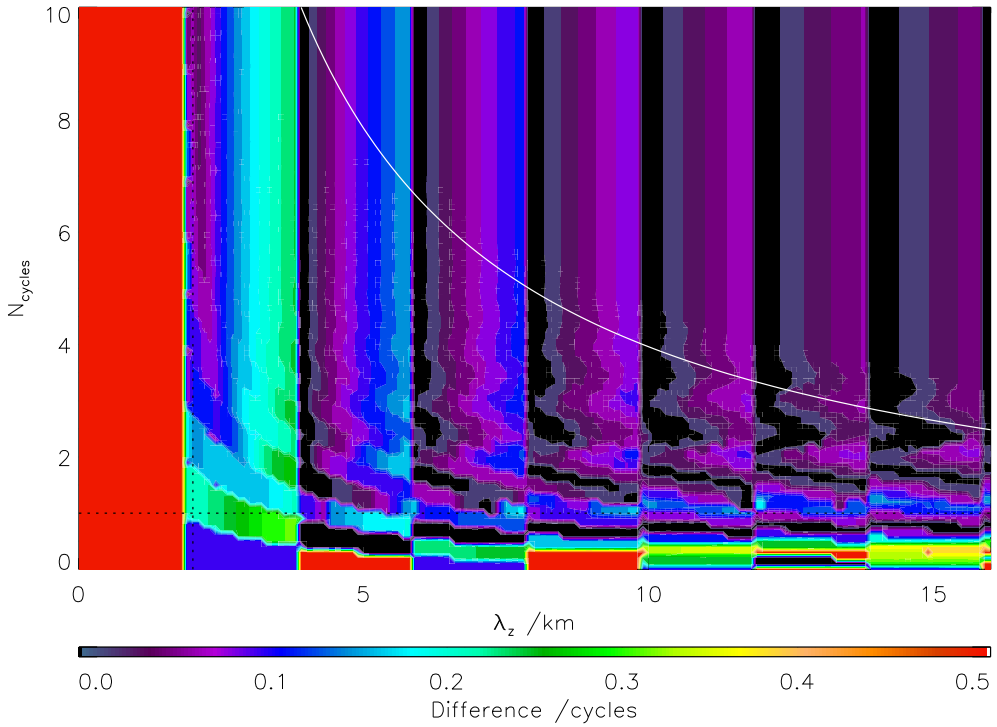


Figure 5.12: *Measured phase differences for a pair of identical wave packets offset in phase by 0.5 cycles, as a function of vertical wavelength and packet size.*

5.4.2 Effects of Phase Variation on Wave Packets

Figure 5.12 shows the phase inaccuracies measured for wave pairs of varying wavelength and packet size centred on an altitude of 40 km, in the same way as previously shown for vertical wavelength and temperature amplitude. Once again, only the absolute phase differences are shown; the two input waves analysed have an offset of 0.5 cycles in each case. The white curve indicates the limit where the wave packet becomes greater than our 40 km range, the horizontal dotted line the one-cycle limit, and the vertical dotted line the 2 km vertical wavelength value.

For waves with wavelengths less than 2 km or packet size less than 1 wavecycle, the measured phase differences are very large, with the phase shift completely undetected at low vertical wavelengths, and only detected for short packets of particular sizes; this is as expected, since waves this short are not Nyquist sampled and hence data at these wavelengths is devoid of meaning. Outside these regions, the plot divides up into vertical stripes of width 2 km, with phase-measurement inaccuracies increasing with increasing wavelength across this stripes, then falling back to zero or near-zero at the end of each stripe. This is once again consistent with our input data resolution; for vertical wavelengths above 4 km maximum phase inaccuracies come

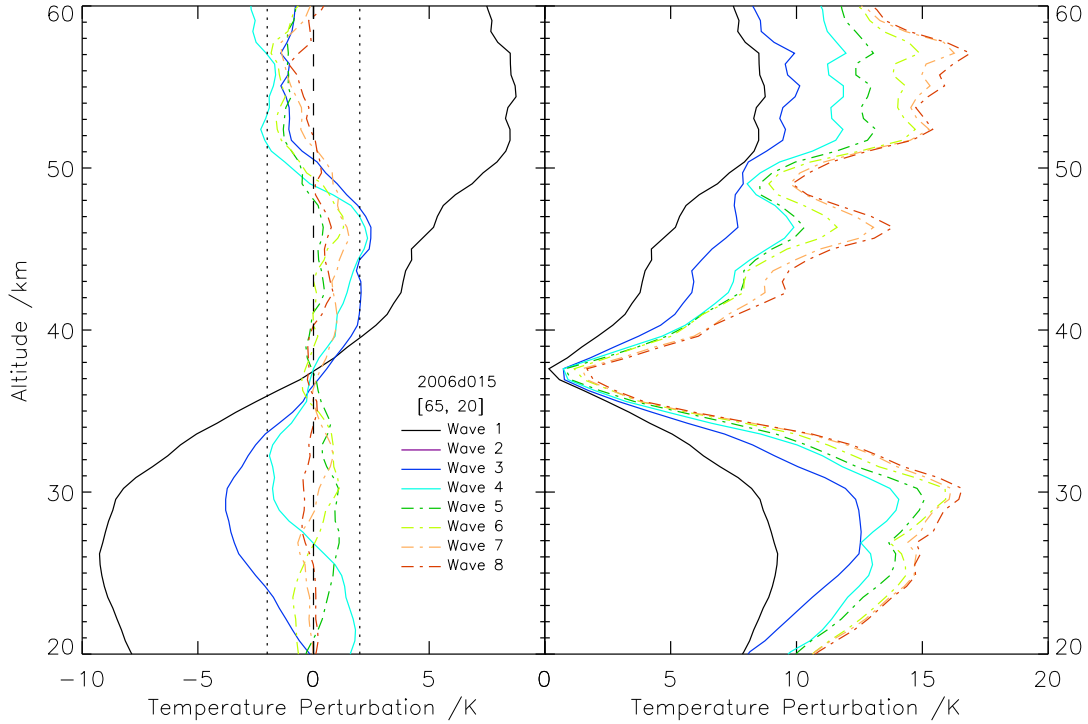


Figure 5.13: *Example illustrating the relative and cumulative magnitude of the first eight planetary wave modes at 65N, 20E on 2006d015 as a function of height.*

to less than 0.2 of a cycle.

5.5 Planetary Wave Removal

As discussed in section 4.5.2, as part of our analysis we remove the background signal due to planetary-scale waves which fit around the Earth up to three times. This section discusses this criterion in further detail.

5.5.1 Examples

Figure 5.13 illustrates the temperature signal associated with planetary wavenumbers 1–8 at 65N, 20E on 2006d015, chosen as it lies in the trough (at the 100 hPa (∼16 km) level) of the planetary wave shown in figure 4.6. The left-hand subfigure show the magnitude of the measured waves as a function of height, and the right-hand figure their cumulative (summed from lowest-order to highest order, that is to say from mode 1 to mode 8) magnitudes as a function of height. At this latitude and time of year, we would typically expect the planetary wave intensity

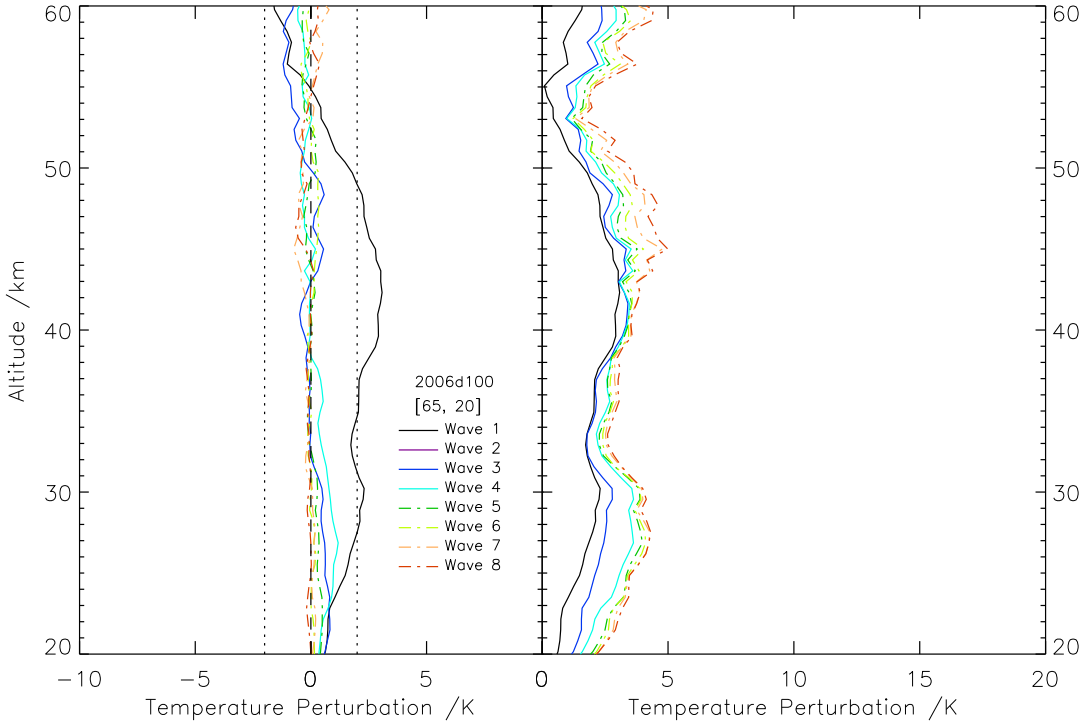


Figure 5.14: *Example illustrating the relative and cumulative magnitude of the first eight planetary wave modes at 65N, 20E on 2006d100 as a function of height.*

to be high, and this example does indeed show large-amplitude planetary wave activity for all modes, with wave-1 magnitude at this location at some heights peaking at nearly 10 K, wave 2 at around 4 K, and wave 3 at 2 K. A trough in wave magnitude appears at around 38 km in altitude for all three of these wave modes: this is due to the geographic location of the planetary wave minimum shifting with altitude; at this altitude, the minimum of the planetary wave happens to lie over the location of the profile.

The remaining wave modes shown, planetary wavenumbers 4–8, have much smaller amplitudes than these three largest wave modes, with peak amplitudes of less than 2 K at all altitudes and their magnitudes as a function of height scattered around the zero line.

Figure 5.14 illustrates the same data for 65N, 20E on 2006d100, i.e. at the same location but during Arctic summer. At this time of year, we would expect significantly lower planetary wave amplitudes, and this is what we observe. All planetary wave signals except for planetary mode-1 are very low, with the majority of the variance once again being in the lower-order modes 1–3.

5.6 Random Noise

5.6.1 Effects on Individual Transformed Profiles

Figure 5.15 illustrates Λ_1, but with statistically random noise overlaid. In the leftmost column, the peak value of this noise is 25% of the amplitude of the input wave, in the middle 50%, and in the rightmost column 100%; in each case, the noise varies randomly between 0% and this value. Only the scaling varies: the form of the noise is identical in each case. As the noise is increased, we observe two key changes in the resultant spectra.

Firstly, high-frequency (low-wavelength) signals are detected at all heights in addition to the primary wave signal at all altitudes. This is as we would expect: essentially, the S-Transform process is attempting to fit waves to the signal, and finding many high-frequency results in the noise. The amplitudes of these additional signals in the S-Transform output are comparable in magnitude to the primary signal, and care must therefore be taken when analysing the results not to treat these as real waves detected. Very little noise is seen at longer wavelengths; this is again consistent with fitting short waves to the random noise but failing to find large-scale structure in the added noise, and is consistent with expectations.

The second effect is a slight shift of the wavelength and height of the peak recorded signal in the spectra, due to the best fit being shifted slightly by the noise. This is more difficult to compensate for, and must be factored into any considerations of the uncertainty on our analysis.

5.6.2 Effects on Amplitude Measurements

Figure 5.16 shows the results of a Monte Carlo analysis, with 10 000 runs, of the absolute difference between the input and measured cross-ST temperature perturbations as a function of noise level, computed by cross-ST analysis of two identical waves. The noise level is defined such that the value shown represents the maximum amplitude of the noise, as a percentage of the input wave amplitude, as with figure 5.15 above, and is analysed in 5% steps.

Wave Λ_3 is used, with an amplitude of 1 K. As discussed above, the accuracy of the measured temperature perturbation is not affected by the input amplitude, so this analysis applies to all equal-magnitude wave pairs of this form; the effects of different wave forms will be discussed below. The zero-noise line is rebased to match the difference measured for this waveform at

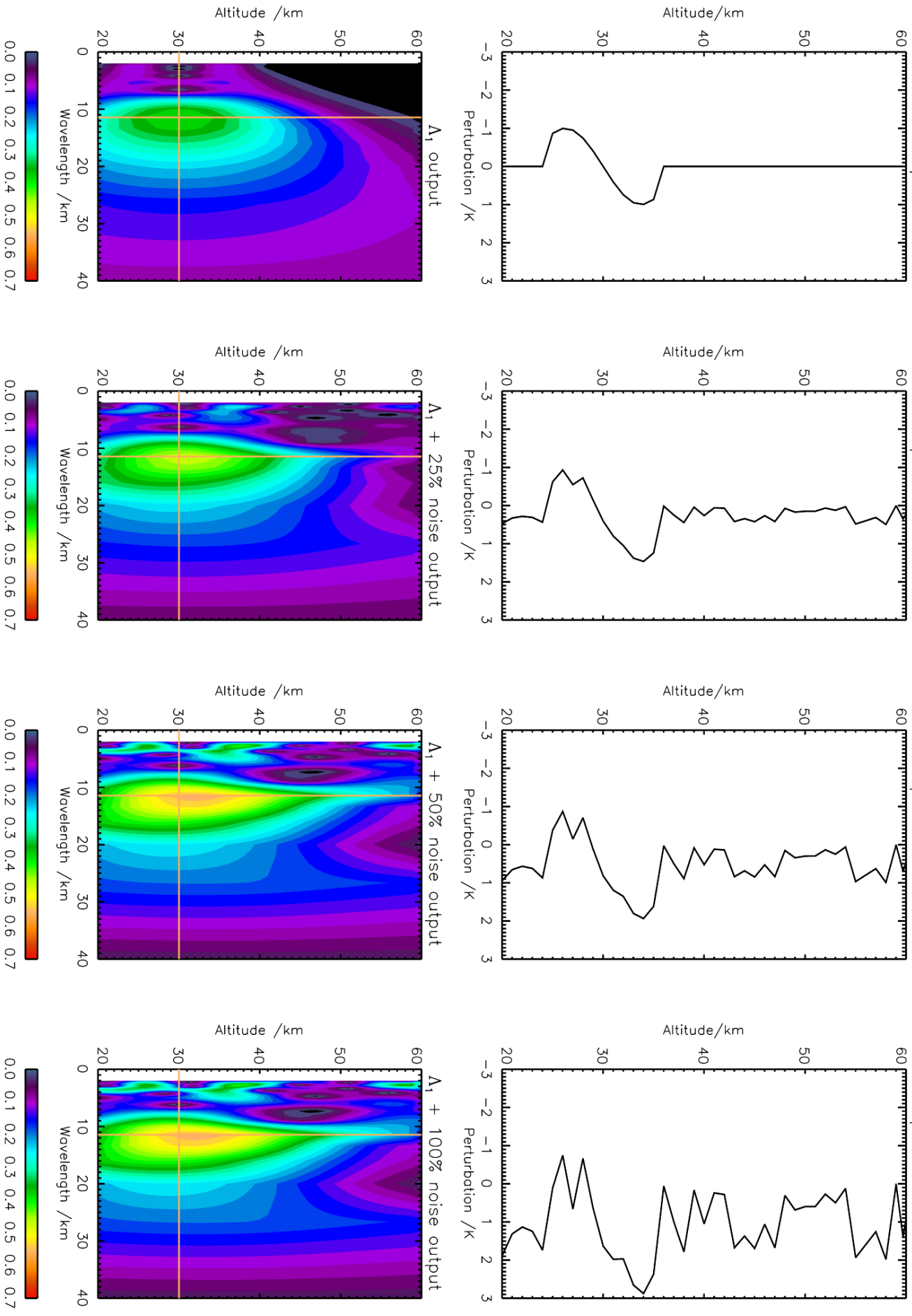


Figure 5.15: *Effects of increasing noise levels on S-Transform detection of a wave signal Λ_1 in a single artificial profile.*

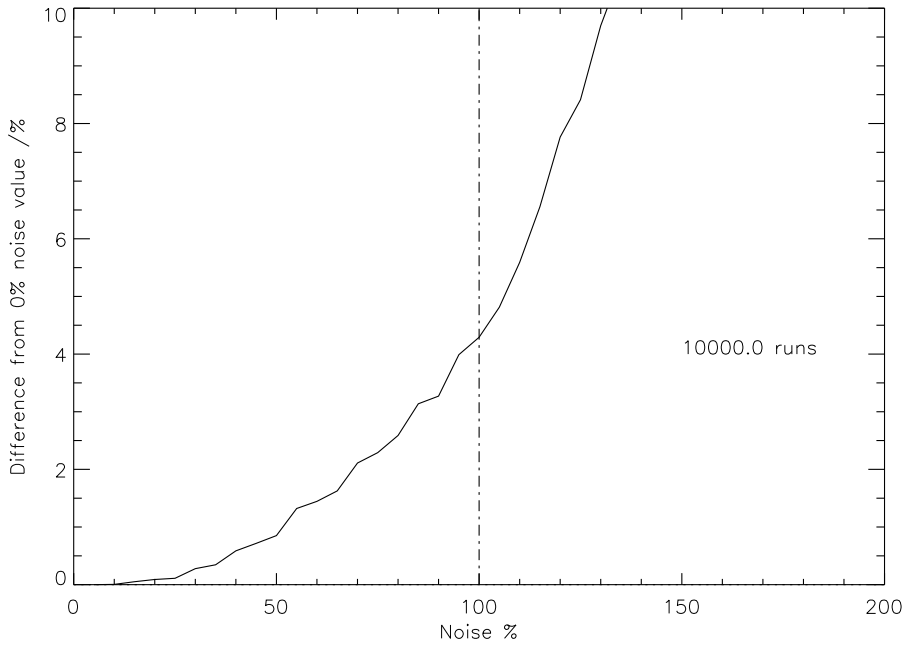


Figure 5.16: *Effects of noise on measured temperature perturbation due to a wave Λ_3. The graph is shifted such that the 0% noise line is at the measured amplitude of the zero-noise wave: as discussed above, this represents a 4% offset.*

zero noise, which as discussed above represents a 4% offset.

As would be expected, the figure shows an increasing difference between the input and measured amplitude of the waves. However, this increase is relatively slow: at 20% noise, the difference between the input and output amplitudes is near-zero, and even at 100% noise, this deviation is still only around 4%, although there is a sharp increase above this point. We hence conclude that, for a given waveform, the impact of random noise on the measured temperature amplitude is relatively low.

5.6.3 Effects on Vertical Wavelength Measurements

Figure 5.17 illustrates an analysis of the effects of noise on the vertical wavelength measured by our S-Transform analysis, for a pair of identical waveforms $\Lambda_i = (\lambda_i\,(30, 40, 50))$. The computed cross-ST peak wavelength data is analysed in wavelength steps of 100 m and noise steps of 10%, with 10 000 runs performed, and the noise level defined as previously.

For the zero-noise line, the results returned are exactly equivalent to those for figure 5.2; significant discrepancies are seen at large wavelengths (far right) due to packet size effects, becoming progressively smaller as the packet size becomes larger relative to the wave. Only

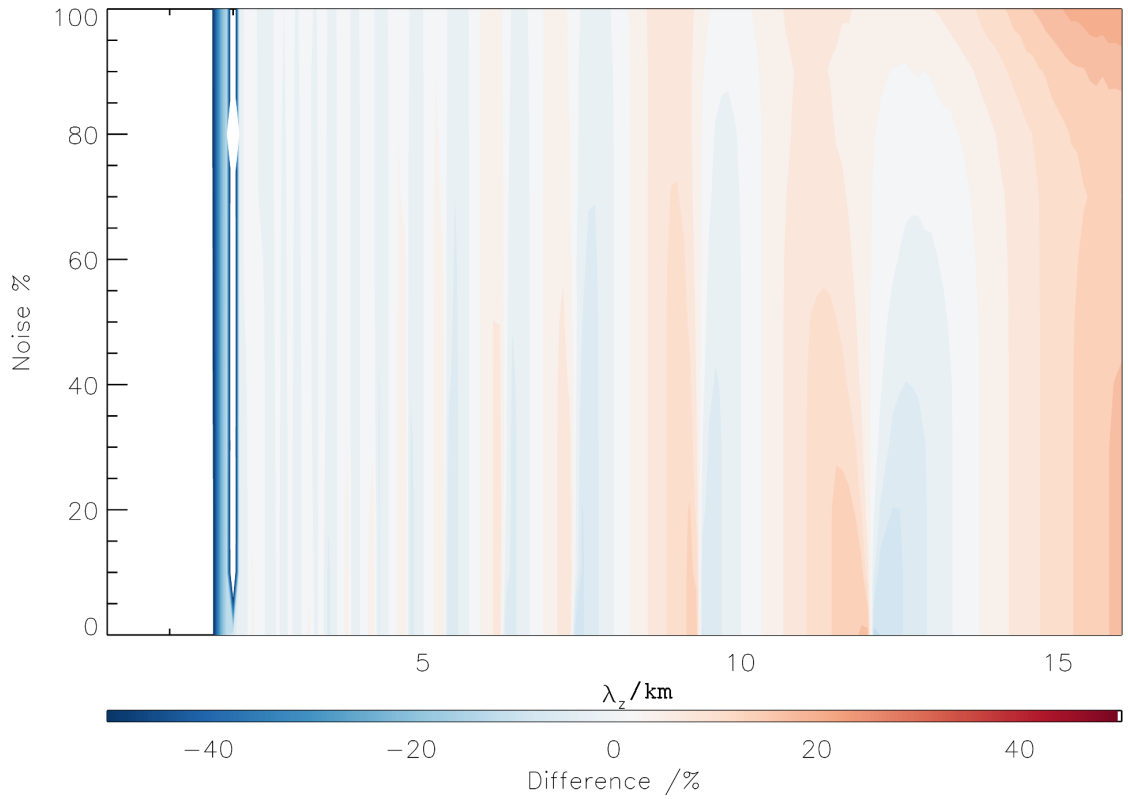


Figure 5.17: *Effects of increasing noise on the difference between measured and input wavelength, as a function of input wavelength for a fixed packet size of 20 km.*

wavelengths up to 16 km are considered; due to the packet size used being 20 km, wavelengths much greater than this will show a substantial deviation regardless of noise level.

As we increase the noise level, measured deviations from the input wavelength remain broadly the same, only increasing significantly as we approach the 100% noise level. This is consistent with our results for temperature, and suggests that the ST-derived vertical wavelengths are also fairly robust with regards to random noise.

5.7 Multiple Waves

Figure 5.18 illustrates two sample waves of differing wavelength and height range, and their sum, together with the S-Transform outputs of each case: $\Lambda_{10} = [10, (30, 35, 50)]$ and $\Lambda_{11} = [20, (40, 45, 60)]$.

As discussed above, the fit to Λ_{11} is weaker due to the smaller number of full cycles present in the input signal. Accordingly, when we combine the two overlapping signals, it becomes difficult to accurately resolve the longer-wavelength wave. This is generally true: in any signal where two waves of equivalent amplitude are present at the same amplitude, the shorter wavelength feature will be better-resolved, subject to the limitations discussed in section 5.2.1. Since as part of our analysis we select for the strongest wave signal at any given height, our results will consequently be biased towards the shorter-wavelength result in any cases of two overlapping waves with similar amplitude.

Figure 5.19 shows an ST analysis of three non-overlapping gravity waves, illustrated in the three plots at the top of the page. The fourth plot shows their sum, and the row below from left to right the S-Transforms of two identical copies of the summed profile and the resulting cospectrum. Finally, at page bottom we see two plots showing the resulting temperature perturbation and vertical wavelength profile, as determined by selection of the peak temperature perturbation at each altitude. In these plots, the regions highlighted by each colour indicate the region covered by the input wave of that colour, and the dashed line the input value for that wave.

Considering first the wavelength plot,we see that the results are in line with the results of our analysis in section 5.2. The short-wavelength wave, in red, has a near-perfect fit to the input wavelength, except in a region towards the bottom of the wave packet where the greater

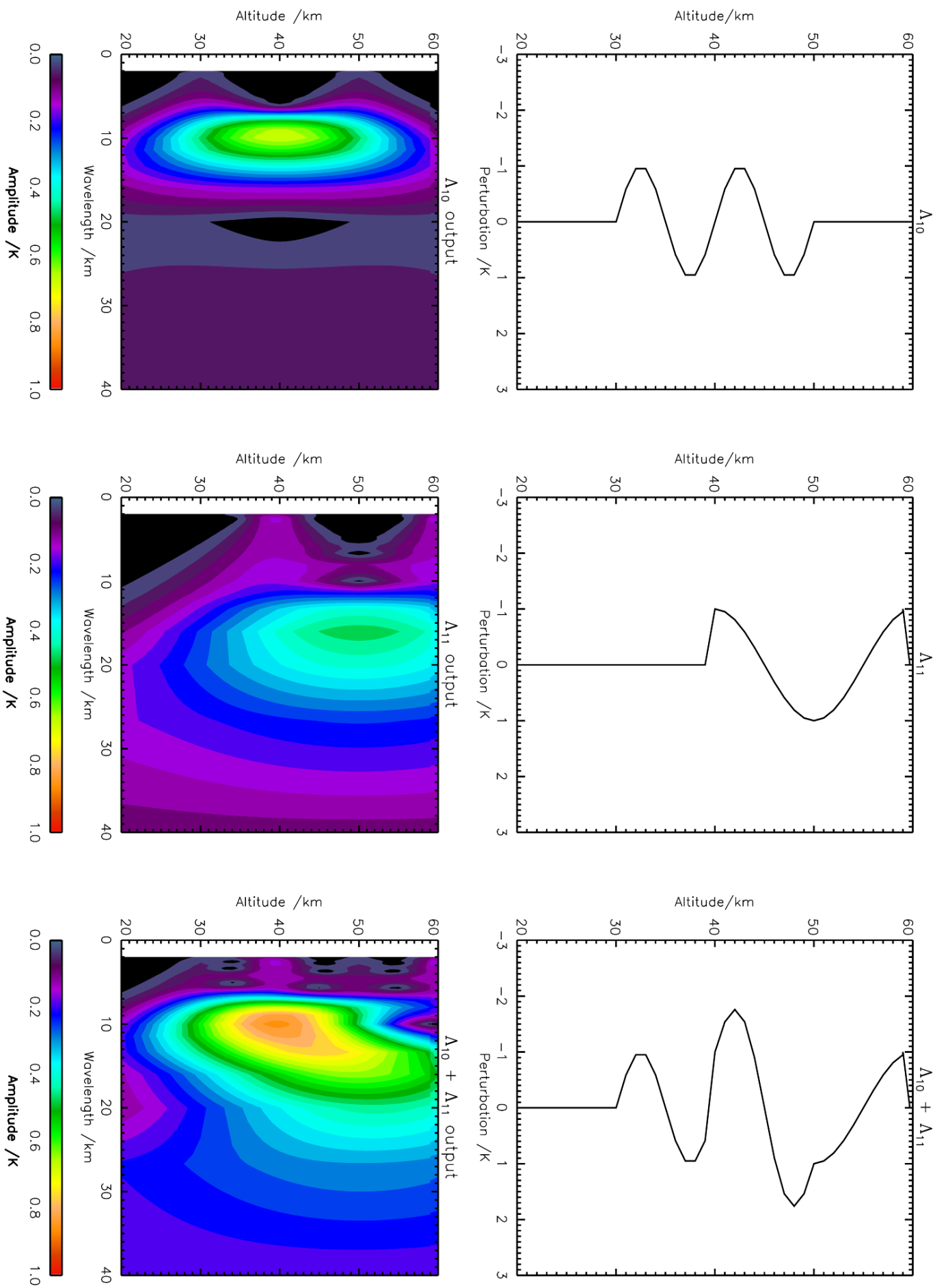


Figure 5.18: *Application of the S-Transform to two artificial waves Λ_{10} and Λ_{11} and the sum of these waves. Upper plots show the input waves, and lower plots the S-Transforms of those waves. The S-Transform of the sum (third column) only weakly resolves the longer-wavelength wave Λ_{11}.*

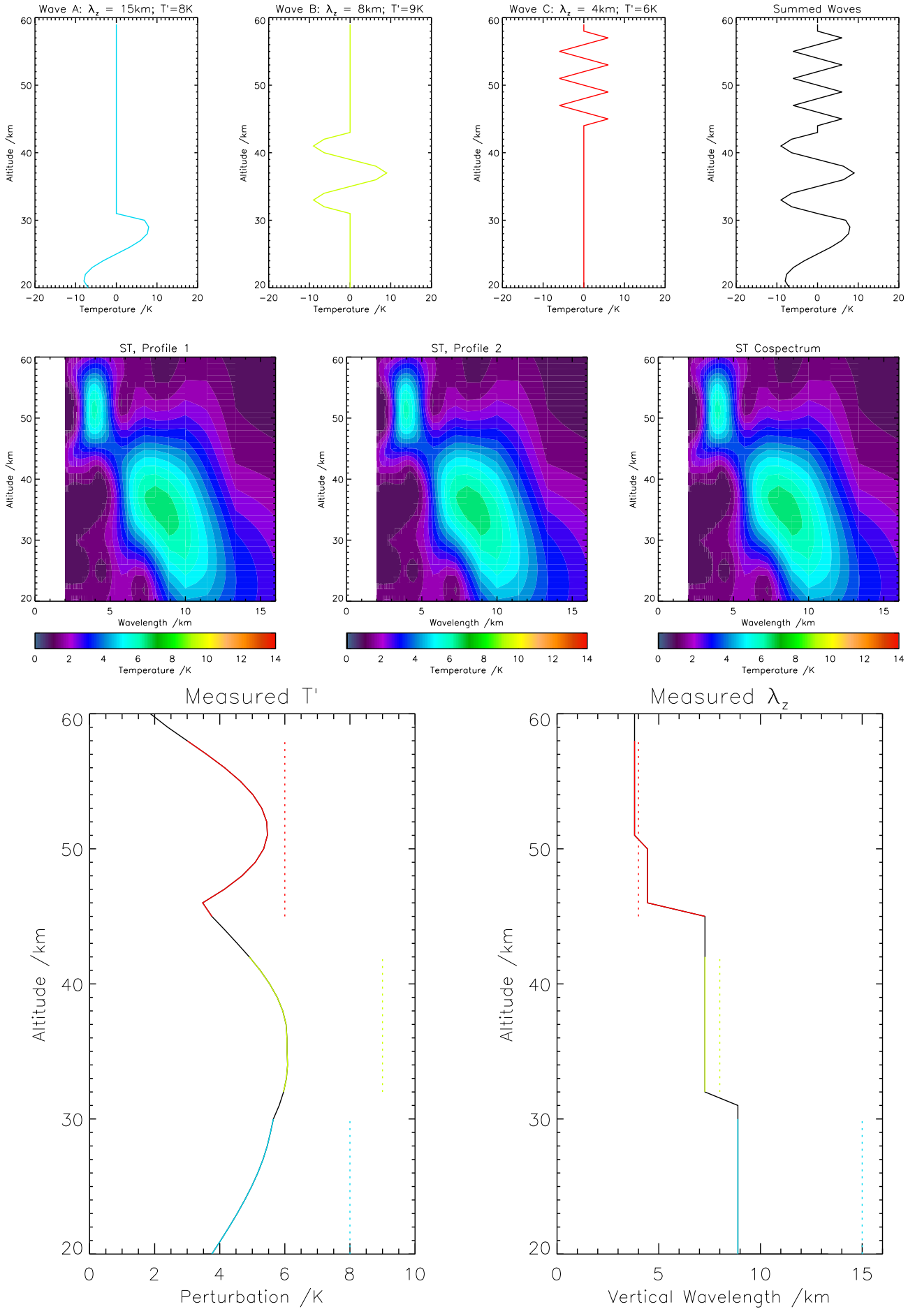


Figure 5.19: *Sample analysis of a profile consisting of three distinct waves*

amplitude of the mid-altitude green wave starts to shift the ST peak. This green wave again has a good fit to the vertical wavelength. Finally, the long-wavelength wave at low-altitude in blue has an extremely poor fit. As the input wave has a packet size of less than a full wave cycle and the wave immediately above it is of greater amplitude, this is the result we would expect.

Considering now the returned temperature amplitudes, we see that the high-altitude short-wavelength wave again has a close fit to the input value, with the value peaking at the centre altitude of the wave packet. The lower two waves, however, have a poor fit to the temperature amplitude. With the lower-altitude, longer wavelength wave this is as expected; the poor quality of the wave fit makes it unlikely that the wave amplitude would be measured accurately, but the result for the mid-altitude mid-wavelength wave is more disappointing. Looking at the ST cospectrum plot, the reason for this becomes apparent: the peaked region associated with this wave forms part of a single continuous peaked region, showing that the relatively small height separation between the waves is insufficient to prevent the lower-altitude wave feature from skewing the result. In all cases, the wave amplitudes are underestimated, consistent with the results of section 5.3.

Figure 5.20 extends this analysis further, considering the more complicated case of three partially-overlapping waves, of the same wavelength and amplitude as those considered in figure 5.19. The temperature perturbation and wavelength plots in this case have had their highlighting removed for clarity, but the two figures are otherwise identical in layout.

The measured wavelengths are once again good, with a good fit to the short-wavelength feature at top in the region where it is independent of overlap, and an excellent fit to the mid-wavelength feature in centre. The low-altitude wave is again poorly resolved due to its small number of complete wavecycles. The measured temperature amplitudes, however, are significantly worse than in the non-overlapping case,with little variation between the waves, showing a value of around 6 K for both the upper two waves, and failing to resolve the lower wave.

In all cases, it should be noted that the S-Transform spectra and cospectrum in the middle row clearly return the input information, providing accurate values for the wavelength and temperature perturbations for all waves, regardless of overlap. The maximal-amplitude selection criterion used in our analysis consequently introduces large uncertainties when applied to profiles made up of multiple overlapping or near-overlapping gravity wave signals, and is only

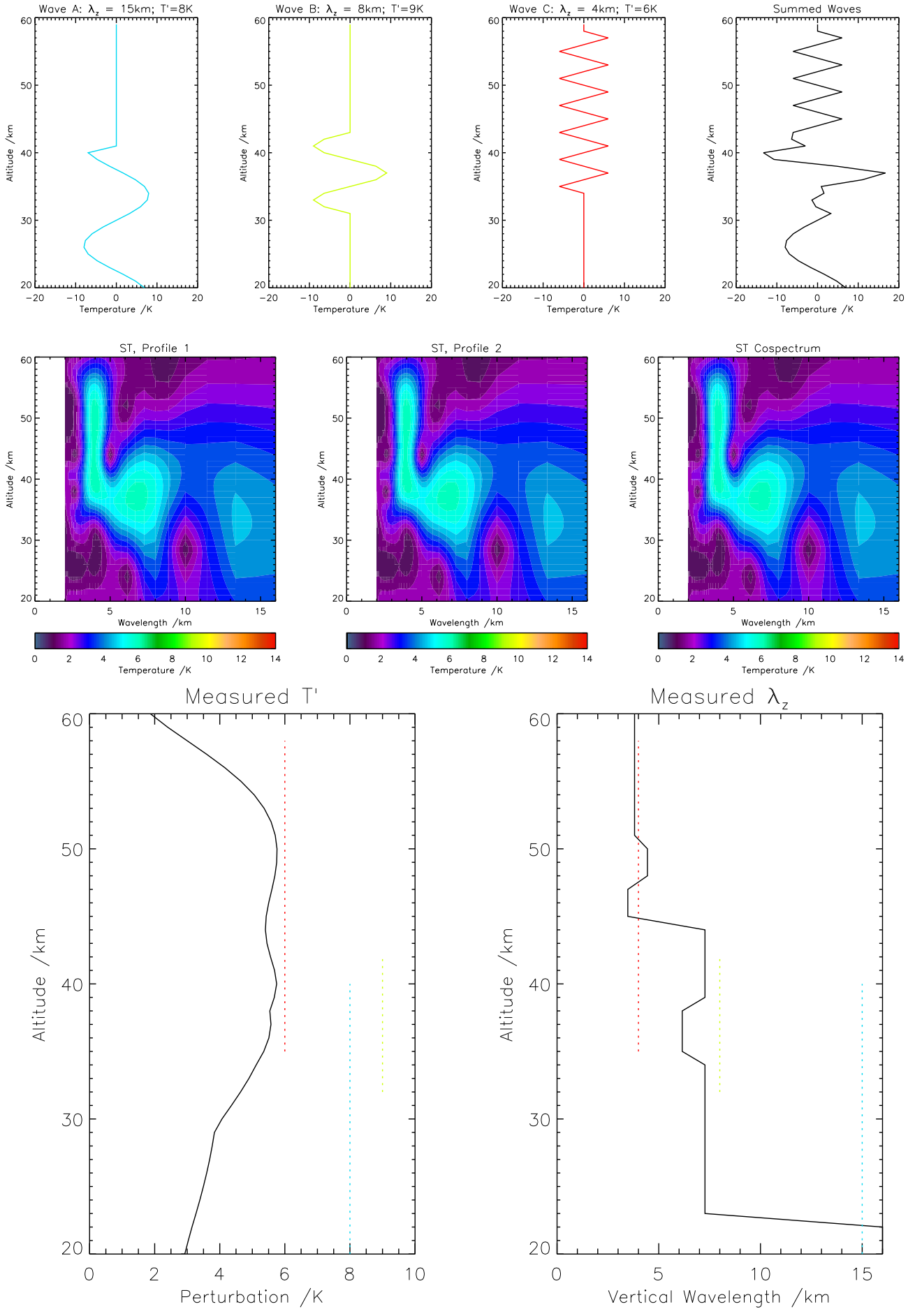


Figure 5.20: *Sample analysis of a profile containing three overlapping waves*

adopted to allow the analysis of the very large dataset in an efficient manner. Future analysis of the HIRDLS gravity wave dataset could focus upon this neglected data to better study these cases.

5.8 Summary and Conclusions

In this chapter, we have discussed several different potential sources of uncertainty in our S-Transform gravity wave analysis.

Considering first the simplest conclusions, we have shown in section 5.5 that the planetary wave signatures removed as the background from our signal are broadly what we would expect climatologically and physically, and in section 5.6 that random noise has a minimal effect on our results even at amplitudes comparable to the gravity wave signal under consideration. The planetary wave results will be validated by comparison to other instruments in chapter 6, below.

We have also considered the effects of input wave amplitude on output results. In the absence of noise, our analysis method underestimates the actual amplitude present in the wave signal, with no dependence of the discrepancy between input and measured amplitude on the input amplitude (top panel of figure 5.5).

The waveform being analysed, however, does have a significant impact, both upon the measured amplitude and measured vertical wavelength. Considering first the amplitude (lower panel of figure 5.5), we find that the amplitude discrepancy depends strongly upon the number of full wavecycles present in the wave packet, with significant discrepancies observed below two and especially below one complete wavecycle. Uncertainties due to this are typically up to $\sim$10% at most wavelengths and cycles-per-packet values, but may rise as high as 50% if only a single wavecycle is present (figure 5.6). We also observe strong amplitude discrepancies for waves below the Nyquist sampling limit(2 km vertical wavelength), consistent with these waves not being resolved. A similar pattern is also observed in the vertical wavelength discrepancy analyses (figure 5.4), but with additional significant ($\sim$10%) uncertainties introduced based on signal wavelengths falling between sampling intervals. Cospectral analysis does not appear to introduce any further errors into the amplitude measurements, provided the wave is well-resolved across both profiles.

We next consider phase results based upon cospectral analysis. Figure 5.12 shows that for

most cases above one complete wavecycle with a vertical wavelength more than twice that of the Nyquist sampling frequency, the errors in computed phase change are relatively small; the reason for the cutoff being at twice the Nyquist frequency is due to the phase being computed across two profiles rather than one.

Finally, we considered the effects on our analysis of profiles containing multiple wavepackets. We concluded that, for wave packets of equivalent amplitude, the detection process will tend to favour the detection of waves with a shorter vertical wavelength as these are better resolved by the transform process. We also concluded that significant problems arise with the detection of multiple overlapping waves due to our analysis method discarding data other than that for the maximum temperature perturbation at each height level.

Accordingly, we conclude that the S-Transform method is

- robust for the detection of waves with strong amplitudes and several full wave cycles
- accurate with some uncertainty (beyond that inherent in the measurements) for waves with low amplitudes relative to background noise or which are represented only by a small number of wave cycles: typically $\sim$10% for both amplitude and wavelength estimates, but potentially rising as high as 50% for amplitude estimates when only a single wave cycle is present in the signal.
- unable to detect waves which present less than a full wave cycle in the data or which are sub-Nyquist sampled, or to distinguish aliases of these waves from real data
- better suited to the detection of shorter wavelengths.

Chapter 6

Validation

6.1 Introduction

As discussed in chapter 2, the HIRDLS Level 2 temperature product has been extensively validated. However, it is still important to validate our gravity wave results against other instruments where possible, both as a test on the HIRDLS data themselves and on our analysis method. This chapter will focus on such validation, comparing HIRDLS small-scale temperature variations against a range of atmospheric sensing instruments, described further in section 6.1.1, and against the expected gravity wave climatology.

Barnett et al. (2008) discuss the cross-validation of HIRDLS and COSMIC data regarding fine vertical structure. In this study, a comparison was made of the correlation coefficients and standard deviations of high-pass-filtered colocated profiles. As a means of validation, we shall also apply these methods here. We shall then analyse our results using S-Transform analysis, comparing wavelike structure measured by each instrument with that observed by HIRDLS. These results will be considered in terms of the correlation/standard deviation analyses as a measure of the performance of our S-Transform analysis. All results will be considered in light of the different vertical resolution capabilities of each instrument.

6.1.1 Instruments and Datasets Used

6.1.1.1 COSMIC

Version Used	2007.3200
Period Analysed	2006d140 – 2007d239
Altitude Range Covered	variable: from near-surface to around 40–50 km
Vertical Resolution	∼1 km above tropopause
Reference	Rocken et al. (2000)

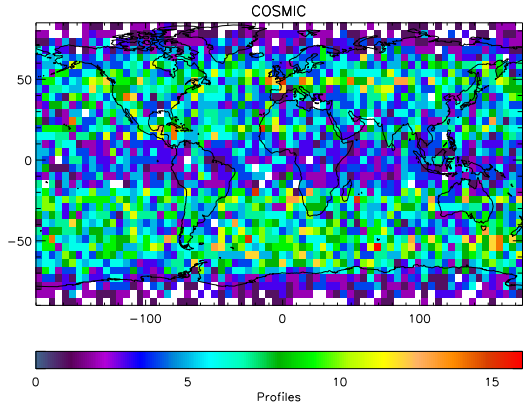
The Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) is part of the joint US/Taiwanese FORMOSAT-3 mission, based on a constellation of six identical microsatellites at an orbital altitude of 800 km. Amongst other instruments, each satellite carries a GPS radio occultation receiver, which is used to intercept GPS signals transmitted through the atmosphere. The phase delay in these signals is then analysed to produce profiles of temperature and moisture (Barnett et al., 2008; Rocken et al., 2000).

Due to COSMIC measurements requiring an optical path between a GPS and a FORMOSAT satellite, the profiles obtained are distributed pseudo-randomly rather than along a regular scan track as with HIRDLS. Figure 6.1(a) illustrates a typical week of such coverage; around 1000–2500 profiles are obtained globally on any given day. Vertical resolution above the tropopause is typically ∼1km (Schreiner et al., 2008). A typical profile is shown as figure 6.2(a).

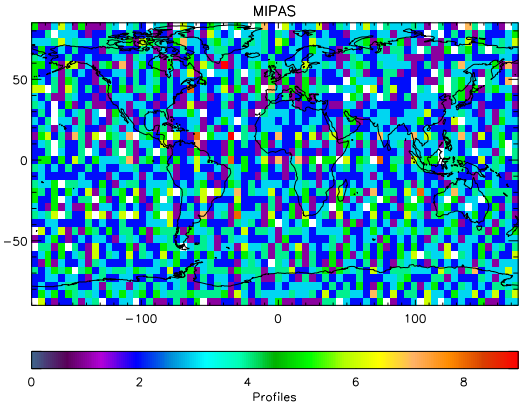
6.1.1.2 MIPAS

Version Used	1.02
Period Analysed	2006d001 – 2007d365
Altitude Range Covered	∼400 hPa – 0.05 hPa
Vertical Resolution	∼ 3 km in stratosphere
Reference	Fischer et al. (2000)

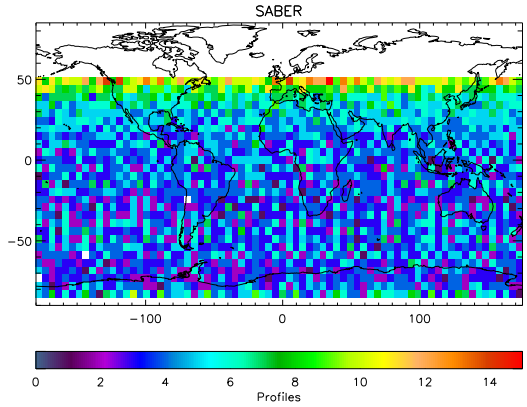
The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) is a Fourier Transform spectrometer on the European Space Agency's ENVISAT mission, launched in March 2002. Designed to take limb measurements of the middle and upper atmosphere, one of the key Level 2 products obtained is temperature. The product used in this thesis is the Level 2 MORSE



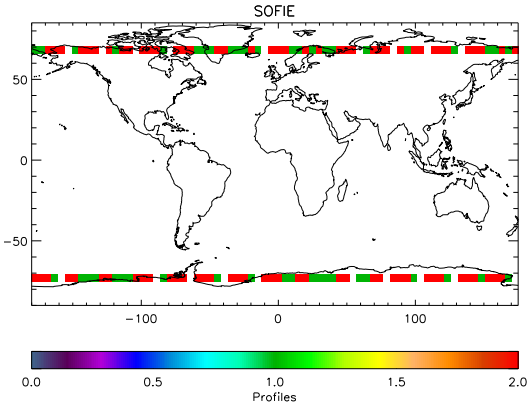
(a) COSMIC



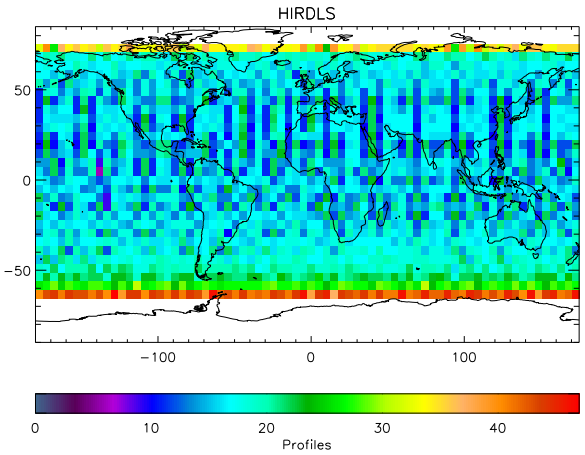
(b) MIPAS



(c) SABER



(d) SOFIE



(e) HIRDLS

Figure 6.1: *Sample coverage charts for the instruments under discussion for a one week period (2007/08/01 – 2007/08/07), binned on a 5 degree by 5 degree scale. Note the different colour scales for each instrument.*

product, retrieved at Oxford.

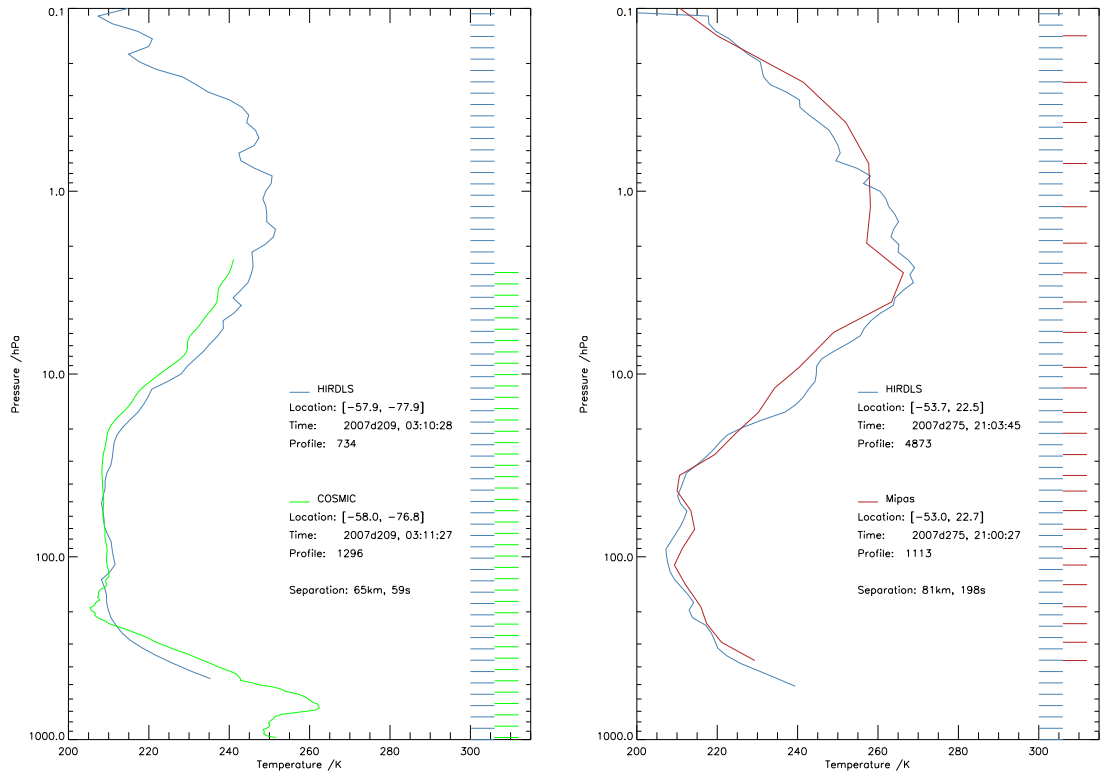
MIPAS operates in several scanning modes, including specialised middle atmosphere and UTLS modes. However, for consistency between profiles across the mission, we shall consider only those profiles obtained in the nominal mode, which is the most commonly used. This takes measurements on 27 height levels between 6 km and 70 km in altitude, with the separation between these levels increasing with height from 2 km in the upper troposphere to around 4 km in the upper stratosphere and lower mesosphere. We shall consider only the height range analysed with 3 km altitude levels, in order to provide a consistent scale for our analysis with a minimal amount of re-interpolation: this region extends from 31 km to 46 km in altitude, well within the range of our HIRDLS measurements.

6.1.1.3 SOFIE

Version Used	v1.022
Period Analysed	2007d135 – 2007d365
Altitude Range Covered	$\sim$400 hPa – 0.001 hPa
Vertical Resolution	$\leq$3.5 km
Reference	Gordley et al. (2009)

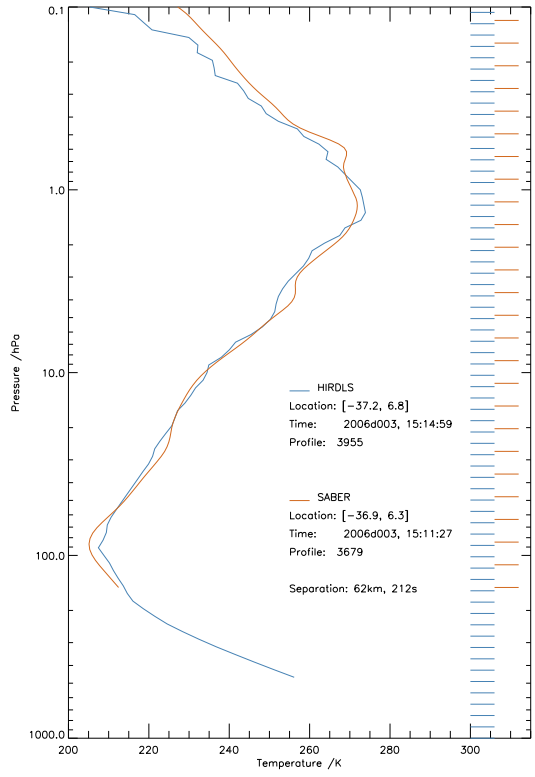
The Solar Occultation For Ice Experiment (SOFIE) is one of three instruments on the Aeronomy of Ice in the Mesosphere (AIM) satellite. While primarily intended to measure polar mesospheric cloud (PMC) formation in the stratosphere, a high-resolution temperature product is also produced as part of the data retrieval.

Figure 6.1(d) illustrates the spatial coverage of SOFIE profiles. This figure illustrates well the key limitation encountered when validating against SOFIE data: the coverage is extremely limited, with only around 30 profiles per day generated by occultations over a small range at high northern and southern latitudes. When combined with the highly limited southern HIRDLS data coverage due to the Kapton blockage, comparisons between HIRDLS and SOFIE temperature can only be made over a very small northern latitude range, and coincidences are unusual.

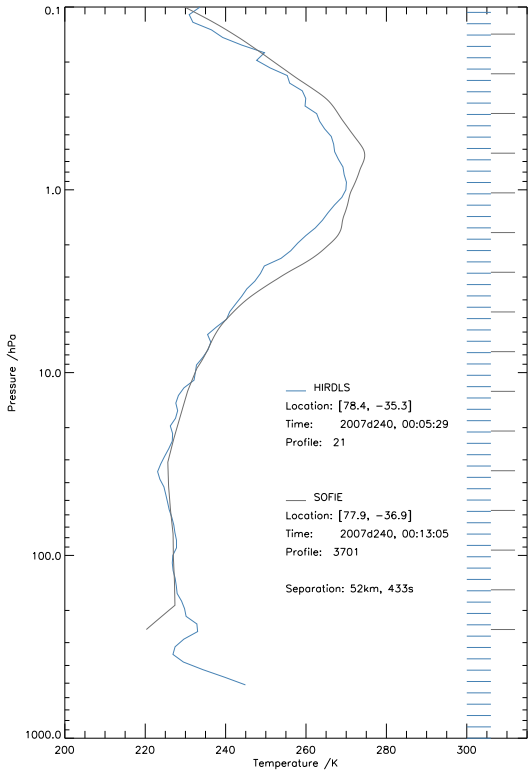


(a) COSMIC

(b) MIPAS



(c) SABER



(d) SOFIE

Figure 6.2: *Sample profiles from the instruments under discussion, shown to illustrate their vertical range and spatial resolution. Profiles are randomly selected, with a colocated HIRDLS profile shown for comparison. Horizontal solid lines at right indicate the published vertical resolutions of the two instruments: see section 6.5.1 for further details on these.*

6.1.1.4 SABER

Version Used	v1.07
Period Analysed	2006d001 – 2007d365
Altitude Range Covered	135 hPa – 0.0003 hPa
Vertical Resolution (max)	∼2 km
Reference	Russell III et al. (1994)

The Sounding of the Atmosphere by Broadband Emission Radiometry (SABER) mission is one of four instruments on NASA's TIMED satellite, intended primarily to measure and characterise the mesosphere and lower thermosphere on a global scale. Unlike the instruments discussed previously, SABER's scanning routine incorporates the TIMED spacecraft's yaw cycle, with the coverage region shifting north and south every sixty days to cover the poles alternately. Accordingly, while the coverage of the instrument in the tropics and at midlatitudes remains constant throughout the year, high northerly and southerly latitudes are only covered for 60 in every 120 days, in a 60-days on, 60-days off cycle. Figure 6.1(c) illustrates weekly SABER coverage for a period during the southerly phase of the instrument's yaw cycle; during the northerly phase this coverage chart is inverted to cover the equivalent northern latitudes. Temperature profiles cover the altitude range from ∼15 km–∼105 km, with a vertical resolution of ∼2 km (Wrasse et al., 2008; Rong et al., 2009).

6.2 Geographic Distribution

Figure 6.3 shows monthly-mean temperature perturbation amplitude maps at 50 hPa for the whole globe for each month of 2006, as measured by S-Transform analysis of HIRDLS data. It is useful to examine these maps by comparison to the expected theoretical sources of gravity waves outlined in chapter 3 in order to be sure that the results obtained are broadly in line with physical expectation.

6.2.1 Orography

We consider first orographical influences on wave generation. As discussed in section 3.3.1, we would expect to see significant gravity wave activity at all altitudes near mountain ranges

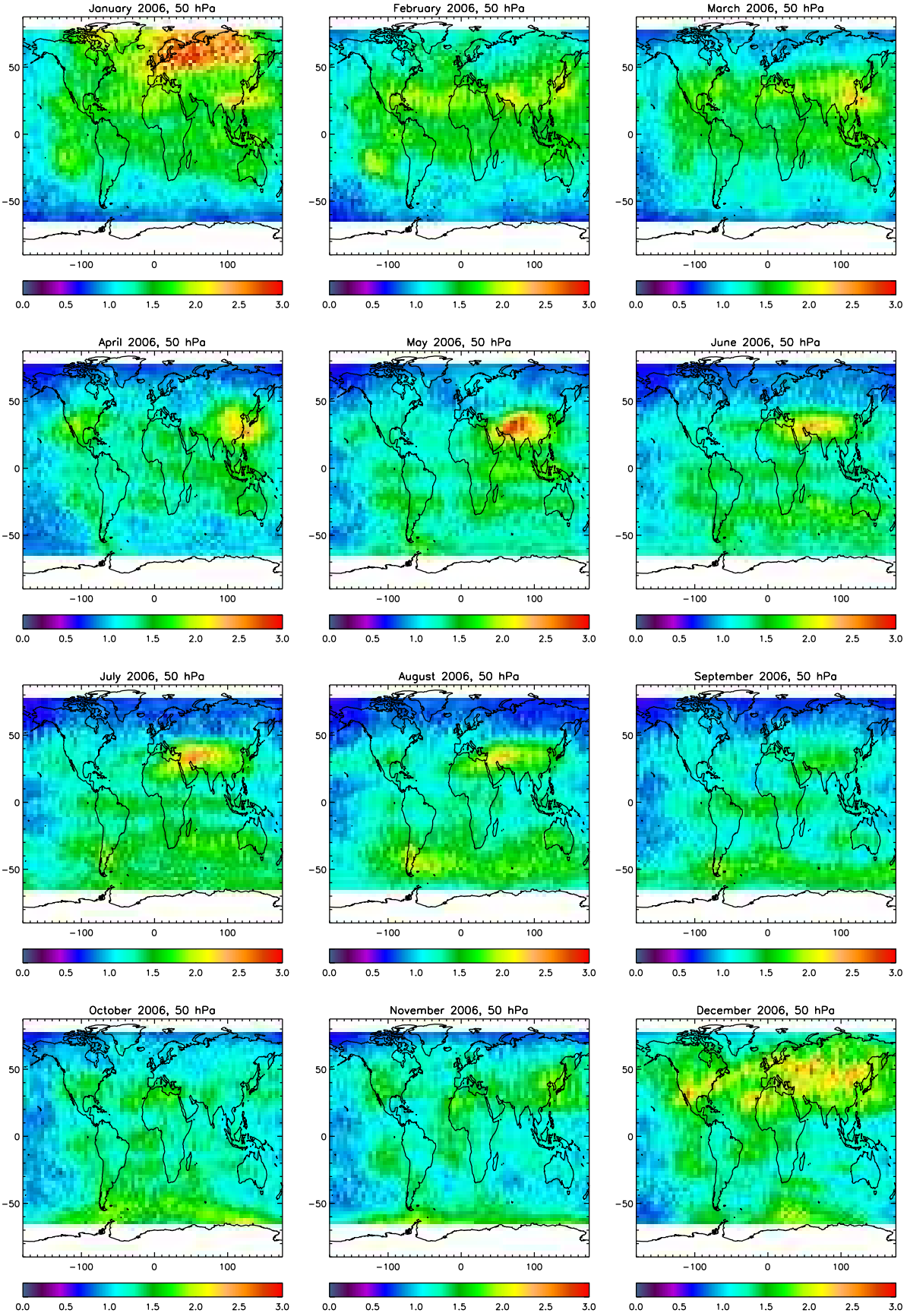


Figure 6.3: *Maps illustrating the monthly average temperature perturbation amplitudes observed at 50hPa for 2006. Scales are in K.*

aligned perpendicular to the prevailing winds, provided the mountain range is comparatively narrow in this direction. These may be offset slightly from their sources, however, due to horizontal propagation between their lower tropospheric sources and our stratospheric measurements. Examining our monthly plots, we see several peaks in the measured wave activity which are generated by this mechanism.

The Andes and the Antarctic Peninsula

Many previous studies (including Eckermann and Preusse, 1999; Jiang et al., 2003; Ern et al., 2004) have pinpointed the region covering the Drake Passage, southern Andes and northern Antarctic Peninsula during Antarctic Winter as a significant source of mountain waves, due to the north-south alignment and narrow width of these mountain ranges acting as a barrier to the prevailing high-magnitude westerly winds in this region. This signal is clearly visible in our data, particularly in the period from July to November. Previous studies have shown this signal to be particularly strong further south along the Antarctic Peninsula: unfortunately, our measurements do not extend this far south due to the Kapton blockage.

The Himalayas

For zonally-propagating winds, the Himalayas do not provide a significant source of mountain waves, due to their east-west alignment. During the period from May to August, however, winds in this region change propagation direction due to the effects of the monsoon, leading to a significant southerly component to winds in this region. This generates mountain waves over the Tibetan plateau and Iran/Afghanistan, which are again clearly observable in our data.

Greenland

The rough and high topography of Greenland provides a significant source of gravity waves. Although we do not observe this in 2006, for reasons discussed further in chapter 7, a significant peak is visible due to this region in 2005 and 2007 (figure 7.9).

The Rockies

Due to their breadth and relatively low height, the Rockies form a much weaker source of mountain waves at stratospheric altitudes than the previously discussed ranges. Nevertheless, a small signal is associated with them in April and December.

6.2.2 Convection

Similarly, several convective sources of gravity waves can be discerned in our monthly-mean maps. The clearest such region is over the South and East China Seas and the Philippine Sea during the period from January to April: this corresponds well with the typhoon season in this region, which provides a strong possible source of such signatures. Generally increased levels relative to high latitudes are also seen over the tropics throughout the year, consistent with the greater cloud convective activity of these regions.

6.2.3 Planetary Wave Modulation

A strong gravity wave signal is observed over western Russia and Scandinavia in January, with a corresponding trough over Alaska and eastern Siberia. This corresponds closely to a strong planetary wave of mode 1 with the same geographic distribution during this period involved in the 2006 Arctic sudden stratospheric warming (see, e.g., figure 5.13 and section 7.2), and hence is probably linked to modulation of the gravity wave spectrum by the winds associated with this wave-1 feature.

6.3 Planetary Wave Validation

The discussion of the planetary wave results in chapter 5 show anecdotally that in two cases the planetary wave results are broadly as expected, but a more systematic study of these results by comparison to other instruments is required to verify that the planetary wave background removed to allow gravity wave analysis is correct. Accordingly, planetary wave data has been computed, via the same method outlined in section 4.5.2, for each of the instruments described above, with the exception of SOFIE due to the limited coverage of this dataset.

6.3.1 Amplitude Time Series

Figure 6.4 shows a sample time series of the computed amplitude of the first three planetary wave modes at 45$\pm$2.5N and at 10 hPa for the period August 2006 - July 2007: this period was selected to maximise instrument coverage. A section has been removed from the MIPAS time series where coverage was poor at these latitudes due to the UTLS mode being used rather than

the nominal: correlations computed for this time series use only the parts excluding this. The data have been smoothed by ten days.

All four instruments show the characteristic climatological pattern of high-amplitudes in (boreal) winter and lower amplitudes during the rest of the year, and generally shadow each other throughout the year: COSMIC deviates markedly from the other instruments in some periods, but this may be due to the irregular sampling pattern of the instrument giving poorer coverage at these latitudes, particularly during late 2006 when the full COSMIC constellation was not yet in place.

Correlations are generally high between the instruments, particularly for the mode-1 planetary wave: for mode-1, SABER shows a 0.97 correlation with HIRDLS, MIPAS a 0.92 correlation, and COSMIC a 0.82 correlation. Correlations are lower, but still high, for the other two planetary wave modes; exceptions are the mode-2 time series for COSMIC and the mode-3 time series for MIPAS. In all cases, the HIRDLS planetary wave data are in agreement with the majority of the other instruments studied.

6.3.2 Spatial Correlation

Figure 6.5 shows an analysis of the same dataset to determine the geographical variability of this similarity. On the left are shown correlation coefficients over the duration of the mission between HIRDLS and each instrument for planetary wave mode 1, and on the right mean planetary wave amplitudes plotted against latitude for each instrument. We observe once again a high degree of similarity between HIRDLS and the other three instruments, with very high correlation coefficients at high boreal and austral latitudes corresponding to the regions of greatest planetary wave amplitude.

The instruments do disagree somewhat in the tropics, in particular COSMIC. However, examining the plot of mean amplitude, we observe a sharp discrepancy between COSMIC and the other three instruments at equatorial latitudes, with much higher wave-1 values measured. This may be due to the differing coverage patterns; as we see from figure 6.1, COSMIC has much poorer equatorial coverage than the other three instruments. Considering the other three instruments at equatorial latitudes and all four at polar latitudes therefore, we conclude that planetary wave amplitudes are resolved similarly by all four instruments.

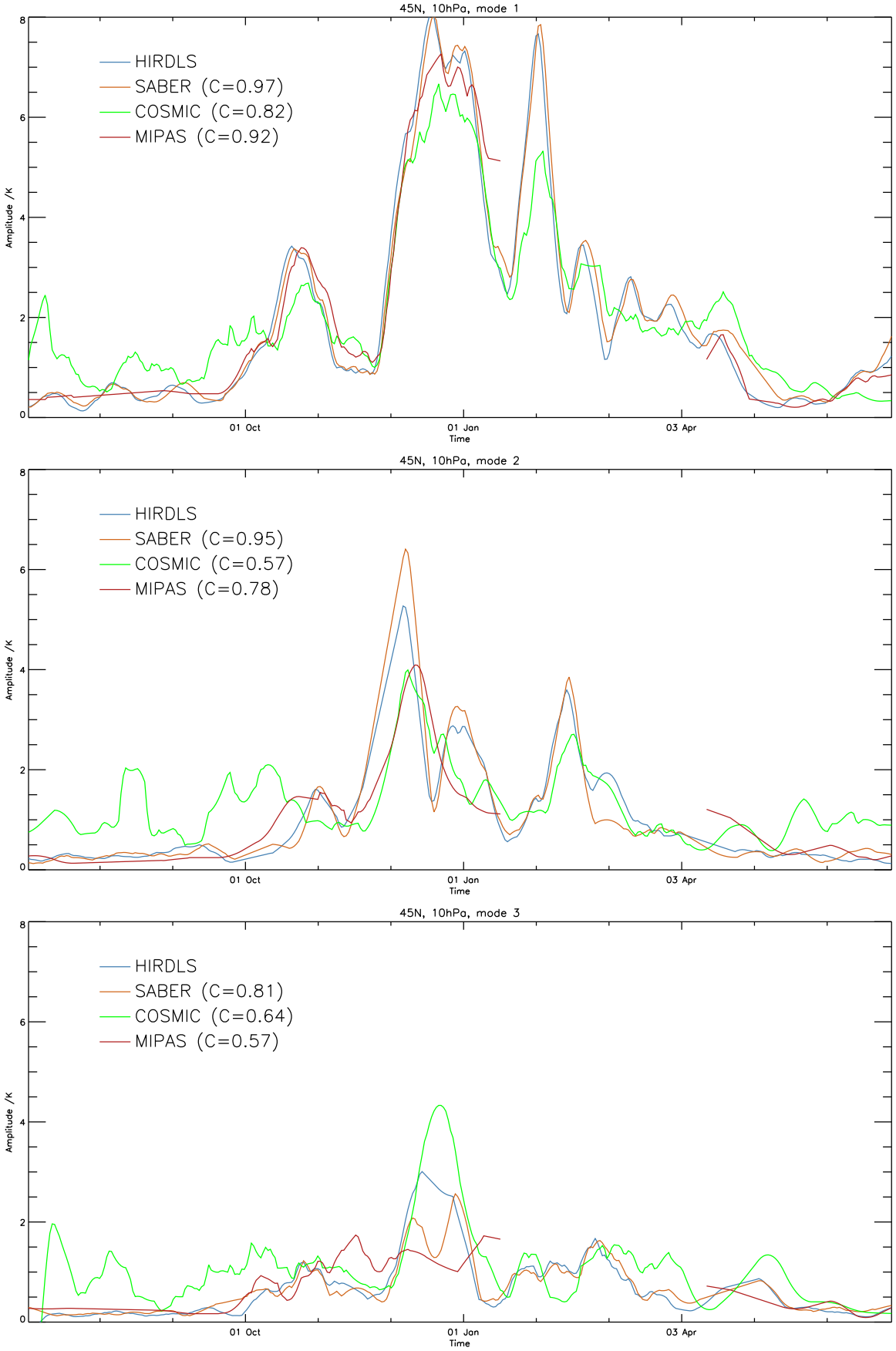


Figure 6.4: *Planetary wave amplitudes at 45N, 10 hPa, for four instruments from August 2006 to July 2007. Correlation coefficients C are computed between each instrument and HIRDLS*

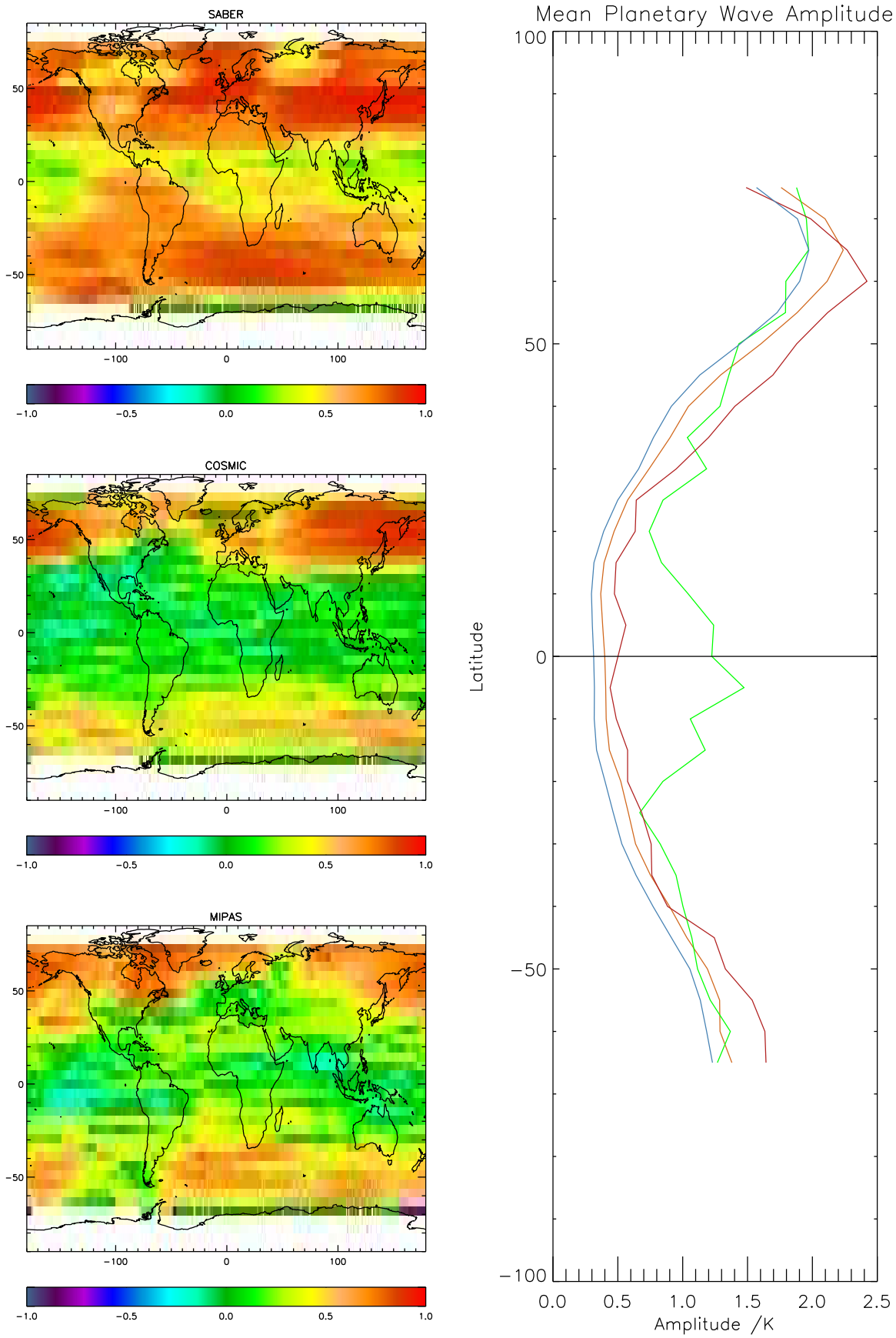


Figure 6.5: *(left) Correlations between planetary wave time series of each instrument and HIRDLs, plotted against latitude and longitude at 10 hPa; (right) mean planetary wave amplitudes for the same instruments plotted against latitude. Colours are as in figure 6.4*

6.4 Profile Colocation

6.4.1 Colocation Criteria

When analysing for transient features such as gravity waves in data from multiple sources, it is important to consider only very closely-spaced profiles. This can be demonstrated by considering the propagation of a midfrequency gravity wave with time.

For midfrequency waves, the dispersion relation

$$\hat{\omega} = N \left| \frac{k_h}{m} \right| = N \left| \frac{\lambda_z}{\lambda_h} \right|, \tag{6.4.1}$$

applies, for intrinsic frequency $\hat{\omega}$, horizontal and vertical wavenumbers k_h and m, and horizontal and vertical wavelengths λ_h and λ_z (Fritts and Alexander, 2003). For the waves we expect to be detected by our S-Transform analysis, typically λ_z $\sim$10 km, λ_h $\sim$200 km, and a typical stratospheric Brunt-Väisälä frequency N of 0.02 s^{-1}, we estimate an intrinsic frequency $\hat{\omega} \sim 1 \times 10^{-3}$ s^{-1}.

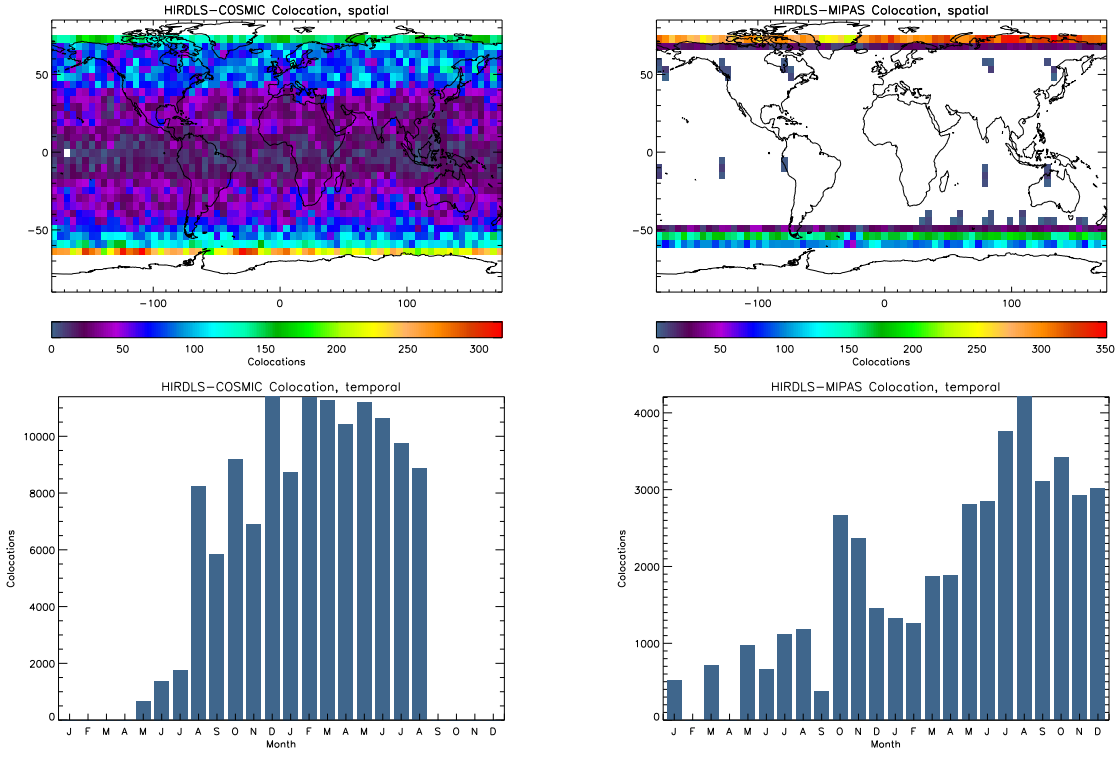
The group velocity of a wave packet in a background flow of velocity $\bar{u_h}$ is

$$[c_{g,h}, c_{g,z}] = (\bar{u_h}, 0) + \left(\frac{\hat{\omega}}{k_h}, \frac{-\hat{\omega}}{m} \right), \tag{6.4.2}$$

where $c_{g,h}$ is the horizontal group velocity and $c_{g,z}$ the vertical. Waves of this type will hence have a horizontal group velocity $c_{g,h}$ $\sim$20 ms^{-1} and a vertical group velocity $c_{g,z}$ $\sim$1 ms^{-1}. If we consider a time period of half an hour (1800 s), such waves will propagate 18 km vertically, well within the height range covered by HIRDLS, and around 360 km in the horizontal during this time. We shall hence adopt this temporal separation criterion of 1800 s, and require colocated profiles to be within 360 km spatially, when considering profiles as pairs.

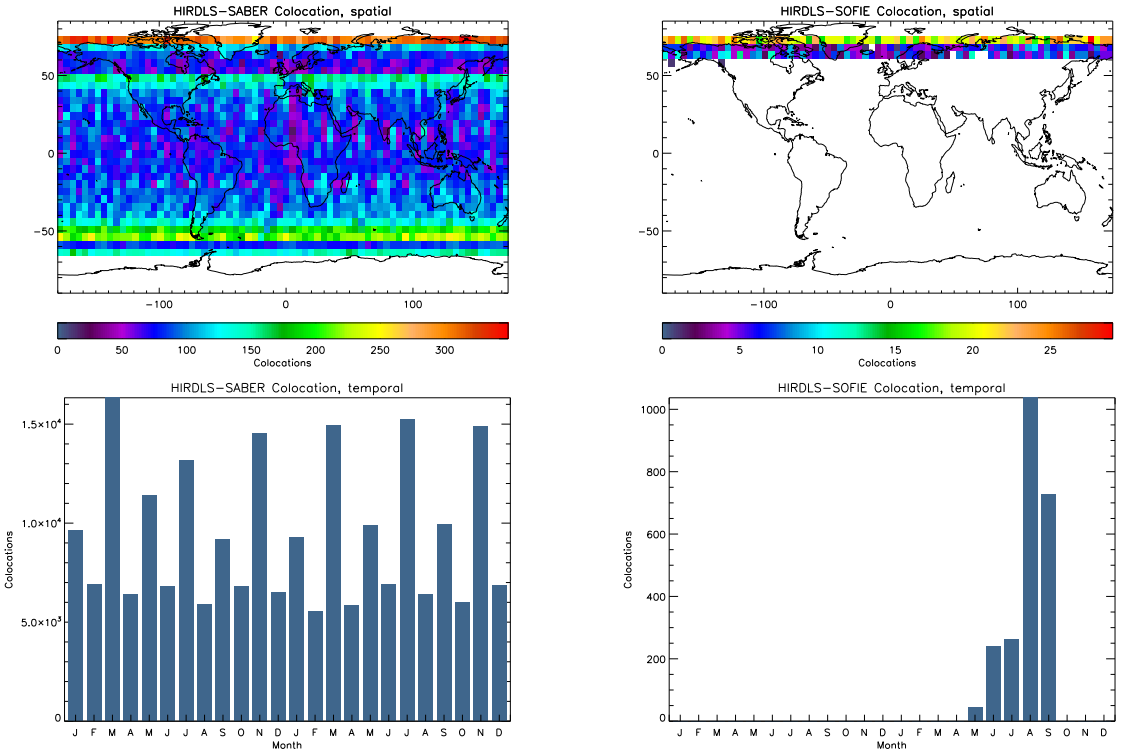
6.4.2 Number and Distribution of Colocations

Figure 6.6 illustrates the distribution of colocations between our instruments both spatially and temporally over the two-year period 2006 and 2007. We shall consider each of these colocation distributions individually.



(a) HIRDLS-COSMIC (127509 colocations)

(b) HIRDLS-MIPAS (44447 colocations)



(c) HIRDLS-SABER (225215 colocations)

(d) HIRDLS-SOFIE (2311 colocations)

Figure 6.6: *Spatial (upper) and temporal (lower) distributions of colocated profiles for each instrument pair. Spatial colocations are shown on a 5-degree grid, and temporal colocations on a monthly scale covering 2006 and 2007.*

6.4.2.1 HIRDLS-COSMIC

Figure 6.6(a) illustrates colocations between HIRDLS and COSMIC. Due to the broad spatial distribution of the profiles produced by both instruments, the geographic profile distribution is generally good across the whole globe and is clearly representative of the input data; the distribution is peaked at the northern and southern turnaround latitudes for HIRDLS, and there are fewer coincidences at low latitudes. A large number of colocations are recorded between the two instruments, and these are fairly evenly distributed across the overlapping time period of the two datasets, ramping up initially over a period of months during the commissioning of the COSMIC constellation before reaching a level of over 6000 coincidences per month.

6.4.2.2 HIRDLS-MIPAS

Figure 6.6(b) shows HIRDLS-MIPAS colocations. These are heavily dependent upon the MIPAS scanning mode used, as we are considering only data retrieved using the nominal mode; MIPAS was run in UTLS mode for a large part of 2006, and accordingly there are significantly fewer colocations in 2006 than 2007. Almost all colocations are at high northerly and southerly latitudes, due to the interaction between the orbits of Aura and ENVISAT.

6.4.2.3 HIRDLS-SABER

Figure 6.6(c) illustrates HIRDLS-SABER colocations. As with the HIRDLS-COSMIC coincidences, these are evenly distributed across most latitudes, as both instruments produce profiles over the whole globe and have frequently-intersecting orbits which, unlike those of HIRDLS and MIPAS, precess relative to each other. This precession leads to colocations occuring over the full range of latitudes covered by HIRDLS. There is an alternating two-month pattern to the number of coincidences per month; as HIRDLS has good coverage at high northern latitudes but not high southern latitudes, this corresponds directly to the SABER yaw cycle discussed above. This is the instrument-pairing with by far the highest number of coincidences (225215).

6.4.2.4 HIRDLS-SOFIE

Figure 6.6(d) shows colocations between HIRDLS and SOFIE. As discussed previously, SOFIE is a solar occulation instrument designed to measure polar mesospheric ice; accordingly, mea-

surements are taken at sunrise and sunset and at high latitudes, leading to a small number of measurements being taken. Hence, any colocations between HIRDLS and SOFIE will only occur when the HIRDLS tangent point crosses the terminator and is spatially close to the SOFIE tangent point. The number of colocations is low, and all are at high northerly latitudes.

6.5 Correlation/Standard Deviation Analysis Plots

6.5.1 Methodology

We shall initially consider our data in terms of the correlation coefficient and standard deviation of colocated profile pairs about their respective smoothed means. In each case, we use the correlation coefficient as a measure of the similarity of the two colocated profiles, and the standard deviation as a proxy to the degree and magnitude of fine structure in the profile. In addition to the basic aim of investigating how well HIRDLS observes fine atmospheric structure relative to other instruments, this will also provide us with a measure of similarity between the datasets with which later to interpret more sophisticated comparisons of colocated profiles.

Due to the different temporal and spatial coverage patterns of our instruments (figure 6.1), we cannot use the profile smoothing method based on planetary wave removal described in previous chapters for all the instruments under discussion; accordingly, we shall use the following method outlined in Barnett et al. (2008), applied to all instruments for consistency.

Each profile is interpolated onto a 500 m altitude scale. The profiles are then convolved with a cosine half-bell filter of 0.5 pressure scale heights ($\sim$3.5 km). The values of this new profile are subtracted from the original[1] profiles to produce perturbation profiles; this process acts as a high-pass filter on the data. Finally, the profiles are reinterpolated onto the published resolution of the coarser instrument: for example, when considering SOFIE and HIRDLS coincidences, a vertical resolution of 3500 m is used, while the HIRDLS-COSMIC comparison uses a resolution of 1000 m. The published resolution as opposed to the vertical resolution of the Level 2 data source is used for the final interpolation step: this distinction is important because all the instruments used except MIPAS are oversampled to some degree in the Level 2 data formats used.

[1] i.e. pre-convolution, not pre-interpolation

Figure 6.7 illustrates this process, using the colocated HIRDLS-MIPAS profile pair shown previously in figure 6.2(b). The top left figure shows the original profiles, the top right the data interpolated onto the 500 m scale with the half-bell-smoothed profiles overplotted with dashed lines, the bottom left the perturbation profiles remaining after the half-bell smoothing, and the bottom right the height region under investigation extracted. In each case, the coloured lines at the right indicate the height levels under consideration at that stage of the analysis, with blue indicating HIRDLS levels and red MIPAS levels. The standard deviations of the perturbation profiles about their means are then computed, and the profiles are correlated using the linear Pearson correlation coefficient.

This analysis is obviously limited in several ways. Firstly, the interpolation of the datasets to new vertical scales inherently alters the structure of the data. It is necessary for the data to be on the same vertical scale in each case, both for the correlation and for the standard deviations to be fully comparable, but in performing this step we are inherently removing useful information from the data.

Secondly, and most importantly, this analysis does not tell us what the fine structure under consideration actually is. All that is being considered is the variability of the small-scale vertical structure profiles and their similarity to colocated data: the analysis cannot tell us whether what we are seeing arises due to gravity-wave type features, other small-scale atmospheric features, or even noise. Accordingly, it can only provide us with useful information when considered alongside other analytical methods.

6.5.2 Comparisons

We shall now discuss the results of the above analysis method as applied to profiles colocated with HIRDLS for each of our analysed instruments. In each case (figures 6.8 – 6.11), we shall consider the correlation as a function of the standard deviation in the profiles from (a) HIRDLS and (b) the matched instrument, and the correlation as a function of the standard deviation in both profiles (c-d). Subfigures (c) and (d) show, respectively, the number density of pairs and the mean correlation at a given standard deviation for each instrument[2].

[2]The standard deviations of the correlation distributions for each point in plot (d) were also considered to see if there was a systematic difference in any region. The results obtained were broadly flat across all the datasets, and have accordingly been omitted for reasons of brevity.

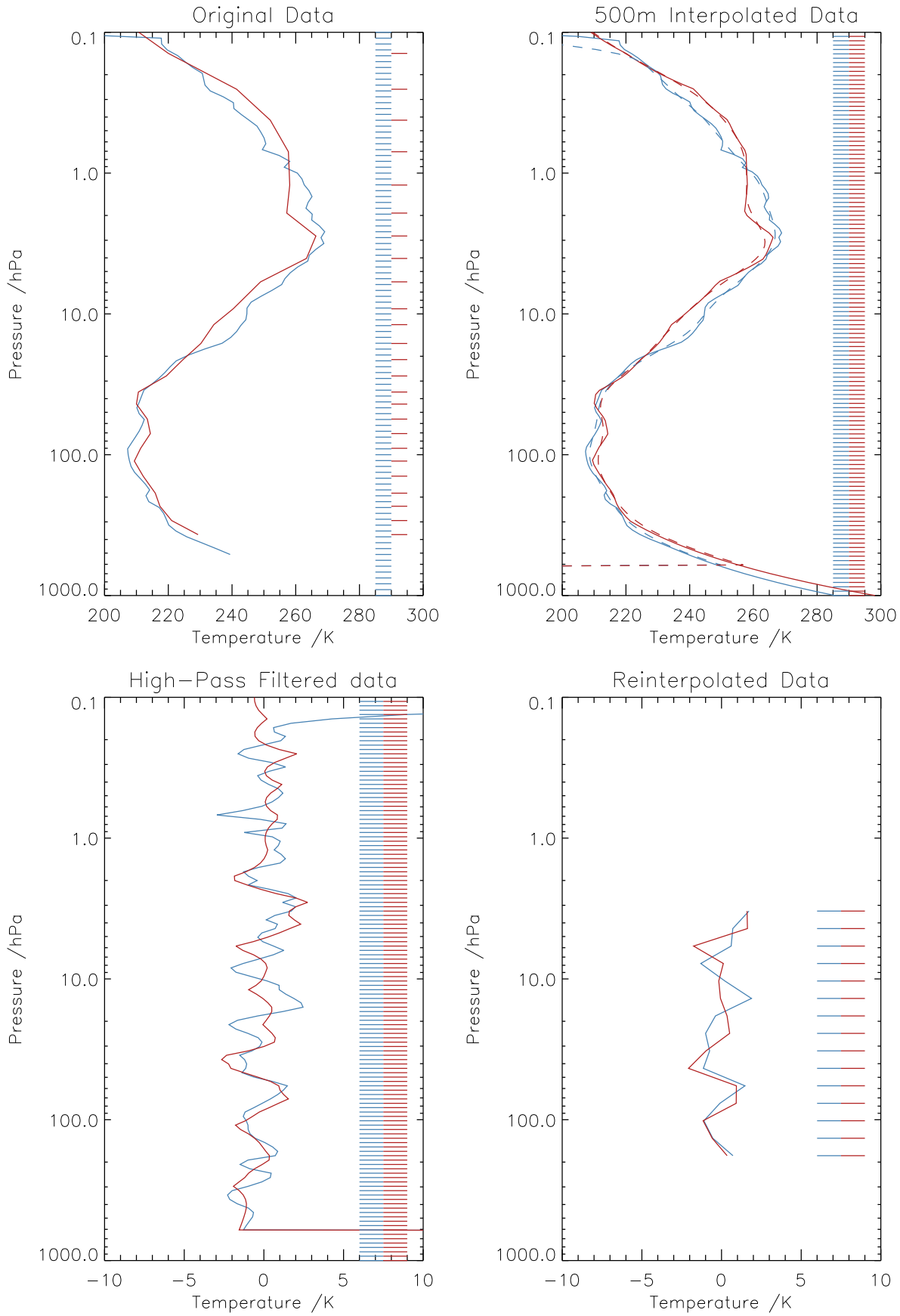


Figure 6.7: *Illustration of the data preparation applied for our correlation/standard deviation analyses, showing a comparison between a HIRDLS (blue) and MIPAS (red) profile. See text for further details.*

These analyses allow us to consider the correspondence between the correlation quality and the degree of fine structure in the signal. Generalising broadly, we would expect real gravity wave signals to be of greater amplitude than background noise, and accordingly would expect profiles with larger amplitudes, and hence larger standard deviations, to be made up to a greater degree of signal rather than noise. Hence, we would expect larger-amplitude profiles to be better correlated.

When considering the results of these analyses, we must bear in mind the effects of instrument vertical resolution upon the measurement of small-scale features. Broadly speaking, we would expect an instrument with finer vertical resolution to better reproduce small vertical-scale temperature peaks which would be averaged out across a channel in an instrument with coarser resolution. Although we have interpolated the data to the same vertical resolution, we do still expect some residual biasing to occur in the standard deviation analyses in favour of the higher-resolution instruments. Assuming that the higher-resolution instruments detect fine structure invisible to the coarser instruments, they will give more widely-spaced extrema to interpolate from, slightly skewing the final results in favour of greater standard deviations. Accordingly, we should expect to see a slight biasing away from the 1:1 correspondence line towards higher standard deviations in profiles from the higher-resolution instrument in each case. This is especially so for high-frequency data near the resolution limits of the finer instrument; as we look towards increasingly large-lengthscale structure, we would expect the standard deviations to become more similar and the correlation between the two instruments to rise.

6.5.2.1 HIRDLS-COSMIC

Figure 6.8 illustrates the results of our analysis for colocated HIRDLS and COSMIC profiles. These have been considered over the height range 200 hPa – 3 hPa, chosen as the region in which both the HIRDLS and COSMIC retrievals are considered to be the most accurate; HIRDLS profiles generally become less accurate below the $\sim$200 hPa height level (Gille et al., 2008), whilst COSMIC profiles begin to suffer severely from noise around the $\sim$3 hPa height level (Barnett et al., 2008).

Considering first the upper two plots, figures 6.8(a) and 6.8(b), we observe that the correlations between the two instruments are generally good: only a relatively small proportion of

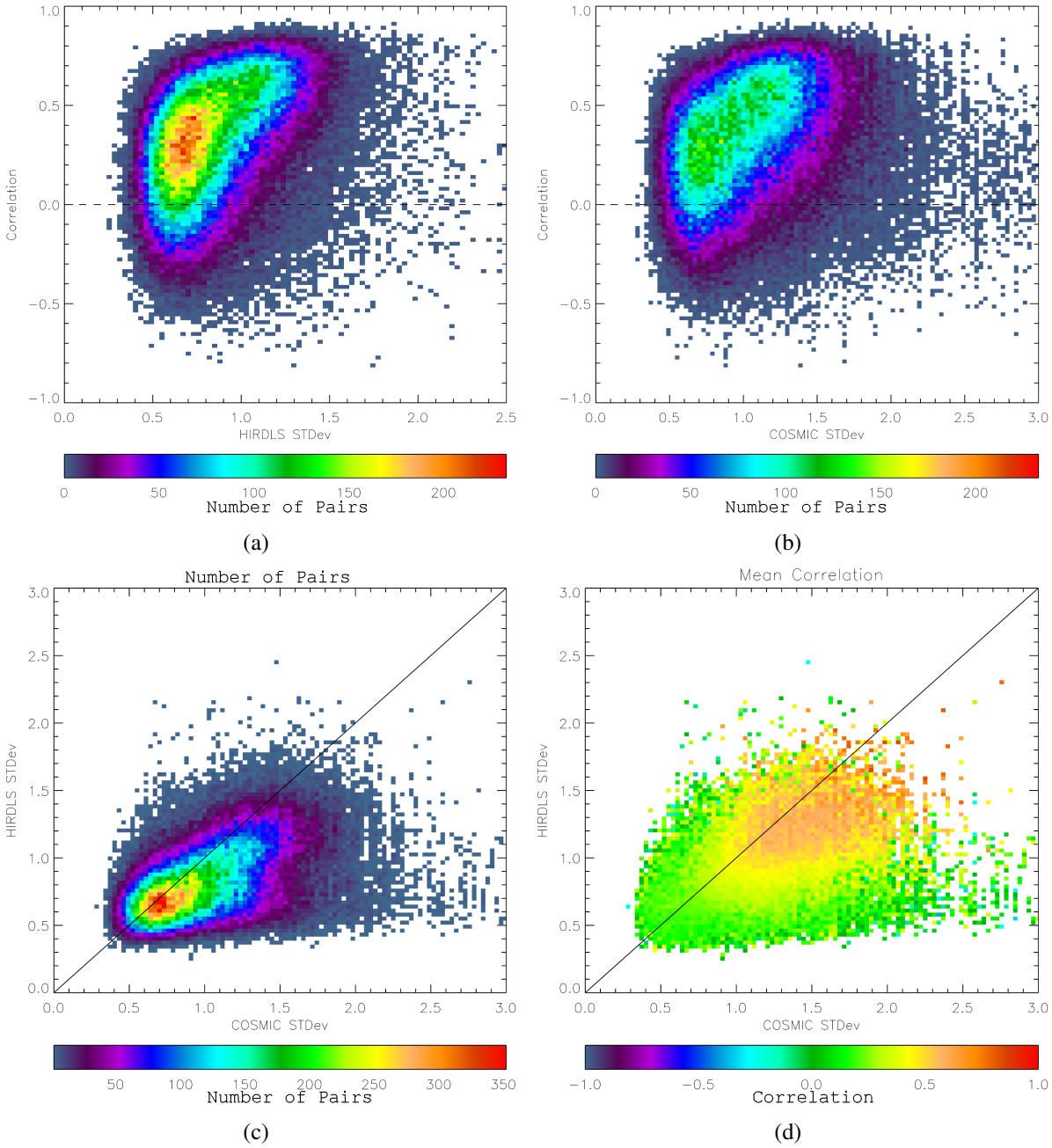


Figure 6.8: *Correlation/standard deviation analysis of HIRDLS-COSMIC colocations*

the profile-pairs considered have low or negative correlations, with the great majority of results lying at mid-to-high positive correlations. In both plots, we also observe a trend of increasing correlation with increasing standard deviation: this suggests that in profiles with more intense activity, we observe greater profile similarity, that is to say that the best-correlated profiles are those with the most activity. We would expect these profiles to include those with significant gravity wave measurements. Concomitantly, profiles with low standard deviations are more likely to represent low-level noise rather than meaningful activity, and consequently would be expected to have weaker correlations.

The distribution of points is slightly shifted towards higher standard deviations in the COSMIC case. This effect is probably due to the different methods utilised for analysis in HIRDLS and COSMIC: whilst the vertical resolution of HIRDLS is limited by the physical size of the radiometer channels, the COSMIC retrieval is based upon atmospheric refraction of a microwave signal and hence potentially allows a slightly higher vertical resolution, capable of reproducing small-scale features which would be averaged out across the channel in HIRDLS measurements, despite the published vertical resolutions of both instruments being broadly similar.

We now consider the lower pair of figures together. Here, the results shown above are plotted against the standard deviations of HIRDLS (left) and COSMIC (right). The solid line indicates the 1:1 correspondence line between the two datasets. Considering first the number density plot in figure 6.8(c), we observe a cluster of high number densities around a cluster at low standard deviations in both instruments, corresponding to a region made up of uncorrelated profiles (figure 6.8(d)). This is consistent with low-structure noise profiles, as discussed above. We also once again see a bias towards larger standard deviations in the COSMIC data, with the distribution as a whole being strongly skewed towards higher COSMIC values. In particular, an extended 'tail' of poorly-correlated profile-pairs with low HIRDLS standard deviations is seen at all COSMIC standard deviation values; this is consistent with the results shown in figure 4 of Barnett et al. (2008), despite the significantly different colocation criteria and different height sampling used. There are few HIRDLS profiles with high standard deviations, and those that do have high values are usually well-correlated with their matching COSMIC profiles: where a HIRDLS profile features a significant degree of fine structure, we also observe this in the matching COSMIC profile, whilst the reverse is not necessarily true. This may be due

either to the resolution considerations detailed above and hence be a tail of small-scale fine structure which HIRDLS cannot resolve, or due to limitations in the COSMIC retrieval, but this distinction cannot be easily made by this analysis.

6.5.2.2 HIRDLS-MIPAS

We next consider our analyses of HIRDLS-MIPAS colocations, as shown in figure 6.9. These have been analysed over the pressure range 200 hPa – 3 hPa; the lower bound is once again selected by consideration of HIRDLS retrieval quality, whilst the upper bound approximates the level at which the MIPAS vertical resolution reduces from 3 km to 4 km.

These results provide a poorer match than our comparisons between HIRDLS and COSMIC, with a greater proportion of results at weak and negative correlations (upper plots). Results are slightly skewed in the direction of higher standard deviations in the HIRDLS data; this is again consistent with the vertical resolutions of the instruments.

A major difference is the distribution of the well-correlated profiles; whilst in the HIRDLS-COSMIC comparison the best-correlated results largely lie clustered about the 1:1 correspondence line at high standard deviations on both axes, the best-correlated results in this comparison are instead scattered across the high standard deviation regions of both instruments, with no noticeable correspondence to the 1:1 relationship line. Hence, although MIPAS and HIRDLS observe small-scale structure of broadly similar magnitude after allowing for resolution differences, it is unsure whether HIRDLS and MIPAS observe the same fine temperature structure in the middle atmosphere.

6.5.2.3 HIRDLS-SOFIE

Figure 6.10 shows our analysis of colocations between HIRDLS and SOFIE. As shown in figure 6.6(d), we only have a small number of such colocations, so it is consequently difficult to draw any significant conclusions from our results.

We consider results from 100 hPa to 0.5 hPa. These results show only very small standard deviations in the SOFIE data analysed, with all the measured profiles having standard deviations in a narrow range around 0.6 K. HIRDLS results are more widely distributed in standard deviation, but are still fairly tightly clustered. Correlations are generally positive, but low. These

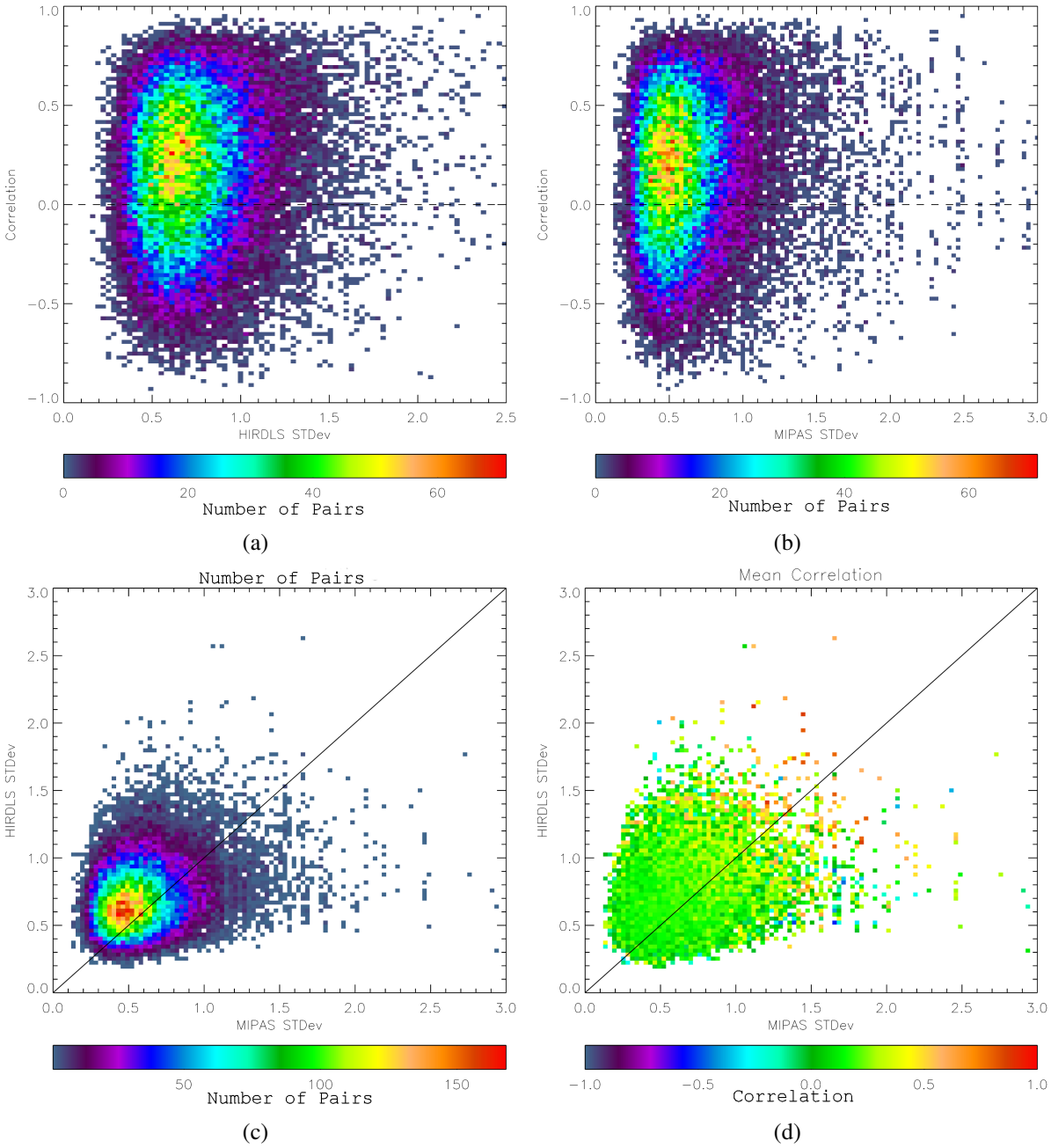


Figure 6.9: *Correlation/standard deviation analysis of HIRDLS-MIPAS colocations*

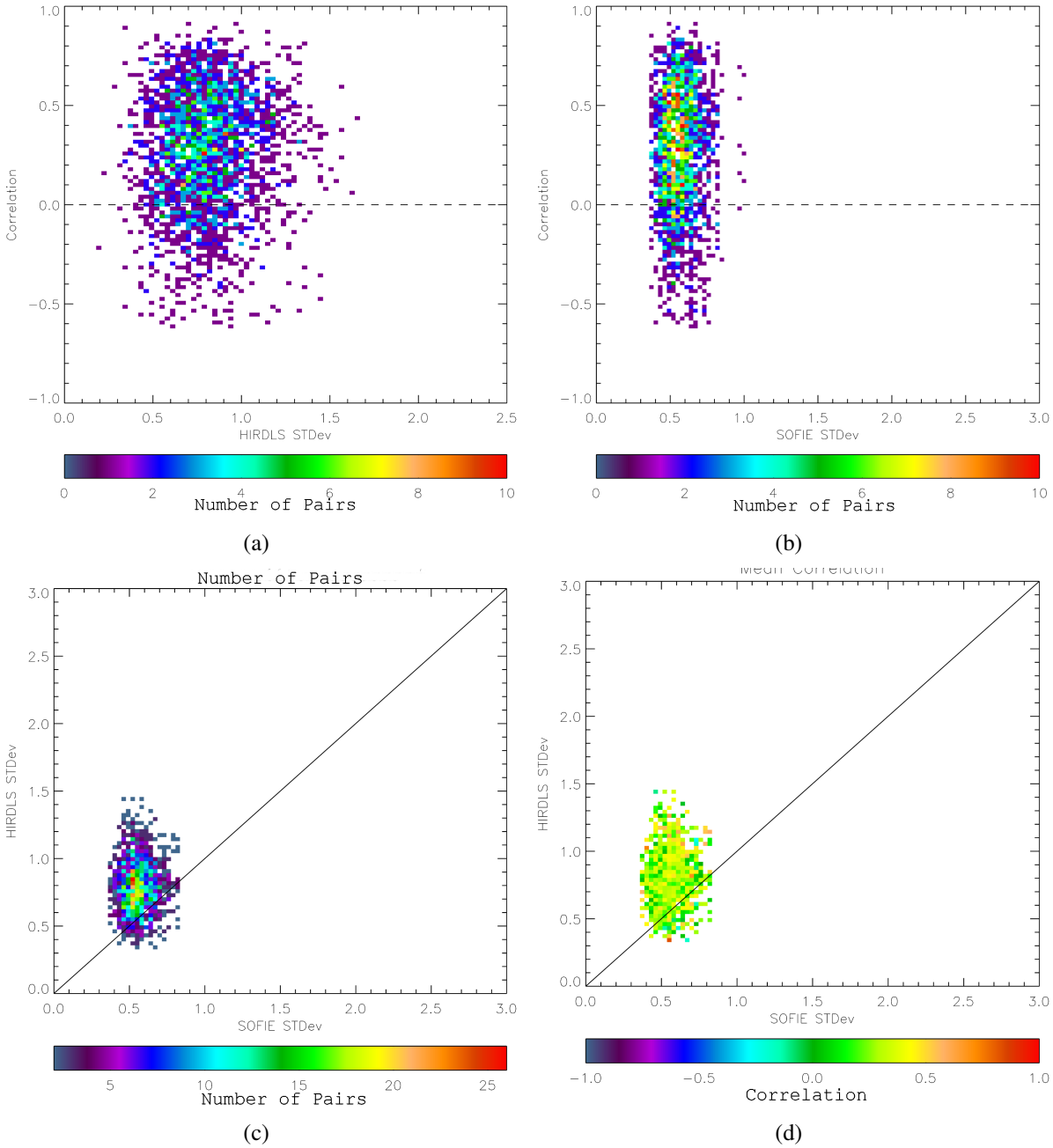


Figure 6.10: *Correlation/standard deviation analysis of HIRDLS-SOFIE colocations*

results are consistent with the significantly lower vertical resolution of SOFIE (∼3.5 km).

6.5.2.4 HIRDLS-SABER

Finally we consider HIRDLS-SABER colocations. For this comparison, we consider data ranging from 50 hPa to 0.5 hPa.

Considering first the standard deviation plots, we once again observe distributions peaked at positive values, with the bulk of the distribution lying at correlation coefficients between 0.2 and 0.6, i.e. fair but not high correlations. In both plots, the peaked region is is centred at broadly the same correlation values as the results for COSMIC, above, but with less of a skew towards higher correlations at higher standard deviations. Standard deviation values are generally broadly-distributed for both instruments.

Moving on now to the lower plots, we once again see the results distributed broadly about the 1:1 correspondence line, in contrast to those for the HIRDLS-COSMIC comparison. The region of high standard deviation on both axes is made up largely of highly-correlated profiles, suggesting as with our COSMIC analyses that those profiles with the greatest amount of fine structure, ideally corresponding to those with gravity wave activity, are the best correlated and that those with low-level noise make up the bulk of uncorrelated profiles.

6.5.3 Conclusions

Whilst the correlation/standard deviation analyses discussed here cannot show unambiguously whether we are correctly detecting gravity wave signals in our HIRDLS data, they do show which instruments reproduce the small-scale structure of HIRDLS well, and accordingly which we would expect to have similar gravity wave structure. This allows us to place the results obtained from the more direct ST comparisons below in their proper context. We conclude from these analyses that the datasets which best reproduce the fine structure in HIRDLS data are COSMIC and SABER, with MIPAS and SOFIE showing much weaker correlation.

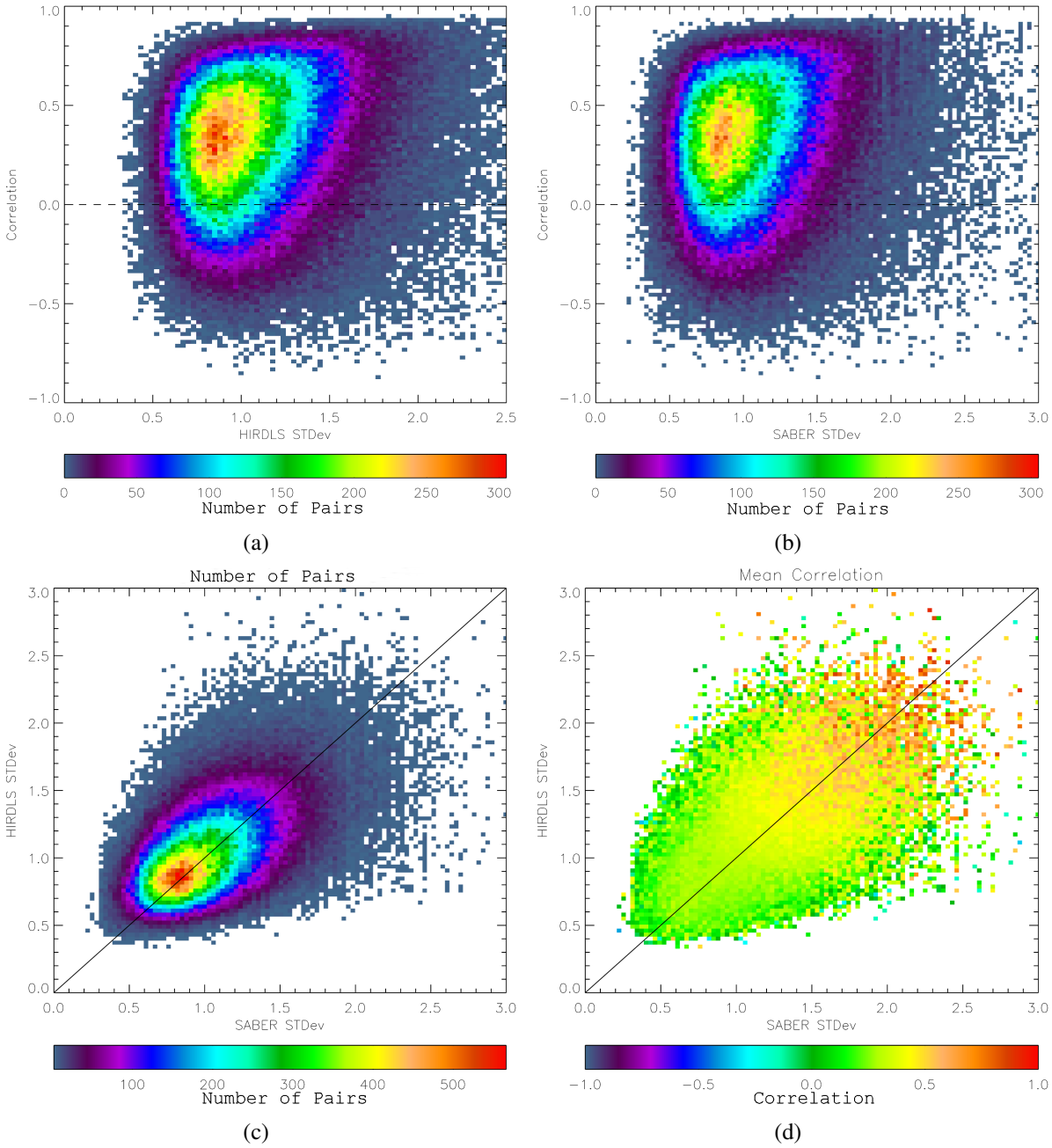


Figure 6.11: *Correlation/standard deviation analysis of HIRDLS-SABER colocations*

6.6 ST Plots

6.6.1 Methodology

In order to validate the results of our S-Transform gravity wave detection, it is useful to compare the results of such analysis over the full range of instruments analysed. Figure 6.12 illustrates the method used to carry out this analysis. Initially, colocated profile pairs are filtered using the half-cosine bell approach outlined above. These yields a colocated pair of perturbation profiles (top left), which are reinterpolated onto a scale representing the published vertical resolution of the coarser instrument. These profiles are then individually S-Transformed (top right).

Since we are considering only individual profiles, we cannot apply the cospectral analysis used for HIRDLS data in the rest of this thesis. Accordingly, we consider only the individual S-Transformed profiles; this yields an estimate of the activity in the single profile rather than that present across both profiles of an adjacent pair, and hence emphasises signals present only in individual profiles, i.e. significantly more spatially-localised gravity waves. We also cannot compute the phase shift of the wave activity between two profiles, preventing us from computing an estimate of the horizontal wavelength of any detected waves.

We next discard any result which will not be Nyquist sampled by the coarser of the two instruments, i.e. signals with vertical wavelengths less than twice the level separation at the coarser resolution. We also discard any signal with a vertical wavelength greater than 16 km, as with our previous HIRDLS analyses. This focuses our analysis onto inertia-gravity waves which are sampled well by both instruments, at the expense of any fine detail which is detected near the resolution limit of the finer instrument; since we are comparing results between instruments in this situation, this is an acceptable tradeoff. We then extract the peak temperature perturbation at each height level (bottom left), and the vertical wavelength of this signal (bottom right). We repeat this analysis for each colocated profile-pair.

Figures 6.13 – 6.16 show the results of these analyses on a whole-dataset scale. We first consider the mean and standard deviation of the amplitudes and vertical wavelengths of the detected waves as a function of height for each instrument pair. In each case, the top left plot shows the mean measured temperature perturbations for the two instruments as a function of altitude, with the dashed lines indicating a one standard-deviation range to either side of this

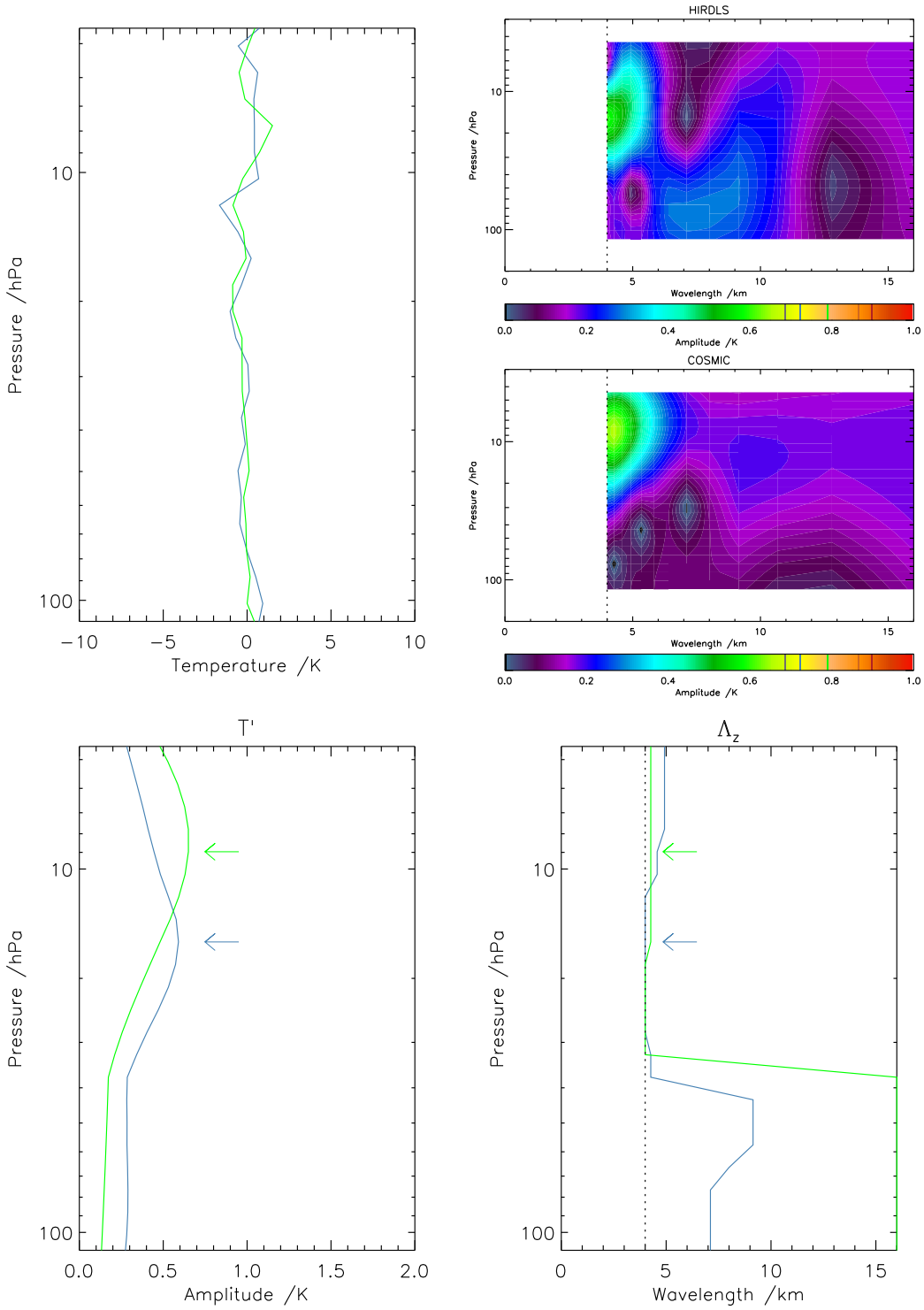


Figure 6.12: *Sample comparison between S-Transform results for our previously-shown colocated HIRDLS (blue) and COSMIC (green) profile, after high-pass filtering with a half-cosine bell filter. Top left illustrates filtered profiles interpolated onto the resolution of the coarser instrument, top right the S-Transforms of these two profiles, bottom left the amplitude of the computed peak temperature perturbations as a function of height, and bottom right the computed vertical wavelengths for these temperature perturbations. The vertical dashed line indicates the resolution limit applied, whilst the arrows indicate the temperature maxima selected from the two profiles and associated vertical wavelengths.*

mean line. The top right plot similarly illustrates measured mean vertical wavelengths and their standard deviations, with the dashed vertical line indicating the resolution cutoff applied to our data[3]. The colours used for the line in each plot are the same as used in figure 6.2; in particular, HIRDLS is always shown in blue.

The two middle plots in each figure illustrate a different approach to this comparison. Consider first the left-hand figure, describing temperature perturbations. On each axis, the coordinates indicates the peak temperature perturbation in each individual profile, as indicated by the arrows on figure 6.12, with results binned on a 0.075 K by 0.075 K scale. These points are not necessarily at the same height level; they indicate only the individual peaks in the profiles. This selection makes sense if considered in terms of an individual gravity wavepacket propagating between the two measurements: in the time between the two profiles being measured, the wavepacket will propagate some distance vertically (and also horizontally), and accordingly a measurement at the same height would fail to detect the corresponding signal. By selecting in this way, any wave packets with temperature amplitude significantly above the local mean will accordingly be associated with each other across the profile pair.

The middle-right plot indicates the measured vertical wavelength of this temperature maximum in each profile; applying the same logic, this should show a good match for high-amplitude wavepackets separated in height but representing the same wave signal. On both of these plots, the solid diagonal line indicates a 1:1 correspondence between the two datasets, and, on the right-hand figure, the solid horizontal and vertical lines indicate the resolution cutoff applied to the data. Due to the discrete height levels, the measured wavelengths do not form a continuous range, but are themselves discretised: accordingly, this figure uses colour to represent the number of results for each wavelength-pair. Note that this colour scale is different for each instrument pair to allow for the significantly different numbers of colocations considered.

Finally, the bottom two figures show the difference in measured vertical wavelength between the two instrument peaks using this technique. For each case, the left-hand figure illustrates the wavelength difference as a function of HIRDLS temperature perturbation, and the right-hand figure as a function of that of the second instrument. A positive difference, above the dashed line, indicates that HIRDLS measures a shorter vertical wavelength at this height, and vice

[3]It should be noted that the pressure ranges covered by the profiles in each following figure are those analysed by this instrument-pair; that is to say, the pressure scales do not cover identical pressure ranges in each plot.

versa. As the measured temperature perturbation peak increases, the magnitude of the signal being measured deviates further from the background mean and is accordingly less likely to be due to noise; hence, we would expect the difference in measured wavelength between the two instruments to decrease with increasing temperature pertubations, as it becomes more likely that we are resolving gravity wavepackets rather than noise. These two figures hence provide a check upon the values shown in the two scatter plots.

It should be noted that, due to the decrease of atmospheric density with height, we would expect the magnitude of temperature perturbations due to gravity waves to increase with increasing altitude. Accordingly, basing our analysis on the largest temperature perturbation present in each profile will inevitably bias our results towards signals at higher altitudes. Nevertheless, this process still compares like-with-like across each profile-pair, and hence provides useful comparative results. In particular, for a gravity wavepacket of significantly greater amplitude than the mean, we should still resolve the same feature at all heights.

6.6.1.1 HIRDLS-COSMIC

Figure 6.13 illustrates the results obtained from the ST comparison of HIRDLS and COSMIC.

Considering first the temperature analyses in the top and middle left figures, we observe a marked bias in the temperature perturbations measured towards COSMIC: the mean temperature perturbations as a function of height (top left) are higher in COSMIC at all height levels, and the profile-peak temperature perturbation scatterplot (centre left) also shows a bias in favour of higher values in COSMIC. As discussed in sections 6.5.2 and 6.5.2.1, this may be residual effects of a potential higher resolution available to COSMIC. The variabilities of the results from the two instruments (dotted lines in upper figure) are broadly similar for the two instruments, suggesting further that the results are fundamentally similar between the two instruments.

We now examine the wavelength plots, at top and middle right. The mean and standard deviations of the measured wavelengths as a function of height are broadly equivalent, with a slight bias towards longer wavelengths being observed by HIRDLS at lower altitudes. The scatterplot for wavelength shows excellent agreement, with wavelength pairings tightly distributed about the 1:1 correspondence line.

Finally, the wavelength difference plots at bottom show the expected distribution: wave-

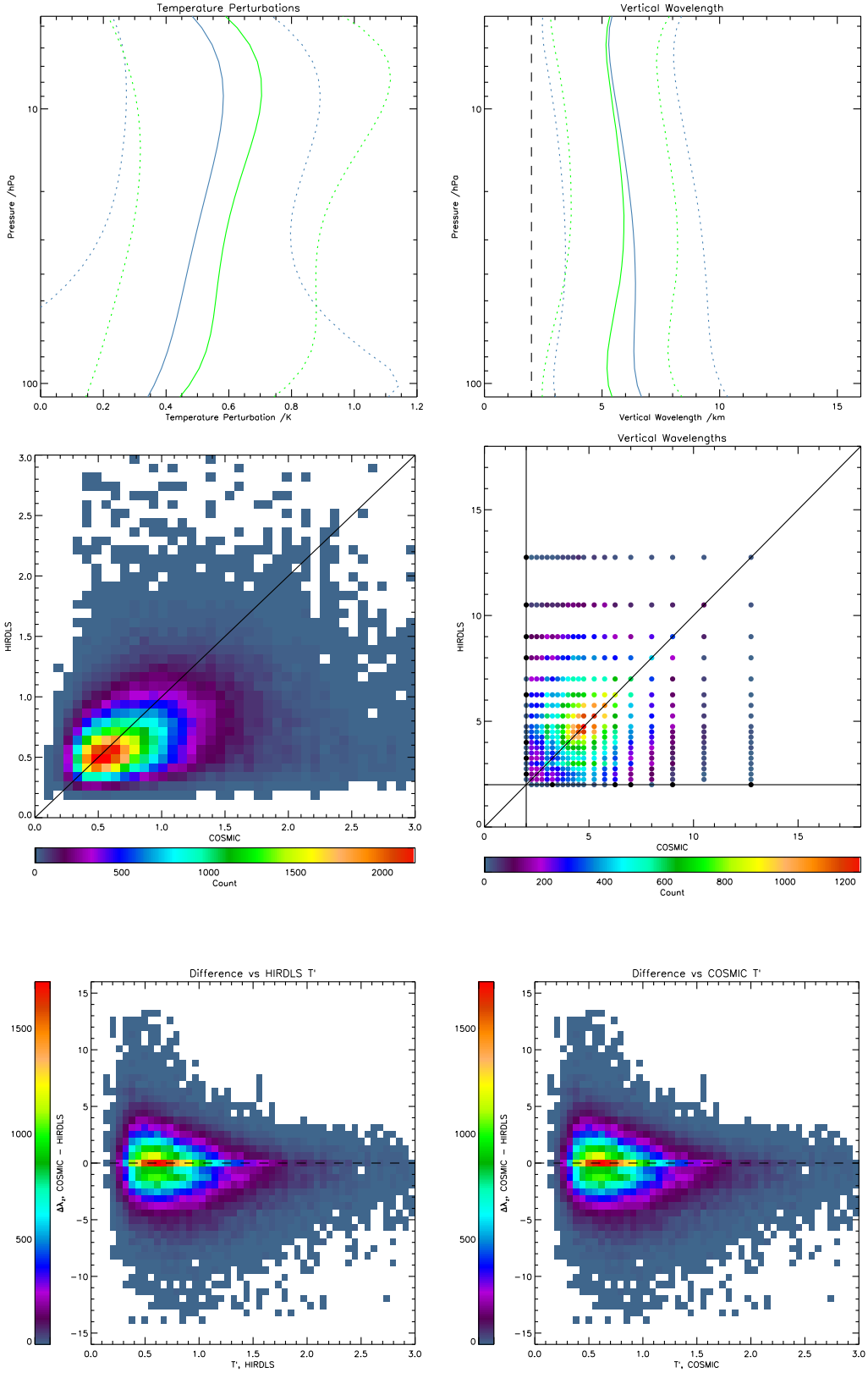


Figure 6.13: *S-Transform distribution analysis of HIRDLS-COSMIC colocations*

lengths are generally closely clustered about the zero-difference line, with the distribution narrowing with increasing temperature perturbation in both instruments, consistent with wave features being detected rather than noise.

6.6.1.2 HIRDLS-MIPAS

We next discuss the HIRDLS-MIPAS ST comparisons, shown as figure 6.14. Once again, the mean temperature perturbation profiles and their variability are broadly similar between the two instruments, as are the mean vertical wavelength profiles.

The temperature perturbation scatterplot shows a bias towards HIRDLS consistent with the higher vertical resolution of this instrument, but the results between the two instruments are not significantly correlated with the 1:1 correspondence line, with the great majority of the results falling into a large cluster at low perturbations: whilst the COSMIC distribution was similarly dominated by a low-amplitude noise region, a linear relationship could be observed after allowing for the probable resolution bias, whilst any similar trend for this pairing is much less apparent. The vertical wavelengths are similarly poorly correlated, with the bulk of results falling into a small range of wavelengths at around 7 km in both instruments and no significant linear trend.

The wavelength difference plots do show the expected distribution at low temperature perturbation, but with a bias in favour of HIRDLS measuring longer vertical wavelengths at all amplitudes.

6.6.1.3 HIRDLS-SOFIE

Figure 6.15 shows the comparisons between HIRDLS and SOFIE. The correspondence between the two instruments is generally poor, with markedly different results observed with all the analysis methods used. In particular, there are substantial offsets in the mean height profiles, no clear correspondence to the theoretical match lines for the scatterplots, and the wavelength-difference plots show large discrepancies and no clear trend.

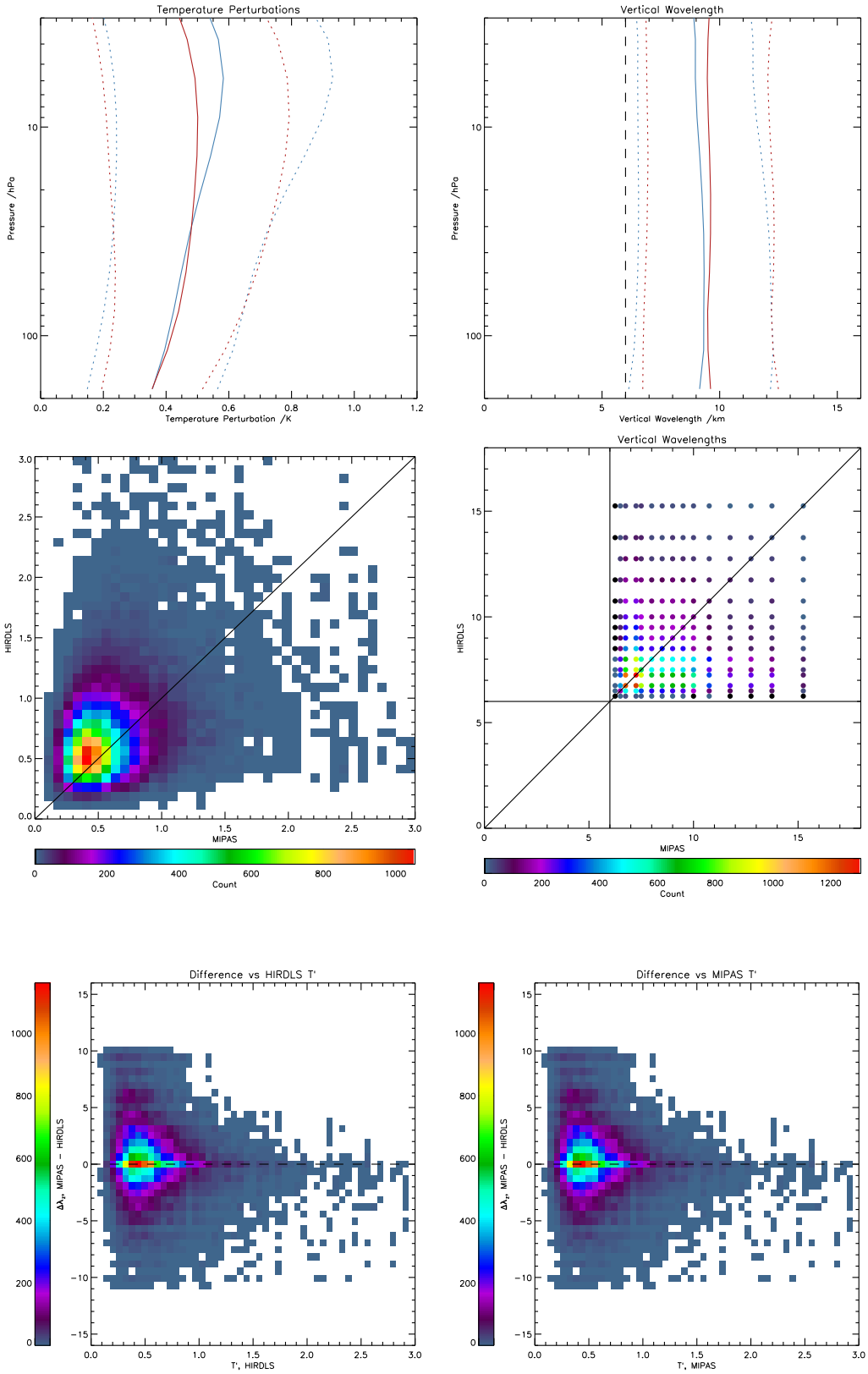


Figure 6.14: *S-Transform distribution analysis of HIRDLS-MIPAS colocations*

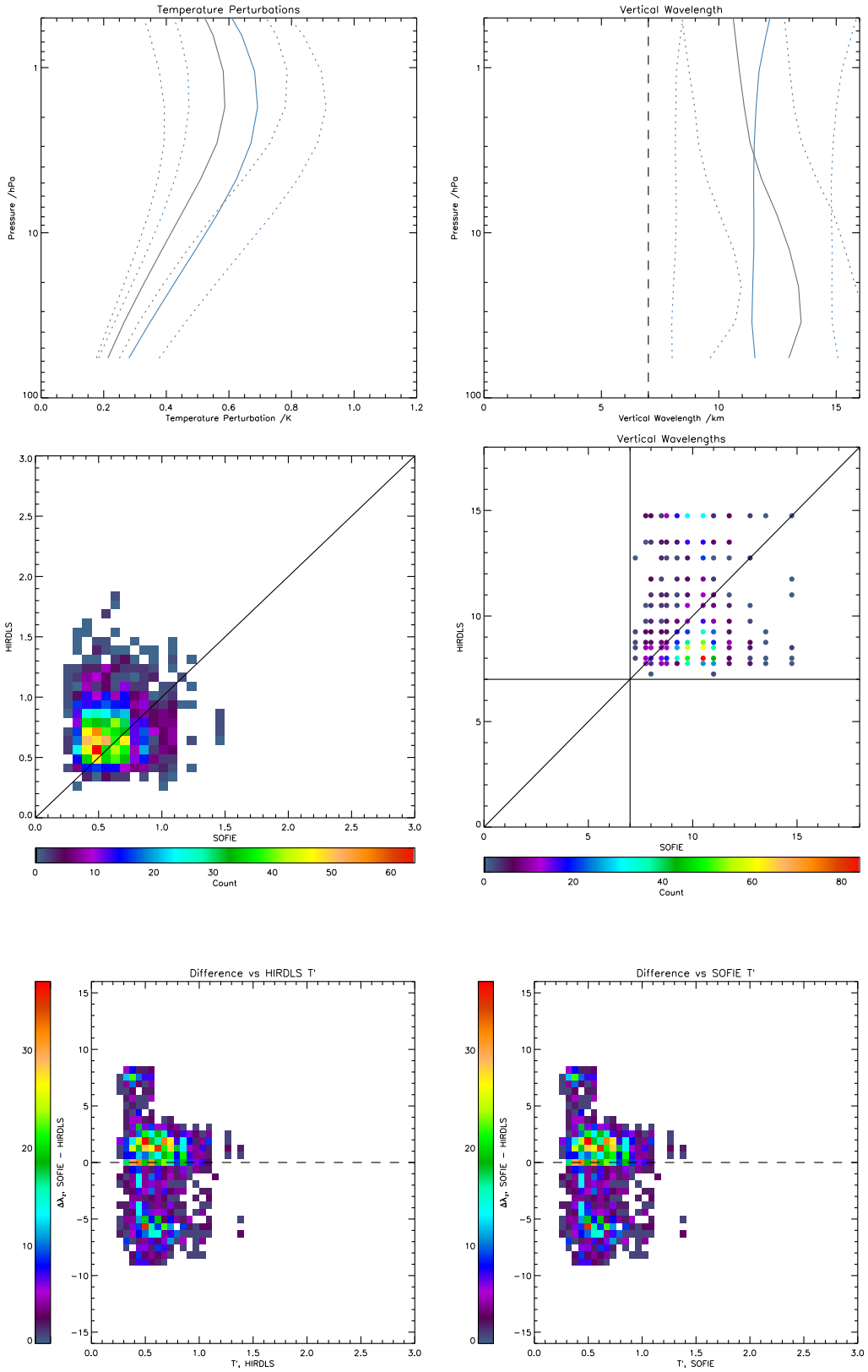


Figure 6.15: *S-Transform distribution analysis of HIRDLS-SOFIE colocations*

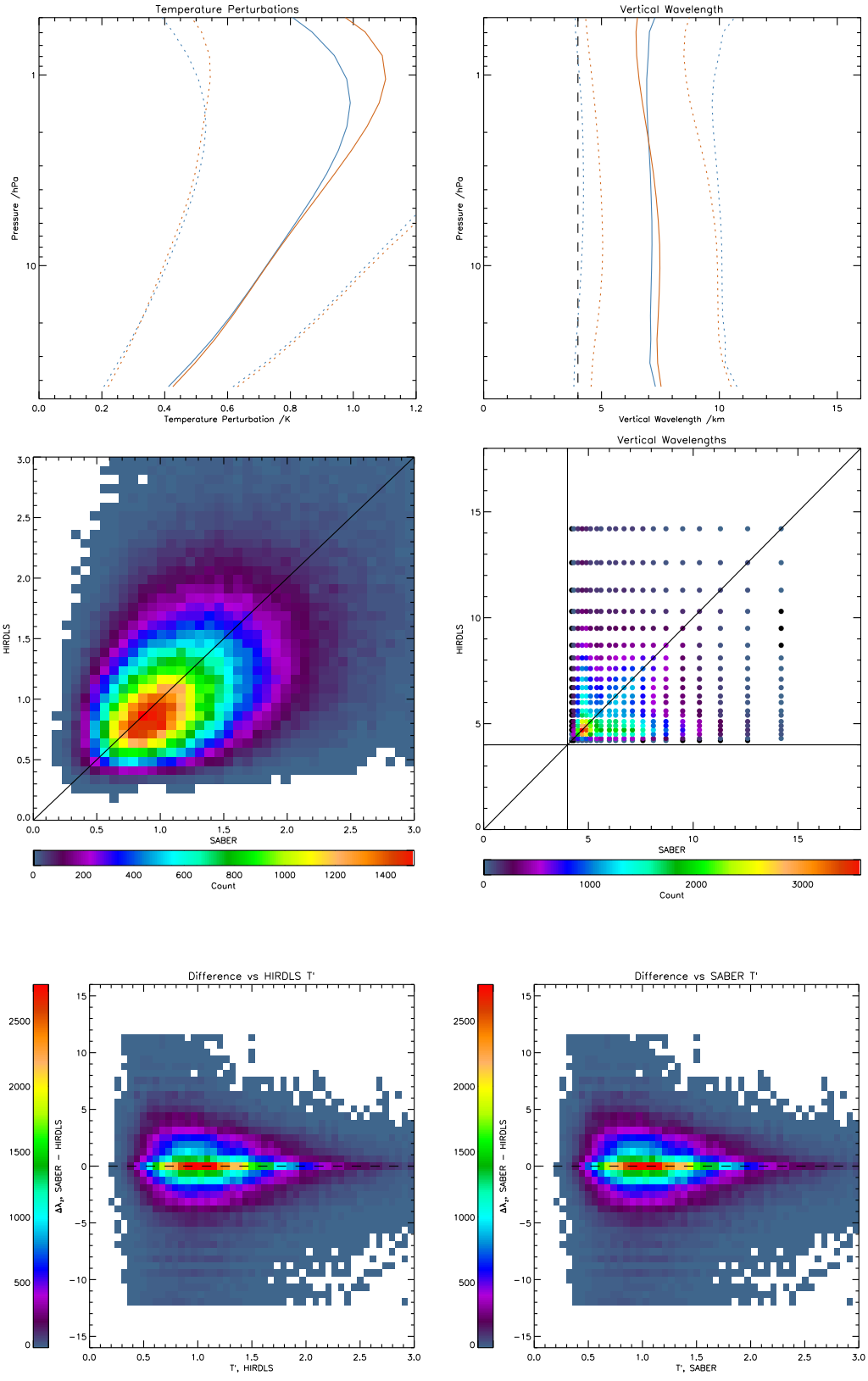


Figure 6.16: *S-Transform distribution analysis of HIRDLS-SABER colocations*

6.6.1.4 HIRDLS-SABER

Finally, we consider the HIRDLS-SABER comparisons shown in figure 6.16. As anticipated by our correlation analyses, these correspond significantly better than did SOFIE or MIPAS, with a very close match observed for the mean profiles except for temperature perturbations at high altitude, which show a bias towards larger values in SABER.

The temperature perturbation scatterplot also shows an excellent correspondence, with the measured results distributed strongly along the 1:1 line, continuing even outside the large cluster presumed to be noise at low T'. The vertical wavelength scatterplot is more interesting, with three distinct clusters apparent outside the main low-wavelength peak at the bottom left. The first of these clusters follows the main 1:1 correspondence line, whilst the other two clusters are distributed parallel to the axes of the graph, indicating profiles in which one instrument observes a longer vertical wavelength whilst the other observes a shorter one. The reasons for these clusters are uncertain, but the results obtained still show a markedly better correspondence those for SOFIE or MIPAS.

6.6.2 Conclusions

Based upon these our S-Transform analyses, we conclude that the instruments which best reproduce the results of HIRDLS are COSMIC (well) and SABER (more weakly, but still well), whilst MIPAS produces some good matches but also many poor ones, and SOFIE fails completely to resolve the same structure. These results correspond well to the vertical resolution of the instruments under consideration, and also agree well with the results of our correlation/standard deviation analyses above, suggesting that our S-Transform analysis method is sound.

Chapter 7

Arctic Winter 2006

7.1 Introduction

Recent studies[1] (Siskind et al., 2007; Manney et al., 2008b) have described a significant vertical displacement of the 2006 Arctic stratopause, with temperatures at 0.01 hPa exceeding 250 K and unusually cold temperatures at the usual stratopause height. Observations from the Microwave Limb Sounder (MLS) instrument on the Aura satellite and SABER, amongst others, showed a steady drop in the polar stratopause height over the course of January 2006, followed by the breakdown of the polar vortex. After around two and a half weeks, a cool high-latitude stratopause reformed at high altitude, remaining separated from that at low and midlatitudes for a further month. Figure 7.1 illustrates this evolution as a series of daily snapshots of zonal mean temperature in the Northern hemisphere with data obtained from SABER, for the period covered by SABER's north-covering yaw cycle phase during this part of 2006.

To more fully understand the processes taking place during this period, Siskind et al. (2007) simulated the middle atmospheric temperature structure at high latitudes for early 2005 and 2006 using the NOGAPS-ALPHA GCM, extended to reach a top model height of 1×10^{-4} hPa and with improved radiative forcing and ozone climatology. Two different models for mesospheric gravity wave drag (GWD) were used, one based on simple Rayleigh friction and one using a subgrid-parameterised orographic gravity wave drag (OGWD) model which also accounted for wave-filtering; they were then compared against a control run with no GWD. Of

[1] please note that this chapter has been published recently as **?**, and accordingly forms part of these recent studies in the literature

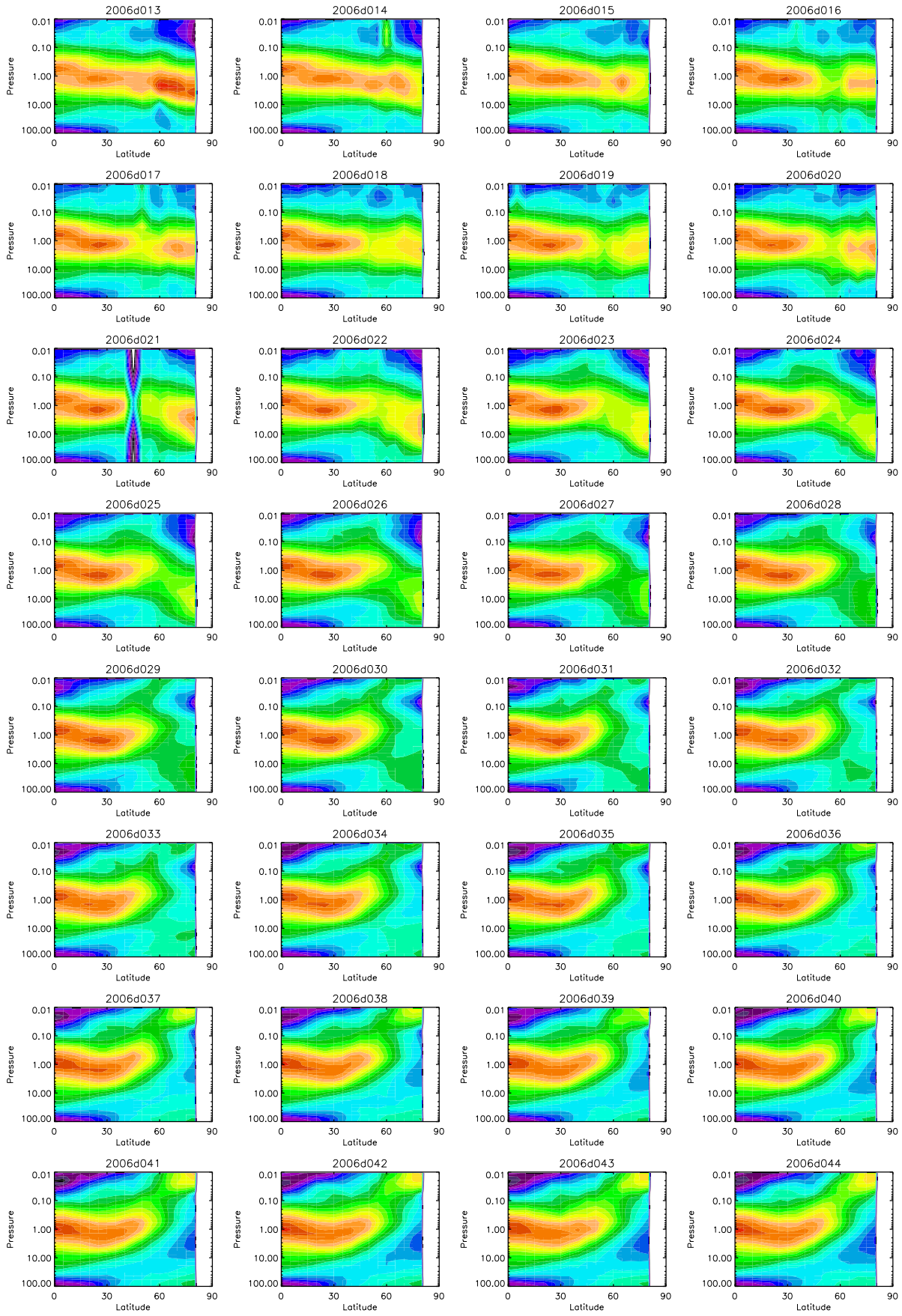


Figure 7.1: *Series of daily snapshots showing the evolution of the polar stratopause in early 2006 as observed by SABER. 2006d059 has been omitted as there was no data for this day. Colour scale is shown after the last figure. (continues)*

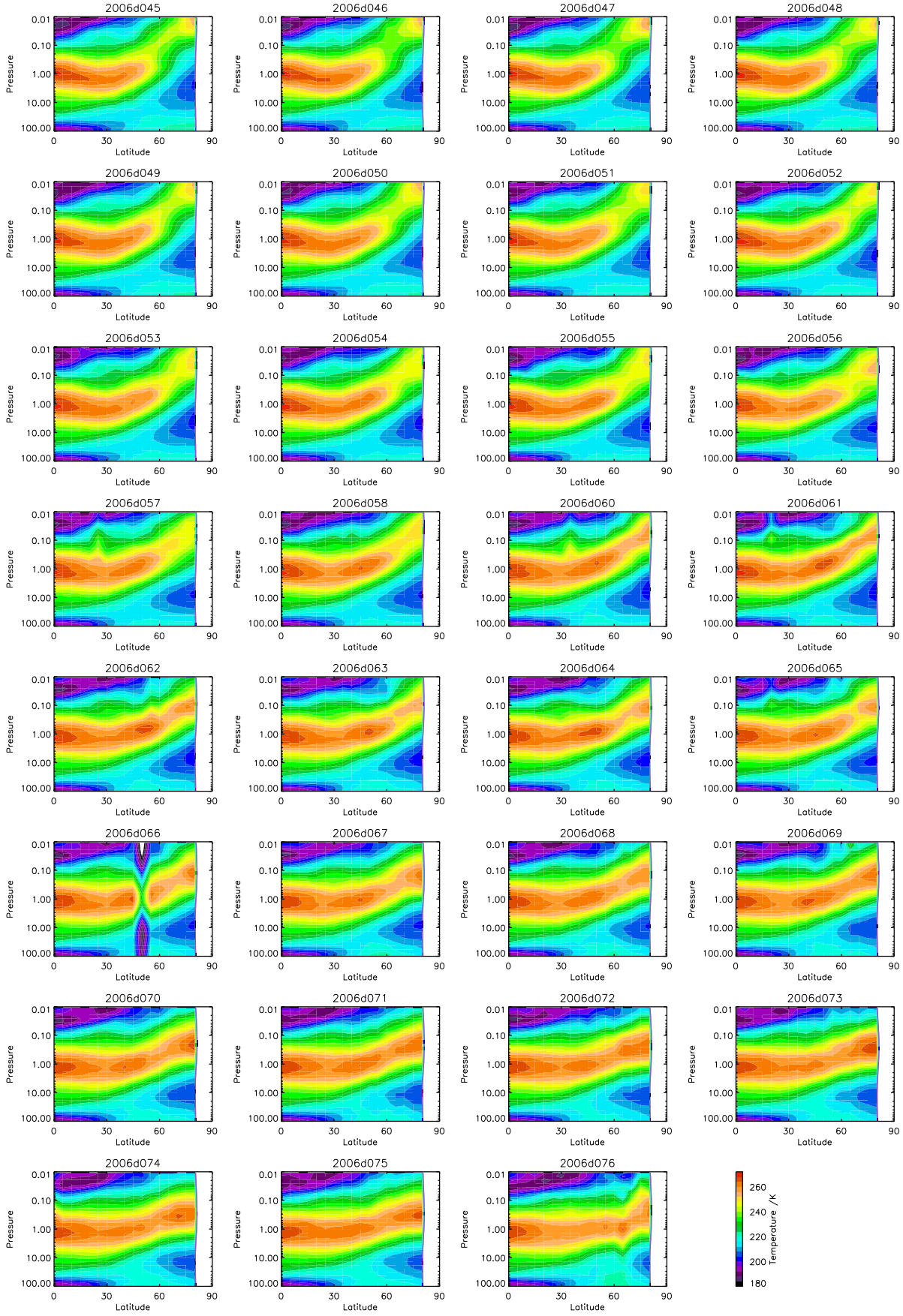


Figure 7.1: *(continued) Series of daily snapshots showing the evolution of the polar stratopause in early 2006 as observed by SABER. 2006d059 has been omitted as there was no data for this day. Colour scale is shown after the last figure.*

these, only the OGWD model captured the key morphology of both years. Further, the OGWD runs suggested that the changed stratopause was related to strongly reduced levels of mesospheric GWD, due to wave-filtering in the stratosphere by atypically weak lower-stratospheric zonal winds.

HIRDLS measurements allow us to directly observe a significant proportion of the gravity wave spectrum of the middle atmosphere. This allows us to analyse HIRDLS gravity wave activity data in the stratosphere during the period of anomalous stratospheric behaviour and compare to other, more typical, years. We can also compare these results to wind data from European Centre for Medium-Range Weather Forecasting (ECMWF) operational analyses, allowing us to examine the stratospheric dynamics during this period and provide possible explanations for the results obtained

As discussed above, HIRDLS began collecting data on the 21st of January 2005 (Barnett and Gille, 2006a), with some intermittency in the temporal data coverage for a few weeks after this. All data discussed in this chapter, both HIRDLS and ECMWF, have been smoothed by 5 days post-processing: this is because the values obtained for momentum flux exhibit significant variation between days and hence require some smoothing to allow conclusions to be drawn.

Profiles are most heavily concentrated around the northern and southern turnarounds at $\sim$80N and $\sim$62S due to the orbital path of the satellite and the way in which the blockage restricts the viewing angle. The orbital path of Aura remains the same throughout the year, in contrast to that of (for example) SABER, which uses a 60-day "yaw" cycle to alternate between coverage of high latitudes in alternate hemispheres. This allows HIRDLS to study the development of selected parts of the high-latitude atmosphere over extended periods.

7.2 Temperature and Wind Trends during Arctic Winter 2005/06

Figure 7.2 shows a time series of zonal mean temperature at 10 hPa, 60N over the three consecutive Northern Hemisphere winters 2004/05, 2005/06 and 2006/07, as measured by HIRDLS. Compared to the other two years shown, a clear difference can be observed in the data for 2005/06. The temperature rises sharply in early-to-mid January 2006, followed by an equally sharp fall in late January, and does not recover to its typical levels until late March-early April. Although it is difficult to validate the momentum flux measurements in such a way as to show

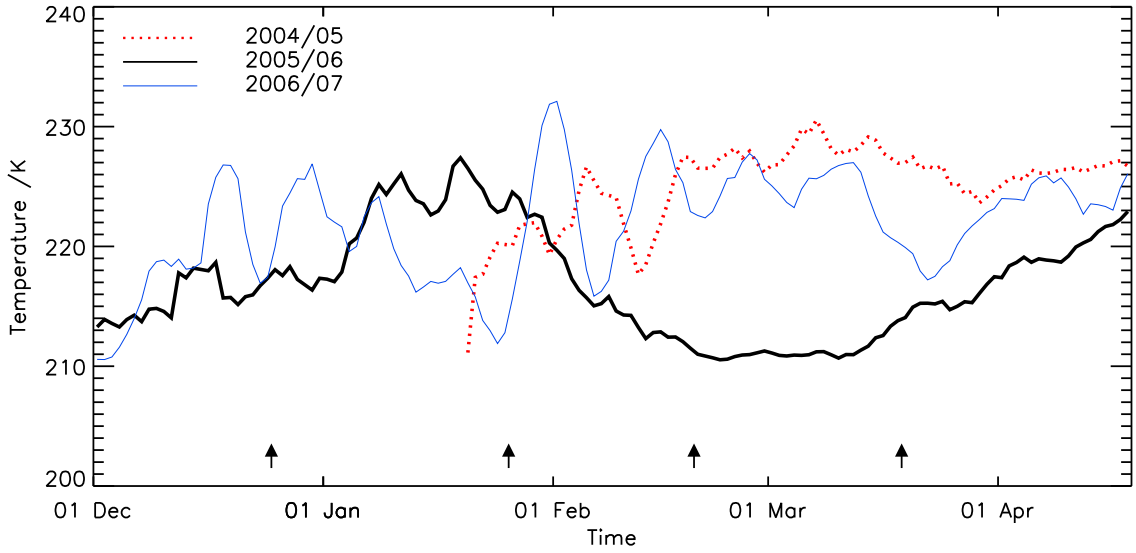


Figure 7.2: *10hPa winter time series of zonal mean temperatures at 60N measured by HIRDLS for 2004/05, 2005/06 and 2006/07. Note the substantially reduced temperatures during February and March 2006 relative to the other two years. Arrow marks in this and subsequent plots indicate the four days detailed in figure 7.6 and discussed in more detail in the text.*

that early 2005 and 2007 are typical years in terms of momentum flux values, we can do so for temperature: the temperatures measured in 2005 and 2007 fall within one standard deviation of the long term monthly-mean temperatures (223.5±3.4 K for February, 223.9±3.7 K for March, derived from ECMWF ERA-40 data for the period 1979–1999), whilst those for late February and March 2006 fall substantially outside this range.

This behaviour is consistent with that described by other studies (e.g. Hoffmann et al. (2007), Manney et al. (2008a), Manney et al. (2008b)) as a manifestation of a major sudden stratospheric warming (defined as having easterly winds down to 10 hPa at 60N) during January 2006, with a strong reversal of 1 hPa zonal winds in mid-January followed around a week later by a reversal of the 10 hPa zonal winds above 60N and a decline in mesospheric wave-one amplitude.

Hoffmann et al. (2007) discuss MF radar and meteor radar observations of mesospheric (70–94 km) winds and temperatures during the same period. In their results, easterly winds descend through the mesosphere during January, followed by a period of intensified westerly winds at all heights in February. This occurred while stratospheric planetary wave-one amplitudes decreased during late January. Easterly winds then once again descend from the upper to lower mesosphere during March and April, and remain at all heights thereafter. Both periods of easterly winds are followed by a prolonged period of reduced gravity wave activity throughout the mesosphere, especially marked in the upper stratosphere-lower mesosphere (55–70 km) region,

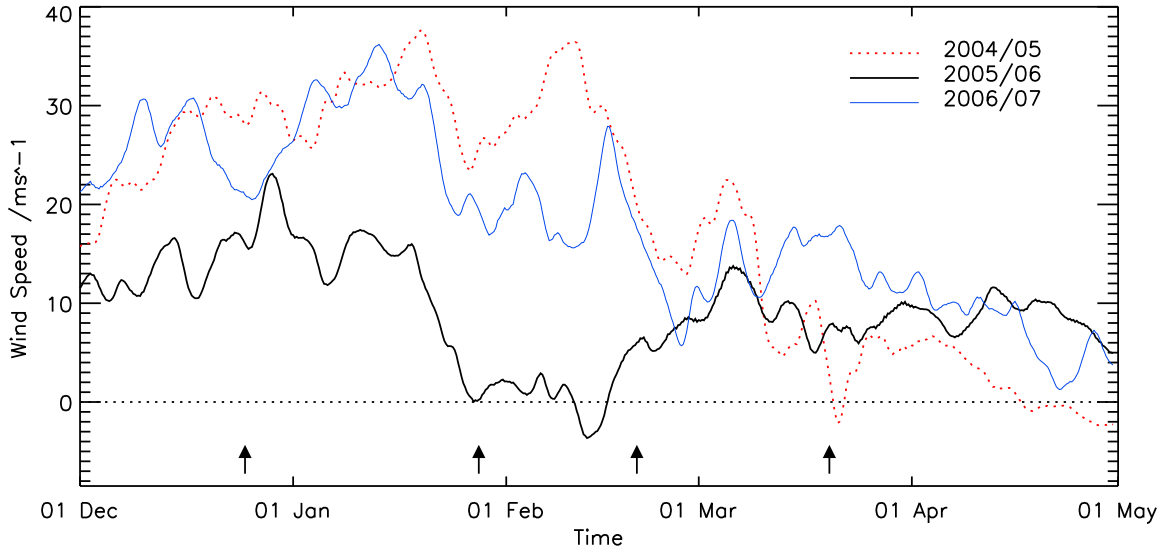


Figure 7.3: *50hPa winter time series of zonal mean zonal winds at 60N from ECMWF. Note the prolonged period of substantial easterly (negative zonal) winds during late January to early February 2006 by comparison to the other two years.*

where activity falls almost to zero. During January, the region of low activity is located almost entirely below 65 km, whilst the more prolonged period beginning in March and April exists at all levels below ∼80 km.

Figure 7.3 shows a time series of zonal mean winds at 50 hPa over the same period, obtained from the ECMWF operational analyses. Whilst the ECMWF data over this period do not fully capture the raised stratopause due to lack of observations at high altitudes (Manney et al., 2008b), at low and mid-stratospheric altitudes the wind speed data is well-constrained by observations. Winds in 2004/05 and 2006/07 show a steady drop with time, consistent with the usual climatological trend. 2005/06 once again appears substantially different: zonal winds are significantly reduced by comparison with 2004/05 and 2006/07 and there is a prolonged period during late January and early February of easterly and near-zero westerly zonal winds. This period of reduced westerlies begins in early January, slightly after the time at which the mesospheric winds observed by Hoffmann et al. (2007) reach the lowest extent of the mesosphere. This suggests continuity with the observed descent, and is likely associated with a change in the temperature gradient caused by the earlier SSW.

Manney et al. (2008b) also discuss the amplitude of planetary wave-one features in the stratosphere and mesosphere. Figure 3 of their paper, drawing from multiple sources, showed that the wave-one amplitude was high at all levels from ∼50 hPa up to ∼0.01 hPa throughout December

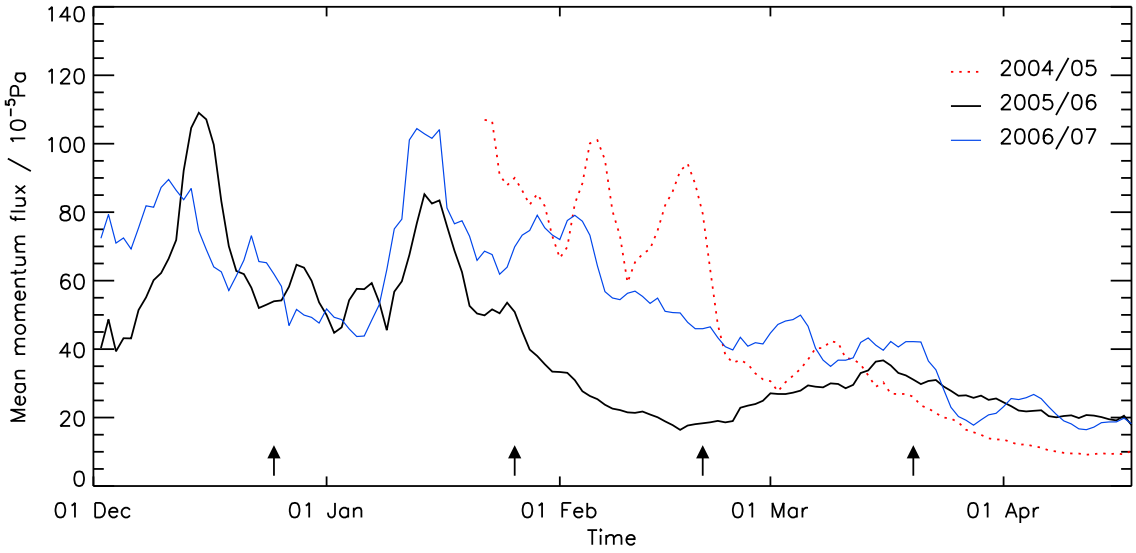


Figure 7.4: *10hPa winter time series of zonal mean momentum flux at 60N measured by HIRDLS for 2005, 2006 and 2007. Note the reduced level between late January and early March 2006 compared to the other two years.*

2005 and most of January 2006, suddenly collapsing at the same time the wind at these altitudes reversed around the 20th of January. A lower-intensity wave-one then formed at high altitudes ($\sim$0.02 hPa), which descended in altitude slowly during February and March. Wave-two amplitudes followed a similar trend in December and January, collapsing around the 20th of January, but did not reform at high altitude with the original intensity.

7.3 Measurements of Momentum Flux in the 2005/06 Arctic Stratosphere

Figure 7.4 shows the zonal-average of the momentum flux (as computed via equation 3.2.30, but neglecting the altitude-dependent term) at 10 hPa, 60N. It should again be noted that this expression describes the absolute value of the along-track component of the full momentum flux, and gives only a constraint on the full vector formulation. The zonal mean is used to allow finer temporal resolution in the data analysis, as not all longitudes are covered on any given day. The two years of data available for the period from late January until early February show a peak; there is then a steady fall-off observed in all three years during February and March to a low-level at which it stays throughout the summer before rising again in late November (not shown). A sharp reduction over the period of interest is seen in 2006 when compared to the

other two years of data. A steady drop in momentum flux levels is observed from mid-January onwards, recovering slightly towards the end of February before merging back into the general seasonal trend by late March. The scale of this decline is significant as a proportion of the mean value in the other years, with measured levels in mid-late February of 0.2 mPa substantially lower than the average for 2005 and 2007 during this period of around 0.6 mPa, and represents a drop in measured momentum flux by two-thirds.

We can compare these results for momentum flux to the stratopause height results of Manney et al. (2008b). Figure 4 of their paper shows a steady descent of the stratopause (calculated as the 'warm point' of the middle atmosphere) at 60N, beginning in the second week of January and dropping steadily until the last week of January. A new stratopause then reforms at $\sim$70 km and begins to descend slowly, stabilising at $\sim$50 km in mid-March. This suggests a strong link between the stratopause height and the reduction observed in the measured momentum flux.

The second row of figure 7.5 shows daily-mean HIRDLS-derived zonally-averaged gravity wave momentum flux in the Northern Hemisphere for the 20th of February of the three consecutive years 2005–2007, as a function of latitude and height. Monthly-mean calculations show broadly similar results, and preliminary processing of a fourth winter (2007/2008) of data, which are still undergoing validation, also serve to confirm that 2006 is anomalous, with broadly similar temperatures and momentum fluxes observed in 2008 to those in 2005 and 2007. As discussed previously, these figures represent lower bounds to the actual momentum flux present in the atmosphere.

In February 2005 and 2007 (left and right respectively), a characteristic pattern is seen of medium-intensity momentum flux at equatorial latitudes, decreasing slightly in the region $\sim$20N–40N before rising at latitudes poleward of 40N. The intensity of the momentum flux drops off steadily but quickly with increasing height, consistent with density-dependent saturation processes in the upper stratosphere: at 100 hPa, values range from $\sim$1.5 mPa at equatorial latitudes to a strong peak of $>$3 mPa at high latitudes (60N–80N), whilst at 10 hPa values ranging from $\sim$0.3 mPa at the equator to 0.6 mPa at 70N.

The distribution for 2006 (centre), however, clearly shows a marked reduction of measured momentum fluxes at all heights for latitudes north of 40N. Momentum flux levels also fall with increasing latitude. At the 100 hPa level, values range from $\sim$1.5 mPa at the equator through

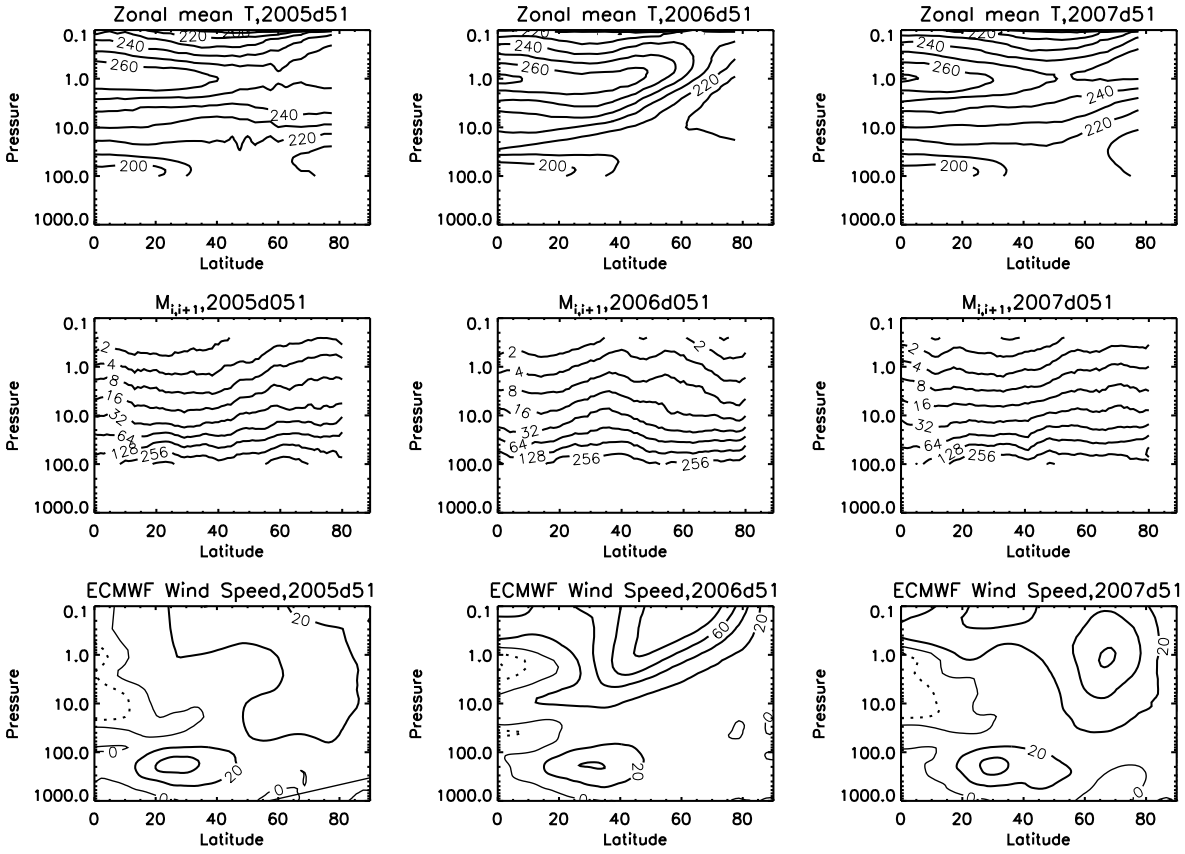


Figure 7.5: *Zonal mean temperature (first row, in Kelvin, from HIRDLS), zonal mean momentum flux (second row, in* 10^{-5} *Pa, from HIRDLS) and zonal mean zonal wind speed (third row, in* ms^{-1}*, from ECMWF) for the Northern Hemisphere on February 20th of 2005, 2006 and 2007. Note that the number of levels in the ECMWF operational analysis model changed at the end of January 2006, giving finer resolution in the vertical after this date.*

a minor peak of ∼3.0 mPa around 35–40N before falling away steadily to a low of 2.5 mPa at 60N, and at the 10 hPa level values rise from around 0.2 mPa at the equator to a high of ∼0.3 mPa at 40N before falling away rapidly to less than 0.15 mPa at 80N. These values are consistent with the conclusions of Siskind et al. (2007) that the level of gravity wave activity, and hence momentum flux due to gravity waves, was low in the Arctic stratosphere during this period, and requires further investigation. The small increase in values at 35N-40N probably represents the edge of the normal region of heightened momentum flux, with regions poleward of this reduced in magnitude.

7.4 Zonal mean temperature, windspeed and momentum flux as a function of latitude

Figure 7.5 also shows Northern Hemispheric zonal mean temperature (top row, from HIRDLS) and zonal mean zonal windspeed (bottom row, from ECMWF operational analyses), for the 20th of February 2005, 2006 and 2007. This day was selected as it lies in the trough of averaged momentum flux at 60N in 2006 (figure 7.4). Several differences can clearly be seen between 2006 (second column) and the other two years. If we first consider zonal mean temperatures (first row), a substantial difference can be observed poleward of 50N, with temperatures at all heights substantially lower than usual. This is consistent with the raised stratopause described in Siskind et al. (2007). There is also, as discussed previously, a strong difference in the recorded momentum flux values (second row), and high-latitude momentum flux values are substantially lower in 2006 at all height levels. The third row, ECMWF zonal mean zonal wind speeds, shows a different polar vortex strength at upper levels but no dramatic difference between the three years below the $\sim$10 hPa surface.

Figure 7.6 shows the same three parameters for four days during late 2005 and early 2006, to illustrate the development of the SSW and subsequent drop in momentum fluxes: the 25th of December (before the divergence of momentum flux from the usual seasonal trend), the 26th of January (the beginning of the trough of zonal mean zonal windspeed at 60N, 50 hPa (figure 7.3)), the 20th of February (as before) and the 20th of March (after levels had merged back into the normal seasonal trend). These days are indicated by small black arrows on each of the relevant time series.

The first column, showing results for the 25th of December 2005, shows a zonal mean temperature profile with a typical separated stratopause. One distinct region stretches from the equator to midlatitudes, associated with adiabatic heating due to solar insolation. The second distinct stratopause, at high latitudes, is due to the gravity wave induced circulation (Dunkerton, 1978; Murgatroyd and Singleton, 1961).

The momentum flux profile is of the form usually observed in the HIRDLS data for Northern Hemisphere winter, with a substantial peak recorded in measured momentum fluxes at high latitudes. The zonal wind distribution on the 25th of December is also typical for this time of

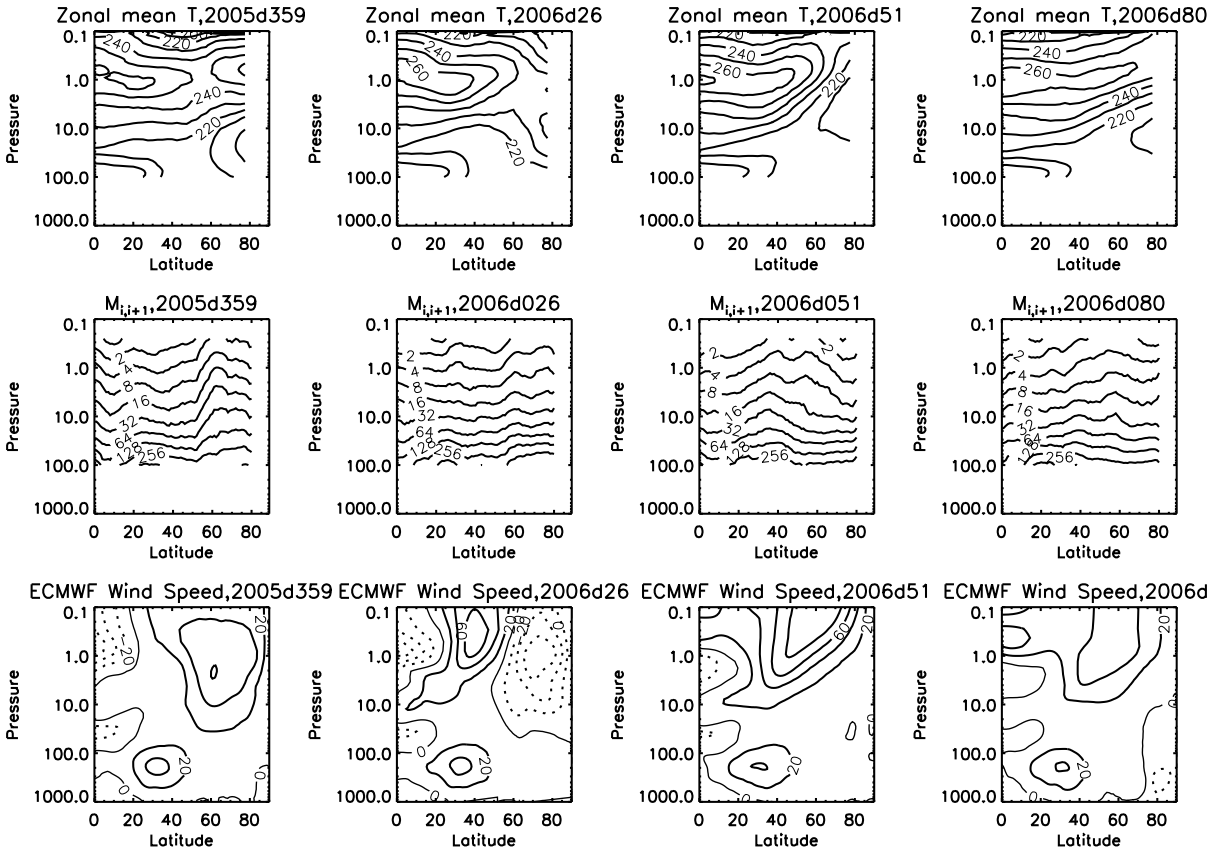


Figure 7.6: *HIRDLS Zonal mean temperature (first row, in Kelvin, from HIRDLS), zonal mean momentum flux (second row, in 10^{-5} Pa, from HIRDLS) and zonal mean zonal wind speed (third row, in ms^{-1}, from ECMWF) for the Northern Hemisphere on four days chosen over late 2005 and early 2006 (December 25th 2005, 26th January 2006, 20th February 2006 and 20th March 2006 respectively). Note that the number of levels in the ECMWF operational analysis model changed at the end of January 2006, giving finer resolution in the vertical after this date.*

year, consisting of a strong westerly zonal wind at high altitudes and latitudes and a smaller peak of easterly zonal winds near the equator.

Distributions for the 26th of January (second column) show a substantially different picture. The most obvious change is in zonal wind speed $\overline{u}$ at high altitude, with a significant peak of easterly winds recorded poleward of 50N indicative of the SSW. Momentum flux values have also changed significantly, with a marked decline from typical values everywhere poleward of around 50N, leaving values at 60N only very slightly above those at the equator and midlatitudes in contrast to their usual comparatively high values. Temperatures, meanwhile, show the early stages of the breakdown in the high-latitude stratopause described in Siskind et al. (2007) and Manney et al. (2008b): this manifests itself as a temperature peak developing at $\sim$10 hPa, representing a significant drop in the stratopause height. This is consistent with a substantial increase in the strength of the Brewer-Dobson circulation during the warming.

By the 20th of February (third column), the breakdown in the high Arctic stratopause has advanced significantly, and the low-altitude stratopause of late January has disappeared. A new higher stratopause has formed at very high altitudes ($\sim$0.05 hPa, see figure 7.1), with significantly lower temperatures than usual throughout the rest of the Arctic stratosphere. Momentum fluxes are further reduced, with high-latitude values less than those at the equator and in midlatitudes. By this point, however, high-altitude polar zonal mean zonal winds have become strongly positive again, heralding the beginning of the return of the Arctic atmosphere to a more normal state. By the 20th of March (fourth column), all three parameters are back to their typical values and distributions for the time of year (taking into consideration that momentum flux values by this point of the winter are usually substantially lower at high latitudes than they would be in mid-late February), and the stratopause has returned to a more typical height ($\sim$0.5 hPa).

7.5 Filtering by the Wind

As shown by figures 7.3 and 7.4, the period of rapid decline in momentum flux during late January and early February 2006 coincided with a period of easterly and strongly reduced westerly zonal mean winds, as described in section 7.2. This suggests a possible factor in the reduction in momentum flux levels.

Gravity waves arise, mainly in the troposphere, from a range of sources, including orography,

convection and wind shear. These waves exhibit a broad range of phase speeds (Smith, 1996). The winds through which they travel, however, affect waves differently, depending upon the propagation direction (Fritts and Alexander, 2003). This can affect both the waves themselves, for example though critical-level wave filtering, and our observational estimate of them, for example by Doppler-shifting the waves out of the instrument's observational filter.

Doppler-shifting represents the effect of (vertical) wind shear on vertically propagating waves (Ern et al., 2004). For example, positive (westerly) wind shear will Doppler-shift easterly travelling waves to lower vertical wavenumbers, m. Conversely, negative (vertical) wind shear will move easterly waves to higher m and toward critical levels. Secondly, the wind can Doppler-shift the waves out of the spectral region observable by HIRDLS. Since gravity waves are inherently broadband phenomena, this can give rise to apparent geographic, seasonal and vertical variations in observed gravity wave effects, without such variations existing (Preusse et al., 2006; Alexander, 1998). Finally, critical-level filtering represents the obliteration of waves as the wave phase speed approaches that of the background flow.

High-latitude tropospheric winds during winter are predominantly either low or slightly westerly (see, for example, the bottom rows of figures 7.5 and 7.6), and hence any filtering of the gravity wave directional spectrum which takes place in the troposphere acts to reduce the proportion of westerly-propagating waves. This leads to an bias in the directional spectrum (relative to the ground) of waves propagating vertically through the stratosphere in favour of easterly propagating (westward) waves. In a typical Arctic winter, such as those for 2004/05 and 2006/07 shown in figure 7.3 and figure 7.5, the prevailing wind direction in the lower stratosphere is also westerly. Since this is the same as the direction in which the tropospheric winds act, no significant additional filtering takes place in the lower stratosphere, leading to the relatively high momentum flux values.

In 2006, however, a significant region of easterly wind exists from the 100 hPa level and upwards during late January and early February, which functions to filter out gravity waves with easterly propagation directions above and beyond the usual westerly filtering taking place below. Since gravity waves propagating in both directions are now being filtered out, this leads to a significant reduction in the overall amount of momentum flux reaching middle and upper stratospheric heights, precisely what we observe in our momentum flux results.

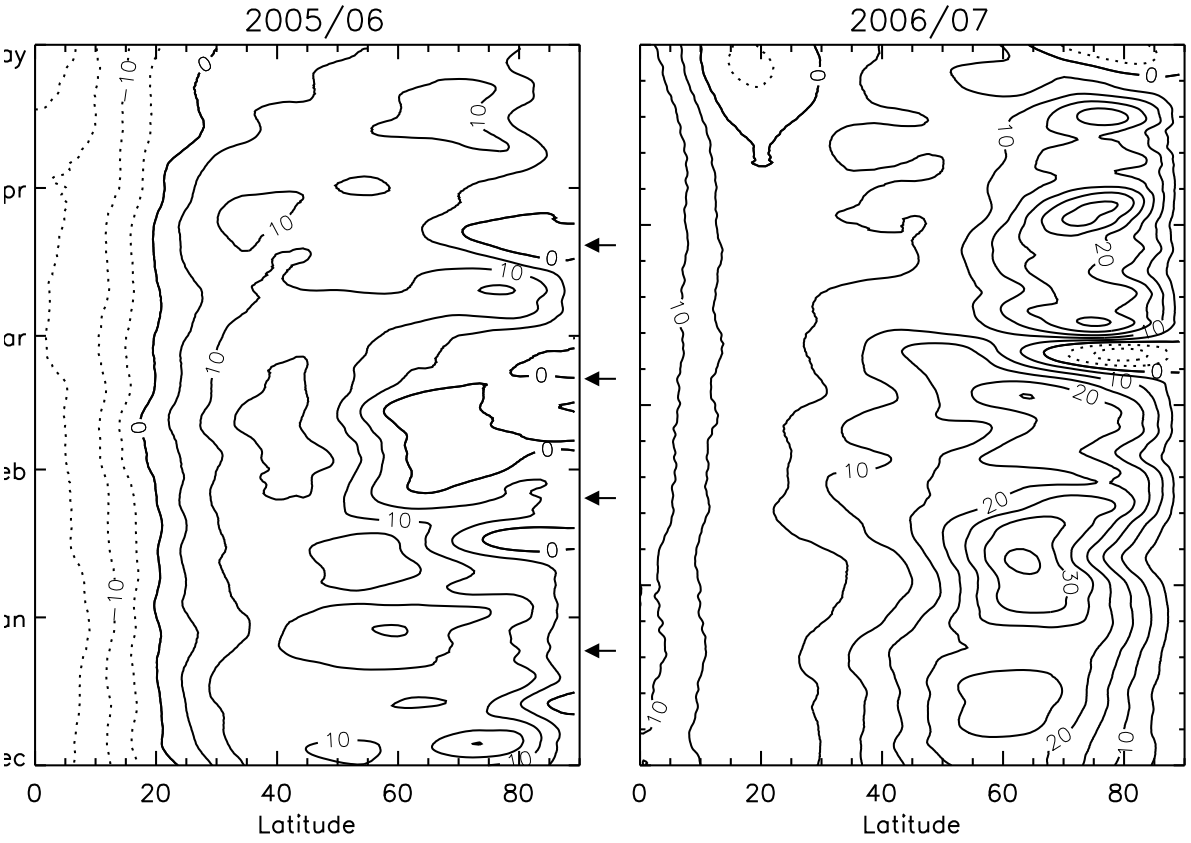


Figure 7.7: *Latitude-time series of Northern Hemispheric ECMWF zonal winds at 50hPa during Arctic winter 2005/2006 (left) and Arctic winter 2006/2007 (right). Negative windspeed contours are dashed.*

Figure 7.7 shows latitude-time series at 50 hPa for $\overline{u}$, obtained from ECMWF operational analyses. 2006/07 shows a fairly typical Arctic winter. Winds poleward of 60N are strong and zonal throughout the winter, slowly decreasing in strength and geographical extent as spring and summer approach. Markedly different behaviour is seen in 2005/06. In December and early January, zonal winds throughout the northern hemisphere are significantly weaker than in 2007, and wind speeds remain below 20 ms^{-1} throughout the entire period shown. Negative zonal winds start to intrude from the pole in late January, reaching as far southwards as 60N by early February, consistent with figure 7.3. This also coincides with the beginning of the minimum of measured momentum fluxes at 10 hPa, (figure 7.4), suggesting a relationship between the low and negative zonal winds and the reduced momentum fluxes recorded.

This is further corroborated by figure 7.8, which shows zonal mean momentum fluxes at 60N as a function of height and time over Arctic winter 2005/2006, with zonal mean zonal winds overplotted. As seen in figure 7.7, wind speeds fall sharply towards the end of January and do not recover until early March, remaining weak, especially in the lower stratosphere. During this time, levels of momentum flux decrease markedly and remain low in the low-mid stratosphere.

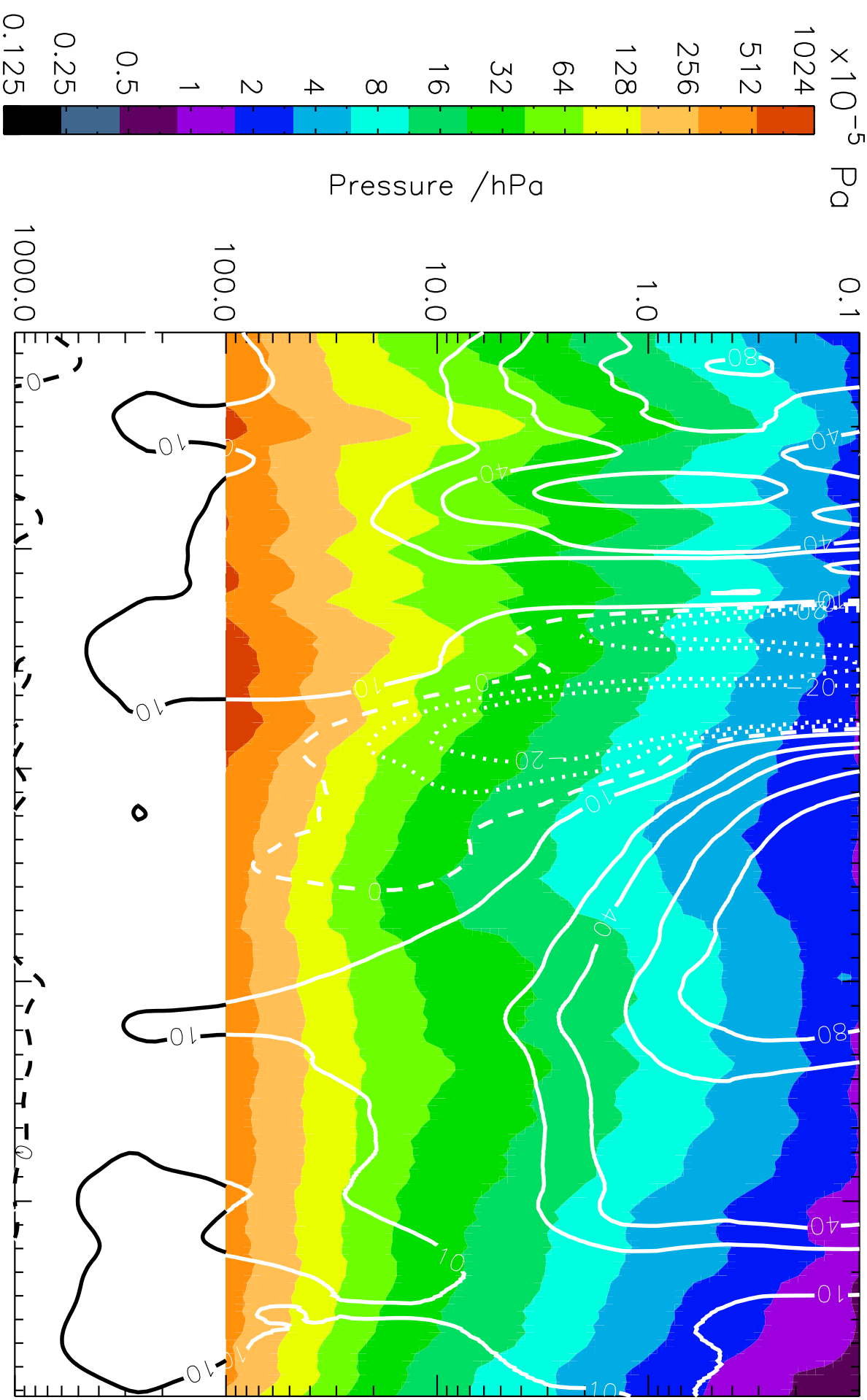


Figure 7.8: *Height-time series of zonal mean momentum flux at 60N during Arctic winter 2005/06, with ECMWF zonal mean zonal winds overlaid. Wind contours are marked at -20,-10, 0, 10, 20, 40, 60 and 80 ms^{-1}. Negative contours are marked as dots, the zero contour as dashes, and positive contours as solid lines.*

This further supports our hypothesis that anomalous easterlies are associated with the reduced momentum flux observed by HIRDLS during this time.

By late February, these easterly winds begin to disappear, although they continue to some degree until late March. Momentum flux values start to return towards more typical values, plateauing after a brief rise to join by late March into the general seasonal trend.

As discussed in section 7.3, there is an increase in momentum flux values at 30-45N in the centre (February 2006) panel of the second row of figure 7.5. If this plot is compared to the latitude-time plot of zonal winds shown in the top panel of figure 7.7, it can be observed that this region is outside the region of strongly reduced zonal winds during February 2006 at 50 hPa. This "bump" in the measured momentum flux field is probably the edge of the normal high momentum flux region, and hence supports our hypothesis of filtering by negative zonal winds in that the region outside the affected area is more normal.

Finally, figure 7.9 shows monthly-mean polar stereographic plots of momentum flux for February 2005, 2006 and 2007, with zonal wind speeds overlaid. This illustrates several significant points. Firstly, reduced momentum flux levels are observed in all locations with a strong signal in 2005 and 2007 throughout the entire Arctic region: that is to say, levels in 2006 are now low throughout the annular region surrounding the Pole. This contrasts with 2005 and 2007, where momentum fluxes are both generally higher and geographically more localised. For example, a maximum in momentum flux occurs around the southern tip of Greenland in 2005 — a known region for orographic gravity waves (see e.g. Leutbacher and Volkert, 2000a) — but this region exhibits the same low values as all other longitudes in 2006. Secondly, the region of reduced momentum flux in 2006 overlaps significantly with the region of easterly and reduced westerly winds. Mean zonal winds for the month are below $10\,\mathrm{ms}^{-1}$ throughout the Arctic, with regions of easterly wind overlying Siberia, Greenland and Northern Canada, whilst the main regions of low measured momentum flux lie at similar latitudes over all longitudes.

It should be noted that reduced planetary wave activity may also have influenced the polar stratopause during these times: planetary wave amplitudes in the stratosphere were also significantly lower than normal during this period (Manney et al., 2008b). We provide evidence to support the idea that reduced gravity wave activity may also have played a role.

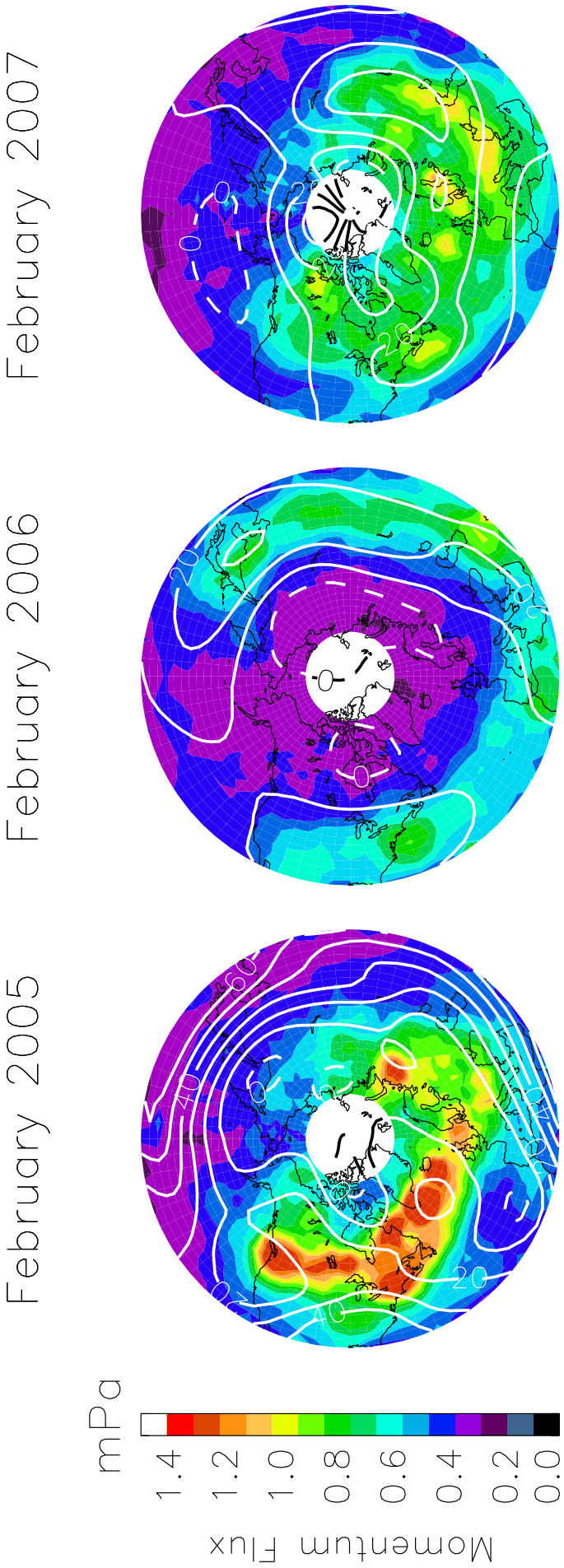


Figure 7.9: *Polar stereographic monthly means of momentum flux at 10hPa for February 2005, 2006 and 2007, with ECMWF-derived zonal winds at 50hPa overplotted. Note the extended region of negative zonal winds at high latitudes in 2006, in contrast to 2005 and 2007.*

7.6 Conclusions

HIRDLS data provide us with an opportunity to directly observe small vertical-scale gravity wave activity on a global scale, allowing us to make detailed estimates of momentum flux activity due to gravity waves throughout the middle atmosphere. This allows us to investigate the underlying gravity wave processes taking place during and after the anomalous change in stratopause height observed around the time of the Sudden Stratospheric Warming of January 2006. We examine the hypothesis of Siskind et al. (2007) that the behaviour observed during this period was due to anomalously low gravity wave activity. Our observations suggest that the gravity wave momentum flux was significantly reduced during this period at latitudes above 50N, starting in mid-January, and remained low throughout the remainder of January and February before resuming the general seasonal trend in mid-to-late March. Furthermore, this trend coincided with the stratopause height variations observed by Manney et al. (2008b).

Analysis of our observations in combination with ECMWF operational analyses of zonal wind speed provide evidence of significant filtering by the background wind of the gravity wave spectrum during this period, contributing to the reduced gravity wave activity observed. We suggest that anomalous easterly and weak westerly winds in the lower stratosphere filter the gravity wave spectrum, reduced the intensity of easterly-propagating gravity waves which reach the middle stratosphere, and hence led to a significant reduction in the overall gravity wave spectrum reaching the middle atmosphere.

Gravity wave driving is believed to exert significant control on mesospheric flow (Holton, 1983). Due to the dynamical nature of the polar winter stratopause (Hitchman et al., 1989), any change in gravity wave activity, such as that observed in Arctic winter 2006, can exert a significant effect on the temperature structure of the atmosphere in this region. As described in Siskind et al. (2007), this further helps to confirm our understanding that the polar winter stratopause is sensitive to gravity wave driving.

Chapter 8

Summary, Conclusions and Future Work

8.1 Summary

In this thesis, we have shown how HIRDLS can be used to study internal gravity waves in the terrestrial stratosphere.

We began in chapter 2 by discussing the background and design of the HIRDLS mission, along with a brief description of the sensing methods used to compute atmospheric temperature from HIRDLS radiance observations. We also discussed the post-launch optical obscuration suffered by the instrument and the methods used to accomodate for this.

Chapter 3 outlined the physical basis behind internal gravity waves. Key gravity wave sources in the terrestrial atmosphere, including mountains and convection, were discussed, together with a brief description of previous studies. We also derived the expected behaviour of the waves, to allow us to better analyse our results.

Chapters 4 and 5 were then used to describe the full details of the detection method used. The Stockwell Transform, used previously for gravity wave studies such as Alexander et al. (2008) and Stockwell and Lowe (2001), was derived, and the application of this method to HIRDLS temperature data discussed in detail, together with limitations on the quality and accuray of the results obtained.

Chapter 6 described validation of the dataset, beginning with a simple geographical comparison of the results obtained to expected gravity wave climatologies. The fine-scale structure of the HIRDLS temperature data were then compared to a range of atmospheric remote-sounding instruments and to ECMWF assimilated stratospheric analyses, and simple S-Transform analy-

sis was applied to these instruments to analyse the performance of this detection method when allowing for the level of similarity between the datasets obtained from the previous analyses.

Finally, chapter 7 described a case study undertaken with the HIRDLS gravity wave dataset on an unusual elevation of the Arctic stratopause during winter 2005/06. In this study, it was shown that gravity wave activity was significantly reduced during this period, corresponding with model-based predictions described in Siskind et al. (2007). It was further shown that this reduction corresponded to a period of strongly reduced zonal winds, providing a possible mechanism for the reduction in gravity wave momentum flux.

8.2 Conclusions

The high resolution and wide global coverage of HIRDLS provides an excellent tool for the detection and characterisation of gravity waves in the terrestrial stratosphere. Using this dataset, we can study these effects and their interaction with other climate processes, such as the anomalous winds during Arctic winter 2005/06. Further study of these and other data may help to further elucidate important gravity-wave related climate processes.

8.3 Future Work

The single case study described in this thesis is just one example of the results which can be obtained from detailed study of the gravity wave data obtained from HIRDLS analyses. Further study of the dataset, particularly in combination with meteorological datasets, should lead to significant new information on the origin and propagation of gravity waves.

8.3.1 Weather Fronts

Weather fronts are a significant source of convectively-generated gravity waves (Fritts and Alexander, 2003), and upward propagation of these waves should produce a measurable effect on stratospheric momentum fluxes. Accordingly, study of HIRDLS gravity wave data above frontal patterns, by contrast with periods without such fronts, should provide interesting results.

8.3.2 Typhoons and the Monsoon

Along with weather fronts, typhoons and the monsoon provide significant tropospheric convective sources which should produce a significant gravity wave signal. This has been previously shown on localised scales by studies such as that of Dutta et al. (2008), and evidence of such effects can be clearly seen in the HIRDLS data - typhoons have previously been mentioned in section 6.2, and clear peaks in momentum flux maps can be seen at subtropical latitudes over Central America, Africa and India at times corresponding with the monsoons of these regions. Further study to tie this to meteorological patterns in these regions should elucidate these links further.

8.3.3 Polar Mesospheric Clouds

Innis et al. (2008) discuss measurements undertaken at Davis, Antarctica, on studies of stratospheric gravity wave intensity in proximity to Polar Meospheric Clouds (PMCs). This study discusses results from a single Rayleigh LIDAR station in Antarctica, and compares them to similar measurements taken over Greenland; studies have suggested a link between strong stratospheric gravity wave activity and reduced PMC strengths. Although HIRDLS coverage does not extend over Antarctica, coverage over Greenland is excellent, and accordingly it may be possible to examine this relationship on a more significant geographical and temporal scale.

Bibliography

Aguilar, D. A. and B. R. Sutherland, 2006: Internal wave generation from rough topography. *Physics of Fluids*, **18**, 066603 1–9.

Alexander, M. J., 1998: Interpretations of observed climatological patterns in stratospheric gravity wave variance. *Journal of Geophysical Research*, **103**, 8627–8640.

Alexander, M. J. and C. Barnet, 2007: Using Satellite Observations to Constrain Parameriza-tions of Gravity Wave Effects for Global Models. *Journal of the Atmospheric Sciences*, **64**, 1652–1665.

Alexander, M. J., J. Gille, C. Cavanaugh, M. Coffey, C. Craig, V. Dean, T. Eden, G. Francis, C. Halvorson, J. Hannigan, R. Khosravi, D. Kinneson, H. Lee, S. Massie, B. Nardi, and A. Lambert, 2008: Global estimates of gravity wave momentum flux from High Resolution Dynamics Limb Sounder (HIRDLS) observations. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD008807.

Alexander, M. J., P. T. May, and J. H. Beres, 2004: Gravity waves generated by convection in the darwin area during the darwin area wave experiment. *Journal of Geophysical Research*, **109**, doi:10.1029/2004JD004729.

Alexander, M. J. and H. Teitelbaum, 2007: Observation and analysis of a large amplitude mountain wave event over the antarctic peninsula. *Journal of Geophysical Research*, **112**, doi:10.10292006JD008368.

Allen, S. J. and R. A. Vincent, 1995: Gravity wave activity in the lower atmosphere: seasonal and latitudinal variations. *Journal of Geophysical Research*, **100**, 1327–1350.

Andrews, D. G., 2000: *An Introduction to Atmospheric Physics*. Cambridge University Press.

Bacmeister, J. T., P. A. Newman, B. L. Gary, and K. R. Chan, 1994: An Algorithm for Forecasting Mountain Wave-Related Turbulence in the Stratosphere. *Weather and Forecasting*, **9**, 241–253.

Barnett, J. J., J. N. Bracken, K. Djotni, C. L. Hepplewhite, J. L. Moorhouse, O. O. Oduleye, C. W. P. Palmer, D. M. Peters, L. A. Rokke, T. W. Walton, R. E. J. Watkins, J. G. Whitney, J. C. Gille, P. I. Arter, T. D. Eden, Jr., and B. Nardi, 2003: Pre-Launch Calibration of the NASA Aura HIRDLS Instrument. *Proceedings of the SPIE*, **5152**, 172–180.

Barnett, J. J. and J. C. Gille, 2006a: High Resolution Dynamics Limb Sounder Earth Observing System (EOS): Data Description and Quality Version 2.00.

Barnett, J. J. and J. C. Gille: 2006b, HIRDLS Overview and Status. *HIRDLS Science Team Meeting, September 2006*.

Barnett, J. J. and J. C. Gille, 2008: High Resolution Dynamics Limb Sounder Earth Observing System (EOS): Data Description and Quality Version 4.00.

Barnett, J. J., C. L. Hepplewhite, S. Osprey, J. C. Gille, and R. Khosravi, 2008: Cross-validation of HIRDLS and COSMIC radio-occultation retrievals, particularly in relation to fine vertical structure. *Proceedings of the SPIE*, **7082**, 16 – 16–7, doi:10.1117/12.800702.

Barnett, J. J., C. L. Hepplewhite, L. Rokke, and J. C. Gille, 2005: Mapping the optical obscuration in the NASA Aura HIRDLS instrument. *Proceedings of the SPIE*, **5883**, I1 – I10.

Carslaw, K. S., M. Wirth, A. Tsias, B. P. Luo, A. Dörnbrack, M. Leutbecher, H. Volkert, W. Renger, J. T. Bacmeister, E. Reimer, and T. Peter, 1998: Increased stratospheric ozone depletion due to mountain-induced atmospheric waves. *Nature*, **391**, 675–678.

Clark, T. L., T. Hauf, and J. P. Kuettner, 1986: Convectively forced internal gravity waves: Results from two-dimensional numerical experiments. *Quarterly Journal of the Royal Meteorological Society*, **112**, 899–925.

Dunkerton, T., 1978: On the mean meriodional mass motions of the stratosphere and mesosphere. *Journal of the Atmospheric Sciences*, **35**, 2325–2333.

Dutta, G., T. Tsuda, P. V. Kumar, M. C. A. Kumar, S. P. Alexander, and T. Koza, 2008: Seasonal variation of short-period (less than 2 h) gravity wave activity over Gadanki, India. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD009178.

Eckermann, S. D. and P. Preusse, 1999: Global Measurements of Stratospheric Mountain Waves from Space. *Science*, **286**, 1534–1537.

Eckermann, S. D., D. L. Wu, J. D. Doyle, J. F. Burris, T. J. McGee, C. A. Hostetler, L. Coy, B. N. Lawrence, A. Stephens, J. P. McCormack, and T. F. Hogan, 2006: Imaging gravity waves in lower stratospheric amsu-a radiances, part 2: Validation case study. *Atmospheric Chemistry and Physics*, **6**, 3343–3362.
`http://www.atmos-chem-phys.net/6/3343/2006/`

Eden, T. D., Jr., J. C. Gille, J. J. Barnett, C. L. Hepplewhite, C. W. P. Palmer, and J. G. Whitney, 2005: HIRDLS Proto-flight Model Radiometric Characterization From Pre-launch Calibration Data. *Proceedings of the SPIE*, **5883**, K1–K13.

Elachi, C. and J. Van Zyl, 2006: *Introduction to the Physics and Techniques of Remote Sensing*. Wiley-Interscience.

Ern, M., P. Preusse, M. Alexander, and C. D. Warner, 2004: Absolute values of gravity wave momentum flux derived from satellite data. *Journal of Geophysical Research*, **109**, doi:10.1029/2004JD004752.

Fetzer, E. J. and J. C. Gille, 1994: Gravity Wave Variance in LIMS Temperatures. part I: Variability and Comparison with Background Winds. *Journal of the Atmospheric Sciences*, **51**, 2461–2483.

Fischer, H., C. Blom, H. Oelhaf, B. Carli, M. Carlotti, L. Delbouille, D. Ehhalt, J. Flaud, I. Isaksen, M. Lopez-Puertas, C. McElroy, and R. Zander, 2000: ENVISAT-MIPAS, the Michelson Interferometer for Passive Atmospheric Sounding; An instrument for atmospheric chemistry and Climate Research ENVISAT Mipas An instrument for Atmospheric Chemistry and Climate Research.

Fritts, D. C., 1984: Gravity Wave Saturation in the Middle Atmosphere: A Review of Theory and Observations . *Reviews of Geophysics*, **22**, 275–308.

Fritts, D. C. and M. J. Alexander, 2003: Gravity Wave Dynamics and Effects in the Middle Atmosphere. *Reviews of Geophysics*, **41**, doi:10.1029/2001RG000106.

Fritts, D. C. and Z. Luo, 1992: Gravity Wave Excitation by Geostrophic Adjustment of the Jet Stream. Part I: Two-Dimensional Forcing. *Journal of the Atmospheric Sciences*, **49**, 681–697.

Gille, J., J. Barnett, P. Arter, M. Barker, P. Bernath, C. Boone, C. Cavanaugh, J. Chow, M. Coffey, J. Craft, C. Craig, M. Dials, V. Dean, T. Eden, D. P. Edwards, G. Francis, C. Halvorson, L. Harvey, C. Hepplewhite, R. Khosravi, D. Kinnison, C. Krinsky, A. Lambert, H. Lee, L. Lyjak, J. Loh, W. Mankin, S. Massie, J. McInerney, M. J., B. Nardi, D. Packman, C. Randall, J. Reburn, W. Rudolf, M. Schwartz, J. Serafin, K. Stone, B. Torpy, K. Walker, A. Waterfall, R. Watkins, J. Whitney, D. Woodard, and G. Young, 2008: High Resolution Dynamics Limb Sounder: Experiment overview, recovery and validation of initial temperature. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD008824.

Gille, J., T. Eden, G. Francis, A. Lambert, B. Nardi, J. Barnett, C. Cavanaugh, H. Lee, C. Craig, V. Dean, C. Halvorson, C. Krinsky, J. McInerney, and B. Petersen, 2005: Development of Special Corrective Processing of HIRDLS Data, and Early Validation. *Proceedings of the SPIE*, **5883**, H1–H11.

Gille, J. C., J. J. Barnett, J. Whitney, M. Dials, D. Woodard, W. Rudolf, A. Lambert, and W. Mankin, 2003: The High Resolution Dynamics Limb Sounder (HIRDLS) Experiment on Aura. *Proceedings of the SPIE*, **5152**, 162–171.

Gordley, L. L., M. E. Hervig, C. Fish, J. M. R. III, S. Bailey, J. Cook, S. Hansen, A. Shumway, G. Paxton, L. Deaver, T. Marshall, J. Burton, B. Magill, C. Brown, E. Thompson, and J. Kemp, 2009: The Solar Occultation For Ice Experiment. *Journal of Atmospheric and Solar-Terrestrial Physics*, **71**, 300 – 315, doi:DOI: 10.1016/j.jastp.2008.07.012, global Perspectives on the Aeronomy of the Summer Mesopause Region, Eighth International Workshop on Layered Phenomena in the Mesopause Region.
http://www.sciencedirect.com/science/article/B6VHB-4T4Y620-1/2/450a1606d6b835e3c8a3d28bff2ee434

Guest, F. M., M. J. Reeder, C. J. Marks, and D. J. Karoly, 2000: Inertia-Gravity Waves Observed

in the Lower Stratosphere over Macquarie Island. *Journal of the Atmospheric Sciences*, **57**, 737–752.

Hamilton, K., 1999: The Gravity Wave Parameterisation Problem for Global Simulation Models. *SPARC Newsletter 12*.

Hepplewhite, C. L., J. J. Barnett, J. G. Whitney, C. W. P. Palmer, O. O. Oduleye, T. W. Walton, M. Dials, J. G. Gille, T. Eden, Jr., and B. Nardi, 2005: HIRDLS Functional Performance in Orbit - A Summary. *Proceedings of the SPIE*, **5883**, J1–J10.

Herzog, A., C. Souprayen, and A. HauchCorne, 2001: Observation and Backward Trajectory of an Inertio-Gravity Wave in the Lower Stratosphere. *Annales Geophysicae*, **19**, 1141–1155.

Hitchman, M. H., J. C. Gille, C. D. Rodgers, and G. Brasseur, 1989: The separated polar winter stratopause: A gravity wave driven climatological feature. *Journal of the Atmospheric Sciences*, **46**, 410–422.

Hoffmann, P., W. Singer, D. Keuer, W. K. Hocking, M. Kunze, and Y. Murayama, 2007: Latitudinal and longitudinal variability of mesospheric winds and temperatures during stratospheric warming events. *Journal of Atmospheric and Solar-Terrestrial Physics*, **69**, 2355–2466.

Holton, J. R., 1979: *An Introduction to Dynamic Meteorology*. Academic Press Inc., 2nd edition.

— 1983: The influence of gravity wave breaking on the general circulation of the middle atmosphere. *Journal of the Atmospheric Sciences*, **40**, 2497–2507.

Houghton, J. T., 1977: *The Physics of Atmospheres*. Cambridge University Press.

Houghton, J. T., F. W. Taylor, and C. D. Rodgers, 1984: *Remote Sounding of Atmospheres*. Cambridge Planetary Science Series.

Innis, J. L., A. R. Klekociuk, R. J. Morris, A. P. Cunningham, A. D. Graham, and D. J. Murphy, 2008: A study of the relationship between stratospheric gravity waves and polar mesospheric clouds at davis antarctica. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD009031.

Jiang, J. H., D. L. Wu, S. D. Eckermann, and J. Ma, 2003: Mountain Waves in the Middle Atmosphere: Microwave Limb Sounder Observations and Analyses. *Advances in Space Research*, **32**, 801–806.

Jones, W. L. and D. D. Houghton, 1972: The Self-Destructing Internal Gravity Wave. *Journal of the Atmospheric Sciences*, **29**, 844–849.

Karoly, D. J., G. L. Roof, and M. J. Reeder, 1996: Gravity wave activity association with tropical convection detected in TOGA COARE sounding data. *Geophysical Research Letters*, **23**, 261–264.

Khosravi, R., A. Lambert, H. Lee, J. Gille, J. Barnett, G. Francis, D. Edwards, C. Halvorson, S. Massie, C. Craig, C. Krinsky, J. McInerney, K. Stone, T. Eden, B. Nardi, C. Hepplewhite, W. Mankin, and M. Coffey, 2009: Overview and characterization of retrievals of temperature, pressure,and atmospheric constituents from the High Resolution Dynamics Limb Sounder (HIRDLS) measurements. *Journal of Geophysical Research*, **114**, doi:doi:10.1029/2009JD011937.

Kinnison, D. E., J. Gille, J. Barnett, C. Randall, L. Harvey, A. Lambert, R. Khosravi, M. J. Alexander, P. Bernath, C. Boone, C. Cavanaugh, M. Coffey, C. Craig, V. Dean, T. Eden, D. Ellis, D. Fahey, G. Francis, C. Halvorson, J. Hannigan, C. Hartsough, C. Hepplewhite, C. Krinsky, H. Lee, B. Mankin, T. Marcy, S. Massie, B. Nardi, D. Packman, P. J. Popp, M. L. Santee, V. Yudin, and K. A. Walker, 2008: Global observations of HNO3 from the High Resolution Dynamics Limb Sounder (HIRDLS): First results. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD008814.

Kuettner, J. P., P. A. Hildebrand, and T. L. Clark, 1987: Convection waves: Observations of gravity wave systems over convectively active boundary layers. *Quarterly Journal of the Royal Meteorological Society*, **113**, 445–467.

Lambert, A., P. L. Bailey, D. P. Edwards, J. C. Gille, B. R. Johnson, C. M. Halvorson, S. T. Massie, and K. A. Stone, 1999: HIRDLS Level-2 Algorithm Theoretical Basis Document.

Lane, T. P., M. J. Reeder, and T. L. Clark, 2001: Numerical Modelling of Gravity Wave Generation by Deep Tropical Convection. *Journal of the Atmospheric Sciences*, **58**, 1249–1274.

Leutbacher, M. and H. Volkert, 2000a: The propagation of mountain waves into the stratosphere: Quantitative evaluation of three-dimensional simulations. *Journal of the Atmospheric Sciences*, **57**, 3090–3108.

— 2000b: The Propagation of Mountain Waves into the Stratosphere: Quantitative Evaluation of Three-Dimensional Simulations. *Journal of the Atmospheric Sciences*, **57**, 3090–3108.

Lindzen, R. S., 1973: Wave-Mean Flow Interactions in the Upper Atmosphere. *Boundary Layer Meteorology*, **4**, 327–343.

Liou, Y. A., A. G. Pavelyev, S. F. Liu, A. Pavelyev, N. Yen, C. Y. Huang, and C. J. Fong, 2007: FORMOSAT-3/COSMIC GPS radio occulation mission: Preliminary results. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 3813–3826.

Luo, Z. and D. C. Fritts, 1993: Gravity Wave Excitation by Geostrophic Adjustment of the Jet Stream. Part II: Three-Dimensional Forcing. *Journal of the Atmospheric Sciences*, **50**, 104–115.

Lynch, A. H. and J. J. Cassano, 2006: *Applied Atmospheric Dynamics*. Wiley.

Manney, G. L., W. H. Daffer, K. B. Strawbridge, K. A. Walker, C. D. Boone, P. F. Bernath, T. Kerzenmacher, M. J. Schwartz, K. Strong, R. J. Sica, K. Krger, H. C. Pumphrey, A. Lambert, M. L. Santee, N. J. Livesey, E. E. Remsberg, M. G. Mlynczak, and J. R. Russell III, 2008a: The high arctic in extreme winters: vortex, temperature, and MLS and ACE-FTS trace gas evolution. *Atmospheric Chemistry and Physics*, **8**, 505–522.
`http://www.atmos-chem-phys.net/8/505/2008/`

Manney, G. L., K. Krüger, S. Pawson, K. Minschwaner, M. J. Schwartz, W. H. Daffer, N. J. Livesey, M. G. Mlynczak, E. E. Remsberg, J. M. Russell III, and J. W. Waters, 2008b: The evolution of the stratopause during the 2006 major warming: Satellite data and assimilated meteorological analyses. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD009097.

McLandress, C., M. J. Alexander, and D. L. Wu, 2000: Microwave Limb Sounder observations of gravity waves in the stratosphere: A climatology and interpretation. *Journal of Geophysical Research*, **105**, 11947–11967.

Meriweather, J. W. and A. J. Gerrard, 2004: Mesosphere Inversion layers and Stratosphere Temperature Enhancements. *Reviews of Geophysics*, **42**.

Millán Valle, L., 2007: Mesospheric studies with MIPAS and HIRDLS. Transfer of Status Report, University of Oxford (Internal Document).

Moorhouse, J. L., 2003: *Calibration of the Field of View of the High Resolution Dynamics Limb Sounder*. D.Phil. thesis, University of Oxford.

Murgatroyd, R. J. and F. Singleton, 1961: Possible meridional circulations in the stratosphere and mesosphere. *Quarterly Journal of the Royal Meteorological Society*, **87**, 125–135.

Nappo, C. J., 2002: *An Introduction to Atmospheric Gravity Waves*. International Geophysics Series.

Nardi, B., J. Gille, J. Barnett, C. Randall, V. L. Harvey, A. Waterfall, W. J. Reburn, T. Leblanc, T. McGee, L. Twigg, A. Thompson, S. Godin-Beemann, P. Bernath, B. Bojkov, C. Boone, C. Cavanaugh, M. Coffey, J. Craft, C. Craig, V. Dean, T. Eden, G. Francis, L. Froidevaux, C. Halvorson, J. Hannigan, C. Hepplewhite, D. Kinnison, R. Khosravi, C. Krinsky, A. Lambert, H. Lee, J. Loh, S. Massie, I. S. MacDermid, D. Packman, B. Torpy, J. Valverde-Canossa, K. Walker, D. Whiteman, J. Witte, and G. Young, 2008: Initial validation of ozone measurements from the High Resolution Dynamics Limb Sounder. *Journal of Geophysical Research*, **113**, doi:10.1029/2007JD008837.

Osprey, S. M., 2002: *Climate Response to a Variable Gravity-Wave Source*. Ph.D. thesis, University of Canterbury.

Pfister, L., K. R. Chan, T. P. Bui, S. Bowen, M. Legg, B. Gary, K. Kelly, M. Proffitt, and W. Starr, 1993: Gravity Waves Generated by a Tropical Cyclone During the Tropical Field Program: A Case Study. *Journal of Geophysical Research*, **98**, 8611–8638.

Piani, C., D. Durran, M. J. Alexander, and J. R. Holton, 2000: A Numerical Study of Three-Dimensional Gravity Waves Triggered by Deep Tropical Convection and Their Role in the Dynamics of the QBO. *Journal of the Atmospheric Sciences*, **57**, 3689–3702.

Preusse, P., M. Ern, S. D. Eckermann, C. D. Warner, R. H. Picard, P. Knieling, M. Krebsbach, J. M. R. III, M. G. Mlynczak, C. J. Mertens, and M. Riese, 2006: Tropopause to mesopause gravity waves in August: Measurement and modeling. *Journal of Atmospheric and Solar-Terrestrial Physics*, **68**, 1730 – 1751, doi:10.1016/j.jastp.2005.10.019.

Rayleigh, L., 1883: Investigation of the Character of the Equilibrium of an Incompressible Heavy Fluid of Variable Density. *Proceedings of the London Mathematical Society*, **14**, 170–177.

Reeder, M. J. and M. Griffiths, 1996: Stratospheric inertia-gravity waves generated in a numerical model of frontogenesis II: Wave sources, generation mechanisms and momentum fluxes. *Quarterly Journal of the Royal Meteorological Society*, **122**, 1175–1195.

Rocken, C., Y.-H. Kuo, W. Schreiner, D. Hunt, and S. Sokolovskiy, 2000: Cosmic system description. *Terr. Atmos. Oceanic Sci.*, **11**, 21–52.

Rong, P. P., J. M. Russell III, M. G. Mlynczak, E. E. Remsberg, B. T. Marshall, L. L. Gordley, and M. Lopez-Puertas, 2009: Validation of thermosphere ionosphere mesosphere energetics and dynamics/sounding of the atmosphere using broadband emission radiometry (timed/saber) v1.07 ozone at 9.6 mm in altitude range 1570 km. *Journal of Geophysical Research*, **114**, doi:10.1029/2008JD010073.

Russell III, J. M., 1980: Satellite solar occultation sounding of the middle atmosphere. *Pure and Applied Geophysics*, **118**, 616–635.

Russell III, J. M., M. G. Mlynczak, and L. L. Gordley, 1994: Overview of the sounding of the atmosphere using broadband emission radiometry (saber) experiment for the thermosphere-ionsphere-mesosphere energetics and dynamics (timed) mission.
`http://link.aip.org/link/?PSI/2266/406/1`

Salby, M. L., 1996: *Fundamentals of Atmospheric Physics*. Academic Press.

Sato, K. and T. J. Dunkerton, 1997: Estimates of momentum flux associated with equatorial Kelvin and gravity waves. *Journal of Geophysical Research*, **102**, 26247–26261.

Schoeberl, S. R., E. Douglass, A. R. Hilsenrath, P. K. Bhartia, R. Beer, J. W. Waters, M. R. Gunson, L. Froidevaux, J. C. Gille, J. J. Barnett, P. F. Levelt, and P. DeCola, 2006: Overview of the EOS Aura Mission. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1066–1074.

Schreiner, B., C. Rocken, D. Hunt, S. Sokolovskiy, and S. Syndergaard, 2008: Cosmic: Constellation observing system for meteorology, ionosphere and climate: Overview, status, and results. HIRDLS Science Team Presentation.

Siskind, D. E., S. D. Eckermann, L. Coy, J. P. McCormack, and C. E. Randall, 2007: On Recent Interannual Variability of the Arctic Winter Mesosphere: Implications for Tracer Descent. *Geophysical Research Letters*, **34**, 1–5.

Smith, A. K., 1996: Longitudinal variations in mesospheric winds: Evidence for gravity wave filtering by planetary waves. *Journal of the Atmospheric Sciences*, **53**, 1156–1173.

Staquet, C. and J. Sommeria, 2002: Internal Gravity Waves - from Instabilities to Turbulence. *Annual Review of Fluid Mechanics*, **34**, 559–593.

Stockwell, R. G., 1999: *S-Transform Analysis of Gravity Wave Activity from a Small Scale Network of Airglow Imagers*. Ph.D. thesis, University of Western Ontario.

Stockwell, R. G. and R. P. Lowe, 2001: Airglow imaging of gravity waves 1. results from a small network of OH nightglow scanning imagers. *Journal of Geophysical Research*, **106**, 17185–17203.

Stockwell, R. G., L. Mansinha, and R. P. Lowe, 1996: Localization of the Complex Spectrum: The S Transform. *IEEE Transactions on Signal Processing*, **44**, 998–1001.

Ungermann, J., L. Hoffmann, P. Preusse, M. Kaufmann, and M. Riese, 2009: Tomographic retrieval approach for mesoscale gravity wave observations by the premier infrared limb-sounder. *Atmospheric Measurement Techniques Discussions*, **2**, 2809–2850. `http://www.atmos-meas-tech-discuss.net/2/2809/2009/`

Vincent, R. A. and M. J. Alexander, 2000: Gravity waves in the tropical lower stratosphere: An

observational study of seasonal and interannual variability. *Journal of Geophysical Research*, **105**, 17971–17982.

Wang, L., M. J. Alexander, T. Bui, and M. Mahoney, 2006: Small-scale gravity waves in ER-2 MMS/MTP wind and temperature measurements during CRYSTAL-FACE. *Atmos. Chem. Phys.*, **6**, 1091–1104.

Waymark, C., 2009: *Joint Validation of HIRDLS and TES*. D.Phil. thesis, University of Oxford.

Wrasse, C., J. Fechine, H. Takahashi, C. Denardini, J. Wickert, M. Mlynczak, J. Russell, and C. Barbosa, 2008: Temperature comparison between champ radio occultation and timed/saber measurements in the lower stratosphere. *Advances in Space Research*, **41**, 1423 – 1428, doi:DOI: 10.1016/j.asr.2007.06.073.
`http://www.sciencedirect.com/science/article/B6V3S-4P5NX7S-3/2/bb4afe2154ef604a6d781bb4bb441ac4`

Wu, D. L. and J. W. Waters, 1997: Observations of Gravity Waves with the UARS Microwave Limb Sounder. *Gravity Wave Processes, NATO ASI Series I: Global Environment Change*, **50**, 103–120.

Yoshiki, M. and K. Sato, 2000: A statistical study of gravity waves in the polar regions based on operational radiosonde data. *Journal of Geophysical Research*, **105**, 17995–18011.