Thermal Structure and Gravity Waves in the Antarctic Middle Atmosphere Observed by Lidar

Dissertation

zur

Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Rostock

vorgelegt von:

Bernd Kaifler

aus Gilching

geboren am 03.06.1983 in Ulm

urn:nbn:de:gbv:28-diss2014-0104-4

Gutachter:

1. Gutachter:
   Prof. Dr. Franz-Josef Lübken,
   Leibniz-Institut für Atmosphärenphysik, Universität Rostock

2. Gutachter:
   Prof. Dr. Ulf-Peter Hoppe,
   Universität Tromsø, Norwegen

Datum der Einreichung: 18. September 2013

Datum der Verteidigung: 24. März 2014

Thermal Structure and Gravity Waves in the Antarctic Middle Atmosphere Observed by Lidar

Bernd Kaifler

Leibniz-Institute of Atmospheric Physics e.V. at the University of Rostock

Thesis submitted for the Degree of Doctor rerum naturalium in the University of Rostock

· 2013 ·

Abstract

The IAP iron Doppler lidar has been operated at the Antarctic station Davis (69°S, 78°E) between December 2010 and December 2012. With more than 2600 hours of lidar observations acquired over a period of 16 months, the iron lidar produced the most extensive high resolution dataset available to date. In this thesis new algorithms have been developed to analyze the molecular part of the lidar return signal and obtain temperature profiles between 20 and 60 km altitude. The resulting stratospheric dataset is combined with Doppler temperature measurements in the mesosphere-lower thermosphere (80–100 km). Seasonal variations in the thermal structure above Davis are studied and wave-induced temperature perturbations are analyzed on different time scales. Gravity wave potential energy densities (GWPED) are computed from lidar observations and compared to a simple gravity wave model. In the stratosphere the GWPED shows a clear annual oscillation with a double-peak in winter, while a semi-annual component dominates in the mesosphere. It is shown that observed seasonal variations in GWPED are caused by selective filtering of zonally propagating gravity waves. The altitude range where most waves are filtered is linked to the reversal of the zonal wind which progresses downward from mesospheric altitudes to the stratosphere during breakdown of the polar vortex at the end of winter. Evidence for wind filtering of gravity waves is found in measured GWPED profiles as well as power spectra of wave-induced perturbations.

Zusammenfassung

Das IAP Eisen-Doppler Lidar wurde von Dezember 2010 bis Dezember 2012 auf der Antarktisstation Davis (69°S, 78°E) betrieben. Über einen Zeitraum von 16 Monaten konnten mit diesem Instrument mehr als 2600 Stunden an Messdaten gesammelt werden. Mit dieser bisher unerreichten Anzahl von Messstunden stellen diese Messungen den größten hochaufgelösten Lidardatensatz dar. In dieser Arbeit wurden Algorithmen zur Auswertung des molekularen Anteils des Rückstreusignals entwickelt um Temperaturprofile im Höhenbereich 20–60 km zu erhalten und damit die Eisen-Doppler Temperaturmessungen in der Mesosphäre/unteren Thermosphäre (80–100 km) nach unten hin zu erweitern. Es werden saisonale Änderungen der Temperatur sowie von Wellen verursachte Temperaturfluktuationen auf unterschiedlichen Zeitskalen untersucht. Aus den Messdaten abgeleitete Schwerewellenenergiedichten (gravity wave potential energy density, GWPED) werden mit den aus einem einfachen Schwerewellenmodell erhaltenen Ergebnissen verglichen. Während die GWPED in der Stratosphäre einer jährlichen Oszillation folgt, dominiert in der Mesosphäre eine halbjährliche Komponente. Es wird gezeigt, dass die beobachteten saisonalen Variationen in der GWPED durch selektive Filterung von in zonaler Richtung propagierenden Schwerewellen verursacht wird. Hierbei hängt der Höhenbereich, in dem die Mehrzahl der Wellen gefiltert wird, eng mit der Windumkehr des Zonalwindes zusammen. Mit dem Zusammenbruch des polaren

Wirbels am Ende des Winters wandert dieser Bereich von der Mesosphäre in die untere Stratosphäre. Hinweise auf die Windfilterung von Schwerewellen liefern sowohl die gemessenen GWPED Profile als auch Leistungsspektren von Temperaturfluktuationen.

Contents

1. Introduction **1**

1.1. Earth's atmosphere 1
1.2. Atmosphere dynamics 3
1.3. Lidar – A tool to study the middle atmosphere 4
1.4. A short history of Project Antarctica 7
1.5. The structure of this thesis 10

2. The Fe lidar system **11**

2.1. The Fe lidar 11
2.1.1. Transmitter 12
2.1.2. Receiver 15
2.2. Preparations for Antarctica 17

3. Temperature retrieval **21**

3.1. Description of the raw dataset 21
3.2. Data reduction and quality control 21
3.2.1. Correcting saturation of photon detectors 21
3.2.2. Background estimation and subtraction 27
3.2.3. Correcting atmospheric transmission 28
3.3. Analysis of the Rayleigh signal 29
3.3.1. Separation of Rayleigh scattering and resonance scattering . . . 30
3.3.2. Temperature conversion 32
3.3.3. Correcting etalon transmission 33
3.3.4. Detecting aerosols 37
3.4. Validation with radiosonde data 39

4. Analysis of lidar temperature data **43**

4.1. Description of the dataset 43
4.1.1. Temporal coverage 44
4.2. Mean temperatures 47
4.2.1. Monthly means 47
4.2.2. Mean seasonal temperature variation 49
4.2.3. Comparison with ECMWF data 51
4.2.4. Comparison with observations in the northern hemisphere 53
4.2.5. Spectral analysis of seasonal variations 55

4.3. Temperature perturbations 56
4.3.1. Data analysis 57
4.3.2. Temperature perturbations in the stratosphere 60
4.3.3. Comparison with other sites 63
4.3.4. Seasonal variations 65
4.3.5. Temperature perturbations in the MLT region 68
4.4. Gravity waves 69
4.4.1. Introduction 69
4.4.2. Gravity wave analysis 72
4.4.3. Variability of gravity wave activity 75
4.4.4. Growth of gravity wave amplitudes with height 79
4.4.5. The effect of the background wind field 83
4.4.6. A simple model to investigate the effect of critical layers 91
4.4.7. Comparison with other datasets 95
4.4.8. Gravity waves in the MLT region and vertical coupling 97
4.4.9. Heat transport by gravity waves 104

5. Summary and Outlook **105**

Appendix A. Heat transport by gravity waves **107**
A.1. Data analysis 107
A.2. Results 109

Appendix B. Filtering of gravity waves by tides **115**
B.1. Tides 115
B.2. Data reduction and analysis 116
B.3. Tides in the stratosphere and comparison with MERRA data 116
B.4. Tides in the MLT region 117

Appendix C. Estimated fluorescence signal gain at 386 nm wavelength **125**

Bibliography **127**

List of Figures **137**

List of Tables **139**

Abbreviations **141**

Acknowledgements **143**

1. Introduction

1.1. Earth's atmosphere

Earth's atmosphere is a gas layer surrounding our home planet which shields the surface from the influence of outer space. It provides breathable air, absorbs deadly cosmic radiation, and regulates the surface temperature by balancing the absorption of solar radiation and the amount of energy the Earth is radiating into outer space. Thus, the atmosphere is one of the key ingredients that make the Earth compatible with life. Its important role for sustaining our human presence on Earth is increasingly discussed in our society. In recent years, this new interest is mainly a result of people experiencing personally the consequences of changes in the atmosphere. Possibly the first widely acknowledged effect was the increased probability of getting sun burns as result of reductions in stratospheric ozone. The link between inorganic chlorine and observed ozone losses above Antarctica was first published in 1985 (*Farman et al.*, 1985), and the term *ozone hole* entered common language soon after. A more recent example is the discussion of record high global mean temperatures occurring in the last decade (e.g. *Hansen et al.*, 2010).

With potentially all people on Earth being directly or indirectly affected by changes in the atmosphere, the question of how to react has gained a political dimension by considering and questioning the anthropogenic influence on the climate. Thus, the discussion in society is subject to political decisions, and the basis of these decisions is a scientific understanding of the atmosphere (*Solomon*, 2007). This includes recognizing the atmosphere as complex and highly dynamic system with many coupled physical and chemical processes. These processes and interactions do not only couple vertical layers of the atmosphere with each other, but also interconnect both hemispheres via global circulations (*Karlsson et al.*, 2009; *Körnich and Becker*, 2010). One important process which couples the atmosphere vertically and horizontally is the exchange of energy and momentum via atmospheric waves (e.g *Becker*, 2012). Waves in the atmosphere may occur on very different horizontal scales ranging from planetary scale waves with wavelengths of thousands of kilometers to gravity waves with typical scales in the order of few tens to hundreds of kilometers, and finally down to acoustic waves and turbulence with characteristic scales of tens of meters and smaller. At present, this large range makes it impossible to study the dynamic aspect of the atmosphere in its entirety because all instruments so far developed are sensitive to parts of the wave spectrum only. Also, studies of the atmosphere based on global circulation models (GCMs) are limited to certain scales because the availability of computing power imposes a natural restriction on the resolution of waves. Atmosphere dynamics is thus still subject to active research despite considerable progress has been made in the last few years. Research presented in this work deals mainly with the study of gravity waves observed by lidar (light detection and ranging).

The atmosphere is commonly divided into several vertical layers, their borders being defined by extrema seen in the thermal structure of the atmosphere. Figure 1.1 shows typical temperature profiles in summer and winter for the location of Davis (69°S, 78°E). Beginning closest to the surface, the layers are called troposphere, stratosphere,

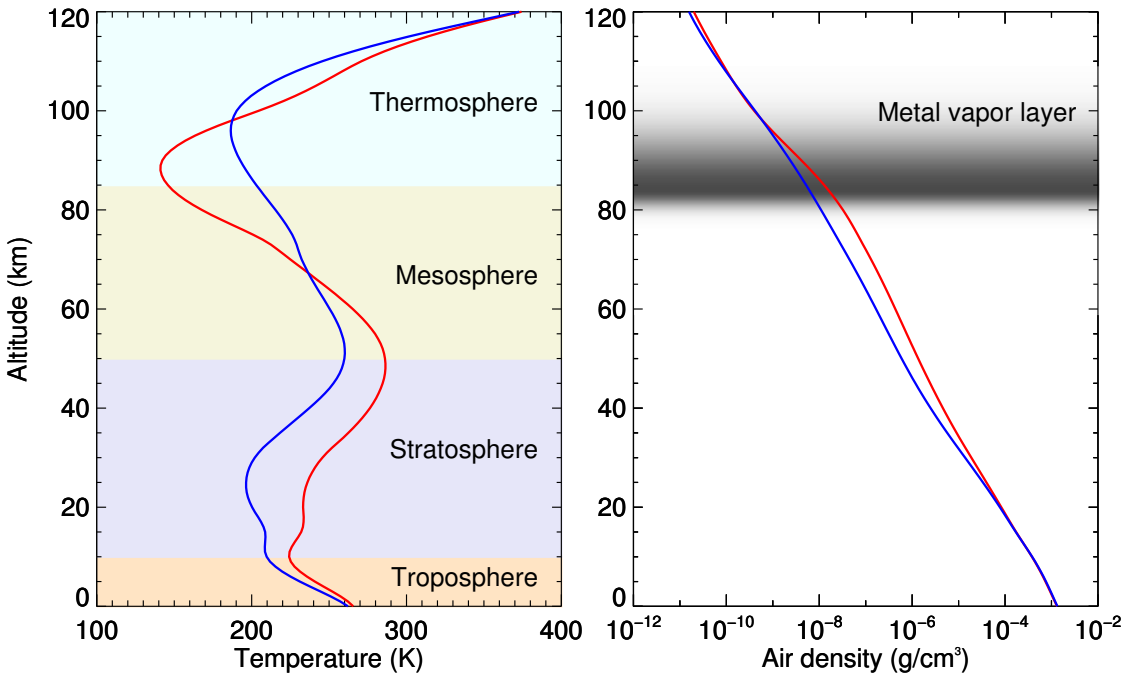


Figure 1.1.: Typical temperature and air density profiles at high latitudes in winter (blue, 1 June 2011) and summer (red, 1 January 2011). The data is taken from MSIS-E-90 model (*Hedin*, 1991) for the location of Davis (69°S, 78°E).

mesosphere, and thermosphere, and are separated by the tropopause, the stratopause, and the mesopause. The region between 50 and 100 km altitude is often referred to as the mesosphere-lower thermosphere (MLT) region. It covers a highly structured region with layers consisting of different materials, e.g. metal vapor layers containing elements like sodium, potassium, and iron, layers containing meteoric smoke, and layers with ice particles at high latitudes, e.g. polar mesospheric clouds (PMC). Of special technical interest are here the metal layers. As will be shown in section 1.3, free metal atoms within the metal layers can be spectroscopically probed by ground-based instruments to determine temperature and winds within the layers. The region of the atmosphere between the tropopause and the mesopause (stratosphere and mesosphere) is called the middle atmosphere.

At high latitudes, the temperature in the stratosphere and the MLT region follows a distinct seasonal cycle. While the stratopause in summer is warmer than in winter, the situation is reversed in the mesopause region (figure 1.1). Record low temperatures occur in summer, followed by much higher temperatures in winter. The correlation is understood to be caused by a summer-to-winter pole meridional circulation which results in upwelling and associated adiabatic cooling at summer high latitudes, and corresponding warming at winter high latitudes through downwelling (e.g *Andrews et al.*, 1987). This circulation is driven by gravity waves interacting with the background mean flow in the MLT region at summer high latitudes (e.g *Becker*, 2012). Gravity waves thus affect atmospheric conditions at global scales. In order for circulation models to accurately reproduce dynamic aspects of the Earth's atmosphere it is important to include a suitable description of gravity wave effects in the model. This work attempts to shed light on the vertical propagation of gravity waves and their interaction with the background flow based on lidar measurements of temperature and vertical wind speed at Davis, Antarctica.

1.2. Atmosphere dynamics

The movement of air parcels in the atmosphere is governed by the Eulerian equation of motion, the equation of continuity, and the equation of energy conservation (e.g. *Fritts and Alexander*, 2003):

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\,\mathbf{u} = -\frac{1}{\rho}\nabla p + \mathbf{g} - 2\boldsymbol{\Omega} \times \mathbf{u} + \mathbf{X}, \qquad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \qquad \frac{\partial \Theta}{\partial t} + \nabla \cdot (\Theta \mathbf{u}) = Q \tag{1.1}$$

Here $\mathbf{u}$ is the velocity, p is the atmospheric pressure, ρ is the density, $\mathbf{g}$ is the acceleration due to Earth's gravity, and $\boldsymbol{\Omega}$ is Earth's rotation vector. The potential temperature, Θ, is defined as

$$\Theta = \frac{p}{\rho\, R}\left(\frac{p_s}{p}\right)^{\kappa}, \tag{1.2}$$

where R is the ideal gas constant, $\kappa = c_p/c_v$ is the ratio of specific heats at constant pressure and constant volume, and $p_s = 1000\,\mathrm{mbar}$ is the reference pressure. The terms $\mathbf{X}$ and Q contain all unspecified forcings such as friction or drag.

Neglecting forcings, the coupled equation system 1.1 can be linearized about a horizontally uniform background state of the atmosphere $\{\mathbf{u_0} = (u_0, v_0, 0)\,, \rho_0, p_0, \Theta_0\}$ with u_0, v_0, ρ_0, p_0, and Θ_0 varying only in the vertical coordinate z (e.g. *Drazin and Reid*, 1981; *Fritts and Alexander*, 2003). It is found that perturbations about this background state of the form of plane waves

$$\left\{\mathbf{u}'(\mathbf{x},t)\,, \frac{\rho'(\mathbf{x},t)}{\rho_0(z)}, \frac{p'(\mathbf{x},t)}{p_0(z)}, \frac{\Theta'(\mathbf{x},t)}{\Theta_0(z)}\right\} = \left\{\tilde{\mathbf{u}}(z)\,, \tilde{\rho}(z)\,, \tilde{p}(z)\,, \tilde{\Theta}(z)\right\} \cdot \exp\left[i\,(\mathbf{kx} - \omega t) + z/2H\right] \tag{1.3}$$

solve the linearized equation system, and propagating modes can exist depending on the properties of the background state and the selection of wave vector $\mathbf{k}$ and frequency ω. The scale height of the atmosphere, H, is defined as the e-folding scale of the pressure $p_0(z)$. In an isothermal atmosphere of temperature $\overline{T}$, the scale height is given by $H = R\overline{T}/g_0$ (*Andrews et al.*, 1987). Thus, for atmospheric conditions typical for the middle atmosphere, H evaluates to $\approx 7\,\mathrm{km}$, i.e. the pressure decreases by one order of magnitude approximately every 16 km.

The exponential decrease in atmospheric pressure has far reaching implications for propagating waves. As the density decreases with height, less mass per unit volume is available to carry the kinetic energy of a vertically propagating wave, and perturbation amplitudes associated with the wave must therefore increase in order to conserve energy. As evident from equation 1.3, the growth rate is proportional to $\exp[z/2H]$.

In addition to mathematical beauty, the fact that perturbation amplitudes grow with height has also practical implications for remote sensing of atmospheric waves. Most instruments are sensitive to the relative change of a measured variable, e.g. the wave induced density perturbation ρ'/ρ_0. Thus, waves are easier to detect in the mesosphere where amplitudes are large. While typical perturbation amplitudes in the lower stratosphere are less than 1%, amplitudes can easily reach 10% in the MLT region. Latter value is, however, significantly smaller than what is expected from the scaling law $\exp[z/2H]$. Wave perturbation amplitudes can not grow indefinitely due to formation of convective and shear instabilities. Moreover, several mechanisms which decrease the perturbation amplitudes of waves do exist (*Fritts and Alexander*, 2003). Of particular interest here is the interaction with the background mean flow (see section 4.4.5).

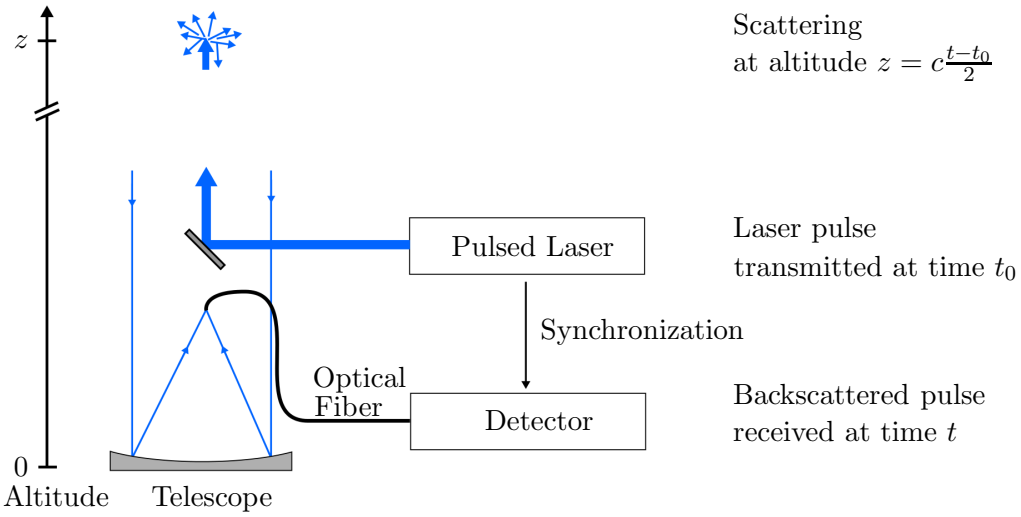


Figure 1.2.: Illustration of the working principle of a lidar. A laser pulse is transmitted into the atmosphere and backscattered light is collected on the ground. The elapsed time between emission of the laser pulse and receiving of the backscattering signal yields the altitude of the scattering event.

1.3. Lidar – A tool to study the middle atmosphere

The acronym lidar (light detection and ranging) refers to an active optical remote sensing technology that can be used for atmospheric soundings. Figure 1.3 illustrates the basic working principle. A powerful laser pulse with a typical pulse length in the range of few nanoseconds to hundreds of nanoseconds is transmitted into the atmosphere at time t_0. On its way through the atmosphere, air molecules and aerosols interact with the light pulse by scattering photons in different directions. Few of these photons are scattered back (scattering angle 180°) and can be collected by a ground-based telescope. A sensitive detector counts the received photons as function of time t, the start time being synchronized with the emission of the laser pulse. Thus, the time delay between emission of the laser pulse and detection of the scattered photons, $t - t_0$, yields the scattering altitude $z = c\,(t - t_0)\,/2$, where c is the speed of light. This ranging capability paired with high sensitivity makes the lidar technique superior to many passive observation techniques, e.g. sun photometers.

As already mentioned, the primary quantity obtained in lidar measurements is the intensity of the backscattered light as function of altitude. This quantity is usually represented as number of photons per height interval because most mesospheric lidars integrate the detected photons over certain altitude ranges. In order to exploit the information present in the lidar return signal, $S\,(z)$, for atmospheric studies, the measured intensity needs to be related to scattering processes. The so called lidar equation (*Fernald et al.*, 1972) is

$$S\,(z) = E\,C\,z^{-2}\,[\beta_{\mathrm{mol}}\,(z) + \beta_{\mathrm{aer}}\,(z)]\,\mathcal{T}^2_{\mathrm{mol}}\,(z)\,\mathcal{T}^2_{\mathrm{aer}}\,(z)\,, \tag{1.4}$$

where E is the intensity of the transmitted laser pulse, C is a calibration constant which includes the receiving aperture, losses in the transmitting and receiving optics, and the quantum efficiency of the detector, z is the range, $\beta_{\mathrm{mol}}\,(z)$ is the molecular volume backscattering cross section, $\beta_{\mathrm{aer}}\,(z)$ is the volume backscattering cross section of

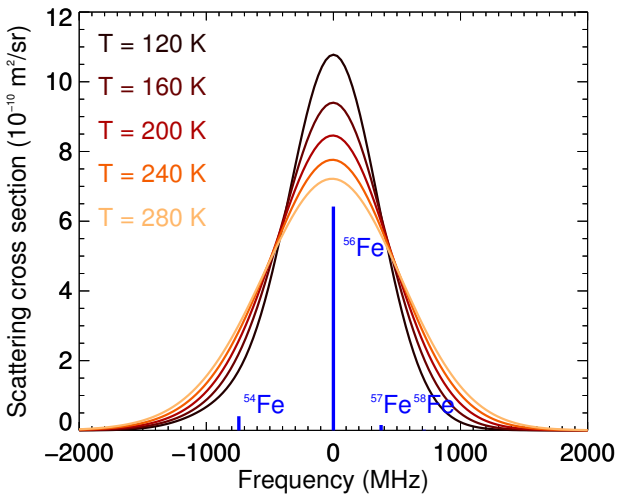


Figure 1.3.: Doppler-broadening of the 386 nm line of free iron atoms (multiplet $a\,^5\mathrm{D} - z\,^7\mathrm{F}$) for different atmospheric temperatures. Also shown is the isotope shift and relative abundance of the four naturally occurring iron isotopes (data are taken from Table 1 found in *Lautenbach and Höffner* (2004), original references are *Kurucz et al.* (1984) and *Kaletta* (1969)).

aerosols, and $\mathcal{T}_{\mathrm{mol}}(z)$ is the molecular transmittance and $\mathcal{T}_{\mathrm{aer}}(z)$ the aerosol transmittance of the atmosphere. In general, both scattering cross sections and transmittance are wavelength dependent, and values need to be computed for the particular laser wavelength.

The volume backscattering cross section $\beta_{\mathrm{mol}}(z)$ is proportional to the molecular number density $n_{\mathrm{air}}(z)$. Thus, equation 1.4 can be used to derive molecular density profiles from the lidar return signal provided aerosol loading and effective scattering cross section are known. However, accurate values are hard to obtain for latter quantities, thus, in practice, the derived molecular density profile is usually restricted to altitude ranges where no aerosols are present. This approximation is valid in between the upper boundary of the stratospheric aerosol layer (approximately 30 km) (e.g. *Thomason et al.*, 1997; *Jäger*, 2005) and the polar stratospheric cloud (PMC) layer at 83 km altitude (e.g. *Fiedler et al.*, 2009). As will be shown in section 3.3.2, based on the assumption of hydrostatic equilibrium the density profiles measured by lidar can finally be converted into atmospheric temperature profiles. Hence, the described lidar technique permits ground-based temperature measurements over most of the middle atmosphere. At present no other sounding method, active or passive, has the ability to provide temperature measurements with sub-kilometer resolution over such a large height range.

Lidar measurements in the MLT region are hampered by very low air densities. Because the lidar return signal is proportional to the air density, the signal decreases exponentially with increasing altitude, resulting in low signal-to-noise ratio at high altitudes. This exponential decay of the air density (figure 1.1) dominates over the z^{-2} dependence of spherical waves formed by scattered light, and limits the maximum altitude where the lidar return signal is analyzable. For large lidar systems such as the ALOMAR RMR lidar the maximum altitude is approximately 85 km (*Schöch et al.*, 2008).

The altitude limitation can be somewhat mitigated by using more powerful lasers and more sensitive detectors. However, since the signal-to-noise ratio is rapidly decreasing

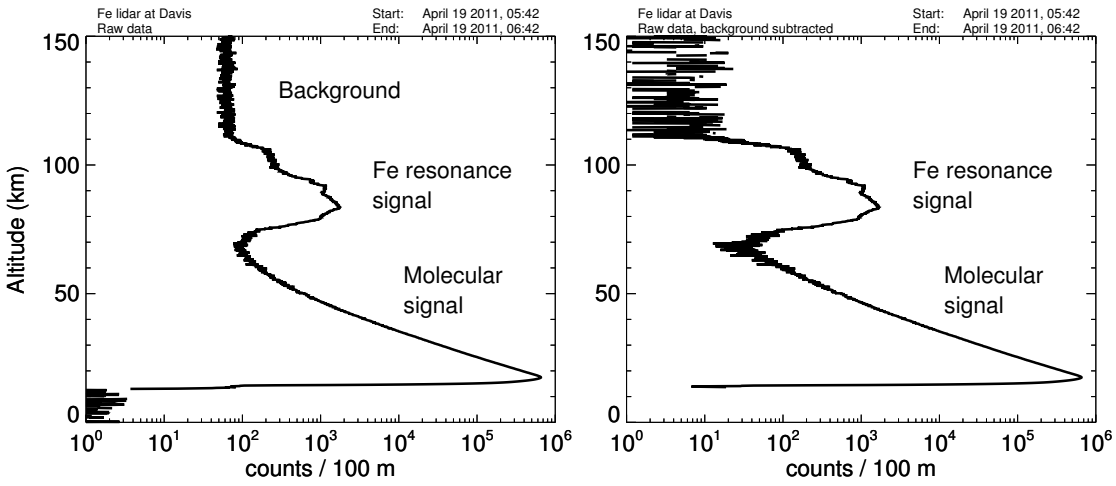


Figure 1.4.: The lidar return signal acquired by the Fe lidar at Davis on April 19, 2011, integrated over one hour. The left panel shows the raw signal, while the background is subtracted from the raw signal in the right panel. During the lower 15 km the signal is blocked by a chopper to prevent overloading of the detector (see section 2.1.2).

with altitude, practical limits concerning the scalability of the lidar system are quickly reached. The sounding range can nevertheless be extended into the mesopause region by exploitation of the much larger scattering cross section of resonant processes such as the fluorescence of iron atoms. A global metal vapor layer exists between approximately 80 and 105 km altitude (e.g. *Plane*, 1991, 2004), and several groups have successfully demonstrated that this metal layer can be probed with ground-based lidar instruments (*Fricke and von Zahn*, 1985; *Gelbwachs*, 1994; *Lautenbach and Höffner*, 2004). Since Doppler-broadening of metal fluorescence lines is related to the kinetic energy of the metal atoms, atmospheric temperature can be retrieved by analyzing the spectral profile of the Doppler-broadened line (see figure 1.3). Precision laser spectroscopy is nowadays a common tool in many optical labs. However, the application of this principle to mesospheric lidars is fraught with many technical challenges, e.g. spectral control of high-power pulsed lasers with peak-powers in the megawatt regime. For this reason, very few atmospheric research groups build and operate these kind of precision lidars.

In case of the fluorescence line of iron at 386 nm the typical net signal gain over the molecular scattering signal at 90 km altitude is approximately four orders of magnitude (see appendix C). It is thus obvious that most lidars designed for temperature studies in the mesopause region operate at resonance wavelengths of metal atoms. This class of lidars is commonly referred to as metal resonance lidars.

Figure 1.4 shows an example raw data profile acquired by the Fe lidar at Davis on April 19, 2011. The enhanced backscatter signal caused by the fluorescence of iron atoms within the metal layer is clearly visible on top of the exponentially decaying molecular signal. A more careful analysis shows that, due to narrowband optical filters in the receiver (see section 2.1.2), only the central part of the molecular backscatter spectrum is recorded. Thus, according to the classification of scattering processes given by *Young* (1981), the molecular signal should be called the Cabannes signal. In this work, however, from now on the term Rayleigh signal is used since it is more common within the lidar community. The analysis of the Rayleigh signal and conversion to temperature is discussed in detail in chapter 3.

Temperature profiles within the atmospheric metal layer can be extracted from the

resonance signal by sampling the Doppler-broadened resonance line at a minimum of two points (*She et al.*, 1990). This technique requires that the Doppler-shift caused by the line-of-sight (LOS) wind (vertical wind in case of a lidar looking at zenith) is zero. The quantities which can be derived from these measurements are atmospheric temperature and atom number density. By adding a third point to the list of sampled frequencies the prerequisite condition of zero LOS wind can be removed. The so called three-frequency method provides LOS wind speed in addition to temperature and Fe number density (*She and Yu*, 1994). This technique has the advantage that it is fairly easy to implement, i.e. it does not require sophisticated equipment to precisely measure the laser frequency. To date most mesospheric resonance lidars which are capable of temperature measurements use the three-frequency method.

Sampling the resonance line at more than three frequencies introduces redundant information to the equation system which defines the three geophysical properties temperature, atom number density, and LOS wind speed. The additional information can be used to determine system parameters such as the line width of the laser. *Lautenbach and Höffner* (2004) demonstrated the derivation of the isotope shift of iron by sampling the 386 nm line with high resolution. The authors used the term "scanning lidar" to highlight the fact that the lidar used in this study repeatedly "scans" across the resonance line as opposed to the two- or three-frequency method which employ fixed sampling frequencies. Having acquired many samples to describe the Doppler-broadened resonance line, the equation system relating the measurements to the three geophysical properties is over-determined and thus cannot be inverted. Instead, the atmospheric temperature is commonly found by fitting the theoretical spectrum to the measured data (*Fricke and von Zahn*, 1985). The lidar data discussed in the present work was acquired with a *scanning lidar*.

While a scanning system provides some advantages over a fixed frequency lidar – some issues were already mentioned above – the downside is an increased complexity regarding data acquisition and laser frequency measurement. For this reason scanning lidars are still in the minority. Problems which arise when using a scanning lidar for Rayleigh temperature measurements are discussed in detail in section 3.3.

1.4. A short history of Project Antarctica

The idea of rocket launches from the Australian Antarctic Station Davis to complement the Rayleigh lidar and Mesosphere-Stratosphere-Troposphere (MST) radar observations dates back to numerous informal discussions between Ray Morris of the Australian Antarctic Division (AAD) and Franz-Josef Lübken of the Leibniz-Institute of Atmospheric Physics (IAP) (*Morris*, 2012). At that time IAP had recently completed a sounding rocket campaign with rocket launches from the British Antarctic Station Rothera (*Lübken et al.*, 2004) and a similar campaign was envisioned for Davis. The possibility of locating the IAP potassium lidar at Davis was first discussed in 2008 when Lübken visited AAD. It was generally agreed that having an IAP metal resonance lidar at Davis would benefit studies of the middle atmosphere and the formation of polar mesospheric summer echoes (PMSE). Lübken indicated his support of an Australian/German collaboration shortly after his visit but suggested the use of the IAP Fe lidar instead of the IAP potassium lidar (*Lübken*, 2012). The Fe lidar was at that time located at the Arctic Lidar Observatory for Middle Atmosphere Research (ALOMAR) in northern Norway and promised better temperature coverage at PMSE altitudes in comparison to the potassium lidar.

Within the AAD Morris then argued a scientific case for positioning the Fe lidar at

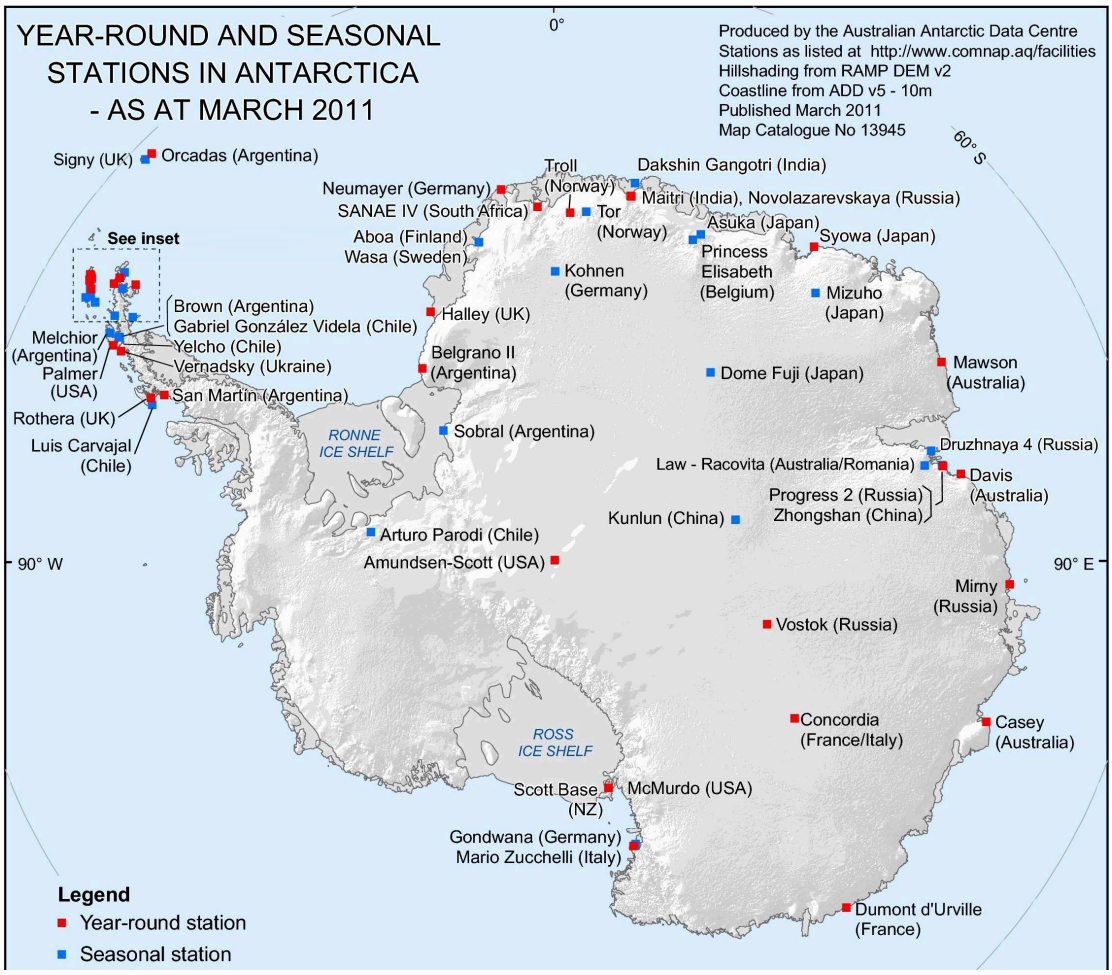


Figure 1.5.: Map of the Antarctic continent. The Australian station Davis is located at the east coast at 69°S, 78°E. The British station Rothera is on the Antarctic Peninsula at 68°S, 68°W. Source: Australian Antarctic Data Center

Davis station which involved discussions with the Program Leader of the Ice, Ocean, Atmosphere and Climate theme of the AAD, the Chief Scientist as well as various other stakeholders within AAD (*Morris*, 2012). These negotiations were successful, and a variation of the Antarctic Research Project 2325 "VHF radar studies of the Antarctic mesosphere, stratosphere and troposphere" was eventually submitted to the Australian Antarctic Advisory Committee by Morris in early 2009. This project was subsequently approved.

A formal discussion of the project between Morris and Lübken took place at the "Layered Phenomenon in the Mesopause Region" workshop in Stockholm 12–15 July 2009 and culminated in signing of the project agreement. However, from IAP point of view the question where to place the Fe lidar was still open. In addition to Davis as proposed by the AAD, the British Antarctic Survey (BAS) had offered Rothera station (see map of Antarctica in Figure 1.5) as base for the lidar. Placing the lidar at Rothera would have allowed IAP to compare lidar observations with earlier co-located rocket based measurements. Davis on the other hand offered the advantage of the co-located VHF radar and Rayleigh lidar. In the end, the deciding factor was the VHF radar, and Davis was officially chosen as site for the lidar on October 8, 2009 (*Lübken*, 2012).

Following the decision of the future location, the Fe lidar was moved from ALOMAR

Figure 1.6.: Aerial photography of Davis station taken in November 2011.

back to IAP in October 2009 for maintenance work and upgrades (see section 2.1). In August 2010 the lidar was shipped to the AAD in preparation for transport to Davis on board the Aurora Australis (Voyage 1, season 2010/2011). The ship arrived at Davis on November 16, 2010. In the following days the lidar container was transported ashore and positioned in between the CPC building and the Australian lidar (see figure 1.6). Following the commissioning phase first lidar observations of the mesosphere were carried out on December 18, 2010. The operational phase of the Fe lidar started in January 2011.

AAD approved the initial science from November 2010 to February 2012. Morris then successfully initiated an extension until November 2012. Finally, a second extension was approved from November 2012 to January 2013. The lidar container was transported back to Australia on board the Aurora Australis (Voyage 3, season 2012/2013) in January 2013.

With more than 2300 hours of lidar observations, the initial science phase was very successful and provided a wealth of data covering a full annual cycle. Only the large number of hours made it possible to study dynamical aspects of the atmosphere such as gravity wave filtering and heat fluxes, both of which require long averaging times in order to smooth out the geophysical variability and obtain meaningful mean values. Initially, it did not look like this goal could be readily achieved. After a good start in January and February 2011, the number of observations per week was reduced, and in the following months the number of observation hours averaged at 3.6 hours per day (see figure 1.7). Weather permitting, nearly continuous lidar observation was resumed in May 2011. Following this increased observation rate the average number of observations nearly doubled and remained at 6.8 hours per day for the remaining part of the initial science phase (figure 1.7). Thus 2343 hours worth of data could be acquired within the first 15 months of lidar operation, making the Davis lidar campaign the most successful lidar campaign to date. The first extension provided another 365 hours over a period of 9 months, followed by 146 hours during the second and final extension. In contrast to

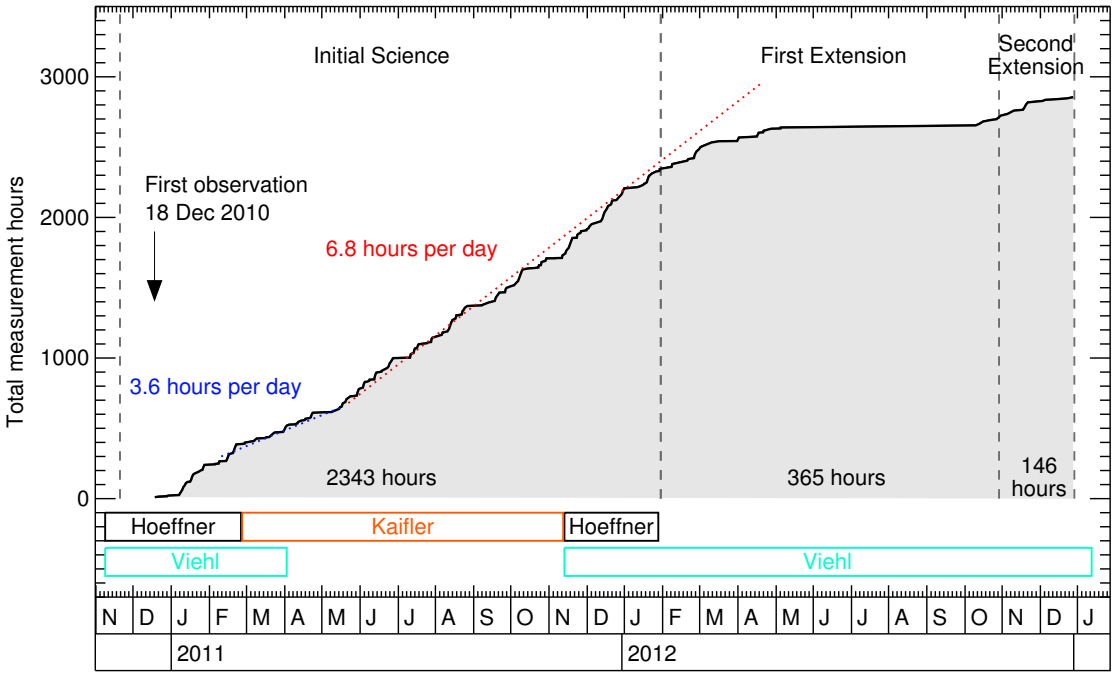


Figure 1.7.: Timeline of the project and number of observations hours. After the initial science phase the project was extended two times. Also shown is the principle lidar operator schedule.

the initial science phase, no winter observations were carried out during the extensions in 2012.

1.5. The structure of this thesis

This thesis comprises three main chapters. Chapter 2 provides a short overview of the Fe lidar system, followed by a brief summary of the modifications which were carried out in preparation of the campaign in Antarctica. Chapter 3 deals with the analysis of the lidar return signal with respect to converting the Rayleigh part of the signal to measurements of atmospheric temperature. Stratospheric temperature measurements as well as temperature measurements obtained from the resonance signal in the MLT region are analyzed in chapter 4. This chapter is subdivided into four sections. Section 4.1 gives an overview of the lidar dataset. Mean temperatures and seasonal changes are discussed in section 4.2. In section 4.3 the temperature data is analyzed for waves with periods in the range 2–12 h. This section is meant to give an overview over amplitudes, frequencies and occurrence of observed temperature perturbations, and no classification is made regarding the nature of the waves (e.g. planetary wave, gravity wave, tide). In contrast, section 4.4 deals explicitly with the extraction and analysis of gravity waves. An attempt to derive heat fluxes from the gravity wave data is described in appendix A.

Although references to gravity waves and their properties are made throughout this thesis, the introduction to this type of waves was deferred to the gravity wave section in order to maintain a logical structure. Readers unfamiliar with gravity waves may therefore consider reading section 4.4.1 first.

2. The Fe lidar system

2.1. The Fe lidar

The IAP Fe lidar system is a world wide unique instrument designed primarily to measure temperature in the mesosphere-lower thermosphere (MLT) region of the Earth's atmosphere 80 to 100 km above ground. Unlike other metal resonance lidars which probe alkali atoms in the visible or near infrared part of the electromagnetic spectrum, e.g. the ALOMAR sodium lidar (*She et al.*, 2002b; *Kaifler*, 2009) at 589 nm and the IAP potassium lidar (*von Zahn and Höffner*, 1996; *Fricke-Begemann et al.*, 2002a) at 770 nm, successful operation of the Fe lidar required the development of a powerful laser for the ultraviolet region of the electromagnetic spectrum. As there are no solid-state lasers available with fundamental wavelengths in the UV, frequency doubling is the method of choice. Iron has two for lidar purposes suitable optical transitions at 386 nm and 372 nm (e.g. *Höffner*, 1990; *Alpers*, 1993). The scattering cross section at 372 nm is approximately two times lager than at 386 nm, and a lidar operating at 372 nm is thus approximately two times more efficient than at 386 nm. However, the longer wavelength 386 nm was easily accessible by adding a frequency-doubling unit to the existing alexandrite laser which was formerly used as light source for the potassium lidar. This laser had a sufficiently large tuning range, and the advantage of using existing and proven laser technology outweighed the benefits of the more efficient lidar wavelength. Hence, 386 nm was chosen as primary wavelength for the new Fe lidar system (*Lautenbach*, 2007).

Conversion of the transportable IAP potassium lidar began in 2004 (*Keller*, 2006). In summer 2008 the lidar system, now called transportable IAP Fe lidar, was moved to Arctic Lidar Observatory for Middle Atmosphere Research (ALOMAR) in Northern Norway in support of the ECOMA[1] rocket campaign which took place in June/July of the same year. The lidar then remained at ALOMAR until September 2009 to measure seasonal temperature variations in the MLT region above ALOMAR. However, the performance of the laser power supply degraded in spring 2009 and failed permanently soon after, thus rendering the lidar inoperable. With the Antarctic project already on the horizon, it was decided not to attempt to repair the lidar on-site and bring it back to IAP instead. Repairs and major upgrades to the lidar as well as the support infrastructure and computer systems were carried out at IAP between October 2009 and July 2010. Finally, in August 2010 the lidar container was shipped to Hobart, Tasmania, in preparation for the transport to Antarctica on board the ship Aurora Australis in October of the same year. A summary of important technical data of the Fe lidar system as used during the Antarctic campaign can be found in table 2.1.

In spite of the so far relatively short existence, the Fe lidar was instrumental to some important observations at ALOMAR. This includes, for example, diurnal variations of the Fe-layer correlated with sunrise and sunset, the discovery of large thermal tides at high latitudes, and detailed observations of the strong mesospheric cooling and the associated reduction in Fe density in response to the major stratospheric warming in

[1] Existence of Charged state Of Meteoric smoke particles in the Middle Atmosphere. More information can be found in e.g. *Rapp et al.* (2011)

Transmitter:	
Power laser	101-PAL Alexandrite ring laser from Light Age Inc. with custom modifications
Wavelength	772 nm, 386 nm
Energy per pulse	110 mJ at 772 nm, 30 mJ at 386 nm
Pulse length	200 ns
Repetition rate	33 Hz
Receiver:	
Telescope	80 cm aperture, 190 cm focal length[1]
Field of view	66 µrad[2]
Detector channels	772 nm: 3 + 1 high spectral resolution, 386 nm: 3
Height resolution	25 m

Table 2.1.: Technical data of the Fe lidar system. [1]*Keller* (2006), [2]*Höffner* (2012)

2009. However, none of these findings were published, and it took three more years before the knowledge became public. *Lübken et al.* (2011) reported on large amplitude tides observed with the Fe lidar above Davis, Antarctica, and diurnal variations of the Fe-layer based on observations with an Fe Boltzmann lidar located at McMurdo were eventually published by *Yu et al.* (2012). As of this writing, a publication dealing with the response of metal layers to the major stratospheric warming is in preparation.

2.1.1. Transmitter

The layout of the laser system used as transmitter for the Fe lidar is illustrated in figure 2.1. The transmitter system comprises the injection seeded alexandrite ring laser, frequency doubling unit, beam expander, and beam guiding optics. Further components are the seed laser, the optical spectrum analyzer which measures the frequency of the power laser with high accuracy, and the optical reference stabilized by Doppler-free saturated absorption spectroscopy of rubidium atoms (not shown).

Laser system

The pulsed power laser is based on components of a commercial pulsed alexandrite laser (101-PAL) from Light Age, Inc. The layout of the resonator follows the setup of a classical ring cavity with three high-reflective mirrors and one output coupler. The beam circulates in clockwise direction, the direction predefined by the optical diode, and leaves the ring cavity to the left. Two flashlamp-pumped alexandrite rods heated to approximately 90 °C serve as active laser medium. Because the laser is operated at a wavelength off the gain maximum, a frequency selective element (Lyot filter) is introduced into the cavity to provide broad-band filtering. The Q-switch prevents amplification of spontaneous emission by increasing losses within the resonator. It can be switched to a low-loss state when the fluorescence maximum is reached and mode-matching is detected, thus allowing the build-up of a signal longitudinal mode. Mode-matching with the seeder is achieved by dithering the length of the cavity with the help of a piezo actuated mirror (M1). A photodiode (PD) detects the superposition of seeder and fluorescence light originating from the alexandrite rods. A maximum signal indicates constructive interference (mode matching), the desired precondition for triggering the Q-switch. A detailed description of the operation of the laser and its components can be found in *Fricke-Begemann* (2004).

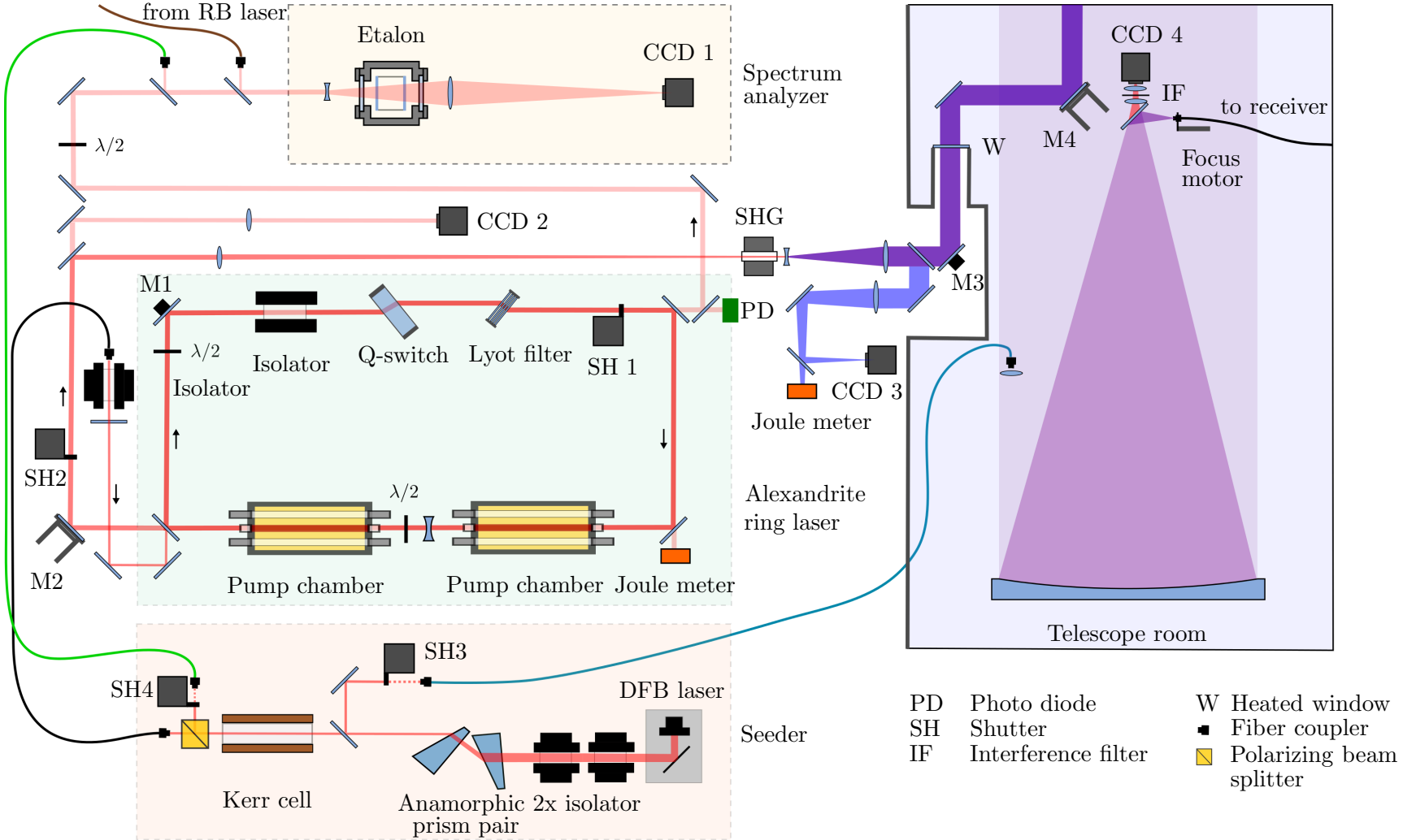


Figure 2.1.: Sketch of the transmitter system (adapted from *Höffner*, 2012).

The beam leaving the ring resonator passes two turning mirrors which direct the light toward the frequency doubling stage (SHG). One of the mirrors (M2) is outfitted with actuators to compensate angle deviations caused by thermal drift of the various laser components. The laser beam is then weakly focused into the frequency doubling crystal (LBO crystal), subsequently expanded and transmitted into the atmosphere via a second set of turning mirrors.

Beam stabilization

Constant overlap of the laser beam with the field of view of the receiving telescope, γ, is crucial for successful operation of the lidar. Because the amount of solar background entering the receiver is proportional to γ^2, efforts have been undertaken in the last few years to increase the signal-to-noise ratio of the lidar signal by making the field of view smaller and smaller. This development was only possible since the advent of sophisticated beam stabilization techniques. Even if perfect beam overlap is achieved initially, without active stabilization the beam tends to tilt against the optical axis of the telescope due to unbalanced thermal expansion of various lidar components and the lidar container itself, as well as vibrations, and finally atmospheric turbulence. While atmospheric turbulence arises on time scales in the order of the pulse repetition rate and thus cannot be predicted by the beam stabilization system, drift due to thermal expansion and low frequency vibrations is slow enough to be compensated for.

The stabilization system employed with the Fe lidar uses two hardware components, a CCD camera, which looks through the telescope and tracks the laser beam in the sky (CCD 4 in figure 2.1), and a set of turning mirrors to steer the laser beam (M3 and M4). With the position of the laser beam known from camera images, the departure from the optimal position (the target) is computed and correction signals are sent to the beam steering system. Adjustments to the beam position are normally made by changing voltages of the piezo actuator of mirror M3. However, when the beam drift becomes large, e.g. during large swings of the outside temperature, the required angular change of the mirror can easily exceed the capabilities of the piezo actuator. In this case, commands are sent to the motorized mirror mount of mirror M4 to move the beam in the direction of largest piezo deformation, thus unloading stresses and bringing the piezo back to equilibrium position.

Timing of the servo loop of the beam stabilization system is governed by the acquisition of beam position data, which in turn depends on the pulse repetition rate of the laser. Hence, in theory, the system is capable of compensating disturbances with frequencies up to approximately 33 Hz. However, the bandwidth of the feedback loop is artificially reduced in order to prevent the system from becoming unstable.

Wavelength measurement and beam diagnostic

The spectrum of the power laser is monitored on a pulse-to-pulse basis with a custom built optical spectrum analyzer which comprises a pressure and temperature stabilized Fabry-Perot interferometer, two lenses, and CCD camera (see figure 2.1). Readout of the camera is triggered synchronously with the pulse repetition frequency of the laser in such a way that each frame contains the Fourier-transformed image of exactly one laser pulse. Information on the centroid wavelength of the laser pulse as well as the spectral width can be inferred from these images by evaluating radius and shape of the interference rings. It is important to note, however, that no information on absolute wavelength can be obtained from these images alone. In order to measure absolute wavelengths, which is the primary objective here, the radius of interference rings needs

to be calibrated with the help of a known reference. For this purpose, a DFB laser (not shown in figure 2.1) is employed, its frequency being stabilized to an absorption line of rubidium using Doppler-free saturated absorption spectroscopy of rubidium atoms in conjunction with the Pound-Drever-Hall (PDH) technique (e.g. *Black*, 2001). Light generated by this laser is coupled into the spectrum analyzer for a short period in between two successive laser pulses from the power laser. Since the frequency of the chosen atomic absorption line is well known, the radius of the corresponding interference ring can be used to define the origin of the wavelength scale on the Fourier-transformed images. Thus, the problem concerning absolute wavelength measurements reduces to differential measurements of the radiuses of two interference rings. In this case, the absolute accuracy of wavelength measurements is expected to be better than 1 MHz. This is sufficient to allow for atmospheric Doppler measurements with a precision of 0.5 K. More information on a spectrum analyzer of similar type can be found in *Kaifler* (2009).

In addition to wavelength measurements, the second function of the spectrum analyzer is the identification of multimode laser pulses which occur at random intervals and account for approximately 1% of all laser pulses. Because the retrieval of atmospheric temperatures is based on the assumption of probing the Fe layer with single-mode laser pulses, successful detection of laser pulses with multimode or otherwise deformed frequency spectra is of great importance. For this purpose the data acquisition software includes several algorithms to analyze the shape of the frequency spectrum (*Höffner*, 2012). Laser pulses with significant deviations from predefined parameters are tagged and subsequently rejected in the temperature retrieval.

Further beam diagnostic elements are a Joule meter which records the energy per laser pulse in the infrared, and a CCD camera (CCD 2) which is used to monitor the beam profile of the alexandrite laser as well as the beam position. A second Joule meter located at a side beam after the frequency doubling stage measures the pulse energy in the ultraviolet. Thus, the ratio of the two energy measurements provides an estimate of the efficiency of the frequency doubling process. Camera CCD 3 is used to monitor the beam profile of the UV beam.

A fiber link (blue line in figure 2.1) runs from the seeder laser to the telescope, thus allowing seeder light to be coupled into the receiver once shutter SH3 is opened. Using the seeder as light source provides an easy way to test the IR part of the receiver without the need to run atmospheric backscatter measurements.

2.1.2. Receiver

The receiver system of the Fe lidar consists of two major parts, the receiving telescope with an aperture of 80 cm and 190 cm focal length (*Keller*, 2006), and the optical bench with optical filters and detectors. Both parts are interconnected through an optical fiber which limits the field of view of the receiver to 66 µrad (*Höffner*, 2012).

Figure 2.2 shows a sketch of the optical bench. Light collected by the telescope enters the optical bench through the fiber from the left. Piezo actuators on the mount allow the fiber end to be precisely positioned in the xy-plane. The mechanical chopper protects the detectors from the intense light scattered back in the lower atmosphere by blocking the optical path for 130 µs (19.5 km round-trip path length of a laser pulse) after firing of the power laser. In order to achieve precise timing of the unblocking, and thus also of the lowest altitude where backscattered light can reach the detectors, timing of the laser pulses needs to be synchronized with the rotation of the chopper. A light barrier measures the phase angle of the chopper, and the firing of the power laser is triggered

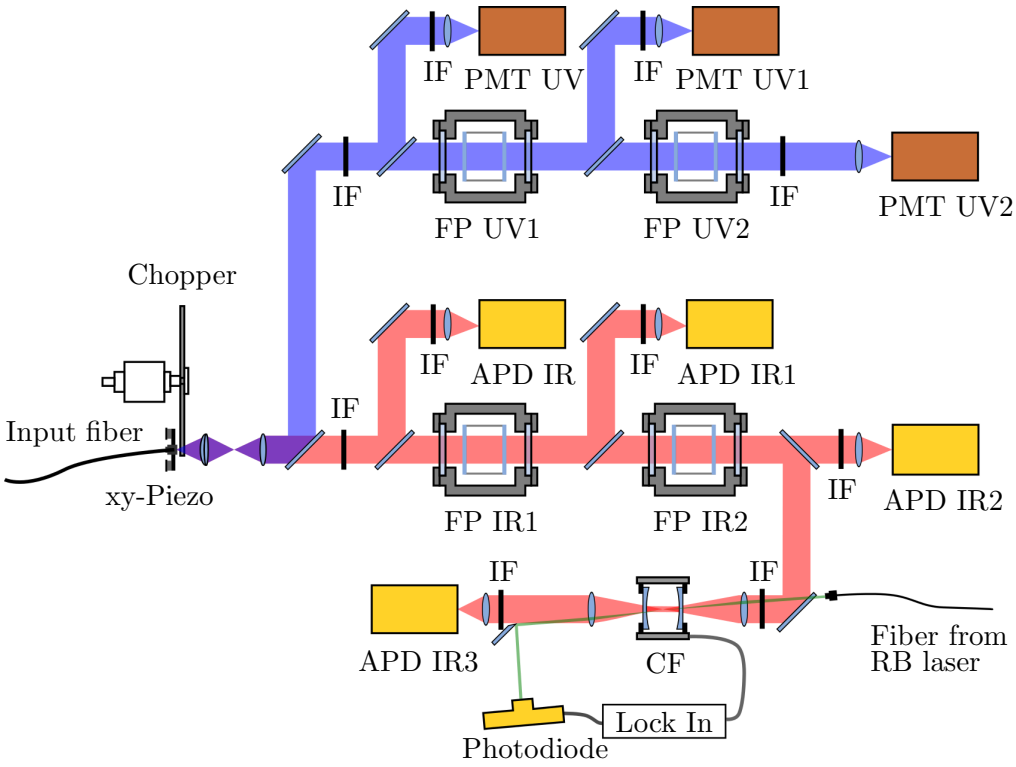


Figure 2.2.: Sketch of the optical bench of the receiver (adapted from *Höffner*, 2012). Infrared beams are drawn in red, ultraviolet beams in blue.

off its signal.

Light leaving the telescope fiber is collimated with the help of an achromatic lens system before the UV-part of the beam is separated by a dichroic mirror. The IR beam continues and passes the first interference filter (IF) which rejects most of the broadband solar background. Two temperature and pressure stabilized narrowband etalons (FP IR1 and FP IR2) narrow the spectral range of the transmitted beam further. Next, the IR beam is split into two main beams. The transmitted beam passes another interference filter and hits finally the main detector (APD IR2), while the other beam is directed into an ultra narrowband confocal etalon (CF). This etalon is locked to the rubidium reference using the Pound-Drever-Hall technique, the light from the reference passing off-axis through the etalon before being picked up by a low-bandwidth photodiode. Due to its extremely narrow transmission band, the etalon rejects most of the Doppler-broadened Rayleigh scattering with only the central part reaching the avalanche photodiode APD IR3. Thus, with the help of the confocal etalon it is possible to discriminate the Mie peak of the backscatter spectrum against the broad Rayleigh background.

Side beams are picked up before and after the first IR etalon (FP IR1) and are directed to their respective detectors APD IR and APD IR1. The signal of APD IR serves as reference for all the other detectors in the IR branch which are located downstream of the first narrowband filtering element (FP IR1). Transmission spectra of the filtering elements can be obtained by taking the ratio of the signals measured before and after the element, e.g. APD IR1 / APD IR yields the transmission spectrum of the etalon FP IR1. The capability to determine the transmission spectrum during lidar operation is important for two main reasons: (1) Tiny leaks in the pressure vessel cause the

Figure 2.3.: Left: The Fe lidar container and chiller hut (black box on the right hand side of the container). The building housing the Australian lidar is visible in the background. Right: View of the telescope in the back of the container.

etalons to drift and retuning is necessary every few days. By measuring the transmission spectrum it is possible to track this spectral drift (section 3.3.3). (2) The transmission bandwidth of the etalons is small enough that the lidar return signal gets spectrally disturbed, and precise knowledge of the transmission spectrum is necessary in order to take these disturbances properly into account in the backscatter analysis. In case of the UV etalons, the process of determining the transmission spectrum and subsequent spectral corrections are discussed in section 3.3.

The UV branch of the receiver is basically a copy of the IR branch adapted for the wavelength 386 nm and without the confocal etalon. Because of the shorter wavelength, photo multiplier tubes are used as detectors instead of avalanche photodiodes. Another specific feature is the 0.3 nm interference filter (IF) in front of the main detector PMT UV2. This filter is mounted on a motorized flip mount and can be retracted when the solar background is low (sun below the horizon), resulting in an approximately 30% increase in transmission of the receiver. The spectrally most sensitive element in the UV branch is the etalon FP UV2 with approximately 2.5 pm FWHM.

Some technical details of an early version of the lidar receiver can be found in *Keller* (2006). Concerning the actual version, specifications may be obtained from principal investigator Josef Höffner.

2.2. Preparations for Antarctica

Between November 2009 and July 2010 a massive refurbishment and upgrade program was undertaken to get the Fe lidar and its infrastructure ready for the upcoming campaign in Antarctica. All components of the lidar system as well as the equipment necessary to operate the lidar are housed in a single 20 foot shipping container (figure 2.3). When the lidar returned from ALOMAR in October 2009, the container was in a bad state. Multiple leaks near the doors allowed melt water to accumulate in the intermediate bottom, causing rotting of the insulation as well as part of the wooden floor. Repairs required stripping of the interior in the forward and aft part of the container,

removal of insulation, and rebuilding of the floor. The old telescope assembly, which was made of steel, was completely removed in this process, and additional bars were installed to strengthen the floor of the telescope room. In anticipation of the cold environment of Antarctica a new telescope was designed around the existing primary mirror. The new superstructure (see right panel of figure 2.3) is made entirely of carbon fibers. This material provides excellent thermal stability in both hot and cold environments, and thus eliminates the need to refocus the telescope during large temperature swings.

Further modifications included rebuilding of the roof hatch and the installation of a new air conditioning system. The old system, which is still retained as backup, comprises two air conditioners mounted near the ceiling of the container. These units operate as air-to-air heat exchanger and thus require one side to be exposed to outside air. Because any exposed void tends to fill up with snow during snow storms, the old system was considered unreliable in polar conditions. Furthermore, the compressors caused vibrations which were transmitted through the structure of the container to optical elements of the lidar system. To improve reliability as well as to reduce vibrations, a new air conditioning system comprising a single powerful air-glycol heat exchanger was installed. A glycol coolant loop transports the heat to a dedicated chiller which is located in a separate chiller outside of the container. The same hut houses a second chiller needed to cool the power laser. The doors of the hut can be closed when the lidar is shut down, protecting the chillers from Antarctic snow storms. In addition to heat traces mounted on the compressor body, two electric heaters were installed in the hut to keep the temperature above $-15\,^{\circ}\mathrm{C}$ when the chillers are not in use. Moreover, the glycol in both coolant loops (air conditioning and laser cooling system) can be preheated by circulating the coolant through tubes lined with heat traces. Because most of the glycol used in the coolant loops is stored outside in the chillers, preheating is important during startup of the lidar in cold weather to prevent circulating glycol from freezing the deionized water in the glycol/water heat exchanger in the primary cooling loop of the laser. The new air conditioning system proved to be very reliable under Antarctic conditions. In comparison to other campaigns at northern hemisphere high latitudes problems with the glycol cooling system caused no significant downtime of the lidar. The chiller hut and coolant piping can be seen to the right of the left panel in figure 2.3.

The electrical switchboard of the container was replaced and most of the electrical systems rewired in May/June 2010. The addition of a second chiller as well as other new electrical components, e.g. installation of a new uninterrupted power supply, necessitated a complete reconstruction of the electrical supply network. In this process, most of the manual switches and old wireless-controlled switching modules were replaced by computer controlled remote power distribution units. The units each comprise six power relays which are controlled by a single ADAM-6066 modbus TCP module. To make the power supply to critical systems such as the electrical heaters that keep the container warm redundant, the remote power distribution units were distributed across three separate electric circuits. The ADAM modules are controlled by two low-power embedded computers running in a redundant set, i.e. both computers are active at all times. However, only one computer assumes command while the other listens passively for system changes, e.g. reconfiguration of the lidar/container systems through commands sent to ADAM modules. To handle all the information and command capability a program called COCOS (Command & Control System) was developed. Not only does COCOS control the ADAM modules in the remote power distribution units, it also collects and aggregates data from various sensor channels, e.g. sensors measuring the coolant temperature at different points, switches signaling the state of the roof hatch, or relays controlling the high voltage for the lidar receiver. In total more than 200 sensor

and data channels are managed by COCOS. While COCOS is primarily intended as compact high-level interface to the many low-level components of the lidar/container system, it also contains a limited set of fault detection algorithms to implement the "fail-safe" concept. In the event of failure of one or more components COCOS responds in a way that will cause no harm to the lidar and prevent further damage. For instance, COCOS will trigger a safeing event if it detects a temperature rise in the secondary cooling loop caused by the failure of a coolant pump. Once safeing is initiated, a predefined sequence of commands is sent to the various subsystems resulting in a controlled shutdown of the lidar and reconfiguration of the container for lidar shutdown. Key events are shutting down the power laser, closing the telescope cover, closing the roof hatch, deactivating the chillers, and activating the heaters. Further development of the fault detection algorithms in COCOS took place in the period March–June 2011. Following this development cycle, the enhanced self-monitoring capabilities made it possible to offload much of the monitoring work from the operator. This was the prerequisite for sustaining very long continuous lidar observations (up to 80 h) in winter 2011.

Upgrades in the period November 2009 to July 2010 were not limited only to the infrastructure of the lidar (the container), but also included work on the lidar itself. Following the breakdown of the laser power supply in early 2009 (see section 2.1), a new unit was installed in summer 2010. The laser was also extensively reworked: The old plastic pump chambers were replaced by two new pump chambers made of steel, resonator mirrors were exchanged, new diagnostic equipment was added (cameras, photo diodes, and energy meters; see figure 2.1), a new optical isolator was tested, a second seed laser was added, and the beam expander was replaced with a larger one to reduce the divergence of the laser beam, requiring the diameter of the downstream optics to be increased to three inch. Eventually every optical element was moved at one time or another, triggered by the need to make space on the laser table. Further changes included the addition of two motorized beam steering mirrors (M2 and M4 in figure 2.1). The ability to adjust remotely the pointing of the two mirrors proved to be very valuable during the Antarctic winter. Due to large swings of the outside temperature the metal structure of the container expanded and contracted, thereby twisting and bending the floor. With the struts fixed to the container, the tiny movements are transmitted onto the laser table, eventually making realignment of optical elements necessary. Heaters were installed on the last two turning mirrors as well as on the optical window in the telescope room. The heaters can be switched on by computer control to prevent the optical surfaces from icing up[2]. Heaters were also added to the motors driving the last beam steering mirror, M4, and the telescope focus motor[3].

Several improvements were made to the receiver. First of all, the focal box sitting in the prime focus of the telescope was completely redesigned (see figure 2.1). The new version includes a motorized stage which allows the fiber to be positioned remotely under computer control. In comparison to the old manual system it is now much easier to find the optimum position, and more important, it can be easily verified by correlating the position information with the strength of the lidar return signal. The power supply and control electronics of the motor driving the mechanical chopper was replaced with a newer model. Also, the power packs supplying current to the avalanche photo diodes haven been integrated in a single unit. Both new power supplies contain a computer interface, thus allowing the systems to be monitored remotely.

[2]Icing was frequently observed at Davis in winter when the sea ice breaks up following a large storm and warm moist air is pushed inland.

[3]Adding heaters to the motors was foresightful. The motors stopped working repeatedly when the temperature dropped below −20 °C.

One can argue whether the increased complexity caused by adding more computer controlled components make a lidar system more fault-prone. On first sight, the reckoning is simple: the more components, the more potential points of failure. On the other hand, a lot of flexibility can be gained by allowing a computer to control the equipment, and new capabilities can be derived from combining data from various subsystems. One example is the monitoring capability implemented in COCOS. The downside of this approach is that the software controlling the equipment becomes increasingly complex, and provisions must be taken to prevent software errors from causing damage to the hardware. Thus, software development and, in particular, quality control become more and more important in the development of modern lidar systems. In case of the Fe lidar at Davis, the adoption of sophisticated software was key to the success.

3. Temperature retrieval

This chapter deals with processing and data reduction steps required to extract atmospheric temperatures from backscatter lidar measurements. Emphasis is put on analysis of Rayleigh backscatter profiles because the development of a dedicated software package for the stratosphere was one of the main topics of this thesis. Temperatures estimated from resonance scattering were provided by the principle scientist Josef Höffner, hence no processing was required.

3.1. Description of the raw dataset

The raw dataset contains single-pulse photon count profiles for each of the seven detector channels as well as certain meta information, e.g. frequency of the emitted laser pulses, pulse energy, and position of the laser beam relative to the field of view of the telescope. 4000 of these records are grouped in a single raw data file. With the lidar emitting 33 pulses per second, this corresponds to approximately two minutes of lidar measurements. Between January 2011 and May 2012, 75807 raw data files have been collected with the lidar instrument totaling more than 2 billion individual backscatter profiles, each of which can be analyzed individually if one chooses to do so. However, in most cases, backscatter profiles are integrated in time and height over a certain range in order to improve the signal-to-noise ratio before the data is analyzed.

The backscatter profiles are stored in the form of photon count profiles with 25 m height resolution covering the height range from ground level up to 200 km altitude. Thus, a single photon count profile comprises 8000 numerical values. Assuming each value to be represented as 8 bit integer (1 byte), the amount of storage space needed to store all the data amounts to approximately 17 terabytes. Though certainly possible on modern computers, the large storage size is rather unpractical to work with. For this reason, a compression algorithm specifically designed to make use of the fact that the photon counting process is a Poisson process, was implemented in the data acquisition software. Making use of this data compression technique cut the size of the memory footprint approximately a hundredfold. Currently, the raw dataset comprising compressed photon count profiles and uncompressed meta data requires 150 gigabytes of storage space.

3.2. Data reduction and quality control

3.2.1. Correcting saturation of photon detectors

Detectors operated in single-photon counting mode, e.g. photo multiplier tubes (PMTs) or avalanche photodiodes (APDs), convert incoming photons into electrical pulses. Because of the high gain and the limited analog bandwidth of these devices the pulse duration is of the order of several nanoseconds. This has important consequences regarding the linearity of the detectors. When a photon is absorbed and the amplification process has started, the detector is blind or *dead* for the duration of the pulse. Consequently, any more photon arriving within this *dead time* is missed. The photon rate as

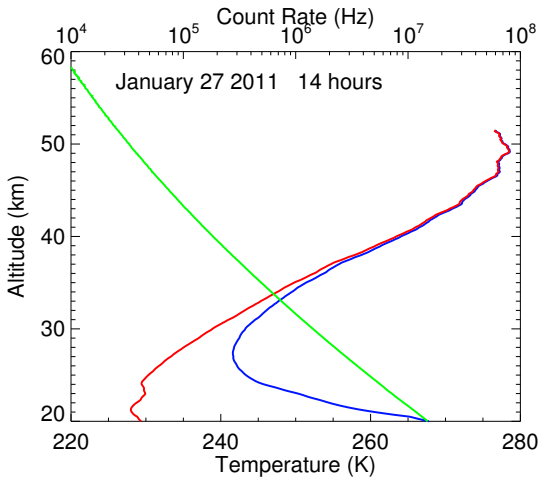


Figure 3.1.: Effect of detector saturation on temperature retrieval. The blue curve shows atmospheric temperature estimated from uncorrected photon count profiles, the red curve is obtained with dead time of 24 ns taken into account (equation 3.1). Also shown is the photon count rate (green line).

detected is lower than the true photon rate.

For low count rates, when the average time between two incoming photons is large compared to the dead time, the probability of two photons arriving less than one dead time apart is small. Thus, the number of counted pulses within a given time interval is proportional to the number of incoming photons. However, if the count rate exceeds a few kilohertz, a significant part of the photons is lost as the detector becomes saturated. This non-linearity poses a problem for lidars because of the large dynamic range required. Figure 3.1 shows the effect of detector saturation on retrieval of atmospheric temperature. The blue curve is obtained if no saturation correction is performed, the red curve in case dead time effects are taken into account. For altitudes above 45 km, corresponding to count rates (drawn in green) of < 100 kHz, there is hardly any difference between the two lines. However, the discrepancy becomes rather dramatic below 30 km altitude where count rates exceed 2 MHz, reaching 40 K at the bottom of the profiles. This is by far the largest temperature shift associated with any effect discussed in this chapter.

For many Rayleigh lidars this problem is mitigated by dividing the optical signal into several cascaded branches with different splitting ratios, each branch supplied with its own detector. In case of the ALOMAR RMR Lidar three detectors with splitting ratios 1:30:900 are used (*Zahn et al.*, 2000). The third detector, which is used for low altitudes, receives only approximately one per mil of the total signal. Hence, the magnitude of saturation effects is greatly reduced.

The Fe lidar was, however, designed as a resonance lidar and for this reason was never equipped with cascading detector channels[1]. So far this did not pose any problem because the resonance signal is low in any case and far from saturating a single detector. An example raw data profile is shown in figure 1.4. Even though the Rayleigh signal was recorded in addition to the resonance signal, it was never analyzed in the past with

[1]The detector channels UV and UV2 may technically be regarded as cascaded channels with a splitting ratio of approximately 1:7 (see figure 2.2). However, detector PMT UV is located in front of the narrowband optical filters (etalons), resulting in high background. With exception of very long integration periods, the low signal-to-noise ratio prevents the use of this detector as low rate channel for the temperature retrieval.

regard to atmospheric temperatures. Hence, saturation at low altitudes was considered a minor problem. Now, when extending the temperature retrieval into the stratosphere, the situation changes. With only a single detector available to cover the large dynamic range, the signal changes by approximately six orders of magnitude from 20 to 70 km altitude, accurate modeling of saturation effects becomes of paramount importance.

Modeling detector saturation

In this work saturation effects are taken into account using a simple model of a non-paralyzing detector

$$N = \frac{M}{1 - \tau M}, \tag{3.1}$$

which relates the true event rate N to the measured count rate M as described in e.g. *Larsen and Kostinski* (2009). In this model the behaviour of a particular detector is characterized by a single parameter – the dead time τ.

Equation 3.1 is only valid if the photon counting process is a true Poisson process, i.e. the distribution of time intervals between two successive events is properly characterized by a Poisson distribution. This assumption is certainly fulfilled in case illumination of the detector does not change over the time span the signal is integrated. With lidars, however, the mean photon rate hitting the detector can vary due to e.g. air turbulence, faint cirrus modulating atmospheric transmission, and fluctuations of laser power[2]. Hence, the time series might be more clustered than a Poisson time series, and the true event rate N is underestimated (*Larsen and Kostinski*, 2009). This causes a positive bias in retrieved atmospheric temperatures. Although the exact magnitude of the bias is unknown, given that the root cause of the bias is a secondary effect one may safely assume that the value is smaller than the initial correction described by the model (equation 3.1).

How to measure the dead time

A common way to estimate the dead time of a particular detector is to compare count rates at different light levels. Two measurements are taken at each level, one with a neutral density filter placed between light source and detector (M_1), and one with the filter removed (M_2). A good value for the optical depth of the filter is 1. However, as long as the measurements are accurate enough almost any filter can be used. If the detector in question is not subject to dead time, one would expect the quotients of all count rate pairs

$$\frac{M_2}{M_1} = k_M \tag{3.2}$$

to be equal[3]. Any deviation is directly related to saturation effects.

An expression for the dead time τ can be found based on the ansatz

$$\frac{N_2}{N_1} = k. \tag{3.3}$$

This expression is similar to equation 3.2 except that measured quantities are replaced by true event rates N_2 and N_1. Substituting N_2 with the model (equation 3.1) expands

[2]The time series belonging to a certain range bin is considered. The laser power is independently measured by Joule meters (see figure 2.1). Typical variations in emitted pulse energy are 20% (1 sigma).

[3]This, of course, is not true for extremely low light levels with photon rates comparable to the dark count frequency of the detector.

equation 3.3 to

$$\frac{M_2}{N_1\left(1-\tau M_2\right)} = k. \tag{3.4}$$

This expression can be solved for the quotient

$$\frac{M_2}{M_1} = \frac{k}{1+k\tau M_1} \tag{3.5}$$

assuming that the count rate obtained with the neutral density filter in place (M_1) is low enough so saturation effects can be safely ignored. This leaves the factor k which is related to the optical density of the filter, and the dead time τ as unknowns. The factor k can be easily eliminated by repeating measurements with the same filter for different count rates M_1. Expanding numerator and denominator with $1-k\tau M_1$ and neglecting the quadratic term yields

$$\frac{M_2}{M_1} \approx k\left(1-k\tau M_1\right). \tag{3.6}$$

This expression can be further simplified making the approximation $kM_1 \approx M_2$. Thus, measurements of the ratios M_2/M_1 follow the linear function

$$\frac{M_2}{M_1} \approx f_\tau\left(M_2\right) = k\left(1-\tau M_2\right) \tag{3.7}$$

and the deadtime τ is readily determined by linear regression.

Dead time estimation from lidar measurements

My goal here is to develop a method for estimating the dead time from lidar measurements. One straightforward possibility would be to periodically insert a neutral density filter into the optical path similar to the procedure laid out in the previous section. This has the disadvantage that lidar measurements are heavily degraded when the filter is in place. A much better solution would be to use count rate variations of the lidar signal itself instead of modulating the signal with a neutral density filter.

As shown in figure 3.1, the count rate changes by approximately one order of magnitude per 10 km. So any two points on the green curve in figure 3.1 spaced roughly 10 km apart can be used to compute the left hand side of equation 3.5. In this case, the change in air density between the two selected altitudes as well as the r^2-dependence of the lidar signal serve as attenuator. Next, the power of the laser beam needs to be modulated somehow in order to enable measurements of the count rate quotient M_2/M_1 at different count rates. Fortunately, a closer look at the lidar return signal shown in figure 3.2 reveals that the signal strength fluctuates randomly, possibly due to turbulence in the lower troposphere and fluctuations in laser power inherent to this laser. These fluctuations are Gaussian-distributed, the standard deviation is approximately 40 percent of the mean. This turns out to be a sufficiently large range and no additional modulation is required.

Following the procedure presented in the previous section, the first step is evaluation of the quotients M_2/M_1. The largest count rates occur at the lower end of the backscatter profile. Thus, M_2 is chosen as the mean count rate determined from the altitude range 20 to 25 km altitude and M_2 from 30 to 35 km respectively. Both count rates are calculated for each single laser pulse within a group of 4000 pulses. Next, laser pulses are sorted according to M_2 and are subsequently binned. Then count rates of all pulses falling into a given bin are averaged to increase the signal-to-noise ratio. Finally, for each bin

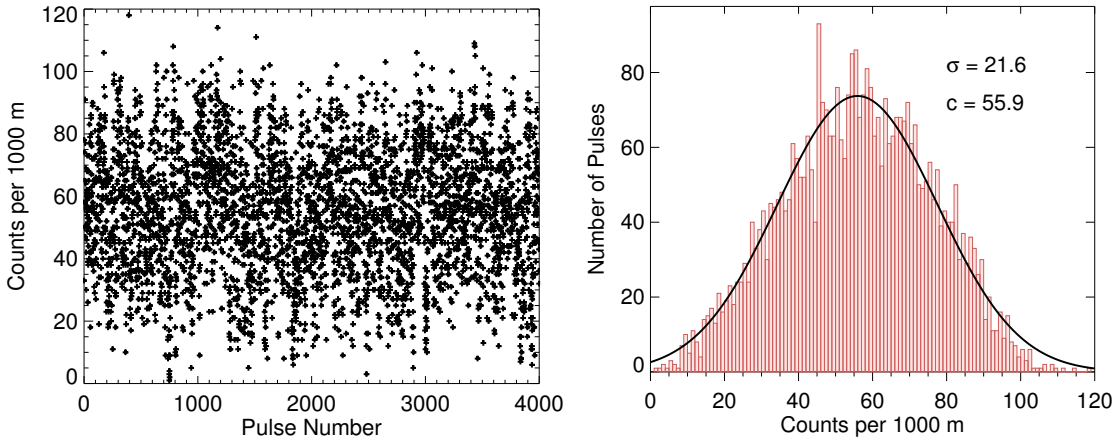


Figure 3.2.: Number of photons detected at 20 km altitude on April 24, 2011, 09:05 UT. The left panels shows the time series of 4000 laser pulses. A histogram of the same dataset is drawn on the right panel. The black curve is a Gaussian with a standard deviation of 21.6 counts/1000 m.

the count rate ratio $\bar{M}_2/\bar{M}_1$ is computed from averaged count rates. Bins with fewer than 40 laser pulses are rejected. It is important to choose the averaging interval as small as possible. 4000 laser pulses correspond to two minutes of lidar observations[4]. If the averaging period is substantially longer, it becomes more likely that atmospheric conditions in the stratosphere change. In this case the assumption that the quotients N_2/N_1 are constant may not be valid any more and derived dead time values become unrealistic.

In a second step all count rates $\bar{M}_1$ and $\bar{M}_2$ obtained from a single measurement series are averaged again in order to further improve the signal to noise ratio. The resulting averages $\tilde{M}_1$ and $\tilde{M}_2$ are then used to evaluate count rate quotients $\tilde{M}_2/\tilde{M}_1$. Figure 3.3 shows results from lidar measurements on April 23–25, 2011. In the left panel the signal ratio $\tilde{M}_2/\tilde{M}_1$ is plotted as function of the count rate $\tilde{M}_2$ together with error bars estimated from the standard deviation of the evaluated quotients. The right panel of figure 3.3 shows the number of values which contribute to the averages $\tilde{M}_1$ and $\tilde{M}_2$.

The last step involves fitting the model given by equation 3.7 to the signal ratio with τ and k as parameters. This can be easily done using all signal ratio values ranging from 0 to 12 MHz count rate. However, it turns out that the precision of the dead time τ can be significantly increased by limiting the data according to two criteria: First, only those data points are included in the fit whose averages comprise at least 10 percent of all available values. In case of the example presented in figure 3.3 this criteria limits the dataset to count rates ranging from 1 to 10.5 MHz, effectively excluding the first and the last data point. Second, only count rates above 4 MHz are considered. On first sight this is an arbitrary criteria which requires further justification. Due to the fact that noise scales with the square root of the signal[5], the relative uncertainty of measurements at low count rates is considerably larger than at high count rates. The counting noise increasingly dominates the error bars below 4 MHz[6] whereas above the main error source is atmospheric disturbances. The latter affects all measurements equally and error bars

[4]The lidar runs at 33 Hz pulse repetition frequency

[5]It is assumed that the photon counting process obeys Poisson statistics.

[6]This is the high count rate in the quotient M_2/M_1. The count rate of the denominator is about ten times lower.

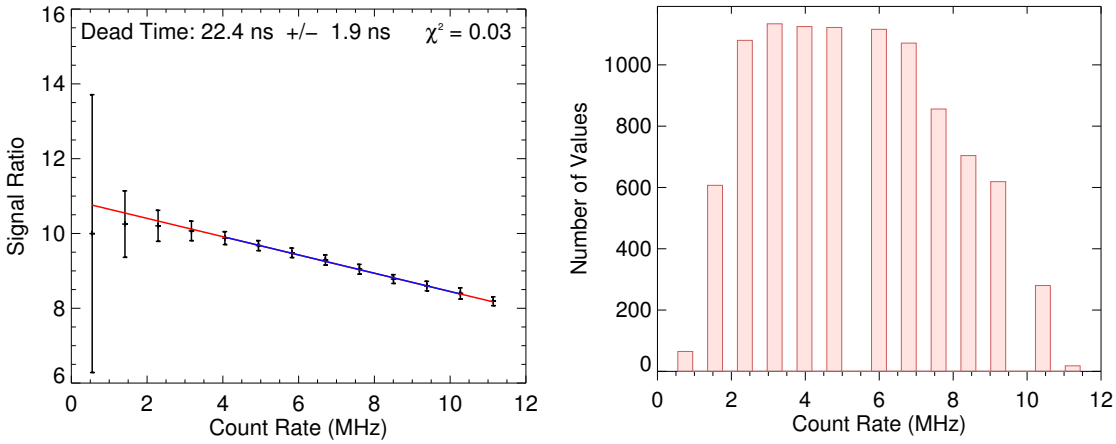


Figure 3.3.: Left panel: Signal ratio of lower $\tilde{M}_2$ and upper channel $\tilde{M}_1$ as function of photon count rate obtained from the dataset April 23–25, 2011. Dead time is estimated from a fit to the model of a non-paralyzing detector (blue line), the red line shows model results for count rate values not included in the fit. Right panel: Distribution of signal ratio values as function of count rate.

are independent of count rate as evident from the left panel of figure 3.3. Both criteria combined limit the dataset to data points marked with the blue line. The dead time determined from the fit is 22.4 ns ± 1.9 ns. The red line in figure 3.3 displays model predictions for data points not included in the fit.

Dead time estimates from all lidar observations exceeding 12 h duration in the time frame January 2011 to January 2012 are summarized in figure 3.4. The left panel shows the histogram, the right panel the values in chronological order. Dead time estimates which result from fits to fewer than four data points are removed as these are considered unreliable (see figure 3.3). Data points plotted in green are well separated from the main peak in the histogram and are therefore classified as outliers. A likely scenario that can explain unrealistically low dead time values is finely structured aerosol layers in the lower stratosphere. The aerosols modulate the count rate M_2 (altitude range 20 to 25 km) whereas the count rate M_1 (altitude range 30 to 35 km) is mostly unaffected due to lack of aerosols at higher altitudes. Thus, the value of the quotient M_2/M_1 is systematically enhanced at high count rates resulting in apparently lower dead time values.

With outliers removed, the mean of the remaining 27 dead time estimates amounts to 23.0 ns ± 0.2 ns. However, this value increases to 24.0 ns ± 0.3 ns if data points contributing to the mean are restricted to uncertainties < 2 ns (blue points in the right panel of figure 3.4). This compares with the pulse-pair-resolution of 25 ns as stated by the manufacturer of the photon counting unit[7]. Because stratospheric aerosols tend to cause a low bias as explained earlier, the larger value of 24 ns is considered more trustworthy even though less points contribute to the mean. Additional confidence is gained when comparing atmospheric temperatures measured with the high rate detector UV2 located behind the etalons and the low rate detector UV which picks up light in front of the etalons (see figure 2.2). This comparison is presented in section 3.3.3.

Based on the arguments mentioned above, $\tau = 24$ ns is chosen as parameter for the saturation model (equation 3.1) and raw photon count profiles are corrected accordingly. Lowering the dead time to 23 ns increases the temperature bias at 25 km altitude

[7]Hamamatsu C9744, prescaler set to 10

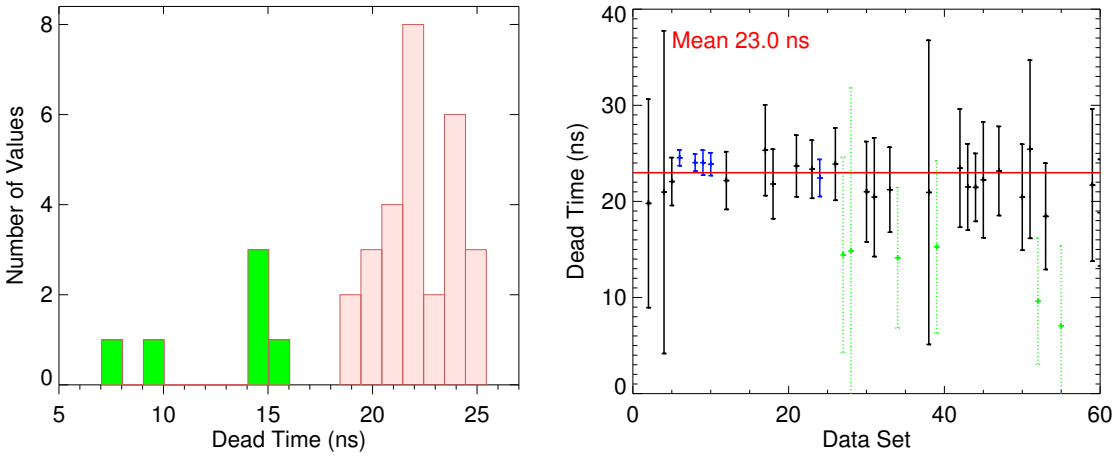


Figure 3.4.: Left panel: Histogram of estimated dead time values. Green data points are classified as outliers. Right panel: Estimated dead time values. The red line marks the mean value of 23.0 ns ± 0.2 ns, outliers (drawn in green) not included. The mean increases to 24.0 ns ± 0.3 ns if selection of data points is restricted to uncertainties < 2 ns (blue points).

to approximately 1.8 K (see figure 3.12). Note that this value represents the average temperature bias observed over a period of approximately one year. The temperature bias of individual measurements can be larger or smaller, depending on the strength of the lidar return signal.

3.2.2. Background estimation and subtraction

There are several light sources which superimpose the backscatter signal of the lidar. The brightest source during the day, of course, is solar radiation scattered by air molecules and aerosols. But also at night there are several light sources which contribute to the total received signal, e.g. scattered moonlight, stars, clouds illuminated by streetlamps, southern lights, and airglow. Most of the light coming from these sources is blocked by narrow-band optical filters. However, the remaining part adds to the backscatter signal and must be removed in the analyzing process. Another noise source which contributes to the total signal are dark counts originating from the photo multiplier tube detector.

All the sources mentioned above have in common that significant intensity variations happen on much larger time scales as the round-trip time of the laser pulse within the atmosphere. Thus, their signals add up to a constant background which can be measured independently of the backscatter signal in between successive laser pulses. For practical reasons, background measurements are usually taken from the lidar profile in an altitude region where no more backscatter photons are to be expected. The altitude range used in this work is 150 to 200 km.

Four different methods are implemented in the analysis software to estimate the background in lidar photon count profiles. The first two methods rely on the analysis of the histogram of counted photons as described in *Müller* (2007). *Müller* fits the photon count distribution to two models, a Gaussian and a Poisson distribution. The background value is then taken from the expected value of the models. Low background with on average less than 10 photons per bin is mostly best described by the Poisson distribution. This stems from the fact that the photon counting process is a Poisson process. When the number of background counts increases, e.g. during sunrise, the dis-

tribution determined from the histogram becomes more and more normally distributed, as expected from the central limit theorem. In this case the Gaussian model describes the histogram best.

The third method involves a linear fit $A_0 + A_1 \cdot z$ to all background bins. The mean background value is then taken from the zero-order term A_0.

Mean values computed for different altitude ranges are the basis of the fourth method. Starting with the altitude range 200 to 150 km, the interval is extended downward bin by bin until the Fe-layer is reached. Thus, the number of bins included in the background calculation increases in each step. For each of the intervals its mean photon count value is computed. Finally, from all these means the one with the lowest value is chosen as final background estimate. This selection criteria guarantees that all bins which contain any backscatter photons (Rayleigh or resonance scattering) are automatically rejected[8]. The advantage of this method is that the altitude range used to estimate the background signal can be safely extended, whereas in case of the other three methods any contamination by resonance scattering would bias the result. A large altitude range is desirable because precision increases with the sample size (number of bins).

The average of the estimated mean background values provided by all four methods weighted by the inverse variance is computed as final result. This ensures a smooth transition from one method to another as main contributor to the average in case the underlying model of the method in question degrades due to changing background at sunrise or sunset. Moreover, one method may be more susceptible to outliers, e.g. electronic noise, then others.

3.2.3. Correcting atmospheric transmission

As the laser pulse propagates through the atmosphere, it is attenuated by atmospheric extinction. This attenuation must be taken into account when atmospheric density profiles are derived from the lidar backscatter signal. Two different mechanisms contribute to atmospheric extinction: scattering and absorption. Scattering can be subdivided into aerosol scattering and Rayleigh scattering. Since aerosols are usually confined to the troposphere and lower stratosphere, extinction by aerosols will be neglected for the moment. This leaves Rayleigh scattering as dominating mechanism responsible for the atmospheric extinction. Molecular absorption in the ultraviolet region of the electromagnetic spectrum is mainly caused by ozone.

Atmospheric extinction is described in terms of the optical depth τ given by

$$\tau(z) = \sum_i \sigma_i \int_0^z n_i\left(z'\right)\,\mathrm{d}z', \tag{3.8}$$

where the index i runs over the involved species, σ_i is the scattering/absorption cross section, and $n_i(z)$ is the height dependent number density of the scattering/absorbing molecules. Including the contributions of Rayleigh scattering and absorption by ozone reduces this equation to

$$\tau(z) = \sigma_{\mathrm{Ray}} \int_0^z n_{\mathrm{air}}\left(z'\right)\,\mathrm{d}z' + \sigma_{\mathrm{O3}} \int_0^z n_{\mathrm{O3}}\left(z'\right)\,\mathrm{d}z'. \tag{3.9}$$

The Rayleigh scattering cross section $\sigma_{Ray} = 1.96 \times 10^{-26}\,\mathrm{cm}^2$ for the wavelength 386 nm is calculated from the analytic formula presented in *Bucholtz* (1995) and the

[8] Photons originating from resonance scattering or Rayleigh scattering add to the background signal and therefore increase the mean number of photons.

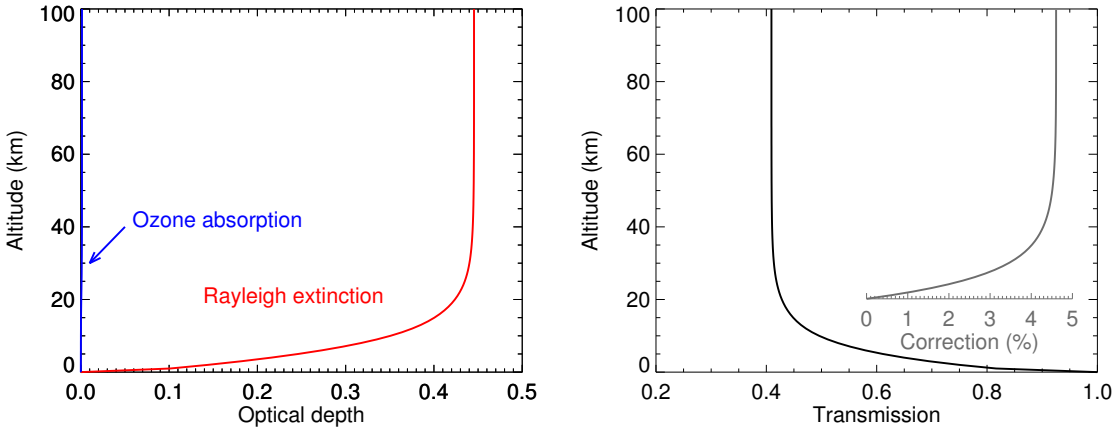


Figure 3.5.: Optical depth of the atmosphere at 386 nm wavelength (left) and transmission of the lidar signal (right). Also shown is the correction which normalizes the height-dependent transmission to the value at 20 km altitude.

number density of the air $n_{\mathrm{air}}(z)$ is taken from the MSIS-90 model (*Hedin*, 1991).

Ozone absorption cross sections for the UV region can be found in *Voigt et al.* (2001). The cross section at 386 nm is quite small because this wavelength lies in the minimum between the Hartley band and the Chappuis band. Values range from $3 \times 10^{-24}\,\mathrm{cm}^2$ at 203 K to $6 \times 10^{-23}\,\mathrm{cm}^2$ at 293 K. Latter value is still approximately 50 times smaller compared to the absorption cross section at 532 nm, the wavelength used by many Rayleigh lidars. Ozone number densities $n_{\mathrm{O3}}(z)$ in equation 3.9 are taken from the ozone climatology published by *Paul et al.* (1998).

Extinction profiles for ozone and Rayleigh scattering are shown in the left panel of figure 3.5. As evident from this plot, the contribution of ozone absorption to the total extinction at 386 nm is negligible. The asymptotic value of the optical depth of the atmosphere is 0.447.

The lidar signal is subject to attenuation on the way up to the scattering volume and again on the way down to the receiving telescope. Thus, the relative atmospheric transmission of the lidar signal scattered at altitude z is

$$\frac{I(z)}{I(0)} = e^{-2\tau(z)}. \tag{3.10}$$

The height-dependent transmission as function of altitude is illustrated in the right panel of figure 3.5. Also shown is the correction factor which normalizes the relative transmission to the value at 20 km altitude. This is the lowest altitude at which the Fe lidar can observe the atmosphere. As evident from figure 3.5, approximately 4.5% of the lidar signal is lost due to atmospheric extinction between 20 and 40 km altitude, and extinction becomes negligible above 60 km altitude.

By applying the correction shown in the right panel of figure 3.5 to measured backscatter profiles, the effect of the height-dependent atmospheric transmission is removed from the backscatter measurements.

3.3. Analysis of the Rayleigh signal

This section explains how the information present in the Rayleigh backscatter signal is used to deduce atmospheric temperatures. There are two basic methods: Doppler

measurements and integration of relative atmospheric density profiles. Doppler measurements rely on the relation of temperature and Doppler-broadening of the backscatter signal. This technique requires spectrally resolved measurements of the Rayleigh backscatter signal of which only very few lidar systems are capable of at present. *Höffner* (2010) reported prove-of-concept studies carried out with the Fe lidar in 2007. Although the system was optimized for Doppler measurements at that time, usable results were obtained only at low altitudes where the signal-to-noise ratio is largest. In Antarctica the lidar was operated in a different mode which is even less suited for Rayleigh Doppler measurements. For this reason no attempt was undertaken to analyze the Doppler-broadening in this work.

A more classic approach is the integration of relative density profiles deduced from the Rayleigh backscatter signal. This method is based on the assumption that the atmosphere is in hydrostatic equilibrium. Hence, it may be considered as less direct than Rayleigh Doppler measurements. However, hydrostatic integration is the method of choice for most Rayleigh lidars because of its simplicity. For instance, no spectrally resolved measurements are required. The theoretical background and the algorithm are presented in section 3.3.2.

Unlike with the more common Rayleigh lidars, two main problems arise when hydrostatic integration is implemented for the Fe lidar. First, resonance scattering in the iron layer superimposes the Rayleigh signal at altitudes above 70 km. Second, scanning of the laser wavelength[9] in combination with narrow-band optical filter leads to distortions in the Rayleigh signal. Both issues are addressed in sections 3.3.1 and 3.3.3 respectively.

3.3.1. Separation of Rayleigh scattering and resonance scattering

The iron layer covers approximately the altitude range 70 to 110 km in winter and 80 to 100 km in summer (*Yu et al.*, 2012). Depending on the power of the lidar, the lower edge of the iron layer resides thus well within the range where Rayleigh scattering is still detectable, e.g. the ALOMAR RMR lidar observes Rayleigh scattering well above 90 km altitude (*Schöch et al.*, 2008). The resonance signal obtained from the lower edge of the iron layer may therefore blend into the Rayleigh signal. This signal overlap poses problems for the temperature retrieval because the analysis algorithms rely on the physical properties of only one scattering mechanism, i.e. resonance scattering or Rayleigh scattering. For this reason special care must be taken to separate resonance scattering from Rayleigh scattering before the backscatter profiles can be further analyzed.

A combined analysis of Rayleigh and resonance scattering is possible from a theoretical point of view. However, the inclusion of Rayleigh scattering involves an additional degree of freedom, and at present the precision of the lidar measurements is not sufficient to determine this parameter with the required accuracy. In this work, the analysis of the Rayleigh backscatter profile is therefore restricted to altitudes below the iron layer.

The problem of superimposing scattering mechanisms is evident from the right panel of figure 3.6. The black curve describes the lidar return signal. Because of the exponential decrease in air density with altitude, pure Rayleigh scattering manifests itself as a straight line in logarithmic representation (lower half of the figure). Above 70 km altitude the contribution of resonance scattering to the lidar return signal becomes increasingly dominant, and the black curve deviates from the extrapolated Rayleigh background (black dashed line). At 77 km the difference in signal strength between resonance scattering and Rayleigh scattering amounts to one order of magnitude. Nevertheless, both components are still visible in the backscatter spectrum shown in the left panel

[9] Fe Doppler measurements require sampling of the Doppler-broadened resonance line of iron.

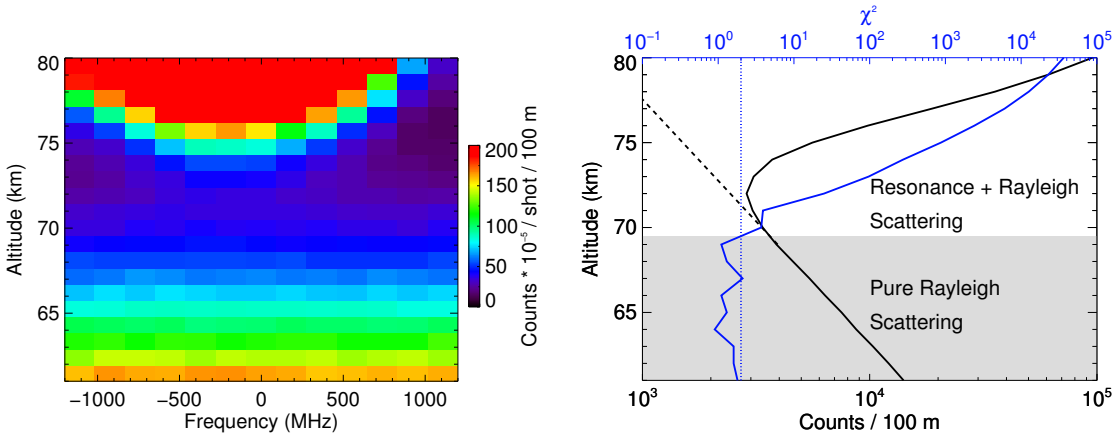


Figure 3.6.: 85 hours of lidar observations on August 13–17, 2011. The left panel shows the backscatter spectrum, the right panel the integrated photon count profile. χ^2-values obtained from linear fits to the spectrum are plotted in blue. The dashed line marks the threshold value $\chi^2 = 2$ used to distinguish pure Rayleigh scattering (no frequency-dependence) from Rayleigh scattering contaminated with resonance scattering (peak around −200 MHz).

of figure 3.6. Rayleigh scattering is independent of frequency within the precision of the lidar measurements. Consequently, Rayleigh scattering appears as horizontal band in the spectrum. On the other hand, resonance scattering shows a strong frequency dependence with significant backscattering being confined to the frequency range of the Doppler-broadened resonance line of iron. The iron line emerges from the Rayleigh background as peak near −200 MHz. The observed differences in the spectral response of the two scattering mechanisms is used to separate resonance scattering from Rayleigh scattering.

Determination of the lower edge of the iron layer

The lower edge of the iron layer, z_1, is defined as the altitude where the contribution of resonance scattering to the lidar return signal drops below the limit of detectability. Because there are no known aerosol layers between the upper edge of the stratospheric aerosol layer (approximately 35 km altitude, *Thomason et al.*, 1997) and polar mesospheric clouds (above 80 km altitude, *Klekociuk et al.*, 2008), the backscatter signal in the vicinity of z_1 comprises only resonance scattering and Rayleigh scattering. Any deviation from the Rayleigh spectrum can be attributed to resonance scattering, which in turn indicates the presence of iron atoms. Thus, the task of determining z_1 reduces to finding the lowest altitude where a deviation from the Rayleigh continuum can be detected in the measured backscatter spectrum.

The algorithm developed in this work comprises three main steps. First, the measured backscatter spectrum is corrected for frequency-dependent etalon transmission. This involves the convolution of the Doppler-broadened Rayleigh line with the etalon transmission function as described in section 3.3.3. Next, all measurements within a given bin of size 200 MHz by 1 km are averaged to increase the signal-to-noise ratio. The uncertainty is estimated from the square root of the number of counted photons. Then a linear model $a_0 + a_1\nu$ is fitted to the binned spectra for each altitude, where ν is the frequency associated with each bin, and a_0 and a_1 are coefficients determined by

the fit. The goodness of the fits is judged from the reduced χ^2 values. Strictly speaking, because the Rayleigh backscatter cross section is independent of frequency within the precision of lidar measurements, the Rayleigh spectrum can be described by the coefficient a_0 alone. An additional degree of freedom was included in the model in order to mitigate small deviations in the transmission model of the double etalon (section 3.3.3), and thus make reduced χ^2 values a more reliable indicator for the presence of resonance scattering.

In case of pure Rayleigh scattering reduced χ^2 values are expected to be close to unity. Figure 3.6 shows the χ^2 profile (drawn in blue) obtained from 85 h of lidar observations on August 13–17, 2011. Below 69 km altitude all values are < 2 and increase rapidly by many orders of magnitude above. Thus, the value $\chi^2 = 2$ is chosen as threshold to separate pure Rayleigh scattering ($\chi^2 \leq 2$) from resonance scattering superimposed on Rayleigh scattering ($\chi^2 > 2$). Starting at the centroid altitude of the iron layer, the lower edge z_1 is found as the altitude of the lowest bin which satisfies $\chi^2 > 2$. In case of the example presented in figure 3.6 z_1 evaluates to 70 km.

Backscatter profiles are then truncated at z_1, and the part above z_1 containing photons from both Rayleigh scattering and resonance scattering is rejected. The "clean" Rayleigh backscattering profiles can then be converted into temperature profiles as described in the next section.

3.3.2. Temperature conversion

The lidar backscatter profile is related to the height dependent volume backscattering cross section through the lidar equation (equation 1.4). In the absence of aerosols, this equation simplifies to

$$S(z) = \frac{E\,C}{z^2}\sigma_{\mathrm{Ray}} n_{\mathrm{air}} \mathcal{T}^2_{\mathrm{mol}}(z)\,, \tag{3.11}$$

where the volume backscattering coefficient has been replaced with the molecular number density of air n_{air} times the Rayleigh backscatter cross section σ_{Ray}. Equation 3.11 shows an important result: If the height dependent transmission $\mathcal{T}^2_{\mathrm{mol}}(z)$ is eliminated by carrying out the transmission correction discussed in section 3.2.3, then the lidar return signal represents the molecular number density of air, n_{air}, scaled by z^{-2} and multiplied by an unknown constant. The density profile can be converted to a temperature profile assuming that the atmosphere is in hydrostatic equilibrium

$$\mathrm{d}p = -M n_{\mathrm{air}}(z)\, g(z)\ \mathrm{d}z \tag{3.12}$$

and obeys the ideal gas law

$$p(z) = k_B n_{\mathrm{air}}(z)\, T(z) \tag{3.13}$$

(*Kent and Wright*, 1970; *Hauchecorne and Chanin*, 1980). The molecular mean weight of the air M is assumed to be constant, which implies a constant mixing ratio of the major atmospheric constituents. This assumption is justified for all relevant altitudes (20–80 km) since a significant decrease in M is observed only above 100 km (*Lübken and Zahn*, 1989).

Integrating equation 3.12 and combining the result with equation 3.13 leads to

$$T(z) = -\frac{M}{k_B}\int_z^\infty \frac{n_{\mathrm{air}}(z')}{n_{\mathrm{air}}(z)} g\left(z'\right)\ \mathrm{d}z'. \tag{3.14}$$

It is important to note that this expression does depend on the function $n_{\mathrm{air}}(z')/n_{\mathrm{air}}(z)$ which describes the relative change in air density between two altitudes z' and z. Thus,

$n_{\mathrm{air}}(z)$ in equation 3.14 can be replaced with any function $n(z) \propto n_{\mathrm{air}}(z)$. An obvious expression for $n(z)$ is obtained by solving 3.11 for n_{air} and dropping all constant factors:

$$n(z) = S(z)\, z^2\, \mathcal{T}_{\mathrm{mol}}^{-2}(z) \tag{3.15}$$

The replacement of $n_{\mathrm{air}}(z)$ in equation 3.14 with $n(z)$ reveals a major shortcoming: In order to carry out the integration, precise knowledge of $n(z)$ is needed at all altitudes. This is not possible for two reasons: First, above 70 km altitude the resonance signal may blend into the Rayleigh signal, thus contaminating the air density measurements as discussed previously in section 3.3.1. Second, the signal-to-noise ratio decreases rapidly with altitude, and numerical problems lead to unrealistic values if the noise content of $n(z)$ becomes large. In order to solve this problem, the integral in equation 3.14 is split into two parts

$$T(z) = -\frac{M}{k_B}\int_z^{z_0} \frac{n(z')}{n(z)} g(z')\, \mathrm{d}z' - \frac{M}{k_B}\int_{z_0}^{\infty} \frac{n(z')}{n(z)} g(z')\, \mathrm{d}z', \tag{3.16}$$

where z_0 is the highest altitude with reliable signal. The second term covering the altitude range with poor or no lidar measurements can now be replaced with the evaluated temperature at z_0 using equation 3.14. This leads to

$$T(z) = \frac{n(z_0)}{n(z)} T(z_0) + \frac{M}{k_B}\int_{z_0}^{z} \frac{n(z')}{n(z)} g(z')\, \mathrm{d}z'. \tag{3.17}$$

The derived temperature profile $T(z)$ becomes reasonably independent of the initialization temperature $T(z_0)$ at altitudes below $z_0 - 2H_n$ (H_n is the density scale height of the atmosphere). An assumed uncertainty of ± 30 K in $T(z_0)$ reduces to ± 3 K at $z_0 - 2H_n$ (*Lübken and Zahn*, 1989). Since $T(z_0)$ can not be estimated from the lidar return signal, it is commonly taken from reference atmospheres (e.g. NRLMSISE00, CIRA86) (*Schöch*, 2007) or other lidar measurements (*Rauthe et al.*, 2008). In this work $T(z_0)$ is taken from the operational ECMWF analysis (Integrated Forecast System version Cy36r1, T1279) for the location of Davis because no other lidar measurements are available in the required altitude range. Depending on the signal-to-noise ratio of lidar measurements and the location of the lower edge of the iron layer, z_0 ranges between 55 km and 75 km. Assumed uncertainties in ECMWF model temperature are ± 20 K in this altitude range.

The values for $g(z)$ are obtained from

$$g(z) = g_0 \left(\frac{R_0}{R_0 + z}\right)^2, \tag{3.18}$$

where $g_0 = 9.8066\,\mathrm{m^2\,s^{-1}}$ and $R_0 = 6356.75$ km (*Atmosphere*, 1976). The mean molecular weight $M = 28.964$ is taken from *Krueger and Minzner* (1976).

3.3.3. Correcting etalon transmission

In the derivation of the temperature profile from lidar backscatter measurements (section 3.3.2) the scaled lidar return signal $n(z)$ (equation 3.15) was assumed to be proportional to the molecular number density of air. This simplification is, however, only valid if the signal is detected by a broadband receiver, i.e. the spectral bandwidth of the receiver is much larger than the Doppler-width of the backscattered laser light. In contrast, all modern lidars with daylight capability employ narrowband spectral filters in order

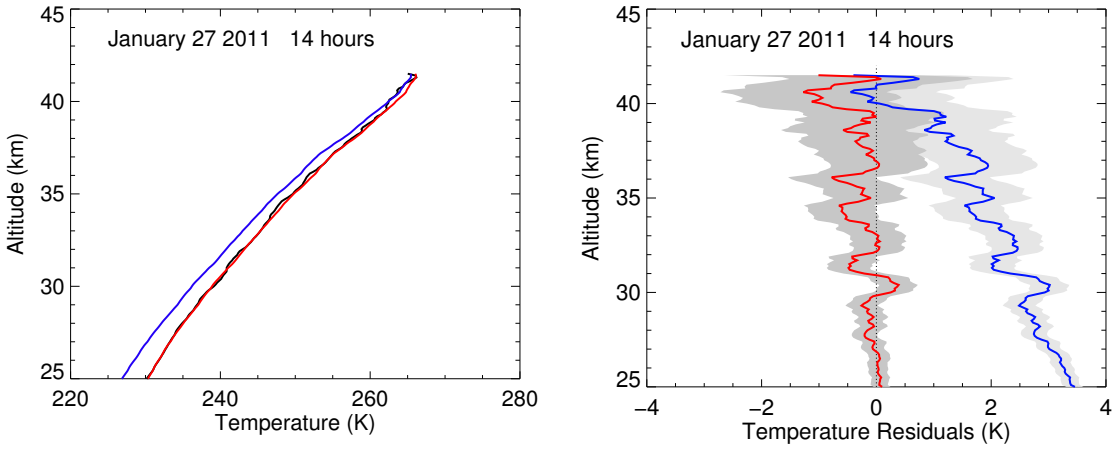


Figure 3.7.: Left: Comparison of temperature profiles; black line: UV detector (picks up light before etalons), blue: UV2 detector (behind etalons), red: UV2 detector with spectral correction. Right: The differences between the UV detector profile and UV2 detector profiles, where the shaded areas indicate the quadratic sum of the temperature uncertainties for both detectors.

to minimize the solar background (e.g. *Chen et al.*, 1993; *McKay*, 1999; *Höffner and Fricke-Begemann*, 2005), and the received spectrum of the backscattered light is thus modified to varying degrees. Since the Doppler-width of the backscattered laser light depends on atmospheric temperature, this may result in $n(z)$ not being proportional to the number density of air. The latter causes the derived temperature (equation 3.17) to deviate from the atmospheric temperature if the modified spectrum is not properly taken into account.

The UV branch of the receiver systems includes several bandwidth limiting elements (see figure 2.2): two interference filters with 1 nm and 0.3 nm, as well as two etalons, the smaller of the two being only 2.7 pm wide. Ideally, all elements are aligned in such a way that their transmission maxima are centered at the resonance line of iron at 386 nm. In this case the resulting transmission spectrum is symmetric and maximum transmission is obtained for the center wavelength.

Both interference filters are broad enough so that their transmission spectra can be assumed to be flat with regard to the Doppler-width (≈ 2 pm) of the lidar return signal. The bandwidth of the etalons is, however, considerably smaller and the influence on the lidar return signal can not be neglected. Spectral filtering by both etalons reduces the transmission of the Doppler-broadened laser line to 82% at 200 K atmospheric temperature and 77% at 300 K respectively. This effect has to be taken into account when temperatures are derived from lidar backscatter measurements. Failure to do so introduces a temperature- and thereby altitude-dependent bias of up to 5 K in retrieved atmospheric temperatures. One example is shown in figure 3.7. The blue curve in the left panel is obtained without any spectral correction. If the filter effect of the etalons is taken into account, the profile shifts to higher temperatures marked by the red line. For reference, the temperature profile computed from spectrally undisturbed backscatter measurements is also given (black line). Latter measurements employ the UV detector which picks up light in front of the etalons (see figure 2.2). However, due to the lower signal-to-noise ratio, the temperature profile appears noisier and becomes less reliable at higher altitudes.

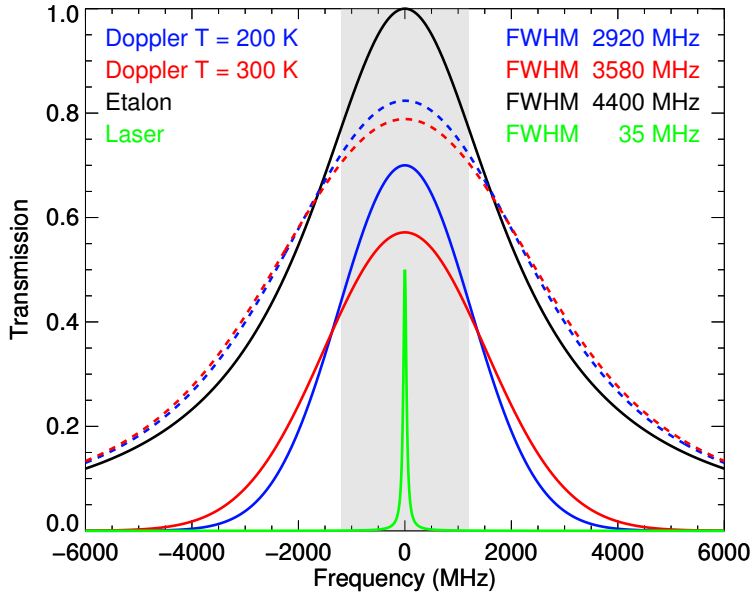


Figure 3.8.: Solid lines: Lorentz profile of the laser (green), Doppler-broadened laser profiles for $T = 200\,\mathrm{K}$ (blue) and $T = 300\,\mathrm{K}$ (red), Airy function of the etalon (FSR 128.9 GHz, Finesse 29.3) (black). The dashed lines result from convolution of the Airy function with Doppler-broadened laser lines. The shaded area marks the scan range of the Fe lidar.

Modeling the transmission spectrum

For simplicity, in this work the combined effect of both etalons is treated as a single bandwidth limiting element with an effective transmission function $\mathcal{T}^*$. Because this function is mainly dominated by the spectral response of the 2.7 pm etalon (FP UV2), $\mathcal{T}^*$ can be approximated as transmission spectrum of a single etalon. In this case the transmitted intensity, I_T, is given by

$$I_\mathrm{T} = I_0 \frac{1}{1 + F \sin^2(\phi/2)} \tag{3.19}$$

(*Demtröder*, 2008), where F is the *coefficient of finesse* and ϕ the phase difference. The maximum of this *Airy* type function is I_0 for $\phi = 2m\pi$, m being an integer. The transmission spectrum of an etalon has a series of peaks spaced by the free spectral range FSR. Thus, the term $\phi/2$ can be replaced with the frequency ratio $\pi\nu/\mathrm{FSR}$. This allows $\mathcal{T}^*$ to be written in the form

$$\mathcal{T}^*(\nu; F, \mathrm{FSR}) = \frac{1}{1 + F \sin^2(\pi\nu/\mathrm{FSR})}. \tag{3.20}$$

The common approximation $F \approx (2\mathcal{F}/\pi)^2$ relates the *coefficient of finesse* F to the finesse $\mathcal{F}$. In this case equation 3.20 becomes

$$\mathcal{T}^*(\nu; \mathcal{F}, \mathrm{FSR}) = \frac{1}{1 + (2\mathcal{F}/\pi)^2 \sin^2(\pi\nu/\mathrm{FSR})}. \tag{3.21}$$

Figure 3.8 illustrates the transmittance of an etalon with $\mathcal{F} = 29.3$ and 128.9 GHz free spectral range, and its influence on the Doppler-broadened laser line $\mathcal{L}_\mathrm{D}$. The laser is assumed to have a Lorentzian profile with 35 MHz linewidth (*Höffner*, 2012).

The spectral transmittance of the lidar receiver

$$\mathcal{T} = \mathcal{L}_{\mathrm{D}} \star \mathcal{T}^{*} \tag{3.22}$$

results from convolution of the Doppler-broadened laser line with the transmission spectrum of the etalon. In normal operations the Fe lidar uses a scan range of 2.4 GHz roughly centered at maximum transmittance. Even though this covers only about half of the bandwidth of the etalon, transmittance of the lidar return signal[10] $\mathcal{T}$ decreases at the edges of the scan range relative to the center by 12.5% at 200 K and 10.7% at 300 K. Hence, $\mathcal{T} = \mathcal{T}(\nu_{\mathrm{L}}; T)$ depends not only on atmospheric temperature T, but also on laser frequency ν_{L}.

The transmittance $\mathcal{T}$ can be computed from equation 3.22 provided the finesse of the etalon, the position of the transmission peak, and the free spectral range are known. How these parameters are determined from lidar measurements is described in the next section.

Determining etalon parameters from lidar measurements

The Fe lidar employs three detectors in the UV branch of the receiver (see figure 2.2). The first detector (UV) is located in front of the etalons and picks up approximately 8% of the incoming light. The remaining beam passes the first etalon (FPUV1), is split again, and passes the second etalon (FPUV2) before hitting the third detector (UV2). Hence, the first detector measures the spectrally undisturbed lidar return signal, while the third detector sees the combined influence of both etalons. In principle, the etalon transmission function $\mathcal{T}$ can be estimated from the backscatter spectrum measured by the third detector alone. However, laser power fluctuations as well as atmospheric disturbances give rise to excessive noise which makes it difficult to interpret the spectrum. In order to mitigate this problem, the signal measured by the third detector is normalized to the signal measured with the first detector. Because laser power fluctuations and atmospheric disturbances affect both signals in the same way, the signal ratio UV2/UV is independent of these effects. The altitude range over which the lidar return signal is integrated is 30 to 35 km. This is to make sure that only Rayleigh scattering contributes to the signal ratio and the spectra are not distorted by aerosols which may be present in the lower stratosphere.

Figure 3.9 shows the signal ratio UV2/UV measured on June 14, 2011. The red line marks the best fit of the model

$$\mathrm{M}(\nu; \gamma, \delta, \mathcal{F}) = \gamma \cdot \mathcal{T}(\nu - \delta, \mathcal{F}; \mathrm{FSR} = 128.9\,\mathrm{GHz}, T = 204\,\mathrm{K}), \tag{3.23}$$

the proportionality constant γ, the frequency offset δ and the finesse $\mathcal{F}$ being free parameters. These three parameter are optimized with the help of the non-linear least-squares fitting package *MPFIT* (*Markwardt*, 2009).

Fitting the model (equation 3.23) to the measured signal ratio is complicated by the fact that the atmospheric temperature T is needed as model input. A deviation of 20 K from the true temperature causes the estimated finesse to deviate by approximately 1 from the true value. The temperature retrieval depends, however, on the finesse and the frequency offset provided by the model. This dependency problem is resolved by iteration: An initial version of the atmospheric temperature profile (blue curve in figure 3.7) is computed from uncorrected backscatter measurements. This temperature profile

[10] Only Rayleigh scattering is considered. The transmittance of the etalon is vastly different in case of aerosol scattering or resonance scattering.

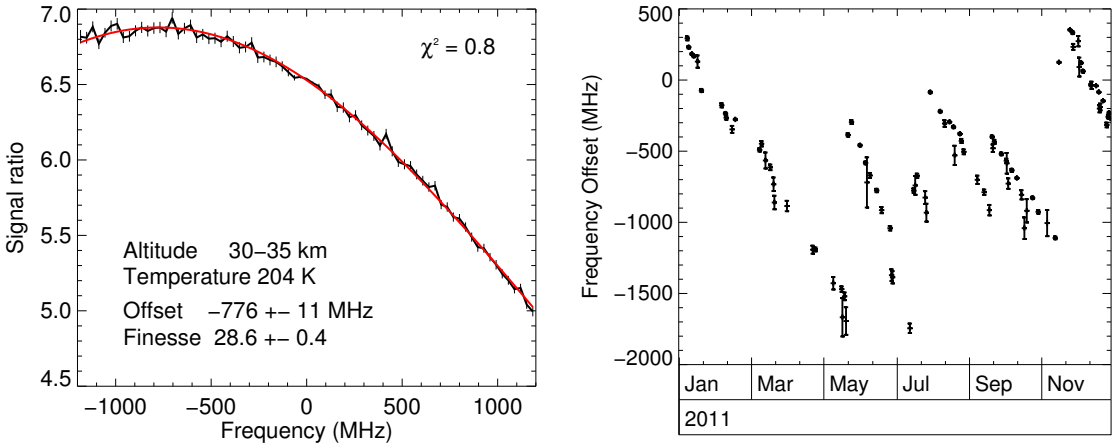


Figure 3.9.: Left: Example of a measured etalon transmission spectrum $\mathcal{T}$. Vertical lines indicate estimated uncertainties; frequency offset and finesse are determined from the fit (red curve). Right: Frequency drift of the etalon over the course of 12 months. Jumps indicate that the etalon was tuned and thereby the frequency of transmission shifted.

is used as input for the spectral model of the double etalon (equation 3.23) for the first iteration. Next, the backscatter profile is spectrally corrected using the transmission function provided by the model and the temperature retrieval is run again. Then the updated temperature profile is employed to update the spectral model in the next iteration. This sequence converges fast. In most cases the change in temperature from one iteration to the next falls below 10% of the estimated temperature uncertainty after four iterations.

The procedure of fitting the transmission model of the etalons to measured lidar data is repeated for every lidar observation. Figure 3.9 (right panel) shows the time series of estimated frequency offsets in 2011. A negative frequency drift in the order of 18 MHz per day is observed throughout the year. Exceptions are jumps when the etalons are retuned to keep the transmission maximum within the scan range of the lidar. The most likely cause for the frequency drift are leaks in the pressure vessel of the etalons. Because the etalons are pressure-tuned, the pressure vessels are partly evacuated. Air leaking into the vessel changes the refractive index between the etalon plates and the frequency of maximum transmission drifts accordingly. Smaller shifts in frequency also occur due to thermal expansion of the optical bench. As the box containing etalons and detectors is mounted directly on the outside wall of the lidar container, thermal gradients inside the box change whenever the outdoor temperature changes. Differential thermal expansion causes the etalons to move slightly and tiny changes in angle alter the length of the optical path between etalon plates. This, in turn, changes the frequency of maximum transmission.

3.3.4. Detecting aerosols

The temperature retrieval discussed in previous sections is based on the central assumption that the scaled backscatter profile $n(z)$ results from pure Rayleigh scattering. This assumption becomes invalid when aerosols are present because scattering by aerosol particles contributes to the lidar return signal. Two aerosol layers are known to exist in the middle atmosphere. While polar mesospheric clouds form above 80 km (*Klekociuk*

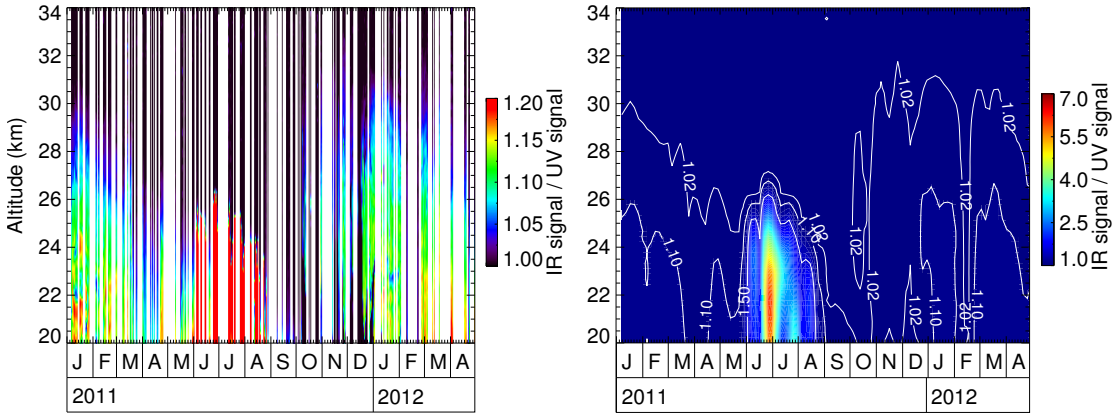


Figure 3.10.: Left: Daily color ratio profiles provide a rough estimate of the aerosol loading of the atmosphere. Note that the color scale is saturated in the period June–August due to the presence of polar stratospheric clouds (PSCs). Right: Contour plot highlighting the PSC signal. Profiles are smoothed with a 15-day Hann window.

et al., 2008; *Fiedler et al.*, 2009) and are thus outside the relevant region, the stratospheric aerosol layer (Junge layer) extends from the tropopause to approximately 30 km altitude (e.g. *Thomason et al.*, 1997; *Jäger*, 2005). The upper part of the Junge layer is therefore well within the altitude range which is observable by the Fe lidar, and methods to detect aerosols need to be implemented if Rayleigh temperature profiles are to be extended into this region. Ignoring the presence of aerosols in the temperature retrieval (section 3.3.2) results in an apparent increase in the number density of air, which in turn causes a low bias in derived temperatures.

The Fe lidar is equipped with additional detectors to allow for detection of aerosols in the infrared (see figure 2.2). The very narrow passband of the confocal etalon (approximately 15 MHz, *Höffner*, 2012) effectively suppresses most of the Doppler-broadened Rayleigh background in the received backscatter spectrum. In contrast, due to the large mass of the aerosol particles, light scattered by aerosols undergoes no significant Doppler-broadening. With essentially the same spectral bandwidth as the laser, the aerosol signal passes through the confocal etalon and is subsequently detected by detector APD IR3. Thus, the signal IR3 contains mainly the aerosol part of the backscatter spectrum while signal IR2 includes both aerosol and Rayleigh scattering. This information can be used to separate aerosol and Rayleigh scattering. A detailed description of this method will be made available in a PhD thesis by Timo Viehl.

In general, the backscatter characteristics of aerosol particles show a complex wavelength dependence (e.g. *Wagner et al.*, 2009). Although the aerosol volume backscatter cross section at the infrared wavelength 772 nm can be determined from Fe lidar measurements with high precision, this information can not be used to separate aerosol scattering from Rayleigh scattering in the UV. It is therefore not possible to retrieve accurate temperature profiles within aerosol layers based on Fe lidar measurements alone. In principle, atmospheric temperatures can be retrieved from the Rayleigh signal measured in the infrared. The infrared channel (APD IR2) is, however, not designed for this purpose, and overall accuracy for the aerosol-corrected temperature profiles retrieved in the infrared is lower than the accuracy of not aerosol-corrected temperature profiles which are obtained in the UV.

In this work the color ratio defined as the atmospheric signal measured at the 722 nm

wavelength divided by the signal measured at 386 nm is used to characterize the aerosol loading of the atmosphere. In case of pure Rayleigh scattering, the color ratio is independent of altitude provided differential molecular extinction can be neglected. However, the exact value still depends on several factors e.g. the ratio of the scattering cross sections at the two wavelengths, losses in the receiver, emitted pulse energies, and atmospheric transmission in the troposphere. The latter are also time dependent. In order to take out this variability, each daily color ratio profile is normalized to unity in the altitude range 35–40 km. The altitude of the normalization window is large enough that the assumption of pure Rayleigh scattering is justified.

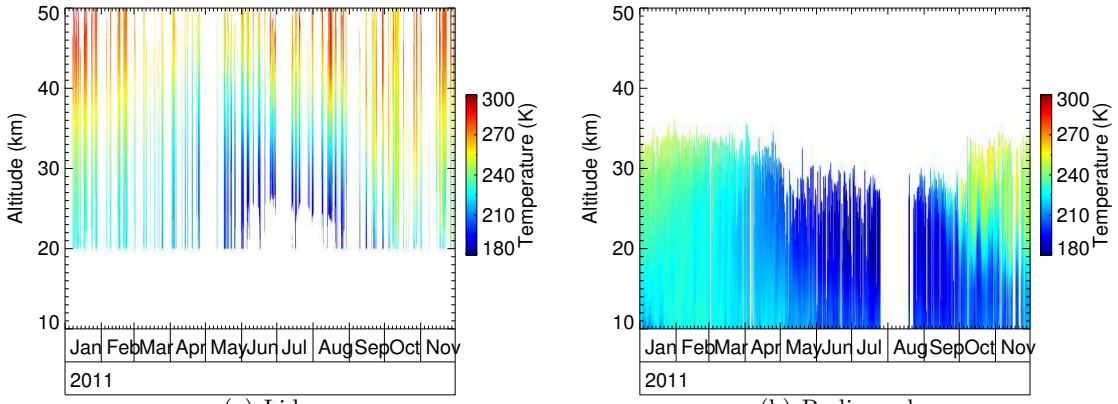
Figure 3.10 shows normalized color ratio profiles for the time frame January 2011 to April 2012. A value of e.g. 1.1 means that the effective volume scattering cross section is 10% larger in the infrared compared to the UV. In general, any value larger than unity is indicative of the presence of aerosol particles. The upper edge of the stratospheric aerosol layer shows a clear seasonal cycle. In summer aerosols can be detected up to approximately 32 km altitude, while in winter the upper edge of the aerosol layer can be as low as 24 km. An exception are the very strong aerosol signatures found in the time frame June–August between 20 and 26 km which result from polar stratospheric clouds (PSCs). PSCs may form only if the temperature falls below 197 K and are thus confined to the cold winter months (*Tabazadeh et al.*, 1994).

As discussed previously, the presence of aerosols causes a low bias in retrieved temperature profiles. In order to limit the maximum deviation, temperature profiles are truncated if the normalized color ratio exceeds 1.25. The implementation of this criteria leads to the exclusion of PSCs, while the rest of the aerosol layer remains mainly unaffected. It is noted that the threshold of 1.25 permits a fairly high aerosol load which may cause temperature deviations of several Kelvin. *Alpers et al.* (2004) report temperature corrections up to 8 K at 28 km altitude for a lidar system operating at 532 nm. However, given that the Rayleigh scattering cross section for the UV wavelength of 386 nm is about four times larger than at 532 nm, the expected temperature deviation caused by the presence of aerosols is smaller in case of the Fe lidar. Based on the comparison with radiosonde measurements (section 3.4) maximum temperature deviations in the order of 8 K occur at 25 km altitude, and less than 4 K at 28 km. Because aerosol loadings fluctuate, mean monthly aerosol temperature corrections are, however, considerably smaller. Hence, the stated values should be regarded as worst case estimates.

3.4. Validation with radiosonde data

The Australian Bureau of Meteorology (BoM) conducts a program of radiosonde soundings at Davis. Radiosondes are usually released twice a day at approximately 1100 UTC and 2300 UTC (*Innis and Klekociuk*, 2006). The balloons carry Vaisala RS80 radiosondes which record temperature measurements every 2 seconds (approximately 10 m vertical resolution). The precision of the temperature measurements in the region of interest is $\leq$0.4 K according to specifications given by the manufacturer.

Figure 3.11b shows temperature profiles recorded by radiosondes launched at Davis between January and November 2011. In summer the radiosondes often reach >30 km altitude before bursting of the balloon, while in winter the maximum altitude ranges between 22 and 28 km. No balloon releases took place during the gap in July/August due to failure of the hydrogen gas generator. Lidar temperature profiles obtained in the same time frame as the balloon soundings are displayed in figure 3.11a. The maximum integration time for individual profiles is 6 h centered around the time of the radiosonde



(a) Lidar (b) Radiosondes

Figure 3.11.: Data sets used in the comparison. The lower limit of the lidar profiles is given by the opening of the chopper (20 km) and aerosol contamination in austral winter (25 km). The height of radiosondes is limited in altitude due to bursting of the balloon at around 30 km altitude. The radiosonde data were provided by the Australian Bureau of Meteorology, and role of the Australian Bureau of Meteorology, Tasmania and Antarctica Regional Office, and the Bureau of Meteorology staff at Davis, Antarctica, is gratefully acknowledged.

launches. Profiles are truncated if PSCs are detected (see section 3.3.4), but otherwise no aerosol correction is applied.

Figure 3.12 shows the temperature residuals $T_{\text{Felidar}} - T_{\text{Radiosonde}}$ for all lidar profiles with corresponding radiosonde observations. Also shown is the number of points as function of altitude, i.e. the number of lidar profiles matching radiosonde profiles. This number reaches a maximum at approximately 25 km altitude, with less matching observations above due to balloon bursts, and less profiles below due to truncation of lidar profiles when PSCs are present. The mean of the residuals shows a small 0.5 K bias over most of the altitude range. This bias is significant since the standard error of the mean is below 0.3 K between 24 and 29 km altitude.

A striking feature of figure 3.12 is the increasing standard deviation of the residuals with decreasing altitude; the value doubles approximately from 2 K observed at 31 km altitude to more than 4 K at 23 km. In order to understand this increased spread, one needs to revisit seasonal changes of the aerosol loading. The left panel of figure 3.13 shows the seasonal variation of the temperature residuals $T_{\text{Felidar}} - T_{\text{Radiosonde}}$ at 26 km altitude with the color ratio (cross section of the left panel of figure 3.10) overlaid. A high negative correlation is observed throughout the year. The temperature residuals are slightly positive during the winter months (May to September) when aerosol concentrations are minimal, while large negative values in the range of -3 to -5 K are observed in summer (January) during maximum aerosol loading (PSCs are excluded in this analysis). This correlation suggests that retrieved Rayleigh temperatures are indeed affected by the presence of aerosols, and profiles need to be corrected below 30 km altitude. The magnitude of the correction can be deduced from the correlation analysis presented in the right panel of figure 3.13. A linear fit to the data (blue line) leads to an empirical correction of -7.7 K per 10% increase in color ratio. Note that color ratios are expected to be larger than 1.0, and smaller values are the result of the infrared signal not being properly corrected for detector saturation and etalon transmission. Hence,

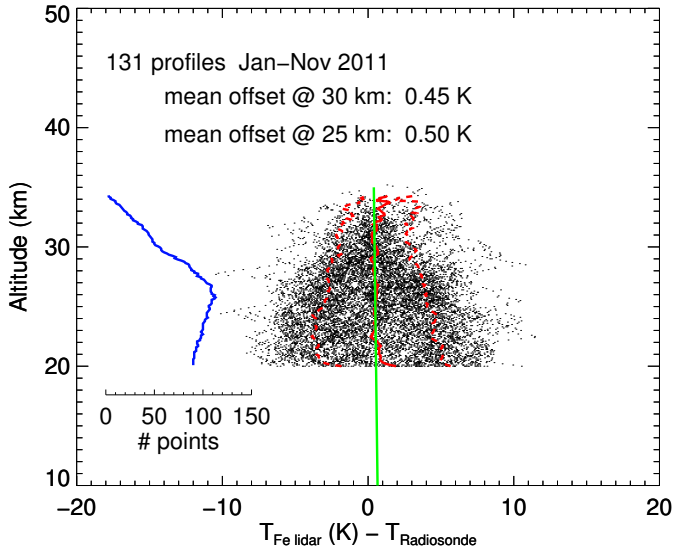


Figure 3.12.: Comparison between lidar observations and temperature measured by radiosondes. The solid red line marks the mean of the residuals, the dashed lines the standard deviation around the mean. The green line represents a linear fit to the mean between 35 and 30 km altitude.

color ratios < 1.0 (gray crosses) are excluded from the fit.

Using the empirical correction factor, the typical temperature bias caused by the presence of aerosols can be estimated from the color ratio measurements displayed in figure 3.10. With exception of the PSC season, color ratios > 1.1 occur only below 26 km altitude, and values > 1.02 are not observed above 31 km. The corresponding maximum temperature biases are -7.7 K and -1.3 K, respectively. Averaged over the entire year, the aerosol induced temperature bias is <5 K above 25 km altitude.

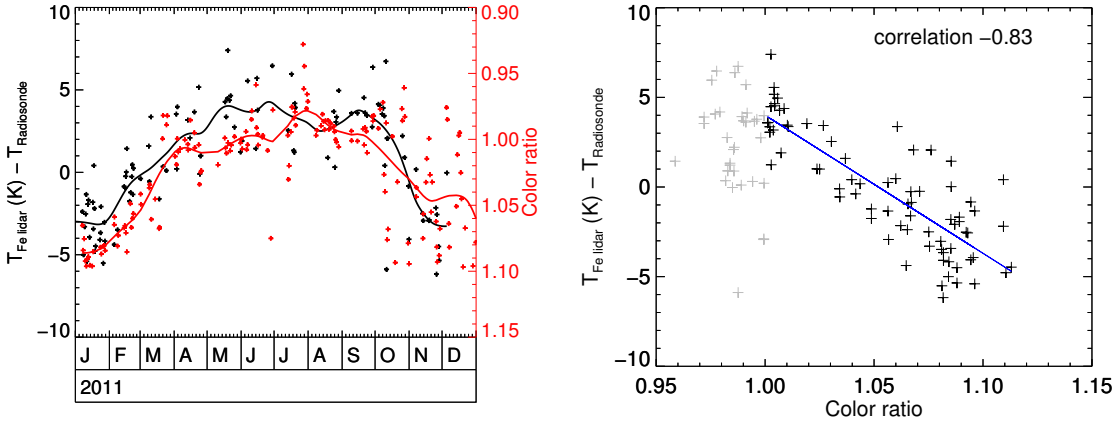


Figure 3.13.: Left: Seasonal variation of the temperature deviation $T_{\text{Fe lidar}} - T_{\text{Radiosonde}}$ at 26 km altitude and color ratio (signal at 772 nm divided by signal at 386 nm). Right: Scatter plot of the data shown in the left panel. The linear fit (blue line) is restricted to color ratios ≥ 1.0. See text for details.

4. Analysis of lidar temperature data

The analysis of the Davis Fe lidar temperature dataset holds the promise of improving the understanding of dynamical processes in the middle atmosphere, as, so far, there are no other instruments capable of providing temperature measurements with comparable resolution in the summer mesosphere at high latitudes. While there have been several lidar observations of the mesosphere-lower thermosphere (MLT) region in the northern hemisphere (e.g. *She et al.*, 2000; *Chu et al.*, 2005; *Höffner and Lübken*, 2007; *Gerding et al.*, 2008), very few do exist at southern mid to high latitudes (e.g. *Gardner et al.*, 2001). This disparity can be mainly attributed to the remoteness of the Antarctic continent and the harsh environment which makes lidar observations much more challenging.

All the lidar projects mentioned above have in common that the number of observation hours per year are in the low hundreds. This is sufficient for climatological studies, e.g. mean temperatures, where data of several years can be binned and averaged. However, this approach does not work anymore when the year-to-year variation of the background atmosphere becomes comparable or even larger then the magnitude of a particular phenomenon. An example is the atmospheric tide discussed in appendix B.1.

With more than 2600 hours of lidar observations, most of which were obtained in 2011, the Fe lidar at Davis produced the most extensive high resolution dataset available to date. The large dataset facilitates detailed studies of seasonal changes including tides and gravity waves (section 4.4). Moreover, the Fe lidar did observe the upper stratosphere in addition to Doppler-measurements in the MLT region, and successful temperature retrieval from stratospheric measurements is demonstrated as part of this work in section 3. This is a novelty for IAP metal resonance lidars, as in the past these lidars were used exclusively for mesospheric studies.

Extending temperature observations into the stratosphere opens up new possibilities to study the atmosphere over a large region of heights including both extreme ranges, the stratopause and the mesopause. The combination of stratospheric and mesospheric measurements is of particular importance for understanding processes that lead to vertical coupling in the atmosphere, e.g. the energy transported by gravity waves from the stratosphere to the mesopause region.

4.1. Description of the dataset

Two temperature datasets covering different regions of heights were available for this study. The resonance temperature dataset derived from iron Doppler-measurements covers the iron layer in the MLT region (approximately 80–110 km) and was provided by principle scientist Dr. Josef Höffner (*Höffner*, 2012). Two versions of this dataset are available: hourly profiles and daily profiles. The altitude resolution in both cases is 2 km. Temperature uncertainty estimates are provided. While reasonable over the large central part of the iron layer, uncertainty estimates appear to be unreliable at the very bottom and top of the profiles. As evident from figure 4.5, there are occasionally large jumps in the temperature toward the bottom and the top of the profiles. These jumps exceed the uncertainty threshold of 10 K and appear to be unlikely from a geophysical

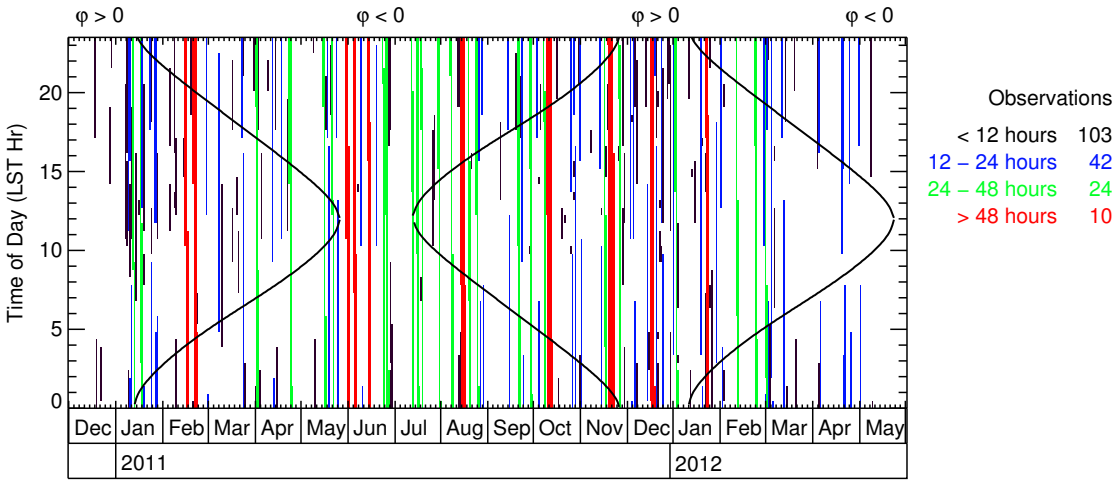


Figure 4.1.: Lidar observations as function of time of day and season. The approximate length of individual observations is indicated by color. The black curve marks points in time with solar elevation angle $\varphi = 0$.

point of view. Most likely, the jumps result from buildup of systematic errors when the iron density and therefore the signal-to-noise ratio is low. Unfortunately, there is no documentation available concerning the temperature retrieval beyond the basic principle discussed in *Lautenbach and Höffner* (2004). Because natural variability of the temperature can be large in the mesopause region (see section 4.2.1), reliable detection of inaccurate temperature profiles can not be implemented with simple algorithms e.g. by analyzing the deviation from the mean. For this reason, conspicuous temperature values are discussed on a case-by-case basis when required.

The second dataset is the Rayleigh temperature dataset produced as part of this work (chapter 3). Several versions with different temporal resolutions exist. The altitude range varies depending on integration time from approximately 20–50 km (1 h integration time) to 20–65 km (24 h integration time). Temperature uncertainty estimates are provided along with the temperature data. It should be noted, however, that two problems regarding temperature values and estimated uncertainties do exist. The presence of aerosols causes an altitude dependent temperature bias, the magnitude increasing from less than −1 K at 31 km altitude to approximately −5 K at 25 km (see section 3.4). Temperatures below 25 km are generally considered unreliable for a variety of reasons, e.g. saturation of the detector (see chapter 3), and data analysis should therefore be restricted to altitudes above 25 km. The second problem involves the upper 4 to 6 km of the temperature profiles. In this altitude range there might still be a significant discrepancy between the retrieved temperature profile and the true atmospheric temperature caused by unfavorable initialization of the integration process. This effect is hard to quantify because no other datasets with the required precision are available for comparison. Estimated deviations range from 3 to 6 K at the top of the profile and decrease rapidly toward lower altitudes (see section 3.3.2).

4.1.1. Temporal coverage

Both datasets, the Rayleigh- and the resonance temperature dataset were acquired with the same instrument. Thus, start and stop times of individual observations listed in both datasets are identical except for minor differences caused by different integration times and grids. Figure 4.1 gives an overview of observation times for the period December

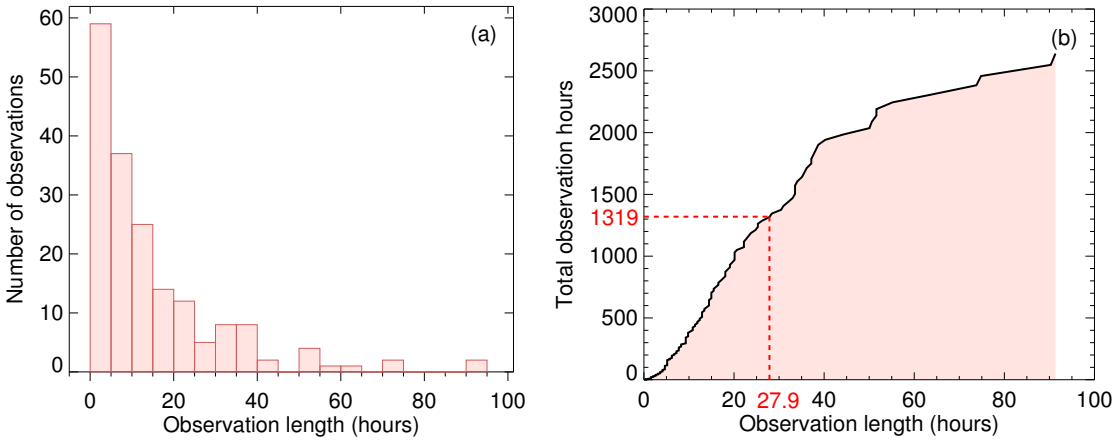


Figure 4.2.: (a) Histogram representation of the number of lidar observations versus observation length. (b) Cumulative distribution of observation hours. Red dashed lines mark the point which represents half of the grand total of all observation hours.

2010 to May 2012. Each lidar observation is marked as vertical bar, where start and end of the bar denotes begin and end of the observation. Hence, the length of a particular bar corresponds to the length of the corresponding lidar observation.

Lidar observations in figure 4.1 are classified into four groups: Continuous runs less than 12 hours in length are marked with black bars. 103 observations fall into this category, thus forming the largest group. The second group is characterized by blue bars and contains all observations between 12 and 24 hours in length. 42 observations belong to this group. 24 observations with periods between 24 and 48 hours make up the third group, where corresponding bars are colored green. Finally, the last group includes all observations with data acquisition periods in excess of 48 hours. These observations are marked with red bars in figure 4.1. Two out of the 10 observations belonging to this group have periods of continuous data acquisition in the order of four days. These data were acquired in August and November 2011. Two continuous observations in excess of 48 hours were obtained in February 2011, followed by three more runs in May/June of the same year, and one each in the months October 2011, December 2012 and January 2012. Since long observations permit precise determination of the background temperature field over long periods, they are of particular value when studying e.g. tides. Most of the long observations are concentrated in austral winter 2011. It is shown in appendix B.1 that clearest tidal signals are also found in months June to August 2011.

The number of observations versus observation length in histogram representation (bin size five hours) is shown in figure 4.2a. Up to the 20 hour mark the number of observations decreases approximately by one third for every 5 hours increase in observation length, while beyond that mark the distribution becomes more random although a general decline is still visible. Continuous lidar observation in excess of 50 hours are singular events. Even though the majority of all observations is shorter than 20 hours, the few observations which exceed 27.9 hours in length account for more than half of all observation hours as evident from figure 4.2b. It is also worth noting that in this cumulative representation the total number of observation hours increases approximately linearly for continuous observations shorter than 40 hours. Thus, if subsets of the lidar data are formed by excluding observations shorter than a particular length τ, the grand

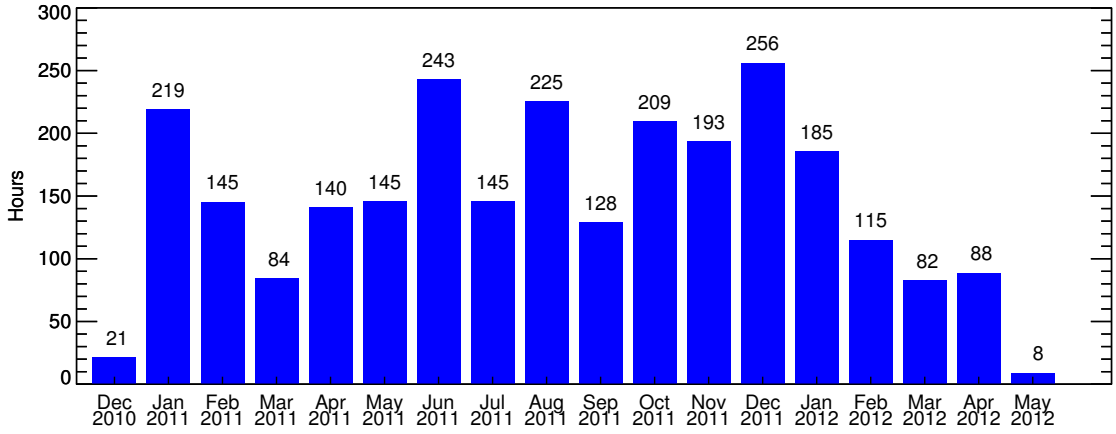


Figure 4.3.: Number of observation hours per month.

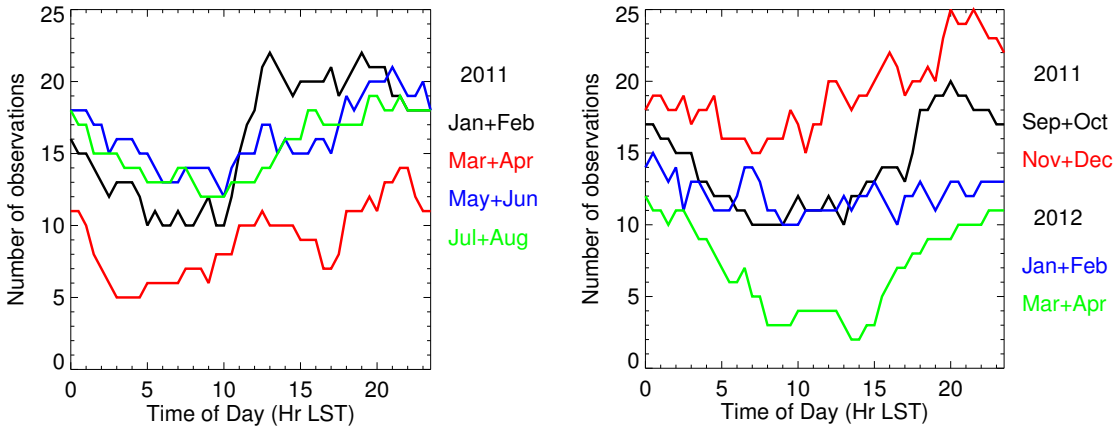


Figure 4.4.: Local time distribution of lidar observations. Each line represents the number of observations gathered in two consecutive months as function of time of day.

total of observation hours included in the subsets decrease approximately linear with increasing τ.

Figure 4.3 illustrates the number of observation hours per month for the combined primary and extended campaign. Months June to October 2012 are excluded from this chart since no lidar observations were obtained during these months. The month with the highest number of observation hours is December 2011 with 256 hours, closely followed by June 2011 with 243 hours, and August 2011 with 225 hours. The least data were acquired in May 2012 and December 2010. Excluding latter two months, the average number of hours per month is 163, which translates into 5.4 hours per day or 22.3%. Thus, averaged over the whole season, the lidar was observing the atmosphere approximately one fourth of the time. This rate is unprecedented for mesospheric lidar instruments. For comparison, the 11 year NLC dataset acquired with the ALOMAR RMR lidar reaches a long-term observation rate of 16.7% over the 75 day NLC season (*Fiedler et al.*, 2009). Outside the NLC season the rate is considerably lower.

The number of observations as function of time of day is shown in figure 4.4. Each line traces the number of measurements aggregated over the period of two months for a given local time. For reasons of better visibility, the data is split over two panels. The striking feature visible in all traces is a minimum occurring in the late morning hours in contrast

to a less pronounced maximum in the evening. Although part of this variation can be attributed to unfavorable weather conditions prohibiting lidar measurements, the daily routine of the lidar operators is certainly another contributing factor. Most observations start during the day and then may go on through the night. Over the course of the night it becomes less and less likely for the operator to begin a new observation as he is typically asleep when the lidar is not running. Also, whenever an observation ends at night due to unfavorable weather conditions or technical problems, it is unlikely that lidar operations are resumed before the next morning. All these factors combined lead to the typical decrease during the morning hours seen in the traces in figure 4.4. This effect is largest in the January/February 2011 data, where the number of observations increases by a factor of two around noon local solar time.

4.2. Mean temperatures

4.2.1. Monthly means

The lidar temperature datasets discussed in section 4.1 can be sorted by month and subsequently divided into 12 monthly subsets to find monthly mean profiles and assess the variability around the means. Temperature profiles selected by this means are shown in figure 4.5. The plots in this figure consist of 232 resonance temperature profiles and 138 Rayleigh temperature profiles obtained between January 2011 and May 2012, where the difference between the numbers is caused by splitting the resonance part of the observations at noon UT. Thus, for instance, a six hour long observation period ranging from 9 am to 3 pm is split into two three hour long observations in the dataset provided by *Höffner* (2012), whereas in the Rayleigh dataset the observation period of six hours is retained as single observation regardless of when the observation took place. All lidar temperature profiles shown in figure 4.5 are restricted to regions of heights where the estimated uncertainty of the temperature does not exceed 10 K.

For comparison, corresponding temperature profiles retrieved from the COSPAR International Reference Atmosphere (CIRA-86) (*Fleming et al.*, 1990) and NRLMSISE-00 (*Picone et al.*, 2002) are presented in figure 4.5 in addition to monthly means calculated from lidar observations. As evident from this figure, CIRA-86 temperatures in the winter mesopause region are in general considerably higher, with maximum deviation of approximately 30 K occurring in June. The temperature of the stratopause is, however, reproduced by CIRA-86 reasonably well for most months although significant differences occur in the height of the stratopause e.g. in June and August.

The NRLMSISE-00 model atmosphere reproduces the thermal structure of the stratosphere much better than CIRA-86. The height of the stratopause matches the lidar observations in most cases, although the temperature of the stratopause is underestimated in the winter months June to August by approximately 13 K. Large differences between observations and model occur mainly in the mesopause region e.g. in July and November ($\Delta T \sim 25$ K).

Looking at the variability as indicated by the spread of individual temperature profiles (black lines), two features stand out clearly: In March, temperature profiles in the mesopause region cluster around two groups at approximately 170 K and 195 K, and a similar fragmentation is observed near the stratopause in October. It will be shown later in section 4.2.5 that these fragmentations occur during the transition from the summer to the winter state of the atmosphere, and vice versa. The transition periods are usually accompanied by rapid changes in temperature, and the fragmentation seen in figure 4.5 hint at the two states.

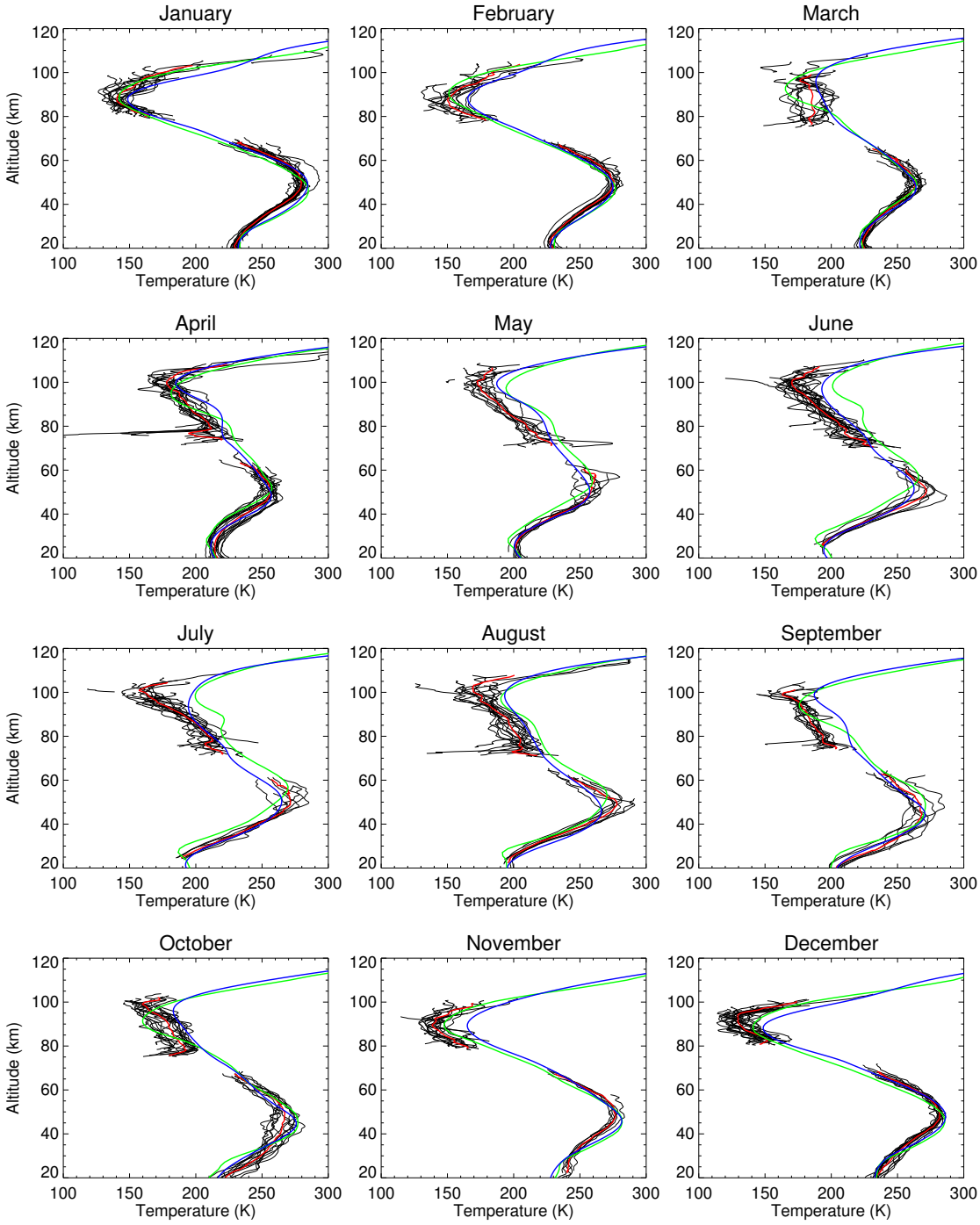


Figure 4.5.: Monthly mean temperature (red line) calculated from individual lidar temperature profiles (black lines). For comparison, temperature profiles retrieved from the COSPAR International Reference Atmosphere 1986 (CIRA-86, green line) and NRLMSISE-00 (blue line) are also shown.

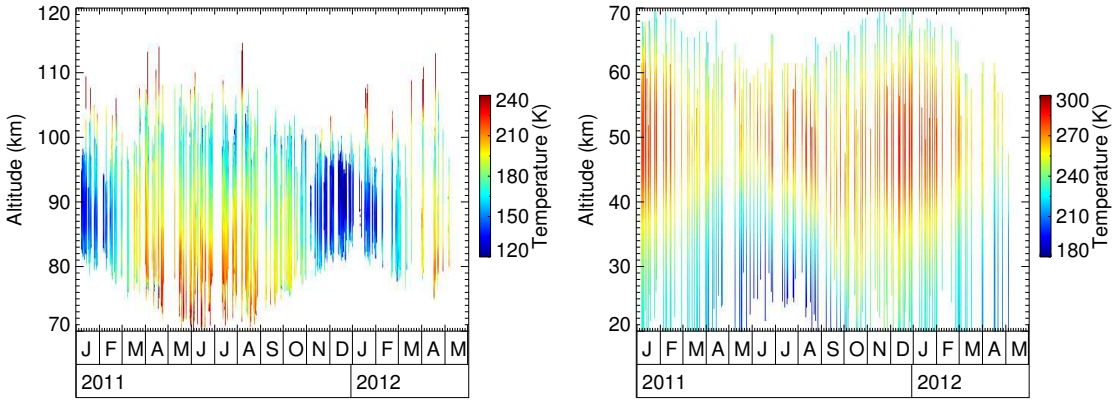


Figure 4.6.: Fe resonance temperatures provided by *Höffner* (2012) (left) and Rayleigh temperature measurements obtained as part of this work (right). Note the different color scales when comparing temperatures.

The second conspicuous feature is the occurrence of sharp drops in temperature near the very bottom of the Fe resonance temperature profiles, e.g. in April, and in some rare cases also near the top, e.g. June and August. These drops, as discussed in section 4.1, are most likely non-geophysical artifacts. Because there is no reliable detection method, it was decided not to attempt to remove the artifacts from figure 4.5.

The variability of the stratospheric temperature during winter is much larger than in the summer months December–February. This is consistent with the idea that most gravity waves in summer are blocked by the stratospheric jet. In winter, the jet reverses direction (the wind structure above Davis is shown in figure 4.31), thus opening the path for eastward propagating gravity waves. With more gravity waves reaching the stratopause region, variability of the temperature is larger in winter. Gravity wave propagation is discussed in more details in section 4.4.

In contrast to the stratosphere no distinct variation of the temperature variability can be observed in the mesopause region. Using the arguments employed for the stratosphere, the absence of any seasonal variation suggests that the gravity wave flux reaching the mesopause should be approximately constant. As will be shown in section 4.4.8 seasonal variations of the mesospheric gravity wave potential energy density is indeed small.

4.2.2. Mean seasonal temperature variation

More insight into seasonal variation of the thermal structure can be gained if temperature profiles are plotted as function of time. Figure 4.6 displays both datasets with the temperature color-coded, each vertical line representing one daily temperature profile. The altitude range covered by the Fe resonance temperature dataset shows large seasonal variations, e.g. temperature profiles ranging from 70 km to 115 km in August and 88 km to 94 km in January 2012. This large variation is linked to the abundance of iron atoms in the atmosphere, as Fe resonance temperatures can only be obtained at altitudes where the atomic iron density is sufficiently large. Since the atmospheric iron layer shows a strong annual cycle (*Gardner et al.*, 2011), this directly affects the altitude range of the Fe resonance temperature dataset.

The maximum altitude of Rayleigh temperature profiles displayed in the right panel of figure 4.6 follows a pattern similar to the lowest altitude of the Fe resonance tem-

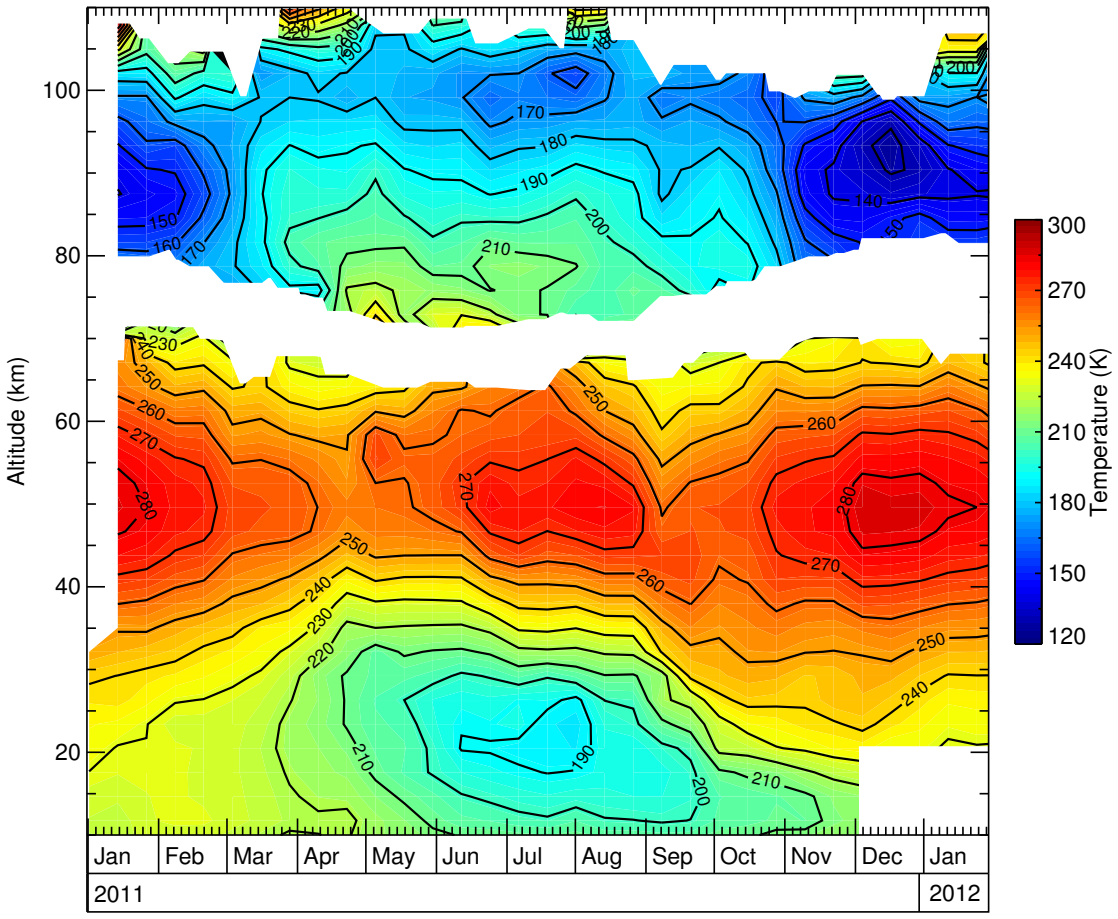


Figure 4.7.: Combination of Fe resonance temperature profiles (figure 4.6 left), Rayleigh temperature profiles (figure 4.6 right), and radiosonde data (figure 3.11) for 2011 and January 2012. Temperature profiles are smoothed with a 20-day moving average filter.

perature profiles. As discussed in section 3.3.1, reliable Rayleigh temperatures can only be obtained from Fe lidar data where there is no iron in the atmosphere. Hence, the expansion of the iron layer to lower altitudes in winter limits the maximum altitude of the Rayleigh temperature profiles. A gap at low altitudes in months May to August is caused by polar stratospheric clouds (PSCs). The backscatter signal from PSCs contaminates the signal originating from molecular scattering and makes it therefore impossible to retrieve Rayleigh temperature profiles in this altitude region.

Combining the two temperature datasets into a single color-coded plot yields the temperature map presented in figure 4.7. The time interval is restricted to the first 12 months when lidar observations are nearly uniform in signal strength and temporal coverage. In order to extend the altitude range down to 10 km, lidar temperature profiles below 30 km altitude are merged with radiosonde data (figure 3.11) which were provided by the Australian Bureau of Meteorology. Thus, the combination of all three datasets, Fe resonance lidar, Rayleigh lidar, and radiosonde, allows the mapping of the thermal structure from the tropopause up to the lower thermosphere.

The most striking feature of the temperature map presented in figure 4.7 is the dipole structure formed by the stratopause and the mesopause in summer (December), i.e. low mesopause temperatures coincide with high stratopause temperatures. This has previ-

ously been noted (e.g. *Gerding et al.*, 2008). Although less distinct than in summer, similar concurrent temperature extrema are also observed in austral winter (July/August).

Another prominent feature in figure 4.7 is the two-level structure of the mesopause. In march, at the end of summer, the mesopause jumps from approximately 88 km to 98 km and stays at this elevated level throughout winter. Then, at the end of October, the mesopause jumps back to the lower altitude, followed by slightly elevated levels in December. The concept of a two-level mesopause has previously been discussed by *She and von Zahn* (1998) and is strongly supported by more recent lidar observations at various latitudes in the northern hemisphere (*She et al.*, 2000; *Fricke-Begemann et al.*, 2002b; *Gerding et al.*, 2008) as well as satellite observations (e.g. *Xu et al.*, 2007). Figure 4.7 confirms this two-level structure for a high-latitude observing site in the southern hemisphere.

Looking at the annual variation of temperatures at constant altitudes, the stratopause shows a quite different behaviour compared to the lower stratosphere or the mesopause region. In the latter case variations are dominated by a strong annual cycle, while the temperature in the stratopause shows a clear semi-annual oscillation. This topic will be elaborated further in section 4.2.5.

4.2.3. Comparison with ECMWF data

The operational model of the European Center for Medium-Range Weather Forecasts (ECMWF) provides temperature analysis every six hours at 0 UT, 6 UT, 12 UT, and 18 UT (Integrated Forecast System version Cy35rl, T799). In order to compare this dataset with lidar observations, the temporal resolution of the ECMWF data needs to be reduced by averaging the four daily profiles to match the lidar data with integration times of up to 24 hours. The comparison of temperature profiles with six hours integration time using the native resolution of ECMWF data was abandoned because, due to the lower signal-to-noise ratio, lidar temperature profiles usually do not reach above the stratopause.

The averaged ECMWF profiles are displayed in the top panel of figure 4.8, and the differential image $T_{\mathrm{ECMWF}} - T_{\mathrm{Lidar}}$ is shown below. Both images are smoothed with the same 20-day averaging filter as used in figure 4.7. As evident from the differential image, ECMWF data agrees well with lidar observations below the stratopause during austral summer. In winter, however, there appears to be a systematic shift in ECMWF data towards higher temperatures, as indicated by the blue colored areas. This mainly affects months May to July, and to a lesser degree August. A similar difference between model data and lidar observations was previously reported by *Schöch* (2007). *Schöch* compared a multi-year dataset obtained with the ALOMAR RMR lidar (69° N) with ECMWF model data and found a mean bias of approximately 3 K for the winter months November–February.

Below 30 km altitude the comparison between lidar data and model data becomes less meaningful due to a possible bias in lidar data caused by the presence of aerosols. Stratospheric aerosol layers increase the strength of the lidar backscatter signal, thus making the temperature apparently lower. Possible signatures of aerosols visible in the differential image (figure 4.8) are regions between approximately 22 and 28 km in November, December and January, where lidar temperatures are slightly lower than ECMWF temperatures (visible as positive temperature difference).

Discrepancies between the two datasets are observed near the stratopause throughout the year. A prominent feature in the differential image is the downward progressing region with negative temperature deviations (higher lidar temperatures) starting

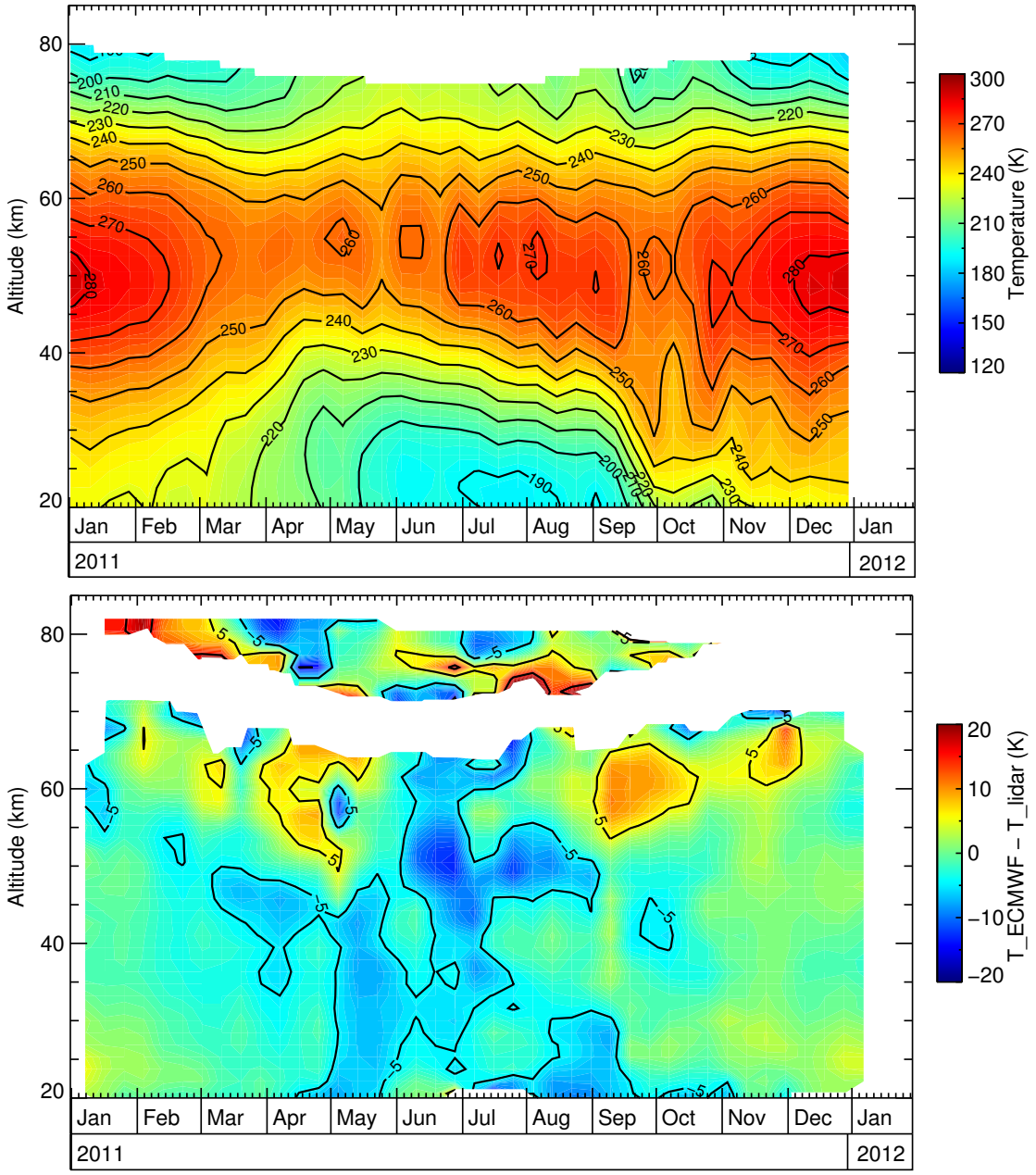


Figure 4.8.: Temperature retrieved from ECMWF model data (top) and differences between lidar mean temperatures and ECMWF temperatures (bottom). Values are smoothed with a 20-day moving average filter.

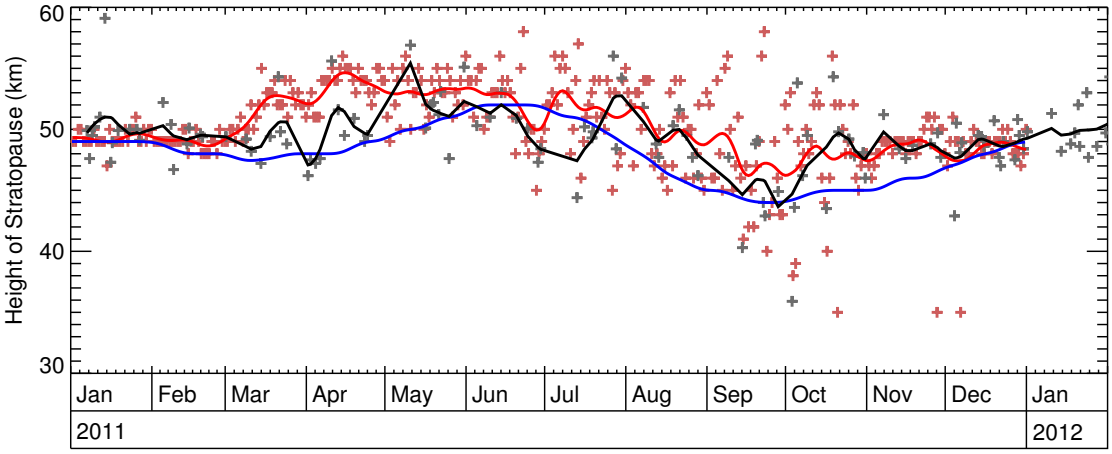


Figure 4.9.: Height of the stratopause as determined from lidar observations (black) and ECMWF temperature data (red). Values are smoothed with a 20-day Hann window (solid lines). For comparison, the trend of the stratopause height determined from NRLMSISE-00 data is also shown (blue line).

at 60 km altitude in January and reaching 40 km in May. This structure is accompanied by regions with positive deviations (lower lidar temperatures) approximately 10 km above, thus forming a dipole which can be explained by a mismatch in the height of the stratopause. As can be seen from figure 4.9, in January and February the stratopause as observed by lidar resides above the stratopause determined from ECMWF data. In March, however, the situation reverses and the average difference in height increases until May. A similar mismatch in stratopause altitudes occurs in July and September. Large temperature deviations are visible in the differential image (figure 4.8) in both cases. This suggests that large differences between lidar temperatures and ECMWF model temperatures in the stratopause region are mainly caused by the inability of the ECMWF model to reproduce the correct height of the stratopause above Davis Station.

4.2.4. Comparison with observations in the northern hemisphere

Figure 4.10 shows annual temperature variations observed at Davis and the ALOMAR temperature climatology derived from Rayleigh lidar measurements at the ALOMAR observatory at 69° N (*Schöch et al.*, 2008). The ALOMAR climatology incorporates 834 individual lidar observations in the years 1997–2005 and can thus be regarded as representative dataset for this high latitude site. It should be noted, however, that lidar observations start at 30 km altitude, and below the climatology is based on ECMWF model data.

The time axis of the Davis contour plot in figure 4.10 is shifted by six months in order to facilitate the comparison of seasonal variations in both hemispheres. Also shown in this figure is the differential image $T_{\mathrm{Davis}} - T_{\mathrm{ALOMAR}}$ which highlights interhemispheric differences. However, it should be noted that the Davis dataset covers only 15 months. Since the atmosphere above Davis is known to show significant year-to-year variability (e.g. *Morris et al.*, 2012; *Kaifler et al.*, 2013), climatological interpretations of the limited Davis dataset should be made with caution. Nevertheless some statements can be made based on the available data: (1) Except for the winter months, there is no significant temperature difference in the stratopause region. (2) The winter stratopause in the southern hemisphere is approximately 10 K warmer. The warming is consistent with an increased gravity wave flux in winter above Davis, resulting in larger heating

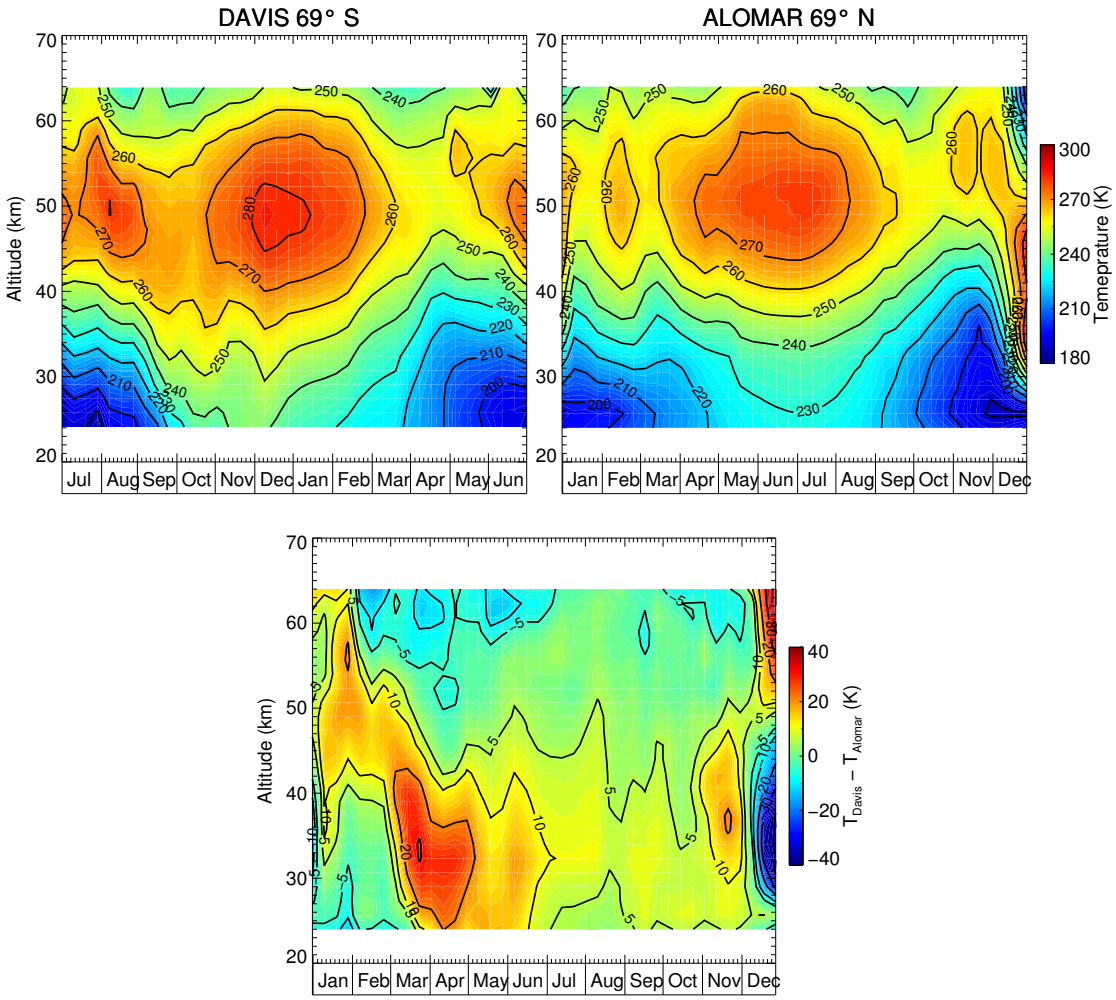


Figure 4.10.: Comparison between stratospheric temperatures in the southern hemisphere (Fe lidar data acquired at Davis, 69° S) and stratospheric temperatures in the northern hemisphere (ALOMAR, 69° N). The ALOMAR lidar climatology is taken from *Schöch et al.* (2008), see text for details.

rates near the stratopause. (3) The upper stratosphere in summer is approximately 5–10 K warmer in the southern hemisphere. On the other hand, the strong warming in September/October around 30 km altitude may be linked to an early breakdown of the polar vortex and represents thus an anomaly rather than the climatological average.

In the northern hemisphere, the higher stratospheric temperatures in December and January are remains of stratospheric warmings. Because of their large occurrence frequency, the signature of these warmings shows up in the climatological mean even though stratospheric warmings are singular events. Stratospheric warmings are usually accompanied by mesospheric coolings, and the resulting dipole structure of the vertical temperature profile can be seen in the December ALOMAR data. This structure transfers inversely into the differential image $T_{\mathrm{Davis}} - T_{\mathrm{ALOMAR}}$, as in general no stratospheric warmings are observed in the southern hemisphere. One notable exception was the major stratospheric warming in 2002 (e.g. *Varotsos*, 2004). No stratospheric warming was observed above Davis in 2011.

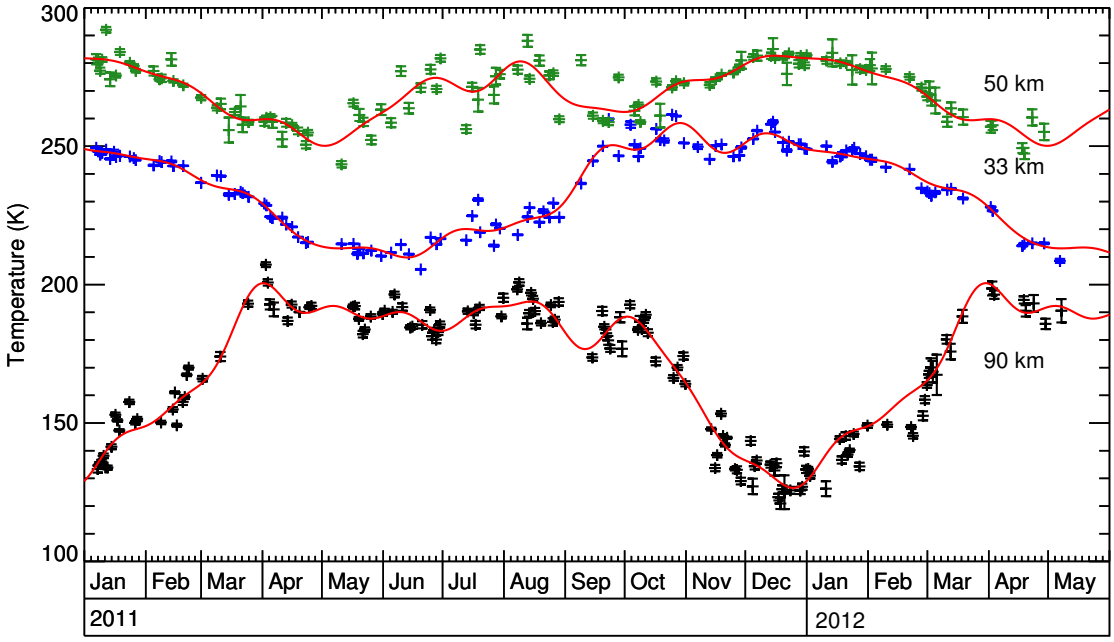


Figure 4.11.: Seasonal variations in temperature for three different altitudes. Solid lines represent non-linear fits with 10 harmonic components.

4.2.5. Spectral analysis of seasonal variations

The annual cycle of temperatures was already briefly mentioned in section 4.2.2 and is now discussed in more detail. Figure 4.11 shows temperature variations for three altitudes 33 km, 50 km and 90 km. Observations are truncated at noon each day, and the maximum integration time is therefore limited to 24 hours. Hence, each observation in figure 4.11 represents the daily mean temperature at the selected altitude. While 90 km and 50 km correspond approximately to the summer mesopause and stratopause respectively, 33 km was chosen because the uncertainty of Rayleigh temperatures (statistical as well as systematic) is believed to be smallest at this altitude. In general, statistical uncertainties become smaller with decreasing altitude because of the increasing signal-to-noise ratio. However, below approximately 33 km altitude other error sources such as aerosol contamination of the Rayleigh signal and detector saturation effects begin to dominate, thus increasing the total uncertainty.

As evident from figure 4.11, the temperature variation in the lower stratosphere as well as in the summer mesopause region is dominated by the annual cycle, while near the stratopause the variation contains a strong semi-annual component. To help quantifying the seasonal variations, temperature time series are decomposed into harmonic components by fitting the function

$$T(t) = A_0 + \sum_{i=1}^{n} A_i \cos\left(\frac{2\pi i}{365}(t - \phi_i)\right) \tag{4.1}$$

to the data, where t is the time in days since January 1, 2011, and ϕ_i are the phases and A_i the amplitudes of the harmonic oscillations. The "oscillation strength" of the harmonics is then given by the amplitudes A_i.

Time series reconstructed from the fit with $n = 10$ are shown in figure 4.11 (red lines), and the corresponding amplitude spectrum as function of altitude is displayed in figure 4.12. In the stratosphere, most of the variance is contained in the first two

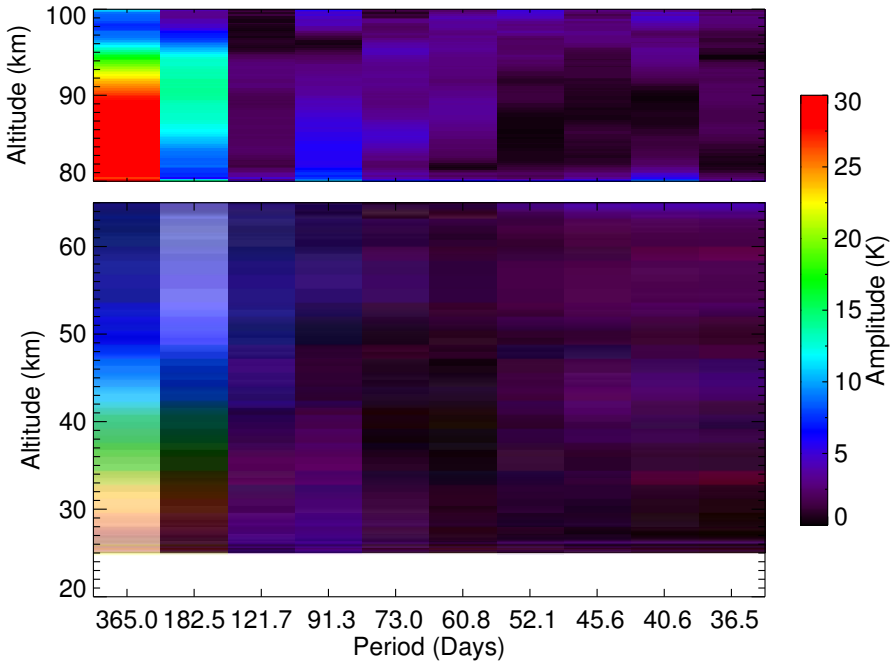


Figure 4.12.: Amplitude spectrum estimated from the components A_i of the non-linear fit.

harmonic components, i.e. in the annual (12 months) and the semi-annual (6 months) oscillations. While the annual oscillation dominates at low altitudes below 40 km, the amplitude of the semi-annual oscillation increases strongly near the stratopause. This trend culminates at approximately 60 km altitude where the temperature time series follows an almost perfect semi-annual oscillation.

In the MLT region the situation is more complex. Like in the stratosphere, there is a strong annual oscillation which dominates over the full altitude range between 80 and 100 km, and the amplitude of the semi-annual oscillation increases towards the altitude of the summer mesopause. In addition, however, there is also a 90 day oscillation with significant amplitude below 88 km. Attempts have been made to link this oscillation to geophysical processes, but no satisfactory explanation could be found. However, it should be noted that between 80 and 84 km there is a gap in the temperature time series in summer (see figure 4.1). Hence, the estimated amplitude of the 90 day oscillation is based on observations which were obtained in winter only.

Using the stratospheric data, a peak close to the 27 day solar rotation signal (e.g. *Dikty et al.*, 2010) is visible in the spectrum when expanding the harmonic decomposition of the temperature time series to higher orders (larger n). This peak is, however, not significant and is therefore not shown.

4.3. Temperature perturbations

Subsequent to the analysis of temperature variations with periods in the order of months which were presented in section 4.2.5, this section deals with temperature perturbations occurring on much shorter time scales. The transition from daily means to hourly profiles makes it possible to resolve perturbations with periods between two hours and the length of the observation which is typically in the order of 24 hours (see section 4.1). A number of geophysical phenomena fall into this time window, e.g. gravity waves,

inertia gravity waves, tides, and in case of longer observations also planetary waves. All these phenomena have in common that the associated temperature perturbations show up as wave-like features in the temperature time series. The difficulty in identifying such features lies in separating the wave-induced variations from the slowly varying background. Before (inertia) gravity waves and tides are studied with more specifically tailored algorithms, this section gives an overview over the characteristics of temperature perturbations without trying to filter for specific types of waves, i.e. no distinction is made between e.g. temperature disturbances caused by gravity waves and those caused by tides.

One common way to study atmospheric waves is to subtract the nightly mean temperature profile from the individual profiles, and the remaining residuals are then identified as wave-induced temperature variations. The mean perturbation amplitude is then found by averaging all residuals for a given height. This method was used in a number of studies (e.g. *Blum et al.*, 2004; *Höffner and Lübken*, 2007; *Rauthe et al.*, 2008). It should be noted, however, that the magnitude of the mean perturbation amplitude depends on the length of the time series. This can easily be seen if one recalls the typical horizontal spectrum of gravity waves which has the spectral dependence $k^{-5/3}$. Waves can only contribute fully to the variance of the time series if their period is in the order of the length of the time series or smaller. Hence, the length of the time series limits the portion of the frequency spectrum the wave analysis is sensitive to. This means, in case of the gravity wave spectrum, that the mean perturbation amplitude increases with the length of the time series.

4.3.1. Data analysis

In this study the time series of temperature observations is analyzed in the frequency domain. Consider an N-element temperature time series $x(t_n)$. Computing the discrete Fourier transform

$$\hat{x}(s_n) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) \exp[-i2\pi nk/N]. \tag{4.2}$$

yields N complex Fourier coefficients $\alpha_n = \hat{x}(s_n)$ at discrete frequencies s_n. Because the time series $x(t_n)$ contains observations and is thus real, α_{N-n} and α_n are related by

$$\alpha_{N-n} = \alpha_n^* \quad \text{for} \quad n = 0, 1, ..., N-1, \tag{4.3}$$

where the α^* denotes the complex conjugate. Moreover, it is $|\alpha_{N-n}| = |\alpha_n|$. Thus, a local power spectral density estimate for the frequency $s_n = n/T$, $n = 1, 2, .., N/2 - 1$, is given by

$$\Psi^\nu(s_n) = \frac{2}{[\nu/2] + 1} \sum_{p=n-[\nu/4]}^{n+[\nu/4]} |\alpha_p|^2, \tag{4.4}$$

where ν is an odd number, and the expression $[z]$ means "largest integer in z". The zero frequency component s_0 is not considered part of the spectrum.

The estimate represented by equation 4.4 is a local average over $[\nu/2] + 1$ squared Fourier coefficients associated with frequencies $s_{n-[\nu/2]}, ..., s_{n+[\nu/2]}$. With the variance σ^2 of the temperature time series defined as $\sigma^2 = N < \Psi^\nu(s_n) >$ follows for the fractual variance contained in the frequency band $s_{n-[\nu/2]}, ..., s_{n+[\nu/2]}$

$$\left\{\sigma^2(s_n)\right\}^\nu = ([\nu/2] + 1)\, \Psi^\nu(s_n). \tag{4.5}$$

Expression 4.5 allows the comparison of how much variance is contained in frequency bands centered at different frequencies s_n. Because atmospheric waves can be characterized in terms of the temperature amplitude, instead of the variance it is often advantageous to look at the mean temperature perturbation defined as the square-root of the variance

$$\left\{T'\left(s_n\right)\right\}^{\nu} = \sqrt{\left([\nu/2]+1\right)\Psi^{\nu}\left(s_n\right)}. \tag{4.6}$$

The question remains on what is a suitable choice for ν. The periodogram is often noisy, and the larger ν, the smoother the power spectral estimate (equation 4.4). On the other hand, the maximum value of ν is restricted by the number of available Fourier coefficients, and therefore the number of samples which make up the time series. The number of samples depends on two factors, namely the length of the observation and the sampling frequency. While the latter value is determined by the integration time of the lidar signal, which is usually one hour, the length of the observation can be freely chosen within limits. On the one hand, the observation period must be longer than the maximum period to be resolved in the periodogram, and on the other hand, as evident from figure 4.2, the choice of the length of the time series limits the number of available observations. Taking these factors into account, the best compromise was found to be 12 hours (or 12 samples respectively), which yields 6 independent complex Fourier coefficients. In this case the choice $\nu = 5$ follows naturally, thus dividing the available baseband into two non-overlapping frequency bands with periods 2–3 hours (band 1) and 4–12 hours (band 2).

Example: Lidar observation on September 23, 2011

On September 23, 2011, the Fe lidar observed the atmosphere for approximately 16 hours. The different steps of the wave analysis are illustrated in figure 4.13. The first step involves subtracting the mean and linearly detrending of the temperature time series. The residuals thus determined, which usually contain signatures of several wave-like features with different periods, are shown in figure 4.13a. It should be noted that individual values in this image with $0.2\,\mathrm{km} \times 0.25\,\mathrm{h}$ resolution are partly correlated. Although temperature profiles are computed every 15 min, the integration time employed in the Rayleigh temperature analysis (section 3.3.2) is 1 h.

Power spectra are computed in the second step. For this, all time series which are longer than 12 h are split into overlapping segments 12 h in length and 1 h time shift. Then the discrete Fourier transform is taken of every segment and the squares of the Fourier coefficients are computed. In case of the example presented in figure 4.13 this yields four periodograms per altitude bin. Next, all resulting power spectra belonging to a given altitude are averaged. This reduces the number of spectra to one per altitude. In order to reduce the variance of the spectra further, neighbouring spectra in the vertical domain within a 2 km height range are also averaged. The final power density spectrum as function of altitude is displayed in figure 4.13b. Two trends are visible: First, the spectral amplitude increases with increasing period (decreasing frequency), and second, the spectral amplitude increases with height. While latter observation is consistent with vertically propagating waves, the decreasing spectral amplitude with increasing frequency is indicative of a classical gravity wave spectrum.

Variances for the two frequency bands are computed from the averaged spectra according to equation 4.5. The resulting variance profiles are plotted in figure 4.13c, the 4–12 h band in blue and the 2–3 h band in green. As expected for a wave spectrum which falls off towards high frequencies, most of the variance is contained in the low frequency band (blue line). For reference, the variance computed in the time domain

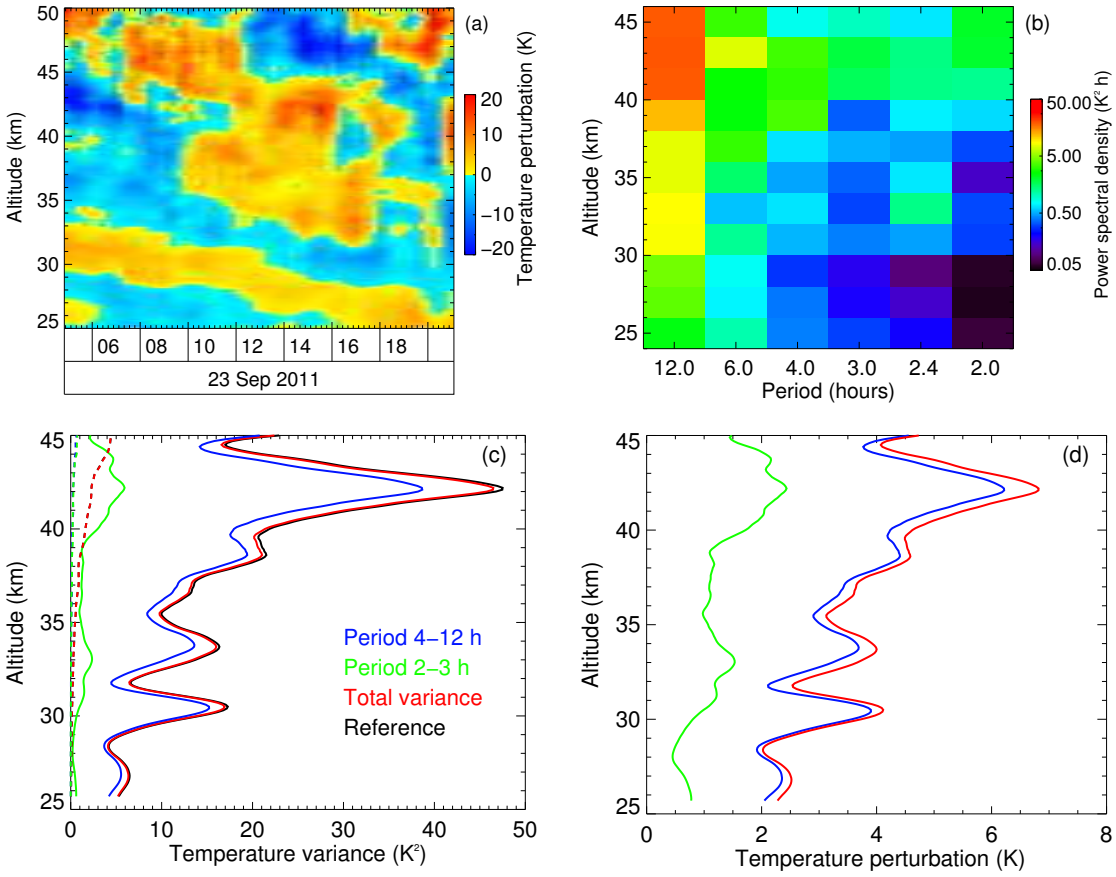


Figure 4.13.: Illustration of the temperature perturbation analysis for the data acquired on September 23, 2011 0500–2100 UT. (a) Temperature time series with mean subtracted, linearly detrended. (b) Power spectrum. (c) Temperature variances estimated from the entire power spectrum (red line) and from two spectral bands: 4–12 h (blue line) and 2–3 h (green line). Dashed lines mark the variance contributed by noise. For comparison, the variance calculated in the time domain is also shown (black line). (d) Temperature perturbation amplitudes calculated from variances shown in (c).

from the original time series of the observations (black line) and the total variance estimated from the spectrum by summing up all squared Fourier coefficients (red line) are also shown. Dashed lines in figure 4.13 represent the estimated fraction of the variance contributed by noise which is subtracted from the "signal" variance. The noise level is estimated as follows: First, a normally distributed random signal is generated with the same number of elements as the time series of the temperature observation. Next, each element of the time series thus created is multiplied with the uncertainty estimate of the corresponding temperature observation. This "uncertainty" dataset is then analyzed in the same way as the actual observations, and the resulting variance profiles are used as estimate for the fraction of the variance caused by noise in the original observations.

Mean temperature perturbation amplitudes computed from the variance profiles are shown in figure 4.13d. The mean perturbation amplitude in the 4–12 h band increases from 2 K at 25 km to approximately 4.2 K at 40 km altitude. The scale height estimated from this increase in amplitude is 10 km, which is larger than the typical density scale height of the atmosphere in this altitude region (≈ 7 km). Three regions with enhanced

wave activity can be identified in figure 4.13.d with peaks at 30, 34 and 42 km altitude. These peaks appear also in the shorter 2–3 h band, the structure is, however, not identical. This is consistent with the picture of several different wave packets crossing the lidar beam at different altitudes and times. The wave spectra, and thus also the perturbation profiles, represent the average condition of the atmosphere with respect to the observation period of 16 hours. It is evident from the perturbation profiles in figure 4.13, and even more so from the power spectrum, that a period of 16 hours is not long enough in order to average out transients caused by individual wave packets. For this reason, in the following sections the focus is shifted on seasonal averages.

4.3.2. Temperature perturbations in the stratosphere

It was shown in the previous section by means of an example that long averaging periods are needed in order to obtain a reliable estimate of the power spectrum of atmospheric waves. On the other hand, averaging periods should be short enough for seasonal effects not to become dominant. Two periods are selected based on these constraints. The first period is designated "winter" and includes 20 observations longer than 12 h between May and August 2011. The average length of the 20 observations is 34.9 h, and in total 697 h worth of data were used for the study. This corresponds to 92% of the total observation time during the four months (see figure 4.3).

The second data subset designated "summer" comprises 24 observations which were obtained in January/February 2011, November/December 2011, and January 2012. With an average observation period of 30.6 h, this amounts to 635 h in total. Observations included in the subset correspond to 61% of the total observation time obtained during this period. This figure is significantly lower than the corresponding figure for the winter dataset. Main reasons are on average shorter observation periods in summer and data gaps caused by bad weather. A single gap up to 3/4 h in length within a 12 h period is filled by linear interpolation in the analysis procedure, while the occurrence of any larger gap causes the observation period to be excluded from the data subset.

Power density spectra computed from the two data subsets are displayed in figure 4.14. The top panel shows power spectral densities as function of altitude for periods 2 to 12 h. As evident from the comparison between the winter spectra and the summer spectra, spectral amplitudes are in general larger in winter. Also, the shape of the spectrum is consistent at all altitudes in winter, i.e. the power spectral density increases with height and decreases with shorter periods. In summer, a local minimum is visible at approximately 33 km altitude. The differences between winter and summer become more apparent when the power spectral densities are illustrated as line plots as shown in the lower panel of figure 4.14. While slopes of the spectra corresponding to different altitudes are nearly identical in winter, a significant change is observed in the summer case. At low altitudes below 35 km, the slope of the spectra is approximately $-5/3$ which is the value expected for the horizontal gravity wave spectrum (*Fritts and Alexander*, 2003). Slopes at greater altitudes become however increasingly shallow and reach -1 at 45 km altitude. Vertical compression of the lines relative to each other indicates that not all of the energy carried by the atmospheric waves is transported vertically to higher altitudes, i.e. energy appears to be lost on the way up. This can be due to dissipation (e.g. waves filtered by critical layers, dynamic instabilities) or refraction. If waves are sufficiently refracted, the energy carried by these waves may be transported over large horizontal distances (e.g. ducted waves). In this case the vertical energy flux carried by the waves is locally not conserved even though no dissipation takes place. Because lidar observations are local observations in this sense, it is not possible to distinguish

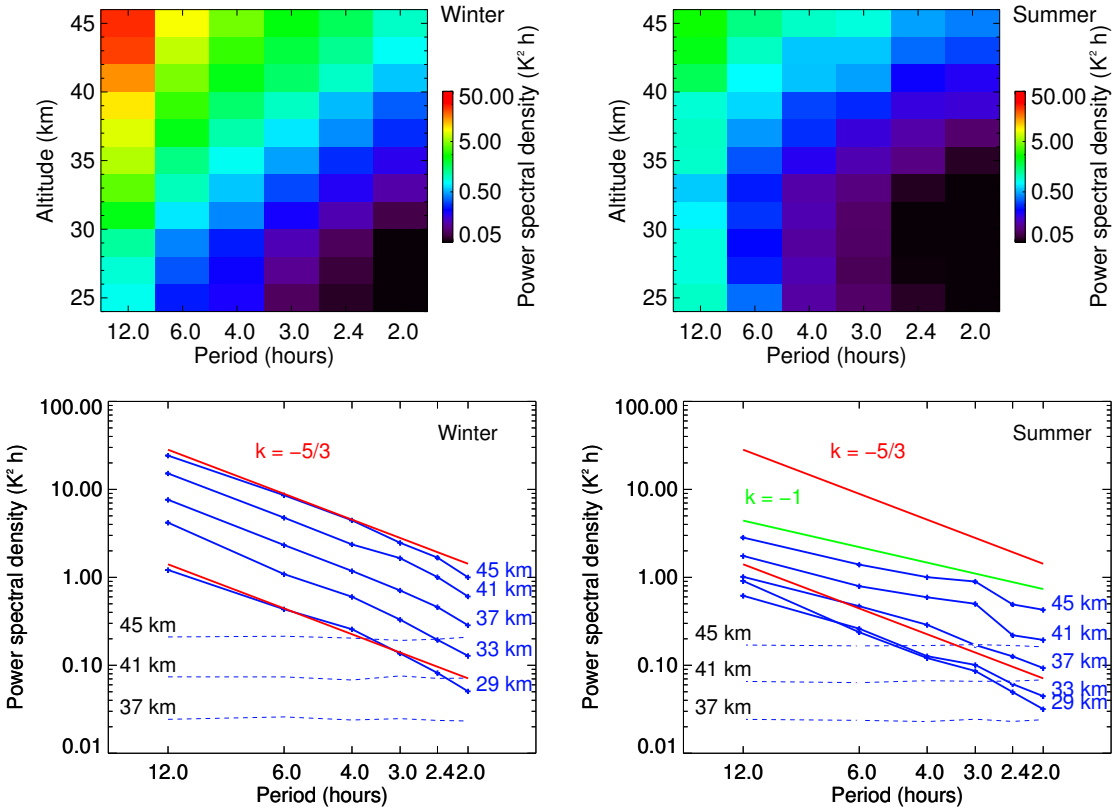


Figure 4.14.: Temperature perturbation spectrum in winter (left) and summer (right). Winter includes months May to August and summer November to February. The dashed lines in the lower two panels mark noise levels.

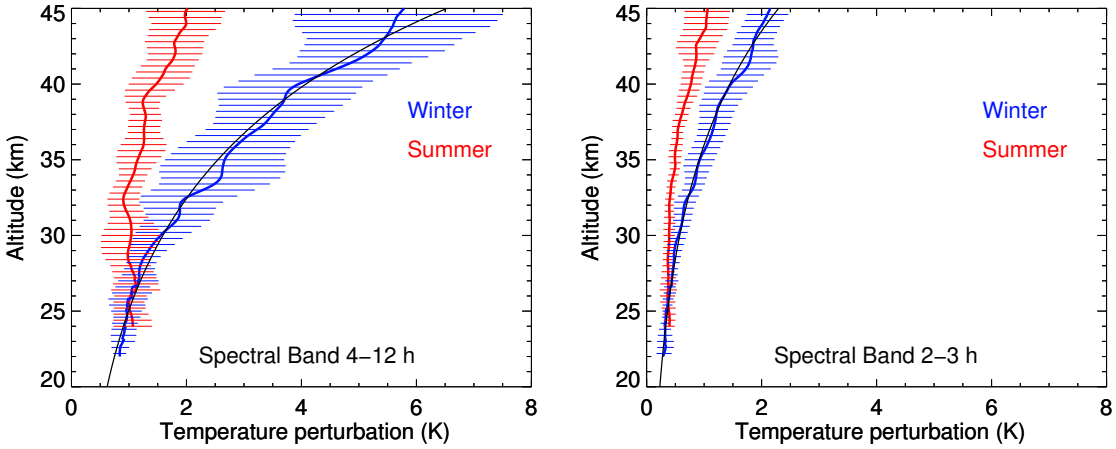


Figure 4.15.: Temperature perturbation amplitude as function of altitude for summer (November to February) and winter (May to August). Horizontal lines indicate the variability estimated from the standard deviation of all profiles. The solid black line marks the exponential $A \exp(z/(2H))$ where $H = 5.5\,\mathrm{km}$ is the scale height estimated from a least squares fit. Note that the constant A is different for different spectral bands.

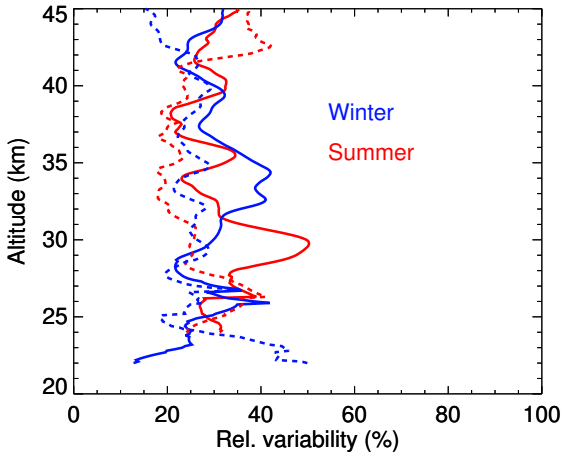


Figure 4.16.: Relative variability of mean temperature perturbations (standard deviation divided by the mean) as function of altitude: 4–12 h spectral band (solid lines) and 2–3 h spectral band (dashed lines)

between dissipation and refraction of waves in lidar data.

Mean temperature perturbation amplitudes computed from the wave spectra are shown in figure 4.15. The amplitude remains approximately constant between 25 km and 35 km in summer, while an exponential increase is observed in winter in both spectral bands. The scale height determined from the growth rate is 5.5 km. This figure is significantly smaller than the density scale height of the atmosphere (approximately 7 km). Thus, temperature perturbation amplitudes grow apparently faster than expected from the argument of energy conservation in a medium with decreasing density, and additional wave sources are needed in order to explain the observed growth rates. Possible sources are atmospheric tides which can be excited at different altitudes by various processes such as absorption of solar radiation (e.g. *Hagan*, 1996), non-linear interaction of planetary waves, and release of latent heat (*Hagan and Forbes*, 2002). The role of the tides will be studied in more detail in appendix B.1.

It is also striking to see that temperature perturbation amplitudes are nearly identical in summer and winter at 25 km altitude. Two possible conclusions can be drawn from this observation: Either there are no seasonal variations in the strength of the wave sources, or there are dissipative/filtering processes in the lower stratosphere which limit perturbation amplitudes in the observed frequency bands to the same level year round. Also, a mixture of both scenarios is possible. However, the available lidar data below 25 km is inconclusive because of various problems with the dataset (see section 4.1).

Another major difference between winter and summer can be seen in the variability of the temperature perturbation amplitudes in both spectral bands. The variability in figure 4.15 is estimated from the standard deviation of all perturbation profiles belonging to a particular season. In general, the variability increases with altitude in winter and to a smaller extent also in summer. This increase becomes most visible in the 2–3 h spectral band and can be attributed to exponentially growing perturbation amplitudes. If the strength of the wave source is modulated, the effect will be largest at high altitudes where wave amplitudes are also largest. Consequently, the impact on absolute perturbation amplitudes is smaller at low altitudes, although the relative change in perturbation amplitudes remains approximately constant at all altitudes as shown in figure 4.16. This allows one to draw two conclusions: (1) The variability of mean perturbation amplitudes observed in the upper stratosphere is caused by processes below 25 km altitude, possibly

Station	Data set		Resolution /	Number of observations	
	start	end	length	winter	summer
ALOMAR	Jul 2004	Jan 2012	1 h / 12 h	18	35
KBORN	Jun 2010	Aug 2012	2 h / 12 h	9	17
KBORN2	Nov 2002	Feb 2012	1 h / 8h	55	0*
Davis	Jan 2011	Apr 2012	1 h / 12 h	20	24

Table 4.1.: Summary of the datasets used in figure 4.17. *The old RMR lidar at Kühlungsborn is not capable of daylight measurements.

by modulation of the wave sources. Processes such as filtering of waves by critical layers (see section 4.4.5) do not alter the variability, i.e. waves which make it through a critical layer show the same relative variability as below the critical layer. (2) The relative variability of perturbation amplitudes has no seasonal dependence, in particular no significant difference in relative variability is observed between winter and summer. This is a rather surprising result given that absolute variability differs by as much as a factor of three at 45 km altitude (figure 4.15).

4.3.3. Comparison with other sites

To assess whether the temperature perturbation amplitudes observed above Davis are representative for mid and high-latitude regions, perturbation amplitudes shown in figure 4.15 are compared with lidar observations at Kühlungsborn (45°N,11°E) and ALOMAR (69°N,16°E). Table 4.1 gives a summary of the datasets used in this comparison. The ALOMAR temperature profiles were provided by *Kaifler and Baumgarten* (2012) and the description of an earlier version of this dataset can be found in *Schöch* (2007). *Kopp and Gerding* (2012) contributed temperature measurements obtained with the new RMR lidar at Kühlungsborn (KBORN). The KBORN2 dataset (old RMR lidar) is an updated version of the dataset discussed in *Rauthe* (2008) and *Rauthe et al.* (2008).

All datasets except KBORN2 were analyzed using the same analysis procedures which are discussed in the previous section. The old RMR lidar at Kühlungsborn has no daylight capability and observation times are therefore restricted to darkness. In addition to other constraints such as weather conditions, the period of darkness limits the maximum length of continuous lidar observations to approximately 12 h in winter and 4 h in summer. In order to retain most of the winter observations, the size of the data window was decreased from 12 h to 8 h for this dataset. Because the resolved spectral band shifts accordingly from 4–12 h to 2.7–8 h, variances estimated from the spectrum need to be scaled in order to make results comparable between different spectral bands. Assuming a spectrum which varies as $\omega^{-5/3}$, the scaling factor which relates variances estimated from the 2.8–8 h band to the 4–12 h band is 1.97.

Figure 4.17 shows seasonal mean temperature perturbation profiles for the three sites ALOMAR, Kühlungsborn, and Davis. The ALOMAR winter profile ranges from 1 K perturbation amplitude at 25 km to approximately 4 K at 60 km. A similar exponential increase is observed in summer, amplitudes are, however, on average 10–20% smaller. This result is consistent with earlier reports by *Schöch* (2007) showing no significant seasonal variation. On the other hand, seasonal variations observed at Kühlungsborn are distinctly larger. Mean amplitudes in summer are in the order of 2 K, while amplitudes in winter reach approximately 5 K at 60 km altitude. Note that the winter profile KBORN comprises few measurements (see table 4.1) and appears thus more noisy compared to other profiles. Nevertheless, taking uncertainties into account there is a good agreement

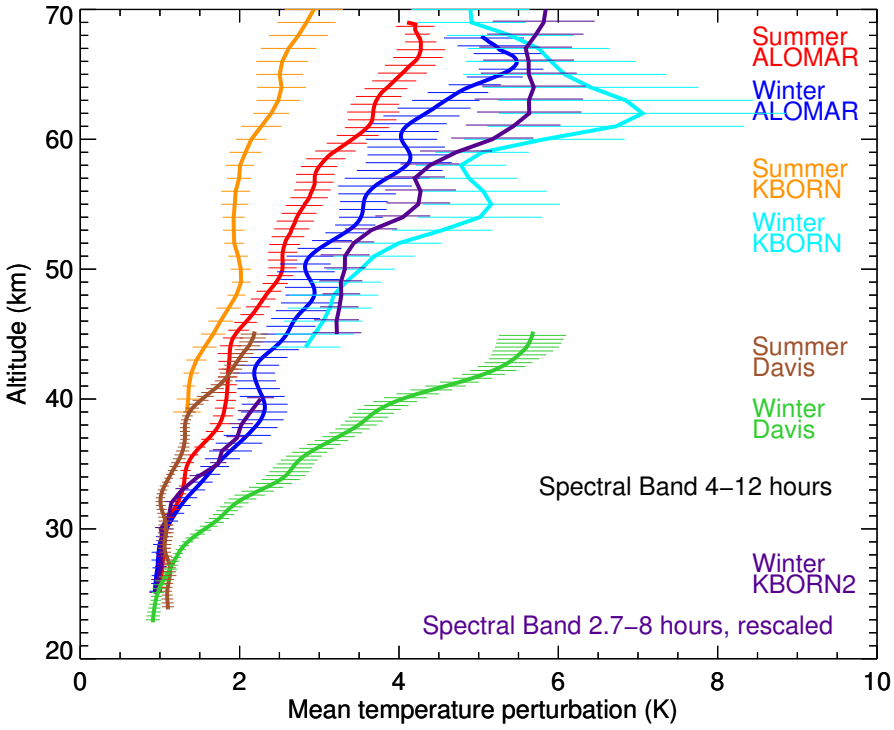


Figure 4.17.: Mean temperature perturbation amplitudes as function of altitude for the sites Davis (68°S,78°E), KBORN (54°N,11°E) and ALOMAR (69°N,16°E). Note that winter observations with the old RMR lidar at KBorn in the 2.7–8 hour band (violet) are scaled to account for the smaller variance content of the spectral band (see text for details). Horizontal lines indicate standard errors.

between the two winter profiles KBORN and KBORN2.

Temperature perturbation amplitudes observed at Davis do not easily fit into the picture as mean amplitudes in winter at 45 km altitude are 2–3 times larger compared to ALOMAR or Kühlungsborn. A large difference between the northern hemisphere and the region of the Antarctic coast is also evident in satellite based momentum flux measurements. *Wright and Gille* (2013) used global temperature data from the High Resolution Dynamics Limb Sounder on board the Aura satellite to calculate momentum fluxes in the stratosphere. Their data show that the Antarctic coast is a large gravity wave hotspot, its magnitude being only second to the hotspot of the Antarctic Peninsula and far larger than any hotspot in the northern hemisphere.

The amplitude of the Davis summer profile in figure 4.17 decreases with altitude below 30 km and converges towards the Kühlungsborn profile at 38 km, while at 45 km altitude the Davis profile lies on top of the ALOMAR profile. It can thus be concluded that wave propagation above Davis is significantly different compared to the northern hemisphere. Because exceptional large perturbation amplitudes are observed in winter when the polar night jet is strongest, dynamic properties of the atmosphere are likely a key ingredient to understanding observed interhemispheric differences. It will be shown in section 4.4.5 that the background mean flow is indeed the most important factor which determines gravity wave propagation in the stratosphere.

Despite the considerable differences in mean amplitudes, spectral observations of the perturbation amplitudes at the sites Davis and ALOMAR share also striking similarities. Figure 4.18 shows mean power spectra observed above ALOMAR in winter and

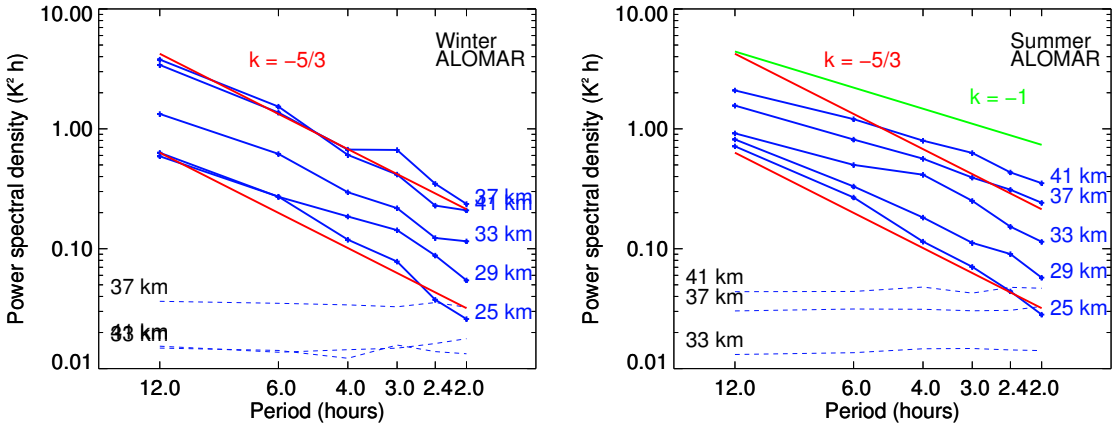


Figure 4.18.: Perturbation spectrum in winter (left) and summer (right) above ALOMAR (69°N,16°E). Summer includes months May to August and winter November to March. The dashed lines mark noise levels. For comparison theoretical spectra with slopes of −5/3 and −1 are also shown.

summer. The slope of the winter spectra is close to −5/3 at all observed altitudes, as it was the case for the Davis data (figure 4.14). Note that the ALOMAR winter spectra appear more noisy compared to Davis because of on average shorter observation periods. The smaller number of observation hours results in residual atmospheric variability showing up as modulation of the seasonal mean spectra. On the other hand, the summer mean spectra comprise nearly twice as many observations (see table 4.1) and the noise component is therefore greatly reduced. As it was the case for Davis, ALOMAR spectra in summer show increasingly shallow slopes. At 25 km altitude the slope is approximately −5/3, while the exponent reduces to −1 at 37 km altitude. For comparison, at Davis a slope of −1 is observed at approximately 41 km altitude.

The fact that the spectrum gets shallower with increasing altitude suggest that waves are being filtered in this altitude region. As evident from the gravity wave dispersion relation (equation 4.7), gravity waves with long periods (small $\hat{\omega}$) have short vertical wavelengths (large m). Thus, these waves are more likely to encounter critical levels than waves with short periods. As a result, waves with long periods are being subject to significant damping in summer, which causes the spectrum to flatten.

4.3.4. Seasonal variations

Seasonal variations in gravity wave filtering can be studied by comparing the variances estimated from different spectral bands. Figure 4.19 displays the ratio $\sigma^2_{4-12\mathrm{h}}/\sigma^2_{2-3\mathrm{h}}$ as function of time, where $\sigma^2_{4-12\mathrm{h}}$ and $\sigma^2_{2-3\mathrm{h}}$ are variances observed in the 4–12 h and 2–3 h spectral band in 35–40 km altitude above Davis. As evident from this figure, small ratios are observed in summer, whereas large ratios occur in winter. The mean summer value of 4.7 is significantly lower than the value of 6.8 that is expected from the theoretical gravity wave spectrum with the exponent $k = -5/3$. Small variance ratios indicate a flattening of the spectrum and can thus be interpreted as a sign of gravity waves being damped, as discussed in the previous section. On the other hand, the large values in winter are indicative of waves propagating freely. According to figure 4.19 the transition from the summer state with waves being filtered to the winter state of freely propagating waves occurs at the end of April and beginning of May. The corresponding winter-summer transition extends over a period of approximately two months in November

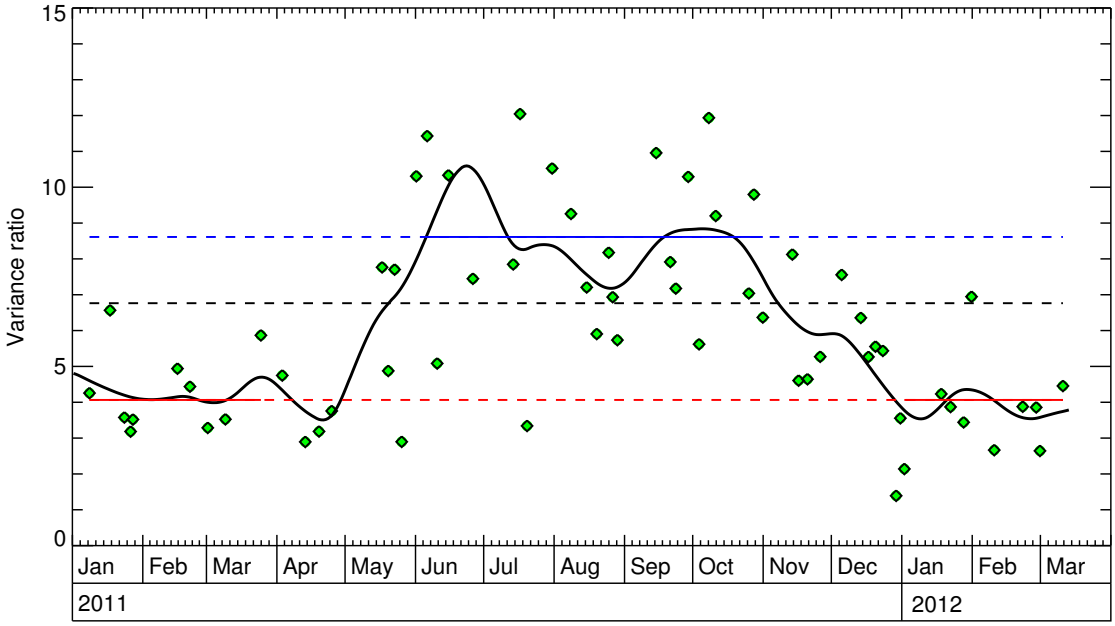


Figure 4.19.: Ratio of temperature variances estimated from the two spectral bands 4–12 hours and 2–3 hours (altitude range 35–40 km). Small values indicate a flattening of the spectrum, large values a steepening. The black solid line represents the observations smoothed with a 60 day Hann window. The blue line marks the mean computed for months June–October (winter), the red line for months January–March. The dashed black line indicates the variance ratio expected from the theoretical $k = -5/3$ spectrum.

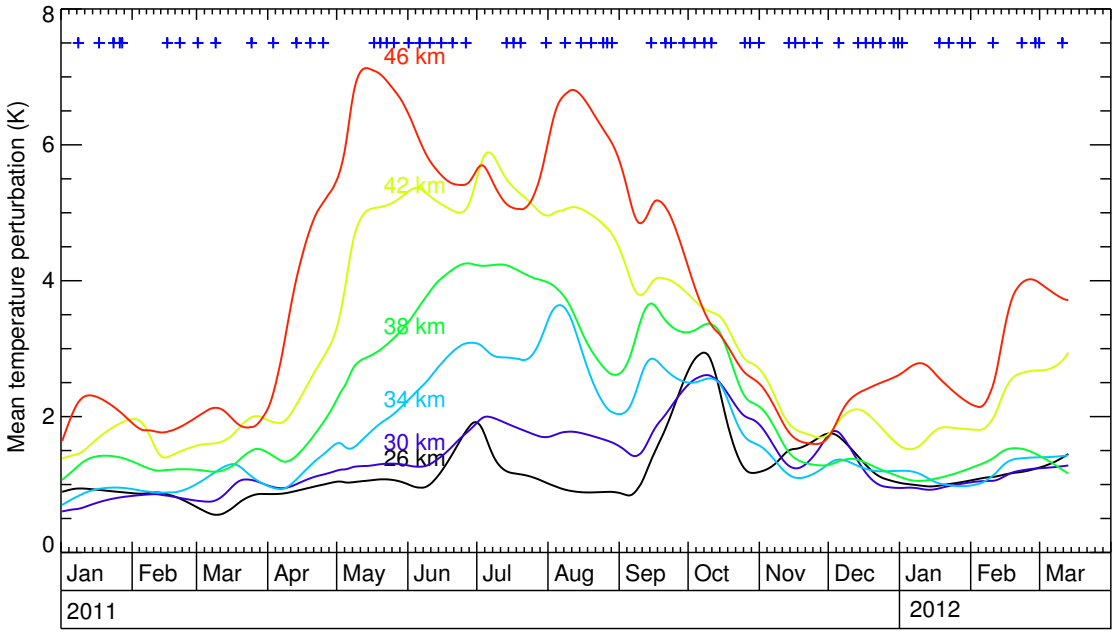


Figure 4.20.: Seasonal variation of temperature perturbation amplitudes in the 4–12 hour spectral band for selected altitudes. The curves are smoothed with a 40-day Hann filter. Observation periods are marked with blue crosses at the top of the panel.

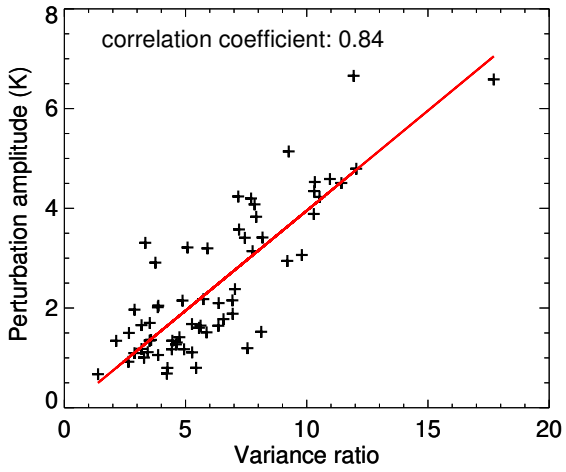


Figure 4.21.: Scatter plot of mean temperature perturbation amplitude at 38 km altitude versus variance ratio in 35–40 km altitude. See text for details.

and December. A more detailed discussion of the vertical propagation of gravity waves can be found in section 4.4.

Seasonal variations of mean temperature perturbation amplitudes are illustrated in figure 4.20. A minimum is observed during the summer months November–February, followed by a sharp increase in amplitude in April. The winter maximum extends from May to approximately September and shows various substructures at different altitudes. A clear double hump is observed at 46 km with a local minimum in June/July, while 4 km below at 42 km altitude no such signature can be identified. The double maximum is visible again at 38 and 34 km altitude, however the local minimum occurs now in August, approximately two months later than at 46 km altitude. The second maximum in September/October becomes the dominant feature in the lower stratosphere, and a smaller secondary peak occurs in June/July. Apart from these two peaks there is no clear seasonal modulation at low altitudes. Note that increasing amplitudes in April are first seen at high altitudes followed by a less steep increase at lower altitudes. At the end of winter, the opposite is happening and the perturbation amplitude decreases first at high altitudes (all lines are close together in October/November). This symmetry can be interpreted as result of the dynamic changes occurring in the atmosphere over the course of one year. In October and November the zero-wind line of the zonal wind descends into the lower stratosphere. Waves are filtered at high altitudes where the wind reversal occurs first. At the end of summer the zero-wind line moves back from the lower stratosphere into the mesosphere. The vertical wind gradient is, however, much flatter and the wind speed fluctuates more compared to spring (see figure 4.31). Waves that make it through the critical layer can propagate freely above, resulting in larger temperature perturbation amplitudes at higher altitudes.

If one compares the seasonal modulation of the temperature perturbations at 38 km altitude (green line in figure 4.20) with the variance ratio displayed in figure 4.19, it becomes clear that the observed perturbation amplitude follows the modulation of the variance ratio, i.e. the mean perturbation amplitude is correlated with the slope of the perturbation power spectrum. A scatter plot of the two quantities is shown in figure 4.21. According to this the slope of the power spectrum can be used as proxy for detecting processes which filter waves. Waves with large periods have generally larger amplitudes than waves with shorter periods and are thus more likely to break. This

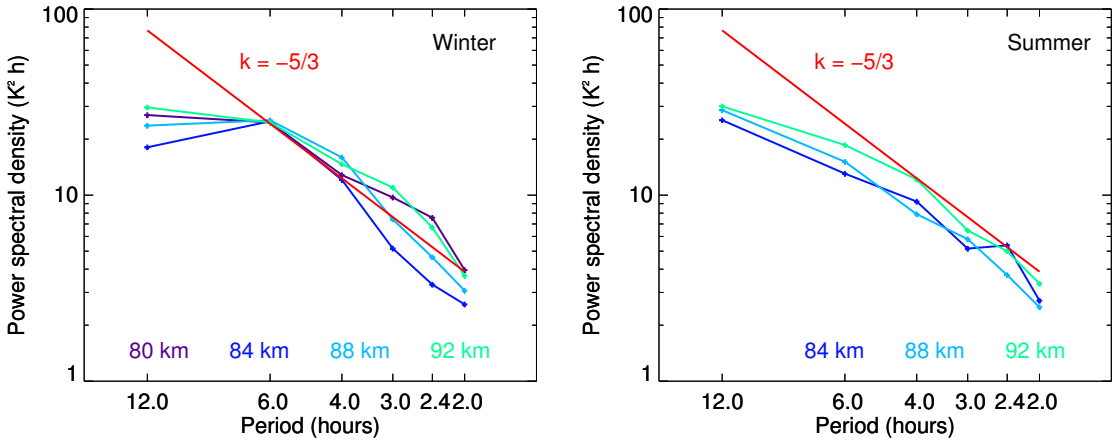


Figure 4.22.: Temperature perturbation spectrum in winter (left) and summer (right) for the mesopause region. Winter includes months May to August and summer months November to February.

selective filtering of waves with large periods results in flattening of the wave spectrum as well as decreasing perturbation amplitudes. The effect of the background wind on wave propagation is discussed in more detail in section 4.4.5.

4.3.5. Temperature perturbations in the MLT region

Temperature perturbation spectra for the MLT region were computed from Fe temperature profiles provided by *Höffner* (2012). In summer, the vertical extent of the Fe layer is considerably smaller compared to winter (*Gardner et al.*, 2011), and the altitude range where temperature measurements are possible is reduced accordingly. For this reason, the altitude ranges for winter and summer differ. Note that MLT spectra and derived results presented in this study are preliminary. Because of the problems with the Fe temperature dataset (see section 4.1) it is not possible to obtain reliable error estimates and confidence intervals. Therefore results presented here should be interpreted with caution.

Power spectral densities observed above Davis in winter and summer are illustrated in figure 4.22. In winter, at periods shorter than 6 h the spectra have slopes of approximately $-5/3$. At low frequencies, however, the spectra flatten and the slope becomes even positive at 84 km altitude. The mean slope of the spectra in summer at periods larger than 6 h may be slightly smaller than $-5/3$. This is consistent with the value $-3/2$ reported in an earlier study by *Dowdy et al.* (2001). *Dowdy et al.* computed power spectra of the zonal and meridional wind variances measured with the Davis MF radar. The data used in the study by *Dowdy et al.* were acquired from mid 1999 to mid 2000.

Mean temperature perturbation amplitudes were derived from integrated power spectra as described in section 4.3. Figure 4.23 shows perturbation profiles for two spectral bands, 4–12 h and 2–3 h. In winter, at large periods the perturbation amplitudes grow weakly with altitude, while a much larger growth rate is observed in summer. At small periods, however, the growth rates are nearly identical. In both cases the growth rate is significantly smaller than the $\exp(z/2H)$ rate that is expected for freely propagating waves. The density scale height H is approximately 6 km in the mesopause region (*Lübken*, 1999). Using the larger temperature perturbation growth rates seen in the 4–12 h band in summer, the scale height evaluates to 14 km. Even larger scale heights

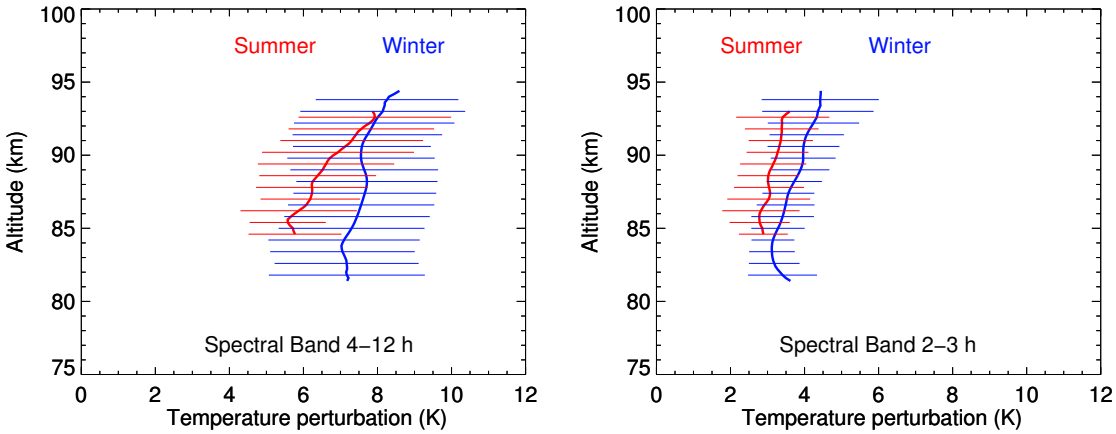


Figure 4.23.: Temperature perturbation amplitude as function of altitude for summer (November to February) and winter (May to August). Horizontal lines indicate the variability estimated from the standard deviation of all profiles.

are obtained for the winter profile. This suggests that a significant portion of the waves is breaking.

Dowdy et al. (2001) found wind variances growing steadily with height between 83 and 93 km altitude. The scale height estimated from their data is approximately 5.5 km. This value is in good agreement with the density scale height. To understand the discrepancy between the radar-based and the lidar-based scale height measurements, one has to take into account the different averaging periods and the background wind field. *Dowdy et al.* (2001) averaged variance profiles for 30 days around the summer solstice. This period coincides with the occurrence of the zonal wind speed maximum at 80 km altitude (figure 1 in *Dowdy et al.*, 2001). Thus, gravity waves are expected to be filtered shortly below the wind speed maximum and free propagation is predominantly observed above. Because the iron density decreases drastically around solstice, few usable variance measurements could be obtained in this period and the averaging window had to be extended to months November–February. However, the wind structure changes significantly within this larger period. As a consequence, the altitude range where waves are predominantly filtered spreads into the observation range resulting in smaller amplitude growth rates and thus larger scale heights.

4.4. Gravity waves

4.4.1. Introduction

Atmospheric gravity waves are oscillations of air parcels. Forces acting on the air parcels are the lifting force of buoyancy and the restoring force of gravity. Any transient disturbance from this equilibrium of forces results in oscillatory motion and generation of waves. The waves propagate horizontally as well as vertically, thus transporting energy and momentum from one place to another. Especially the vertical transport is of great importance because it couples different atmospheric layers. Therefore gravity waves play an important role in defining the thermal structure and variability of the atmosphere. This is in particular true for high latitudes where the thermal structure is driven far from radiative equilibrium (*Becker*, 2012).

Gravity waves are generated on a global scale mainly in the troposphere by innumer-

able individual sources. These sources are usually classified according to the dominating process which leads to generation of gravity waves. *Becker* (2012) lists four generation processes which are thought to be most important for gravity waves observed in the middle atmosphere: flow over topography, convection, shear instability, and geostrophic adjustment. A comprehensive description of these processes can be found in the review by *Fritts and Alexander* (2003). The following summary is based on *Fritts and Alexander* (2003) and references therein:

Flow over topography gives rise to the so-called mountain waves with horizontal wavelengths in the order of tens to hundreds of kilometers. These waves are mostly relevant for the troposphere and the stratosphere and do usually not penetrate into the mesosphere. Gravity waves generated by convection typically cover the full range of possible wavelengths with no predominant phase speed. This type of wave source plays a major role at low latitudes, as deep convection is mainly confined to the tropics. It is known, however, that low-frequency waves originating from these sources can propagate horizontally over large distances.

The process of gravity wave generation in shear instabilities is still not well understood at present. Typical wavelengths are thought to be in the order of few to tens of kilometers. This estimate is based on the assumption that the so-called envelope radiation is the predominant excitation mechanism.

The term geostrophic adjustment describes the relaxation of an initially unbalanced flow to a new balanced state by redistributing mean energy and momentum. In this process excess energy is radiated away as inertia-gravity waves. These waves have sufficiently long periods (several hours) for the Coriolis force to become an important contribution to the restoring force in addition to the buoyancy force. Typical horizontal scales are several hundred kilometers.

Dispersion relation

The geophysical fluid dynamics equations (equation 1.1) contain in their linearized form solutions of the form of plane waves. Substituting the plane wave solution (equation 1.3) into the linearized set of equations allows the determination of the coefficients, and the gravity wave dispersion relation

$$\hat{\omega}^2 = \frac{N^2\left(k^2+l^2\right)+f^2\left(m^2+\frac{1}{4H^2}\right)}{k^2+l^2+m^2+\frac{1}{4H^2}} \tag{4.7}$$

is obtained after some algebraic manipulations (*Fritts and Alexander*, 2003). The dispersion relation relates the components of the wave vector $\mathbf{k} = (k, l, m)$ to the intrinsic frequency of the wave $\hat{\omega} = \omega - ku_0 - lv_0$. The frequency may be modified by the properties of the background atmosphere through the buoyancy frequency $N = \sqrt{g\partial \ln \Theta / \partial z}$ and the horizontal components of the background mean flow u_0 and v_0. The Coriolis parameter, f, is defined as $f = 2\Omega \sin\phi$, where Ω is the Earth rotation rate and ϕ is the latitude. For the location of Davis station (69°S) f evaluates to $1.36 \times 10^{-4}\,\mathrm{s}^{-1}$ or the period of 12.8 h. For comparison, the buoyancy frequency in the middle atmosphere is in the order of $0.02\,\mathrm{s}^{-1}$ or 5 min.

A closer look at the dispersion relation (equation 4.7) reveals that the range of possible intrinsic frequencies spans the entire range between the Coriolis parameter and the buoyancy frequency. Three cases are typically distinguished: high frequency waves with $\hat{\omega} \gg f$, medium-frequency waves with $N \gg \hat{\omega} \gg f$, and low frequency waves with $\hat{\omega} \sim f$. Latter waves are commonly referred to as inertia-gravity waves because the rotation of the Earth has an important influence leading to a three-dimensional helical

structure of the wave perturbation. An approximation to the dispersion relation valid for both low- and medium frequency waves is

$$\hat{\omega}^2 = N^2 \frac{k_h^2}{m^2} + f^2, \tag{4.8}$$

where $k_h = \sqrt{k^2 + l^2}$ is the horizontal and m the vertical wave number (*Fritts and Alexander*, 2003). An even simpler relation

$$\hat{\omega} = N \left| \frac{k_h}{m} \right| \tag{4.9}$$

can be obtained if the frequency range is restricted to $N \gg \hat{\omega} \gg f$ (*Fritts and Alexander*, 2003). Equation 4.9 is called the mid-frequency approximation to the gravity wave dispersion relation. Substituting the intrinsic horizontal phase speed $\hat{c}_h = \hat{\omega}/k_h$ into this equation yields

$$|m| = \frac{N}{|\hat{c}_h|}. \tag{4.10}$$

Because the intrinsic phase speed is also related to the horizontal mean background wind, $\hat{c}_h = c_h - \overline{u}_h$, equation 4.10 can be used to asses the dynamic stability of waves in a vertical shear flow. As the horizontal wind speed $\overline{u}_h$ approaches the horizontal phase speed c_h, the wave is compressed and its vertical wavelength shrinks. Compression of the wave can not continue indefinitely, however. Once gradients become too large, instabilities emerge and the wave is dissipated. Hence, the level where $c_h = \overline{u}_h$ is called critical level.

For high-frequency waves ($\hat{\omega} \gg f$) the contribution of the Coriolis force to the restoring force can be neglected. If the vertical wave number is sufficiently large, m^2 dominates over the term $1/4H^2$, and the dispersion relation simplifies to

$$\hat{\omega}^2 = \frac{N\left(k^2 + l^2\right)}{k^2 + l^2 + m^2} \tag{4.11}$$

(*Fritts and Alexander*, 2003). Because waves with periods shorter than approximately twice the integration time ($\sim$ 1 h) can not be resolved in lidar data, high-frequency waves are invisible to the Fe lidar.

Gravity wave energy

The kinetic gravity wave energy per mass as well as the potential energy per mass can be calculated from the respective wave-induced perturbation amplitudes:

$$E_k = \frac{1}{2}\left(u'^2 + w'^2 + v'^2\right) \qquad E_p = \frac{1}{2}\frac{g}{N^2}\left(\frac{\rho'}{\rho_0}\right)^2. \tag{4.12}$$

The sum of the two quantities forms the total energy density of gravity waves, $E_t = E_k + E_p$.

Many waves may exist in the atmosphere at any given point in time and space. Because remote sensing instruments record the combined effect of all waves within the sampling volume, perturbation amplitudes associated with single waves are usually not determinable from observation data. Temporal averages of the measured perturbation amplitudes are then used instead of calculated gravity wave densities.

The gravity wave potential energy density is an important quantity because it de-

scribes the available energy of the dynamic system. For this reason it is often also called the *available* potential energy.

4.4.2. Gravity wave analysis

Extracting wave-induced temperature perturbations

In this work gravity waves are characterized by the wave-induced temperature perturbation $T'(z,t)$. This approach is complicated by the fact that $T'(z,t)$ is not directly accessible since the lidar measures the superposition of wave-induced temperature perturbations and the background temperature $T(z,t) = T'(z,t) + T_0(z,t)$ only. Thus, in order to extract perturbation profiles $T'(z,t)$ a way must be found to estimate the background temperature $T_0(z,t)$ from lidar observations.

Several methods for the derivation of wave-induced perturbation profiles from lidar-obtained temperature or relative density profiles have been used in the past. *Gerrard et al.* (2004) for example computed background density profile based on a third-order polynomial fit to the measured atmospheric density profile. The fit was then subtracted from the original profile and the residuals were low-pass filtered. Finally, the resulting profile was added to the fit to form the background profile. A similar approach was taken by *Yamashita et al.* (2009) except that a fifth-order polynomial was used. *Rauthe et al.* (2008) characterized gravity waves based on temperature deviations from the nightly mean. *Duck et al.* (2001) and more recently *Alexander et al.* (2011) estimated the background temperature from a series of fits with a third-order polynomial, where the altitude window used in each fit is slightly shifted vertically. The weighted average of all fits is then taken as background profile. This method was considered most promising given that subtracting the nightly mean does not account for variations of the background temperature with periods comparable or longer than the observation window (due to e.g. tides, planetary waves). The fit of the observed profile with a single polynomial was not deemed an appropriate representation of the background because the tail regions of the fit may wag substantially.

Following the approach taken by *Alexander et al.* (2011) the resolution of the initial temperature profiles was chosen to be one hour by two kilometers. This has important consequences for the gravity wave analysis because the resolution limit defines the lower cut-off points of the observable two dimensional gravity wave spectrum. Spatial and temporal resolution can be traded off to a certain extent, and the spectral bands shift accordingly. However, the signal-to-noise ratio of the temperature measurements imposes a fundamental resolution limit on the high-frequency part of the spectrum. Because the gravity wave spectrum falls off at high wavenumbers, the signal may vanish in the noise floor before the Nyquist frequency is reached. Choosing the same resolution as *Alexander et al.* facilitates a direct comparison between earlier results and results obtained as part of this work. The gravity wave study by *Alexander et al.* is based on Rayleigh lidar observations at Davis during the 2007/2008 season, whereas the Fe lidar observations commenced in 2010.

The $1\,\mathrm{h} \times 2\,\mathrm{km}$ temperature profiles $T(z,t)$ range typically from 20 to 54 km altitude and mean statistical uncertainties ΔT of the Rayleigh temperature retrieval (section 3.3.2) are 2.8 K at 50 km and 0.5 K at 30 km altitude. It should be noted that biases in the retrieval due to inaccuracies in the treatment of the optical filters and atmospheric transmission can be substantially larger than the statistical uncertainties which arise from the photon count process. These biases are likely to vary on time scales in the order of tens of hours, which is much longer than typical gravity waves periods (few hours). Thus, biases can be assumed to modify the background temperature field $T_0(z,t)$, and

they do not appear in perturbation profiles $T'(z,t)$ because the background field is subtracted from lidar observations.

The signal-to-noise ratio at a certain altitude is not necessarily constant over the duration of a particular lidar observation. For example, faint clouds advected through the laser beam attenuate the lidar signal. Data is rejected if the statistical temperature uncertainty exceeds twice the standard deviation of all temperature profiles obtained during the observation. Rejection results in a decreased altitude range.

Temperature perturbation profiles $T'(z,t)$ are extracted from lidar observations $T(z,t)$ based on the method described in *Duck et al.* (2001) and *Alexander et al.* (2011). $T(z,t)$ is split into overlapping segments of length 20 km, where the start of each segment is shifted by 2 km in altitude relative to the previous segment. Next, each segment is fitted independently with a third-order polynomial. The background temperature field at each altitude $T_0(z,t)$ is then calculated from the weighted average of all fits, the weights being distributed as follows: While the central four bins of the fits are weighted fully, the contribution of the lower and upper tails to the resulting average $T_0(z,t)$ is smoothly reduced to zero. This is accomplished by decreasing the weights exponentially with an e-folding length of 3 km in the lower and upper three bins of each segment. Reducing the weights guarantees that discontinuities, which may arise from fits "wagging" at the endpoints, are sufficiently suppressed. The resulting $T_0(z,t)$ is subsequently smoothed vertically by applying 6 km averaging filter. In a last step the temperature perturbations $T'(z,t)$ are found by subtracting $T_0(z,t)$ from the observations $T(z,t)$. It was shown by *Duck et al.* (2001) that this method extracts gravity wave amplitudes accurately enough so that gravity wave energies can be determined to within 20% for vertical wavelengths between 4 and 15 km.

Although the sensitivity of the gravity wave extraction procedure drops significantly for vertical wavelengths λ_z larger than 20 km, the resulting perturbation profiles may still contain contributions from tides (typical $\lambda_z > 30$ km) or planetary waves. In order to retain only the gravity wave part, temperature perturbation profiles $T'(z,t)$ are zero-padded and filtered spectrally in the vertical domain with a low pass filter. Choosing a cut-off wavelength of 20 km retains most gravity waves while other background effects are sufficiently suppressed.

Figure 4.24 illustrates the basic steps of the gravity wave extraction procedure. The temperature profile $T(z)$ observed at 00:01 UT on August 25, 2011 is shown in figure 4.24a. Overlaid in red is the background temperature $T_0(z)$ produced by the fitting algorithm discussed above. Subtracting $T_0(z)$ from $T(z)$ yields the gravity wave-induced temperature perturbation drawn in figure 4.24b. The progression of the temperature disturbances on August 24/25 2011 can be seen in figure 4.24c. This color coded plot comprises 144 temperature perturbation profiles of 1 h × 2 km resolution and 15 minutes shifted from another. The stripe pattern indicates a quasi monochromatic wave train with an average vertical phase speed of -0.3 m/s and observed period of 9 h. While the wave structure can be clearly identified in the second half of the measurement, the first half appears to be distorted with notable temperature enhancements at 35, 39, 43 and 41 km altitude. This is likely the result of interference with another wave propagating in a different direction. Because wave packets are finite in the time domain as well as spatially, a complex intersection region is created which also varies in time and space. Interference pattern may thus appear and vanish at random time intervals although a coherent wave train may be observed for extended periods.

The superposition of all hourly temperature perturbation profiles obtained from the lidar observation on August 24/25, 2011, is shown in figure 4.24d. Red lines mark mean and R.M.S. values. The mean is close to zero at all altitudes as expected from the long-

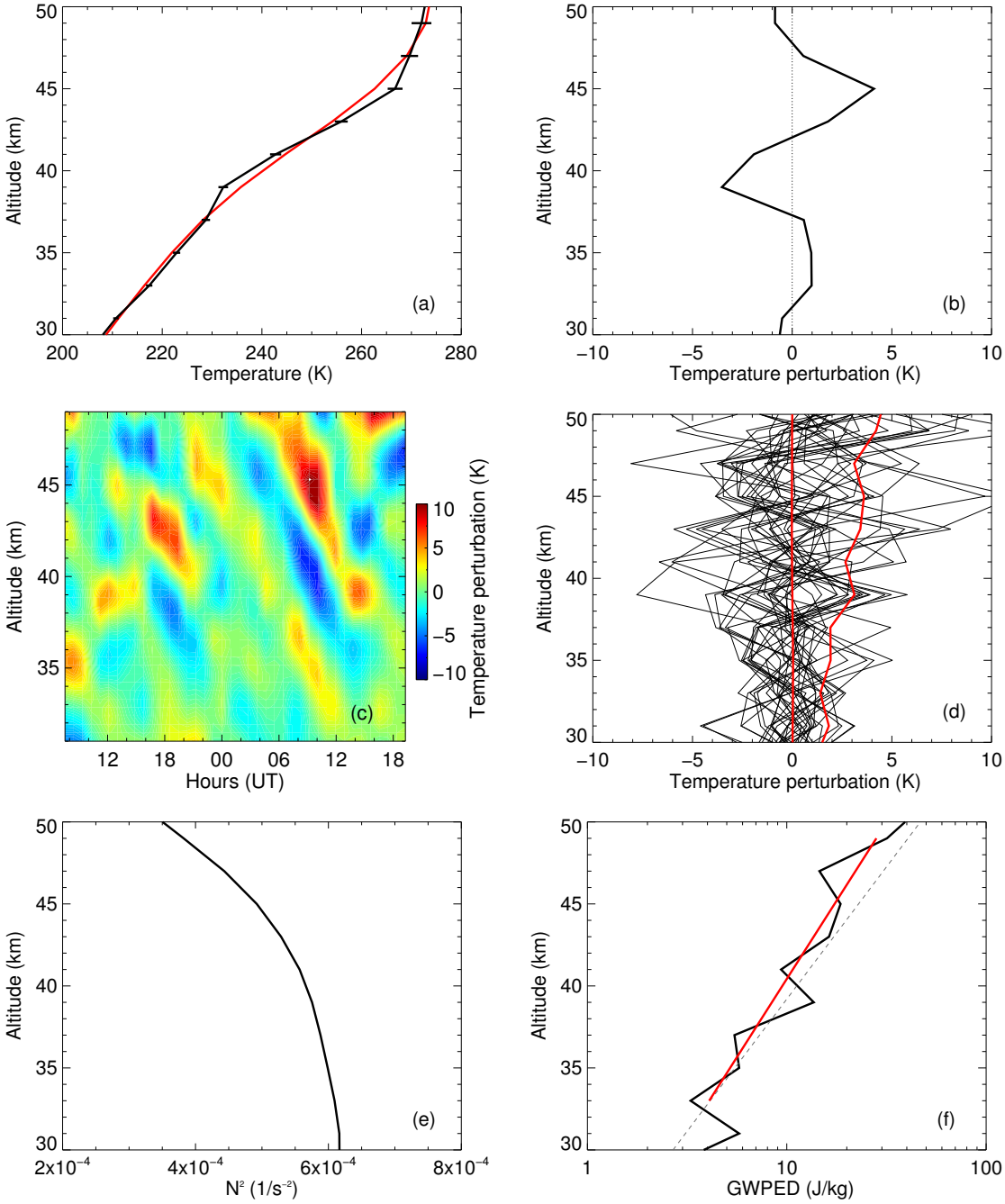


Figure 4.24.: Illustration of the gravity wave analysis process: (a) Measured temperature profile $T(z)$ (black line) with uncertainties (horizontal bars) at 00:01 UT on August 25, 2011, and estimated background temperature profile $T_0(z)$ (red line). (b) Perturbation profile $T'(z) = T(z) - T_0(z)$. (c) Temperature perturbation $T'(z,t)$ on August 24/25, 2011. (d) Hourly perturbation profiles $T'(z,t)$ (black lines). Red lines mark mean and R.M.S perturbation amplitudes. (e) Mean Brunt-Väisälä frequency. (f) Mean gravity wave potential energy density. The red line marks the estimated mean, the dashed line the expected growth rate of a conservative wave.

term average of a clean wave-induced perturbation field. Thus figure 4.24d confirms that the gravity wave extraction procedure does not introduce any bias which could lead to altitude dependent over- and/or underestimation of gravity wave amplitudes. The mean gravity wave amplitude characterized by the R.M.S temperature perturbation $\overline{T'}(z)$ grows with height from 1.5 K at 30 km altitude to 4.5 K at 50 km altitude. This is approximately consistent with a freely propagating gravity wave. In the absence of dissipation and refraction, gravity wave amplitudes are expected to increase exponentially according to $e^{z/2H}$ due to reasons of energy conservation, where H is the scale height of the atmosphere.

Gravity wave energy

When the amplitudes of the gravity waves are known, the (available) potential energy density E_p can be calculated from equation 4.12. Two normalizations of E_p are common in literature: energy density per mass and energy density per volume. If not stated otherwise, in this work the potential energy is normalized by mass and the value is computed according to

$$E_p(z) = \frac{1}{2} \frac{g^2}{\overline{N^2(z,t)}} \overline{\left(\frac{T'(z,t)}{T_0(z,t)} \right)^2}. \tag{4.13}$$

Equation 4.13 is an adapted version of the definition of E_p (equation 4.12) where the normalized density perturbation, ρ'/ρ_0, is replaced with the normalized temperature perturbation T'/T_0.

The squared Brunt-Väisälä frequency $N^2(z,t)$ is calculated from measured temperature profiles using the relation

$$N^2(z,t) = \frac{g}{T_0(z,t)} \left(\frac{\partial T_0(z,t)}{\partial z} + \frac{g}{C_p} \right), \tag{4.14}$$

where g is the gravitational acceleration, equal to approximately $9.7\,\mathrm{m\,s^{-2}}$ in the stratosphere; C_p is the specific heat at constant pressure, equal to $1004\,\mathrm{J\,K^{-1}\,kg^{-1}}$. Overbars in equation 4.13 denote temporal averages with regard to the duration of the lidar observation.

Figure 4.24e shows the mean N^2 profile obtained from lidar observation on August 24/25, 2011. It is approximately constant below 40 km altitude due to the steady increase in background temperature (see figure 4.24a), and falls off towards the stratopause. The gravity wave potential energy density (GWPED) profile of the same observation is displayed in figure 4.24f. In the absence of dissipation and refraction one would expect a constant growth rate as indicated by the dashed line. Comparing the expected growth rate with an exponential fitted to the GWPED profile (red line) reveals that approximately one fourth of the energy is lost between 33 and 50 km altitude. The peak at 31 km is likely caused by noise and is therefore excluded from the fit.

4.4.3. Variability of gravity wave activity

In this section the gravity wave activity is characterized by means of the gravity wave potential energy density calculated from lidar observations as described in section 4.4.2. More than 2600 hours of data were acquired during the Antarctic lidar project in the time frame December 2010 to May 2012. A detailed description of the dataset can be found in section 4.1. In order to obtain energy density estimates of comparable

x	4	6	8	10	20	40
$\overline{q_x}$ at 45 km altitude (%)	41	38	38	37	39	47
$\overline{q_x}$ at 35 km altitude (%)	58	56	53	53	57	64

Table 4.2.: Comparison of the mean relative variability $\overline{q_x}$ at two different altitudes. The index x specifies the averaging period in days. Mean time interval between two observations is 5 days. See text for more details.

reliability, two main quality criteria were implemented for the gravity wave study: (1) individual observations are required to be longer than 6 hours, and (2) the monthly number of hours needs to be larger than 80 hours. The first criteria reduces the effect of transient events (e.g. single waves with large amplitudes) by guaranteeing a minimum number of hourly profiles to be averaged, while the second criteria makes sure that the monthly time intervals contain a minimum number of data points. Rejecting months with less than 80 hours of observations results in continuous subset spanning January 2011 to April 2012 (see figure 4.3). In this time frame 94 lidar observations totaling 2293 hours were found to satisfy criteria (2). The mean number of hours per observation in this data subset is 24.3 h.

Figure 4.25 shows calculated gravity wave potential energy densities (GWPEDs) for the January 2011 to April 2012 time frame at two altitudes 35 and 45 km. Each dot represents the average of one lidar observation, the length t_{obs} is indicated by color: Black dots mark measurements shorter than 12 hours, red dots measurements longer than 48 hours. Time intervals of 12 to 24 hours and 24 to 48 hours are indicated by blue and green color, respectively. Lidar observations may include data gaps of up to 6 hours which are, however, not included in t_{obs}. Thus, for example, a particular measurement with $t_{\mathrm{obs}} = 36$ h is guaranteed to comprise 36 hourly profiles although the actual observation period may be as long as 42 hours.

Short-term variability

As evident from figure 4.25, there is significant short-term variability, i.e. from one observation to the next few days later. On this time scale, E_p values can change by as much as half an order of magnitude. This variation is much larger than typical uncertainties estimated from error propagation of instrumental noise. The magnitude of noise induced uncertainties depends, amongst others, on the length of the lidar observation and amounts to less than 10% in most cases. Hence, noise does not play a major role. This is also evident from the comparison of E_p values obtained from long observations with those from short observations. For instance, in May/June 2011 observation with $t_{\mathrm{obs}} > 48$ h (red dots) show similar variations in E_p as compared to observations with $12\,\mathrm{h} \leq t_{\mathrm{obs}} < 24\,\mathrm{h}$ (blue dots). It is thus concluded that the variability seen in figure 4.25 is indeed for the most part due to geophysical effects.

In order to quantify the variability on short time scales, a new quantity called "mean relative variability" $\overline{q_x}$ is introduced, where the index x specifies the averaging period in days. $\overline{q_x}$ is computed as follows: A time interval of x days is formed for each lidar observation centered at the observation time. Next, mean values $\overline{E_p}$ and standard deviations σ_{Ep} are computed for all time intervals. Finally, $\overline{q_x}$ is obtained by taking the average over all ratios $\sigma_{Ep}/\overline{E_p}$.

Table 4.2 lists $\overline{q_x}$ values for two altitudes 35 and 45 km. The most striking feature is that short-term variability is significantly larger at lower altitudes, e.g. 37% at 45 km versus 53% at 35 km (10 day averaging period). Although the relative variability depends

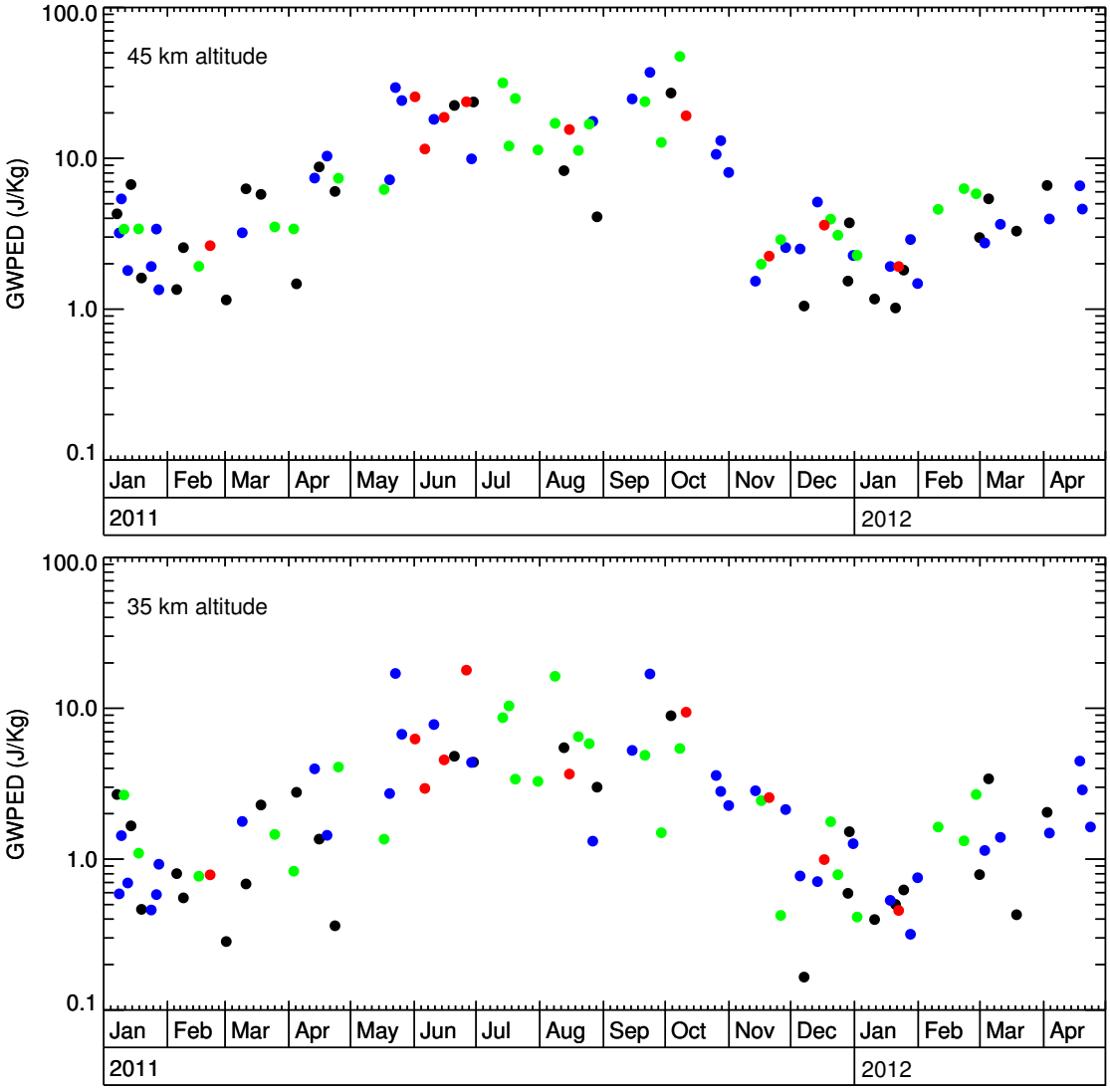


Figure 4.25.: Gravity wave potential energy density at 45 km altitude (top) and at 35 km altitude (bottom). Each dot represents the average of one observation of variable duration. The length of the observation t_{obs} is indicated by color: $6\,\mathrm{h} \leq t_{\mathrm{obs}} < 12\,\mathrm{h}$: black; $12\,\mathrm{h} \leq t_{\mathrm{obs}} < 24\,\mathrm{h}$: blue; $24\,\mathrm{h} \leq t_{\mathrm{obs}} < 48\,\mathrm{h}$: green; $t_{\mathrm{obs}} > 48\,\mathrm{h}$: red. Observations may include gaps of up to six hours which are not included in t_{obs}. Typical uncertainties due to instrumental noise are less than 10%.

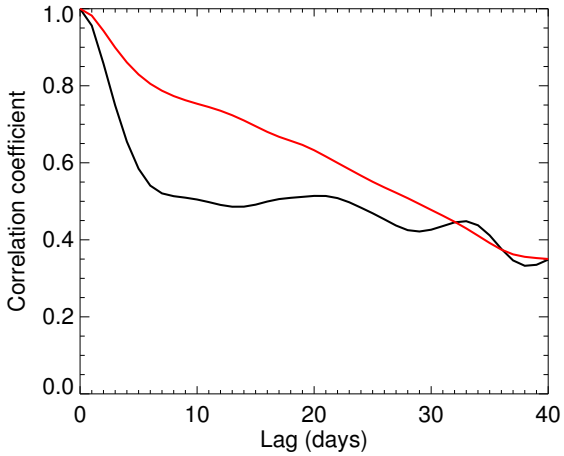


Figure 4.26.: Comparison of the autocorrelation of the gravity wave potential energy density time series at 35 km altitude (black) and at 45 km altitude (red).

slightly on the averaging period, the difference in variability between the two altitudes is approximately constant (20 days 18%, 8 days 15%). This suggests some sort of filter mechanism which reduces the variability at higher altitudes on time scales of few days. It will be shown in section 4.4.5 that the process of wind filtering is indeed responsible for taking out part of the variability of the gravity wave activity.

Comparing the $\overline{q_x}$-values with each other at one altitude reveals another interesting property: Starting at short averaging periods, the relative variability decreases with increasing averaging intervals and reaches a minimum between 8 to 10 days. Beyond 10 days the variability increases again. While latter increase can be attributed to large amplitude seasonal variations becoming more and more dominant over short-term variations, the initial decrease is likely a statistical effect. In case of 4 days averaging period the mean number of lidar observations is 2.2, i.e. for the majority of all four-day intervals there are only two observations per interval[1]. Since the variability is not strictly normally distributed, a significant bias is introduced for small sample populations, which in turn leads to larger estimates for $\overline{q_x}$. Hence, most reliable $\overline{q_x}$-values characterizing the short-term variability of gravity waves are found for the 10 day averaging period.

A similar conclusion is reached when looking at the autocorrelation of the GWPED time series. Figure 4.26 displays correlation coefficients as function of time lag. As expected, the time series at 45 km altitude (red line) shows larger correlation coefficients (less short-term variability) than the time series at 35 km altitude (black line). A notable exception are time lags larger then 30 days when correlation coefficients of both time series are approximately equal. On the other end of the spectrum, the fast decay of the correlation for time lags up to 5 days is caused by linear interpolation of the irregular spaced GWPED time series[2].

Annual cycle

In addition to short-term variations, there is a clear annual cycle visible in figure 4.25, with low GWPED-values in austral summer and high gravity wave activity in austral winter. In order to extract this cycle, GWPED-values in two altitude bands 30 to 40 km and 40 to 50 km are averaged assuming a log-normal distribution, and monthly means

[1] Intervals with less than two observations are rejected because the variability can only be estimated from two or more observations.

[2] 5 days is the mean time between two successive lidar observations

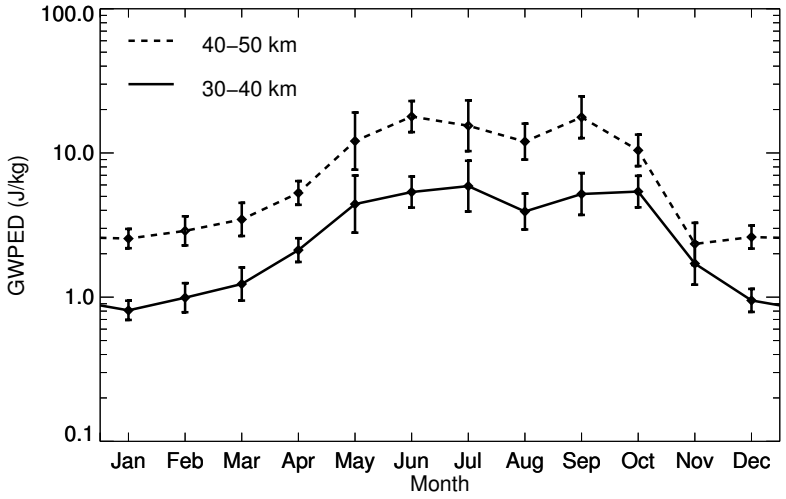


Figure 4.27.: Monthly mean gravity wave potential energy density in 30–40 km (solid line) and 40–50 km altitude (dashed line). Error bars mark standard errors.

are formed. The resulting climatology of gravity wave potential energy density is shown in figure 4.27. Lowest energy densities (30 to 40 km altitude) are found in December and January with 1.0 J/kg and 0.8 J/kg, respectively. Then the energy density increases through austral autumn and reaches a first prolonged maximum in May–July. After going through a small dip in August, a second maximum is reached in September/October before the energy density falls off again in late austral spring. Maximum energy densities are reached in July (5.9 J/kg) and October (5.4 J/kg). By comparison, the local minimum in August with 3.9 J/kg is rather distinct (note the logarithmic scale in figure 4.27). A similar double humped structure is also seen in the 40 to 50 km band. Peak energy densities occur in June (17.9 J/kg) and September (17.7 J/kg), which is one month earlier than at lower altitudes. The dip (12.0 J/kg), however, is located at the same place, and the magnitude of the relative drop in energy density (-33%) is comparable to the drop observed in the lower altitude band.

The magnitude of the variation in the annual cycle is slightly larger in the upper altitude band. Monthly energy densities in 40–50 km altitude in winter increase by a factor of 7.6 with respect to the minimum observed in summer, while energy densities in the lower altitude band vary by a factor of 6.6. q_x values for the upper and lower altitude band are 77% and 70%, respectively. The comparison with $\overline{q_{10}}$ (table 4.2) suggests that seasonal variability exceeds short-term variability by approximately 25%. For comparison, seasonal variability of the mean temperature at 33 km altitude is approximately one order of magnitude larger than the corresponding short-term variability (see figure 4.11).

4.4.4. Growth of gravity wave amplitudes with height

It was shown in section 1.2 that gravity wave amplitudes grow with height due to conservation of energy. In the absence of dissipation growth is approximately $\propto \exp(z/2H)$, where H is the scale height of the atmosphere. This translates into $\propto \exp(z/H)$ if the gravity wave energy is chosen as observable instead of the amplitude and linear wave theory is applicable (i.e. the energy of the wave is proportional to the amplitude squared). Thus, a growth rate $\propto \exp(z/H)$ for energy densities is indicative of a freely propagating (conservative) wave. The reverse is, however, not necessarily the case, namely that a non-conservative growth rate implies energy dissipation. Under

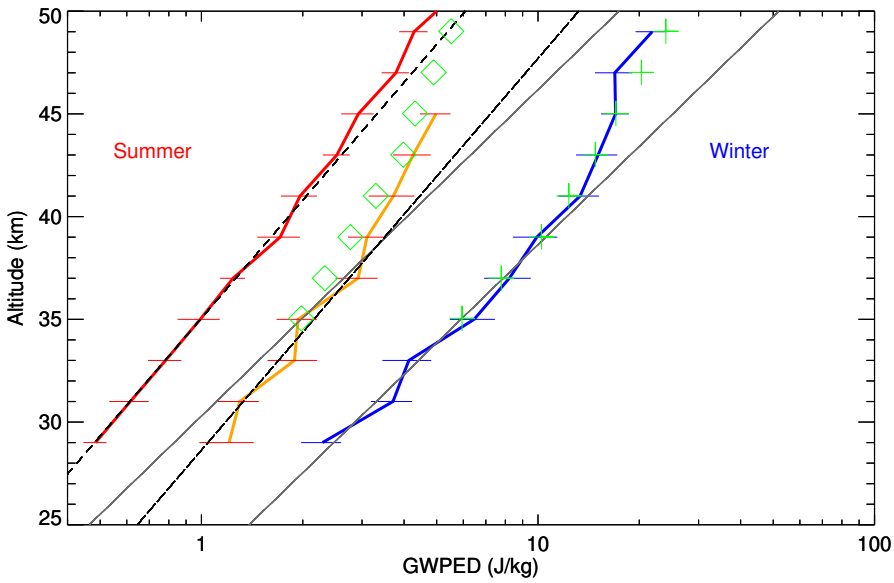


Figure 4.28.: Mean gravity wave potential energy density (GWPED) as function of altitude for summer (January and February, red line), winter (May to August, blue line) and autumn (March and April, orange line). Horizontal lines indicate standard errors. Values marked by green crosses and diamonds (seasons 2007 and 2008) are taken from *Alexander et al.* (2011). Dotted lines mark conservative growth rates $E_\mathrm{p} \propto \exp(z/H)$, dashed lines growth rates $E_\mathrm{p} \propto \exp(z/1.2H)$.

certain background conditions waves can be refracted and channeled in "atmospheric waveguides" (*Fritts and Yuan*, 1989). Gravity waves may propagate freely inside these waveguides, while at the same time the wave energy density decreases directly above the waveguide. Therefore caution should be taken when interpreting gravity wave energy density profiles.

Another complication results from the fact that lidar observations are point observations. Without spatial coverage it is usually not possible to observe the evolution of a single gravity wave. Many waves may intersect the laser beam at one altitude or another, travelling in different directions. Thus, lidars always observe the superposition of the wave field, and energy density profiles derived from lidar measurements should be treated in a statistical sense as measurement of the wave field rather than single gravity waves.

Seasonal GWPED profiles are displayed in figure 4.28. The mean winter profile (blue line) includes months May to August, the summer profile (red line) months November to February, and the autumn profile (orange line) months March and April. The latter two profiles span two years, 2011 and 2012, while the winter profile covers austral winter 2011 only as no lidar observations were carried out in austral winter 2012. More information on the number of observation hours per season can be found in section 4.1.

In winter, the growth in energy density matches the conservative growth rate (dotted line) up to 41 km altitude. This suggests that nearly all the energy is carried up to this level with no or very little dissipation between 29 and 41 km. However, in the following 8 km approximately half of the energy is lost due to gradual saturation of the gravity wave spectrum. Saturation effects and their manifestation in the vertical wavenumber spectrum are discussed in more detail in the following section.

Conservative growth rates in austral winter have previously been reported by *Alexan-*

der et al. (2011). *Alexander et al.* studied gravity waves observed by a Rayleigh lidar located at Davis station. Comparing their results (green crosses in figure 4.28) with Fe lidar measurements yields no significant difference within stated uncertainties. This is a rather surprising result given the data used by *Alexander et al.* were obtained in 2007 and 2008, three years before Fe lidar winter observations. The absence of any significant difference in GWPED between seasons 2007/2008 and 2011 suggest that the year-to-year variability in winter is small. This does not only include the region with freely propagating waves between 29 and 41 km, but also the altitude where the energy density departs the conservative growth rate.

The summer potential energy density profile (red line in figure 4.28) is characterized by a constant growth rate between 29 and 43 km altitude. The scale height estimated from the growth rate is approximately 20% larger compared to the winter state. At first sight one is tempted to conclude that wave propagation in summer is clearly non-conservative. However, taking into account seasonal variations in background temperature mitigates drastically the magnitude of the departure from the conservative growth rate. Computing mean background temperatures averaged over the altitude range 29–41 km reveals that the atmosphere in the region of interest is 29 K warmer in summer. This translates into a 860 m or 13% increase in scale height as calculated from relation $H = kT/Mg$, where $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T the atmospheric temperature, $M = 4.81 \times 10^{-26}$ kg the mean molecular mass of air, and $g = 9.70\,\mathrm{m/s^2}$ the acceleration due to gravity. Including this scale height variation in the comparison of potential energy density growth rates reduces the energy loss per scale height in summer relative to winter to 7%. This value is well within the uncertainty of the measurements. It is thus likely that conservative or nearly conservative growth rates are observed in winter as well as in summer. Judging by the spread of the observations, summer GWPED values in figure 4.28 appear to be more robust than winter values, the larger part of the 7% discrepancy therefore may be associated with winter GWPED observations. This comes as no surprise given that there are more observations in summer than in winter (see section 4.1).

Potential energy densities at the 31 km level are 3.7 J/kg in winter and 0.6 J/kg in summer. Thus, the energy density changes by a factor of six. This value compares to the peak-to-peak variation seen in the annual cycle (figure 4.27). The autumn GWPED profile (orange line) in figure 4.28 takes a special position in between the summer profile and the winter profile. The bottom part follows approximately a conservative growth rate, while above the 37 km level the profile converges towards the summer profile. The point where the profile departs the constant growth rate is thus considerably lower compared to the summer/winter profile. Values above 45 km are not shown because of large statistical uncertainties in the gravity wave analysis caused by low lidar signal in March/April 2012.

Also shown in figure 4.28 is the autumn GWPED profile published by *Alexander et al.* (2011) (green diamonds). Again, this profile matches surprisingly well Fe lidar observations although differences are larger in comparison to the winter case. Most of the differences, however, can be attributed to a shift of the two profiles relative to each other. Given that there are relatively few measurements in March (see section 4.1), Fe lidar observations in autumn are likely biased towards the winter state. This shifts the profile to larger GWPED values. The shift may explain observed differences between the two autumn profiles. Thereby it seems reasonable to conclude that there is very little if any year-to-year variation.

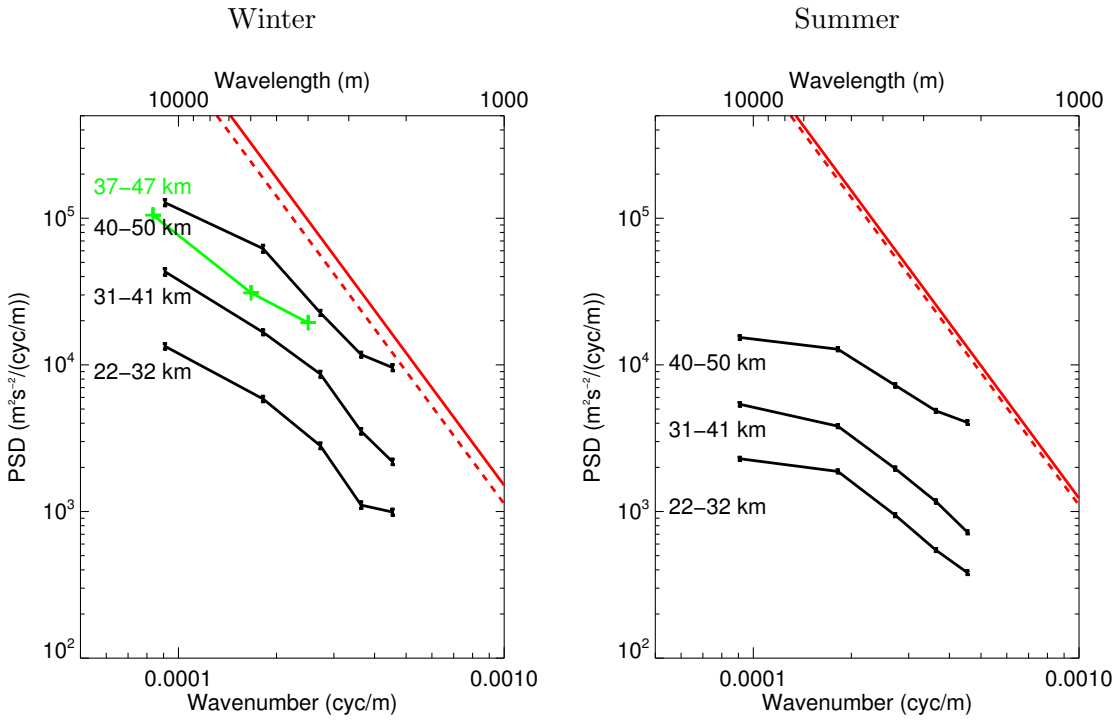


Figure 4.29.: Mean vertical wavenumber potential energy density power spectral density for winter (left, months June and September) and summer (right, months January and February) at three different altitude ranges. The green spectrum is taken from *Alexander et al.* (2011) and covers the combined winter season 2007 and 2008. Also shown is the saturation limit $\frac{1}{(2\pi)^2}\frac{N^2}{10m^3}$ at 25 km (red line) and 45 km altitude (red dashed line).

Vertical wavenumber spectra

Potential energy density power spectral densities are used to determine whether the non-conservative GWPED growth rates observed above 40 km altitude are connected to saturation of the gravity wave spectrum. The spectra are calculated as follows: Hourly fractional temperature perturbation profiles with 1 km vertical resolution are extracted from lidar measurements as discussed in section 4.4.2. These individual profiles are then scaled by $g/N\,(z)$ and split into three slightly overlapping altitude ranges before calculating vertical wavenumber (m) power spectra, following *Whiteway et al.* (1997) and *Alexander et al.* (2011). All spectra belonging to the same altitude range within a period of two months are subsequently averaged in order to decrease the variance of the resulting mean spectrum and increase the confidence in the spectral amplitude estimates. The winter and summer mean spectra are illustrated in figure 4.29. Also included in the winter panel is the 2007–2008 mean potential energy density power spectrum published by *Alexander et al.*. The spectral amplitudes found by *Alexander et al.* for the 37–47 km altitude range are smaller compared to values found in this study for the 40–50 km range. This difference is probably due to two reasons: First, since gravity wave amplitudes grow exponentially with height, the higher altitude range used in this study results in larger spectral amplitudes. Second, *Alexander et al.* averaged spectra over the whole winter season including months with lower wave activity.

Also shown in figure 4.29 are model spectral amplitudes for saturated wave spectra of the form N^2/m^3. As described in *Smith et al.* (1987), the basic idea of the saturation

theory leading to the result N^2/m^3 is that gravity waves become convectively unstable if the perturbation amplitude causes the total lapse rate to become larger than the adiabatic lapse rate. Emerging instabilities limit the maximum amplitude of propagating waves and the gravity wave spectrum becomes saturated.

It is evident from figure 4.29 that all measured spectra are below the theoretical saturation limit. In winter, however, the 40–50 km spectrum approaches the limit at the high-wavenumber end of the spectrum, and the slope in the mid-wavenumber region is comparable to the form predicted by *Smith et al.* (1987). Thus, it can be concluded that the observed gravity wave spectrum is increasingly saturated above 40 km altitude. This is consistent with non-conservative GWPED growth rates in figure 4.28. Note that the calculation of gravity wave spectra involves averaging over an altitude range of 10 km. Estimated spectral amplitudes are therefore not expected to reach the theoretical saturation limit unless the transition from a non-saturated spectrum to saturation happens below the averaging range.

The summer spectra are well below the theoretical saturation limit. Hence, saturation of the gravity wave spectrum can be ruled out as explanation for non-conservative GWPED growth rates in summer. It will be shown in the next section that the apparent energy dissipation visible in the summer profile of figure 4.28 is consistent with gravity waves being filtered by critical layers.

4.4.5. The effect of the background wind field

Gravity wave propagation is related to the background wind field through Doppler-shifting and filtering of gravity waves. Doppler-shifting occurs whenever there is a non-zero component of the wind vector in the direction of wave propagation. Because the vertical wind speed is in the order of centimeters per second (e.g. *Hoppe and Fritts*, 1995) and thus usually much smaller than the horizontal wind component, the main contribution to the total frequency shift arises from Doppler-shifting in the horizontal plane, and the effect of the vertical wind can be neglected. Assuming zero vertical wind, medium-frequency gravity waves ($N \gg \hat{\omega} \gg f$) with observed frequencies ω are shifted relative to intrinsic frequencies $\hat{\omega}$ as

$$\hat{\omega} = \omega - \bar{u}_h k_h = (c_h - \bar{u}_h)\, k_h, \tag{4.15}$$

where u_h is the horizontal wind speed in the direction of wave propagation, c_h the horizontal phase speed, and k_h the horizontal wave number. The intrinsic phase speed is

$$\hat{c}_h = c_h - \bar{u}_h. \tag{4.16}$$

Substituting expression 4.15 into the dispersion relation (equation 4.9), the vertical wavelength $\lambda_z \equiv 2\pi/m$ becomes

$$\lambda_z = \frac{2\pi}{N}\,(c_h - \bar{u}_h)\,. \tag{4.17}$$

It is evident from this relation that for waves with positive phase speeds increasing the horizontal wind speed leads to compression of the vertical wavelength ($c_h > u_h$, $u_h \to c_h$). Conversely, waves with negative phase speeds are compressed by horizontal winds with negative wind speeds.

Waves encounter a critical level where the horizontal wind speed $\bar{u}_h$ approaches the horizontal phase speed c_h, and the vertical wavelength shrinks to zero. However, this is a pure theoretical limit which is never realized in the real atmosphere because instabil-

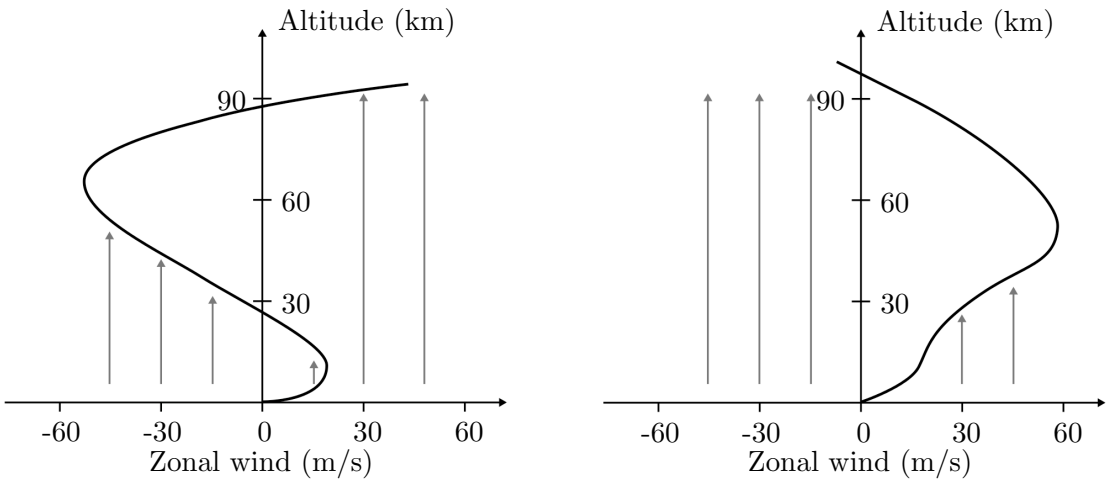


Figure 4.30.: Illustration of the filtering of gravity waves for typical high latitude zonal wind profiles in summer (left) and winter (right). Adapted from *Becker* (2012); based on *Lindzen* (1981).

ities and dissipation processes become important well below a critical layer (*Fritts and Alexander*, 2003). The amplitude of the wave is therefore significantly reduced before the wave encounters the critical layer defined by the condition $c_h = u_h$.

Gravity waves penetrating into the middle atmosphere have typical phase speeds which are well within the range of zonal wind speeds. Thus, filtering of waves can occur where the vertical gradient of the horizontal wind is nonzero (see figure 4.30). In the non-stationary case where the wind speed varies over time, gravity wave filtering leads to a time-dependent modulation of the gravity wave flux. It is therefore important to take the effect of the background wind into account when interpreting observed variations in gravity wave energy density.

Background conditions and link to gravity wave propagation

The background wind field above Davis is characterized by a strong stratospheric jet in the zonal direction and small wind speeds with fluctuating signs in the meridional direction. Insight on the annual variations of the wind structure can be gained from model results published by the European Center for Medium-Range Weather Forecasts (ECMWF). Figure 4.31 shows the model output of the zonal and meridional component of the wind field for the year 2011. A westerly circulation develops in the stratosphere in autumn and strengthens over the winter months. This so called polar night jet is strongest in the upper stratosphere where wind speeds can reach over 80 m/s. The jet is important because it acts as a barrier between the cold air in the region inside of the jet (the polar vortex) and the warmer air outside, thus effectively preventing any air exchange. This isolation of the polar air promotes ozone loss during winter months since ozone-rich air from the mid latitudes cannot be transported into the polar vortex. Also, the isolation facilitates the development of very low temperatures inside the vortex. A visible sign of these low temperatures is the frequent occurrence of polar stratospheric clouds (PSCs) which can only exist below 197 K (*Tabazadeh et al.*, 1994).

With the breakdown of the polar vortex at the end of winter, the wind structure changes dramatically. While wind speeds in excess of 40 m/s at the 40 km level prevail throughout most of the time span April to October, the jet slows down in early November, followed by a wind reversal few days later. This wind reversal marks the

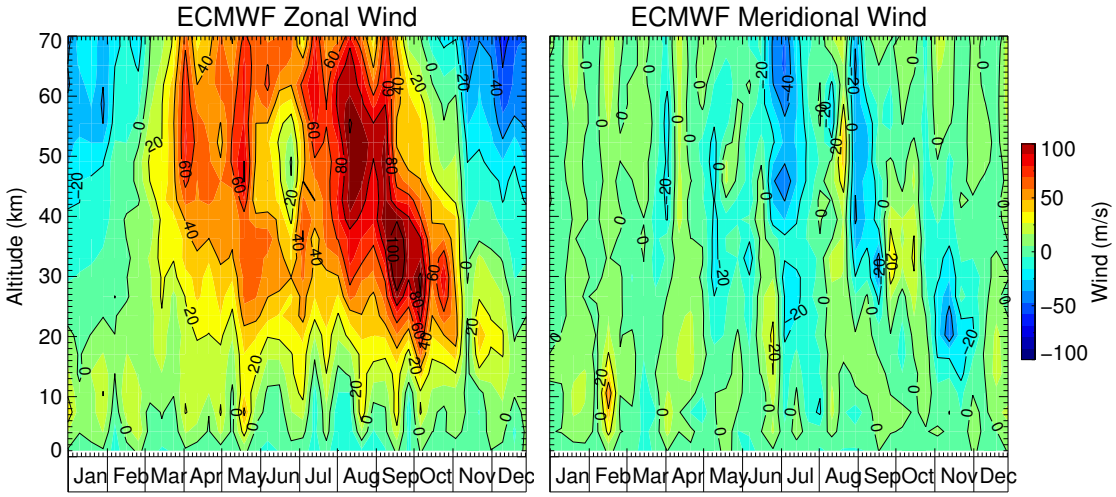


Figure 4.31.: Zonal and meridional wind above Davis for 2011 extracted from the ECMWF model dataset.

winter-summer transition in the atmosphere. The actual date of this transition can vary from year to year (*Kaifler et al.*, 2013), but this fact remains of secondary importance to this study since Fe lidar observations in winter were carried out in 2011 only.

Figure 4.32 shows seasonal zonal wind speed profiles extracted from ECMWF data. In the winter state the wind speed increases steadily from the tropopause up to the 35 km level where observed mean wind speeds are in the order of 50 m/s. Eastwards propagating waves are thus filtered with the exception of very fast waves with large phase speeds, while westwards propagating waves can penetrate undisturbed into the upper stratosphere. It is in particular the freely westwards propagating gravity waves that are responsible for the large gravity wave potential energy densities observed throughout winter (see figure 4.28).

In summer, the zonal wind speed reverses direction in the lower stratosphere. This causes nearly all gravity waves with low phase speeds to be filtered in the lower stratosphere. Because the source spectrum of gravity waves is believed to peak around zero phase speed (e.g. *Alexander and Vincent*, 2000), the removal of gravity waves with low phase speeds from the gravity wave spectrum is expected to have a large impact on observed potential energy densities in the upper stratosphere. Indeed, as evident from figure 4.28, the energy density observed in summer at 35 km altitude is approximately eight times lower than in winter.

The autumn wind profile in figure 4.32 is in between the winter state and the summer state. Large variations in wind speed are observed around this time of the year, though values remain positive throughout the stratosphere. The corresponding energy density profile (figure 4.28) resembles the winter profile at low altitudes, while above 35 km altitude the profile converges towards the summer profile. This comes at no surprise because the wind reversal (winter-summer transition) begins at high altitudes and progresses downward over time. The energy density profile in autumn can therefore be considered as a mixed state. In the upper stratosphere the winter-summer transition is already in progress, while at the same time the lower stratosphere is still, for the most part, in the winter state.

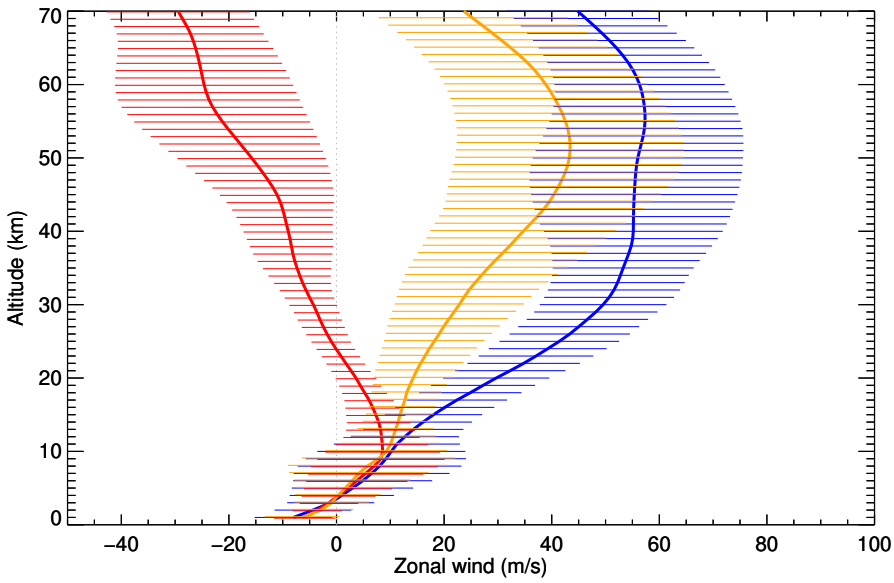


Figure 4.32.: Seasonal mean ECMWF zonal wind profiles for the same periods as in figure 4.28: summer (January and February, red), autumn (March and April, orange), and winter (May to August, blue). Horizontal lines indicate the variability of the wind estimated from the standard deviation.

Comparison of gravity wave potential energy density profiles and zonal wind profiles

More compelling evidence for gravity wave filtering being responsible for the observed large seasonal variation in gravity wave potential energy density (GWPED) is found when comparing GWPED profiles with zonal wind speed profiles on short time scales (few weeks). Although gravity wave sources are believed to be mostly isotropic and in the troposphere, few of the gravity waves propagating in meridional direction can penetrate into the stratosphere. The meridional wind with its low speed and fluctuating sign acts as effective barrier for these waves. Thus, zonally propagating waves are believed to carry most of the energy, and, in first approximation, the effect of gravity wave filtering in the stratosphere can be studied by considering zonally propagating waves only. Consequently, the influence of the meridional wind on gravity wave propagation is neglected in the following discussion.

Figures 4.33 and 4.34 show monthly GWPED profiles and monthly zonal wind profiles. In January, the atmosphere is in the summer state. The wind profile is approximately linear and ranges from 10 m/s at the tropopause to −25 m/s at 50 km altitude. Thus, most of the zonally propagating gravity waves are filtered and minimum potential energy densities are observed. Gravity waves which reach the stratopause must have sufficiently high phase speeds in order to being able to cross the stratosphere. That these fast waves indeed propagate freely can be be seen from the GWPED profile which matches the conservative growth rate marked by gray lines. Note that in order to facilitate easy comparisons, the mean density scale height was used for computing conservative growth rates (gray lines) and seasonal effects are neglected. Due to seasonal changes in temperature, actual conservative growth rates in summer and winter are approximately 5% smaller and 5% larger, respectively.

The potential energy density increases over the period February–April and reaches the winter plateau in May. This increase is made possible by the formation of the polar night jet with its strengthening westerly winds which open up the path for low phase speed

gravity waves propagating eastwards. In March the zonal wind speed is nearly constant throughout the lower stratosphere, but increases slightly in the upper stratosphere. As a result, eastwards propagating gravity waves can reach the stratopause undisturbed. Fast westwards propagating waves can also penetrate into the lower stratosphere, but are increasingly filtered in the upper stratosphere where the wind speed increases. The net effect produces energy density profiles which deviate from the conservative growth rate at high altitudes.

In June, the wind speed gradient is steep enough such that all low speed gravity waves which propagate eastwards are already filtered below the 25 km level, and waves which make it up to this level can propagate freely above. As result, one would expect conservative growth rates for the energy density. As evident from figure 4.34, this is indeed the case between 25 and 40 km altitude. The growth rate above 40 km is, however, limited by saturation of the gravity wave spectrum (figure 4.29).

The winter dip in the annual cycle of the GWPED (figure 4.27) occurring in July/August is possibly also related to changes in the stratospheric wind field. Compared to other winter months, in July the mean zonal wind profile (figure 4.34) shows smaller wind speeds above 40 km altitude. Moreover, the wind speed is much more variable with frequent excursions below 25 m/s in the upper stratosphere. The month August stands out because of the mean zonal wind speed reaching zero near the tropopause, which results in noticeable reduction of the gravity wave flux entering the stratosphere. It should be noted that the corresponding observed GWPED profile (figure 4.33) apparently contradicts this statement. However, GWPED values below 35 km altitude are possibly enhanced by additional wave sources (e.g. ducted waves) not visible at higher altitudes.

The polar night jet begins to slow down in the stratopause region in September. This breaking process is accompanied by the downward progression of the zonal wind maximum from 37 km in September to approximately 20 km in December. With the wind speed now decreasing, the threshold where gravity waves are being filtered by the zonal wind is moving back into the range of the slow-speed gravity waves, and, starting at high altitudes, a significant reduction in the gravity wave potential energy density is observed. While the GWPED profile in September is characterized by conservative growth rates up to 45 km and saturated gravity wave amplitudes above, the October profile shows smaller values in the stratopause region.

What dramatic influence the background wind can have on gravity wave propagation becomes most apparent in November and December. In November the GWPED decreases in the range 29–43 km altitude, while conservative growth rates are observed above. In December decreasing GWPED values are found up to 33 km altitude and conservative growth rates follow. By comparing these critical altitudes with monthly wind profiles (figure 4.34), it is found that gravity waves can propagate freely when the mean zonal wind speed crosses -5 m/s. This suggests that most gravity waves have phase speeds below 5 m/s.

Link between gravity wave filtering and zero-crossing of the zonal wind

Figure 4.35a and figure 4.35b show height ranges where the absolute value of the wind speed is < 2 m/s (blue areas). The gray shaded area marks the altitude ranges where the Fe lidar can measure GWPED. Assuming that most gravity waves have low phase speeds in the order of 5 m/s or less, it is evident from figure 4.35 that zonally propagating low-speed gravity waves can propagate freely within the observation range in winter (months April–September) and summer (January). On the other hand, meridionally propagating

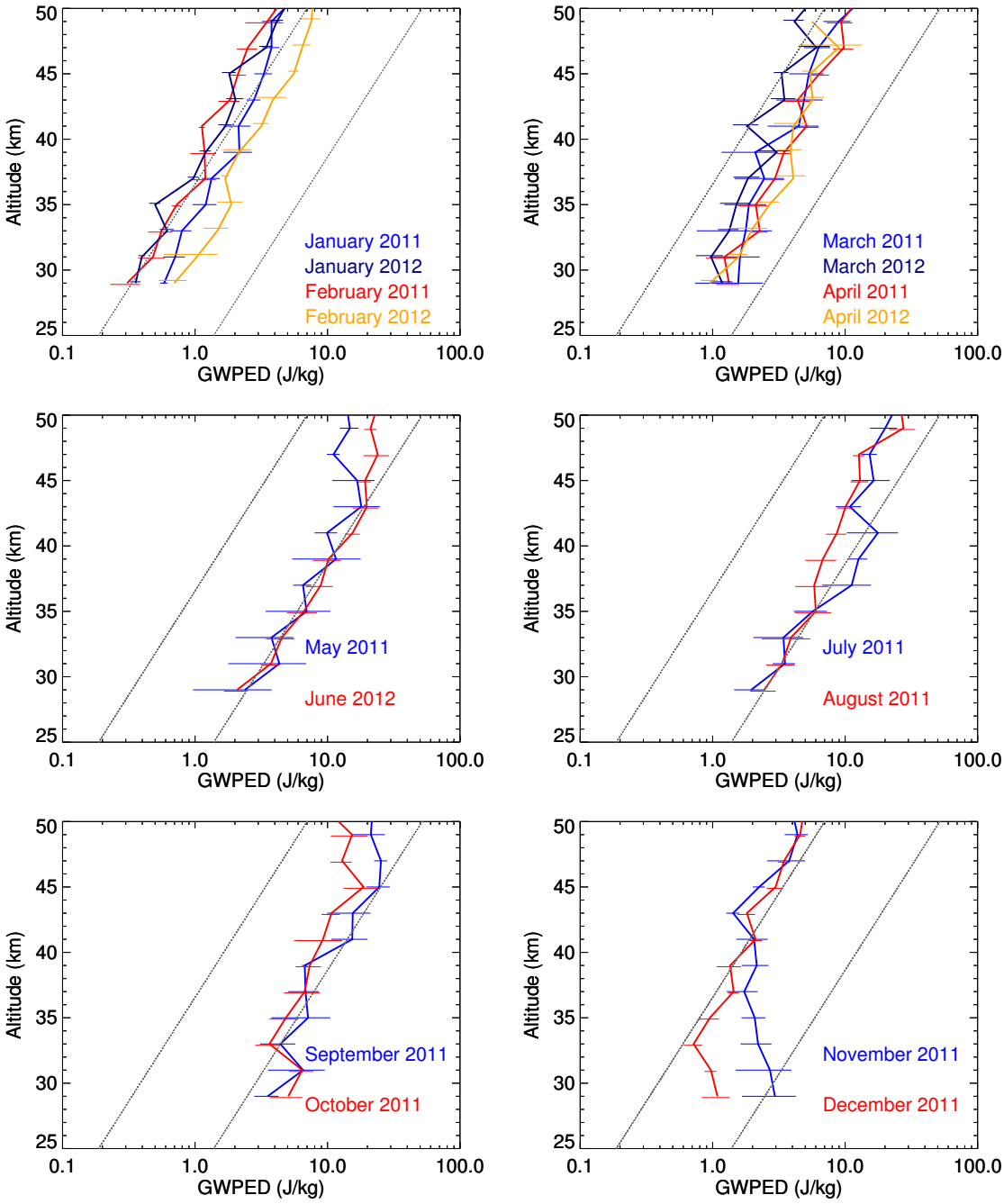


Figure 4.33.: Sequence of monthly gravity wave potential energy density profiles. Error bars indicate standard errors, and gray lines mark conservative growth rates ($E_\mathrm{p} \propto \exp(z/H_\mathrm{mean})$). Note that mean density scale heights are assumed here and seasonal effects on scale heights are neglected.

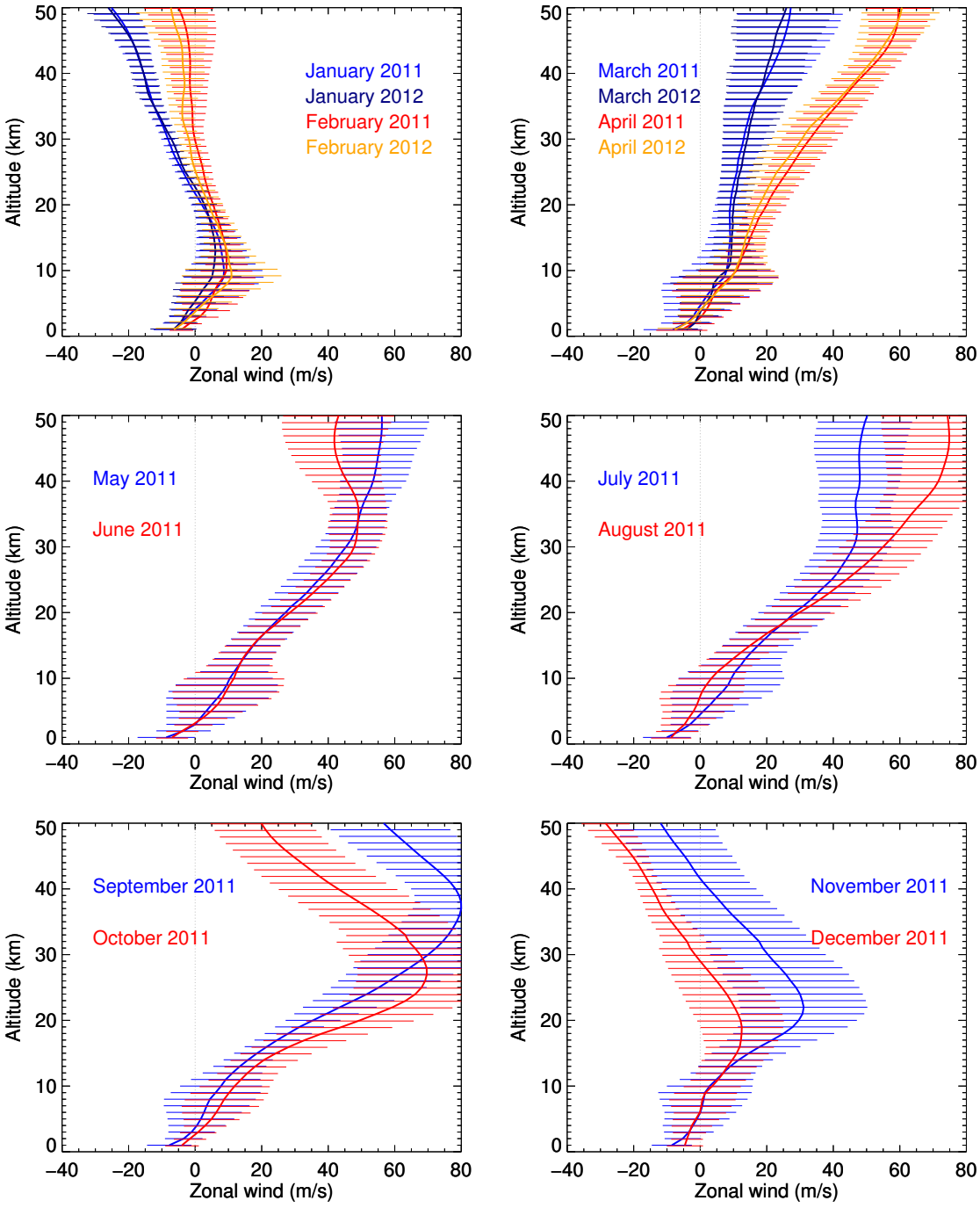


Figure 4.34.: ECMWF mean zonal winds for the same periods as in figure 4.33. Horizontal lines indicate the variability estimated from the standard deviation.

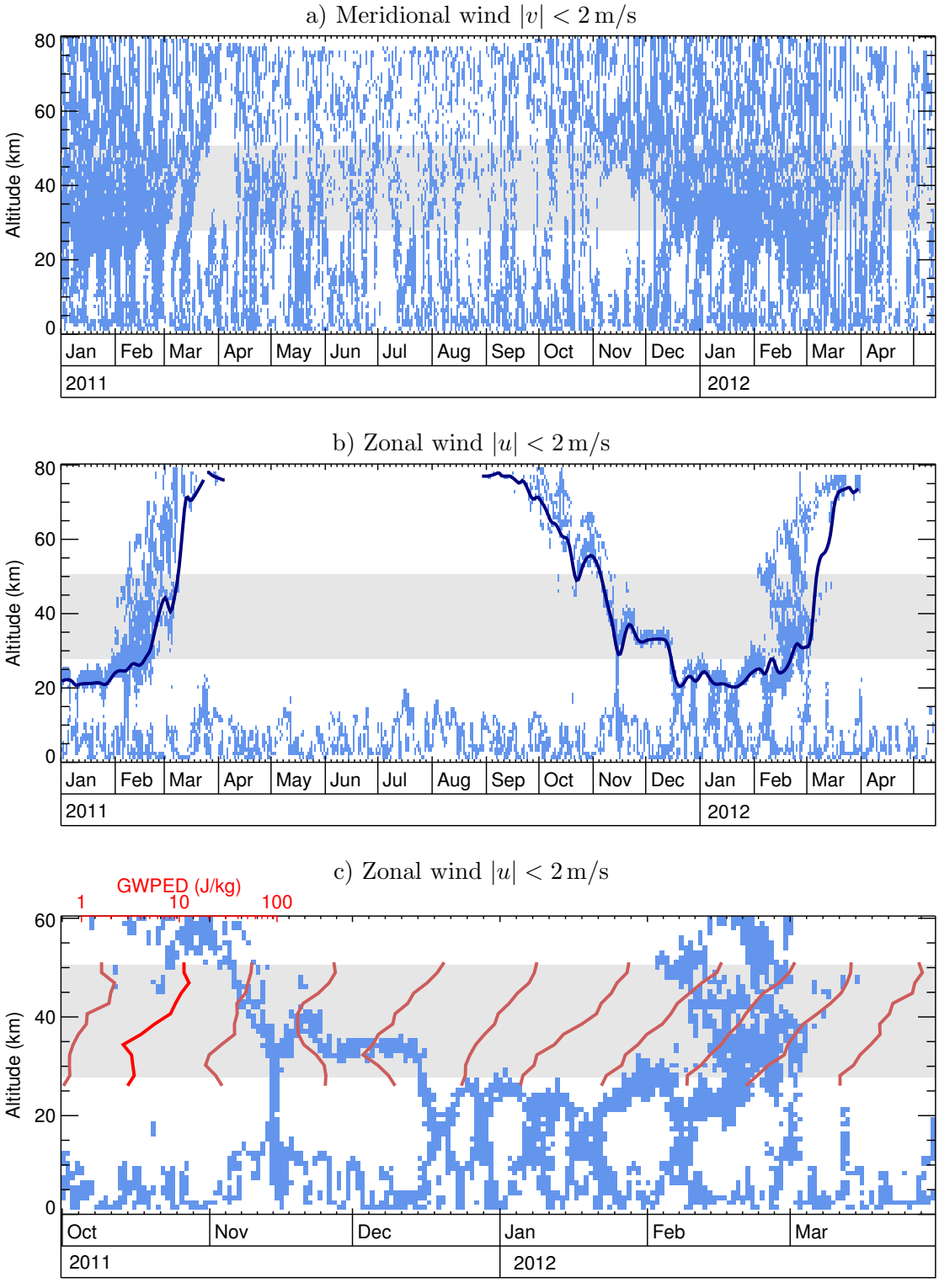


Figure 4.35.: Interaction of gravity waves with the background wind field (wind data taken from ECMWF): a) Meridional wind, and b) zonal wind. Blue shaded ares mark regions where the wind speed is $< 2\,\mathrm{m/s}$ and wave filtering is to be expected. The dark blue line in b) shows the mean height of the lower boundary of the wind reversal (see also figure 4.31). Gravity wave measurements by lidar are possible in gray shaded areas. c) Detail of b) with GWPED profiles overlaid.

gravity waves encounter wind speeds $|v| < 2\,\mathrm{m/s}$ year-round and are therefore expected to be subject to wind filtering in summer as well as in winter.

The summer months with GWPED measurements overlaid are shown in figure 4.35c. GWPED profiles are 14-day averages and appear thus more noisy than the monthly profiles discussed before. Nevertheless it is possible to study the effect of the zonal wind on gravity wave propagation in detail: Apart from noise, the two profiles in October show approximately conservative growth rates up to 45 km altitude where scattered spots with low wind speeds first appear. The zero-crossing of the zonal wind as well as the point where the GWPED profile departs the conservative growth rate progresses downward in November. A clear link between reduced GWPED values and low wind speeds is evident from the second November profile. In this profile, the position of the "bulge" matches closely the altitude range with wind speeds $< 2\,\mathrm{m/s}$. A similar correlation can be seen in the first half of December. Here, the region with low wind speeds is vertically more confined and a kink appears at the same altitude in the corresponding GWPED profile.

The altitude where the zonal wind reverses falls below the observing range of the Fe lidar in the second half of December. However, the effect on gravity wave propagation is still visible in the form of reduced GWPED values at the very bottom of the profile. Gravity waves which make it successfully through this critical altitude are expected to propagate freely above, and as result approximately conservative GWPED growth rates are observed. Within the following eight weeks the situation does not change much because the reversal of the zonal wind remains shortly below the observing range. Hence, within the altitude range the lidar is capable of measuring GWPED, free gravity wave propagation is observed in winter and in summer (see figure 4.28). The different GWPED values at common heights – values in summer are approximately six times smaller than in winter – are being caused by filtering of gravity waves with low phase speeds.

When comparing GWPED profiles with the zonal wind structure shown in the bottom panel of figure 4.35 it should be kept in mind that observations are compared with model results. While GWPED profiles are based on lidar measurements, zonal wind data is taken from the ECMWF model. Data input into the ECMWF model is very limited in the Antarctic region, and the accuracy of the model output is therefore degraded. This is especially true for altitudes above 40 km. Taking these limitations into account, it is surprising how consistent appears the relationship between GWPED and zonal wind.

4.4.6. A simple model to investigate the effect of critical layers

Description of the model

In order to study the effect of the zonal wind on gravity wave propagation more quantitatively, a simple model was developed. The basic steps are illustrated in figure 4.36. Gravity waves with a given energy distribution in phase-speed space are launched at the 12 km level and propagate upwards. It is important to note that the total amount of energy per unit phase speed is the primary quantity in this model regardless of the actual number of waves involved. On first sight this appears to be a rather crude oversimplification since it is well known that the more waves are generated, the more energy can be carried. However, as we are interested in the total energy as function of altitude, it makes no difference whether this energy is distributed between many low-energy waves or few high-energy waves. Leaving out the question of how much energy is carried by a single wave greatly reduces the complexity of the model.

For a real atmosphere, the question of how the gravity wave energy is distributed in

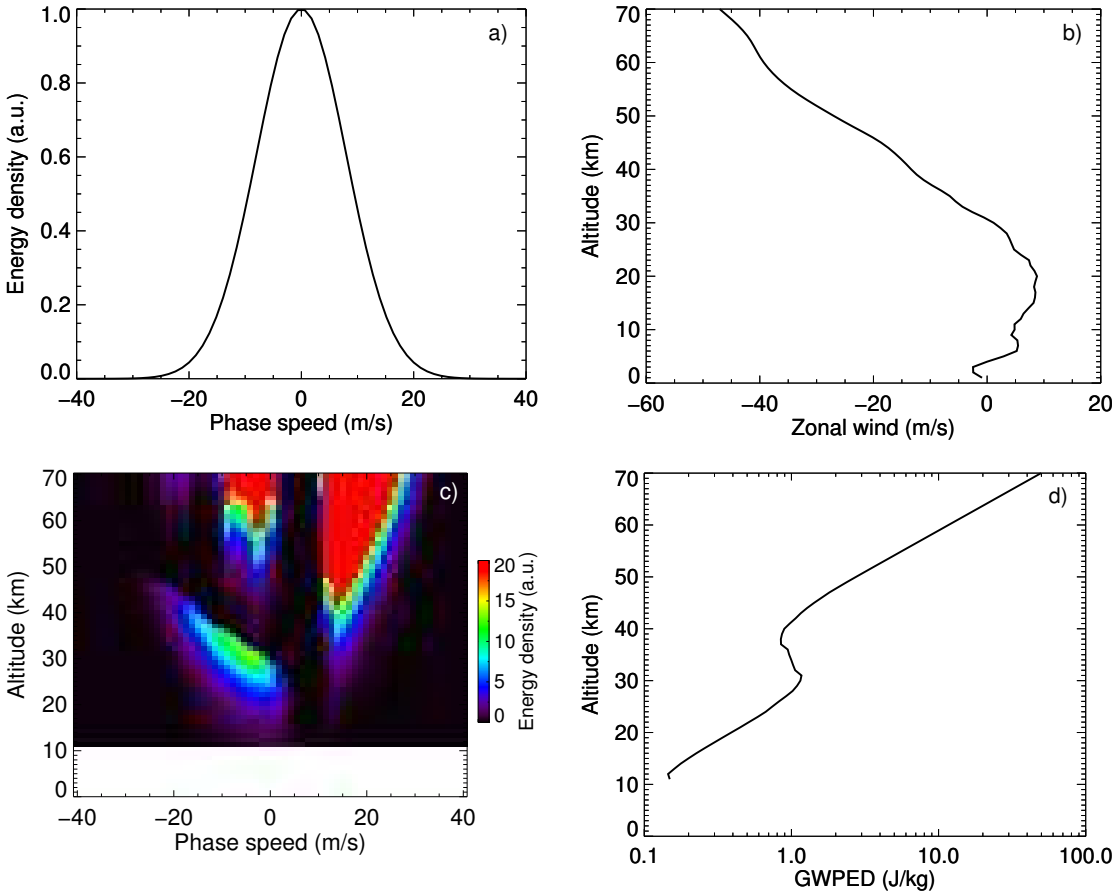


Figure 4.36.: Modelling gravity wave propagation in a background wind field: a) Assumed energy distribution of gravity waves at 12 km altitude; b) mean ECMWF zonal wind on December 15, 2011; c) energy distribution after critical layer filtering; d) resulting GWPED profile.

phase speed space is still not satisfactorily addressed in literature. In the absence of a deeper understanding of the processes involved in generating gravity waves, assuming a Gaussian dependence on wave phase speed is often the method of choice. For instance, several versions of the Whole Atmosphere Community Climate Model (WACCM) use a Gaussian distribution for the standard gravity wave source spectrum (*Beres et al.*, 2005). In this work, a Gaussian phase speed spectrum $\propto \exp\left(-(c/\sigma)^2\right)$, $\sigma = 8\,\mathrm{m/s}$, is used (figure 4.36a). The spectrum in the range of $-40\,\mathrm{m/s}$ to $40\,\mathrm{m/s}$ is partitioned into 100 bins approximately 0.8 m/s wide, and each bin is represented by a single "model gravity wave". These 100 waves then propagate upwards. The vertical resolution of the model is 1 km.

The effect of the zonal background wind is taken into account by comparing the phase speed c of the "model waves" with the zonal wind speed u extracted from ECMWF model runs. A sample wind speed profile is shown in figure 4.36b. If $|c-u|$ is larger than 5 m/s at a given altitude, the "model wave" is assumed to be propagating freely and the energy it carries is multiplied by $\eta = 1.1560$. Latter factor results from the conservative growth rate $\propto \exp(z/H)$ calculated for a 1 km interval. Conversely, the wave is assumed to be close to a critical layer if $|c-u| < 5\,\mathrm{m/s}$, and amplitudes are damped by reducing the energy multiplication factor η. The closer the wave gets to the

Condition	η
$\lvert c-u\rvert \geq 5\,\mathrm{m/s}$	1.156*
$2\,\mathrm{m/s} \leq \lvert c-u\rvert < 5\,\mathrm{m/s}$	1.000
$1\,\mathrm{m/s} \leq \lvert c-u\rvert < 2\,\mathrm{m/s}$	0.500
$\lvert c-u\rvert < 1\,\mathrm{m/s}$	0.100

Table 4.3.: Energy multiplication factors η employed in the model. The factors determine the energy changes a "model wave" undergoes when propagating vertically over the distance of 1 km. *Conservative growth rate

condition $|c-u| = 0$, the more the wave will be damped. A list of η-values used in this model is shown in table 4.3.

Figure 4.36c shows the resulting energy distribution as function of altitude. The energy density decreases rapidly where the "model waves" encounter critical layers. However, if the vertical extension of the critical layer is sufficiently small, i.e. the vertical gradient of the zonal wind speed is large, waves can penetrate the critical layer although the energy they carry is greatly reduced. In the example shown in figure 4.36, this happens to waves with phase speeds between approximately 0 and $-10\,\mathrm{m/s}$. In contrast, waves with phase speeds above $15\,\mathrm{m/s}$ do not encounter critical layers at all and therefore reach very large amplitudes at high altitudes since no limiting processes such as saturation of the gravity wave spectrum or convective instabilities are included in the model.

Finally, the energy spectrum (figure 4.36c) is integrated over phase speed to produce a GWPED profile similar to what the Fe lidar measures. The resulting profile needs to be scaled properly in order to convert the arbitrary energy units used in the model to the GWPED unit J/kg. The scaling factor is found by comparing the annual GWPED cycle of lidar measurements with model results. Note that this factor stays the same for all profiles belonging to one model run, i.e. it does not change from one month to another. In case of the example presented in figure 4.36 the scaled GWPED profile is shown in figure 4.36d.

Model results and comparison with lidar measurements

In order to compare model results with lidar measurements, model-based GWPED profiles are computed for each hourly lidar profile. Input data for the model are ECMWF zonal wind profiles which are available with six hour time resolution. The ECMWF profile which is closest in time to a particular lidar observation is chosen as representative wind profile to be used in the gravity wave filtering algorithm. Hence, up to six independent lidar observations share the same wind profile in the model. However, when comparing monthly averages, the large number of observation hours guarantees that reasonable number of independent wind profiles contribute to the average.

Figure 4.37 shows model results and corresponding lidar observations. In general there is a good agreement between model GWPED profiles and lidar measurements regarding the shape of the profiles. Also, seasonal variations of the mean GWPED value at a given altitude are reproduced by the model reasonably well. This allows one to conclude that the filtering of gravity waves due to the background wind is indeed the main process which defines the GWPED in the stratosphere.

In June (figure 4.37a) observed GWPED values are on average slightly larger compared to predicted values. This may be due to less wave filtering taking place in the

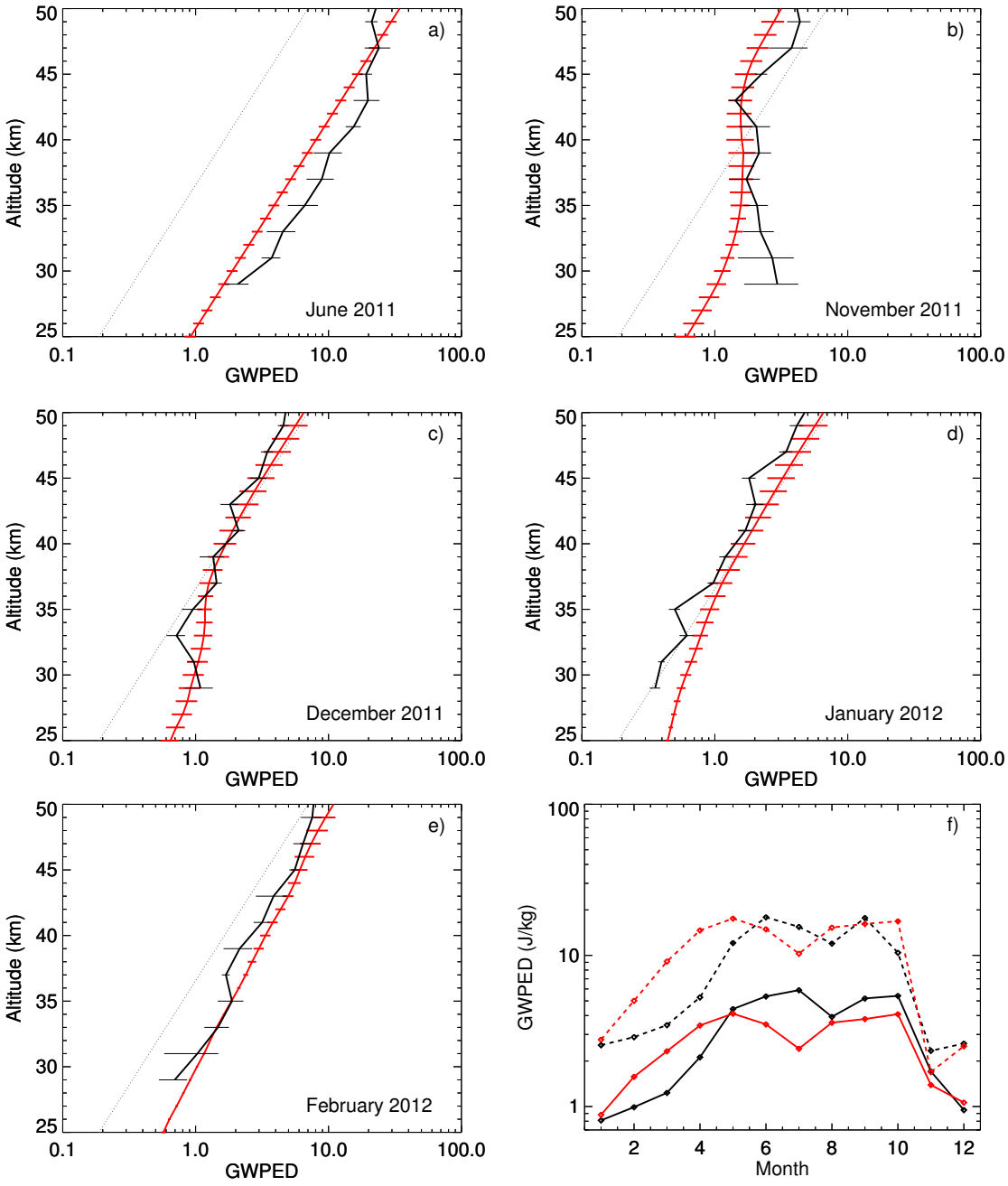


Figure 4.37.: Comparison between gravity wave potential energy densities observed by lidar (black lines) and model results (red lines). Panels a)–e) show monthly means. Horizontal lines mark standard errors, and gray lines mark conservative growth rates $\propto \exp(z/H_{\mathrm{mean}})$. f) Annual cycle in 35–42 km altitude (solid lines) and in 40–50 km altitude (dashed lines).

troposphere or changes in the strength of the gravity wave sources. Also note that there is no equivalent of the saturated gravity wave spectrum implemented in the model. Thus, gravity wave amplitudes can grow indefinitely within the model, and the observed GWPED growth rates above 43 km altitude, which appear to be limited by saturation, can not be reproduced. On the other hand, the effect of the descending zero-wind line on gravity wave propagation can be seen in lidar measurements as well as in model results. The model reproduces the altitude range where the energy density decreases or is approximately constant with altitude in good agreement with the observations (figure 4.37b and 4.37c). Note that shape of the simulated GWPED profile depends sensitively on width and shape of the assumed phase speed distribution of gravity waves. The fact that the model reproduces measured GWPED profiles well supports the choice of a Gaussian phase speed spectrum with $\sigma = 8\,\mathrm{m/s}$. The large discrepancy between observation and model towards the bottom of the November profile (figure 4.37b) appears to be striking on first sight. However, in this particular altitude range the standard errors of the observations are also largest.

Model results for January (figure 4.37d) show non-constant growth rates (wave filtering) below 35 km altitude which are not seen in observational data. This model behaviour can be explained when taking into account a small bias in the zonal wind profiles. It is evident from figure 4.35 that the zero-wind line remains shortly below the observational range throughout the summer months. Also, the vertical gradient of the zonal wind speed near the zero-wind line (not shown) is small. Thus, a small bias of few meters per second in wind speed can shift the zero-wind line in altitude, resulting in gravity wave filtering taking place near the bottom of the lidar profile. A good agreement between model and lidar measurements is again observed in February 2012 (figure 4.37e).

Figure 4.37f shows the seasonal cycle estimated from lidar measurements and model data at two different altitudes. The double humped structure is well reproduced by the model, but the characteristic dip in winter occurs one month earlier. As stated already in section 4.4.5, the dip is possibly related to changes in the zonal wind speed in the lower stratosphere, and these changes may occur in the ECMWF model earlier than in the real atmosphere. The gravity wave model is very sensitive to variations in wind speed near the launch level at 12 km if the mean wind speed is close to the maximum of the assumed phase speed distribution of gravity waves, i.e. the mean wind speed is close to zero. A small bias in mean wind speed in the lower stratosphere can therefore significantly affect GWPED estimates produced by the gravity wave model.

Another significant difference between model and observations is visible in late summer and spring (months February to May) when the GWPED estimated from the model increases faster than the observed GWPED. Two possible explanations are obvious. First, the model does not include the troposphere. Thus, if the troposphere becomes more opaque to gravity waves, the GWPED in the stratosphere decreases independently of the wind structure in the stratosphere. Second, gravity wave sources are assumed to be constant throughout the year. Any seasonal variation may therefore directly affect the energy density observed in the stratosphere.

4.4.7. Comparison with other datasets

Few extensive studies of stratospheric GWPED measurements at Antarctic latitudes have been published in literature so far. *Alexander et al.* (2011) present GWPED values based on lidar measurements with the Australian Rayleigh lidar at Davis. Published values are daily averages at 37 and 49 km altitude covering winter (March–October) in 2007

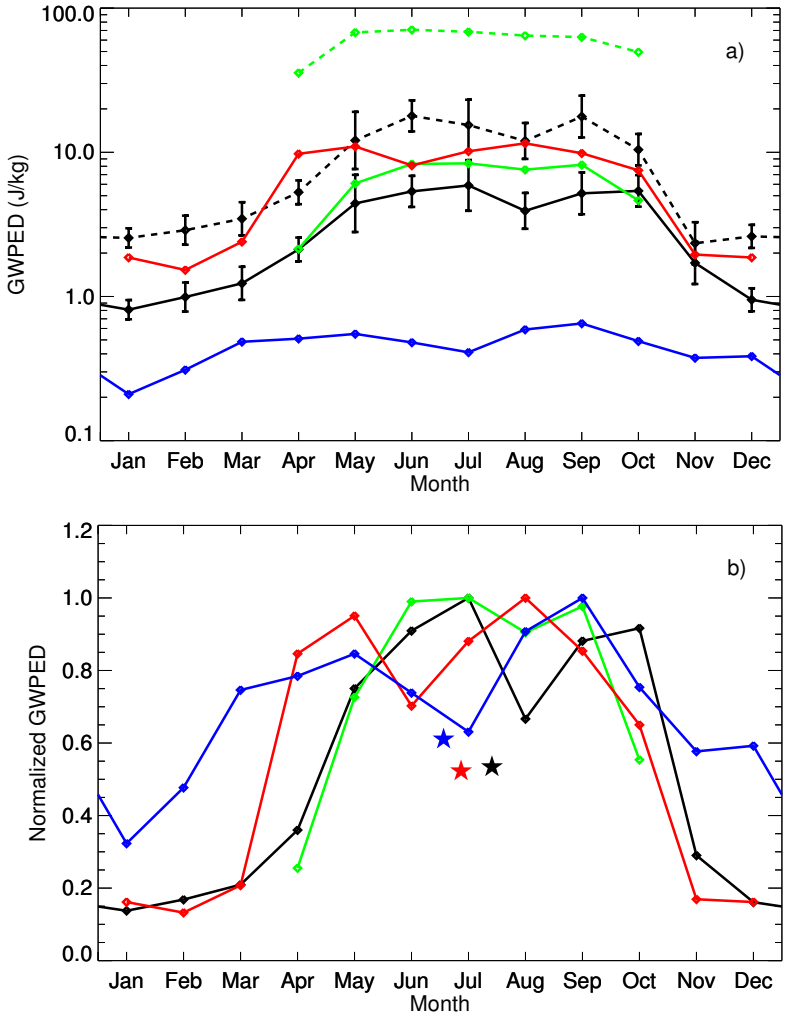


Figure 4.38.: a) Comparison between monthly mean gravity wave potential energy densities (GWPED) in 30–40 km altitude (black solid line) and 40–50 km altitude (black dashed line). Also shown are mean energy densities at 37 km (green solid line) and at 49 km altitude (green dashed line) computed from measurements published by *Alexander et al.* (2011). For comparison, two datasets measured at Rothera are added: Mean GWPED in 30–45 km altitude based on *Yamashita et al.* (2009) (red line), and mean GWPED in 15–22 km taken from (*Moffat-Griffin et al.*, 2011) (blue line). b) Data sets of a) with maxima normalized to unity. Stars mark the center of gravity of the curves with their baselines subtracted.

and 2008. *Yamashita et al.* (2009) show weekly averages for Rothera (67.5°S, 68.0°W). This dataset is also based on lidar measurements (December 2002 to March 2005), the GWPED values are, however, averaged over the altitude range 30–45 km. While *Alexander et al.* (2011) estimated GWPED values from temperature perturbations, the data presented in *Yamashita et al.* (2009) is based on measurements of relative air density. Also, the algorithms used in these two studies are sensitive to different, though overlapping, parts of the gravity wave spectrum. *Moffat-Griffin et al.* (2011) published seasonal variations of gravity wave activity over Rothera based on an 8 year series (2002–2010) of high-resolution radiosonde soundings. GWPED values presented in this study are monthly means, the altitude range is 15–22 km.

Figure 4.38a shows the combination of all four available datasets. In order to facilitate comparisons, monthly values were computed from the data presented in *Alexander et al.* (2011) and *Yamashita et al.* (2009). Mean energy densities reported in *Moffat-Griffin et al.* (2011) were divided by two in order to account for the additional contribution of the kinetic energy which is not observed by lidars. All three datasets with year-round coverage show similar seasonal variations: a minimum in summer and a broad maximum in winter. Also, the double humped structure appears to be a common feature.

Comparisons of the curves in figure 4.38a are complicated by the fact that the datasets represent GWPED measurements at different altitudes and altitude ranges. To account for GWPED values increasing exponentially with altitude, in figure 4.38b the curves are normalized such that maxima equal unity. Also, the energy scale is changed to a linear scale to highlight the magnitude of seasonal variations. Several things can be learned from the comparison: 1) Lidar-based GWPED measurements at Davis and Rothera show similar peak-to-peak variation, GWPED values in winter being approximately six times larger than in summer. On the other hand, the difference between summer and winter seen in radiosonde data (blue curve) amounts to a factor of two. *Wright and Gille* (2013) identified a gravity wave hot spot over the Antarctic Peninsula using data from the HIRDLS (High Resolution Dynamics Limb Sounder) instrument on NASA's Aura satellite. As Rothera is located near the central part of the Antarctic Peninsula, strong forcing of gravity waves is expected. Observing larger GWPED values in summer at lower altitudes is consistent with gravity waves being filtered predominantly in the lower stratosphere (section 4.4.5). 2) The winter season at Rothera appears to be shifted relative to Davis. This shift is evident from the position of the center of gravity (marked by stars in figure 4.38b) as well as the location of the winter dip which occurs one to two months earlier at Rothera. A well-founded explanation for the early onset of the summer/winter transition at Rothera has so far not been found. Since Rothera is located at the Antarctic Peninsula, the influence the Peninsula has on large-scale wind systems may play a major role.

4.4.8. Gravity waves in the MLT region and vertical coupling

Gravity wave analysis

Gravity wave analysis in the MLT region is complicated by the fact that temperature profiles are limited to the extent of the iron layer. For the gravity wave analysis described in section 4.4.2 vertical temperature profiles of at least 20 km in length are needed to extract gravity wave-induced temperature perturbations. However, the vertical extent of the iron layer shows considerable seasonal variation, and in summer the average height range where temperature can be measured is often well below the 20 km threshold (*Gardner et al.*, 2011). Gravity wave potential energy density measurements in summer are therefore limited to periods with iron density enhancements caused by sporadic

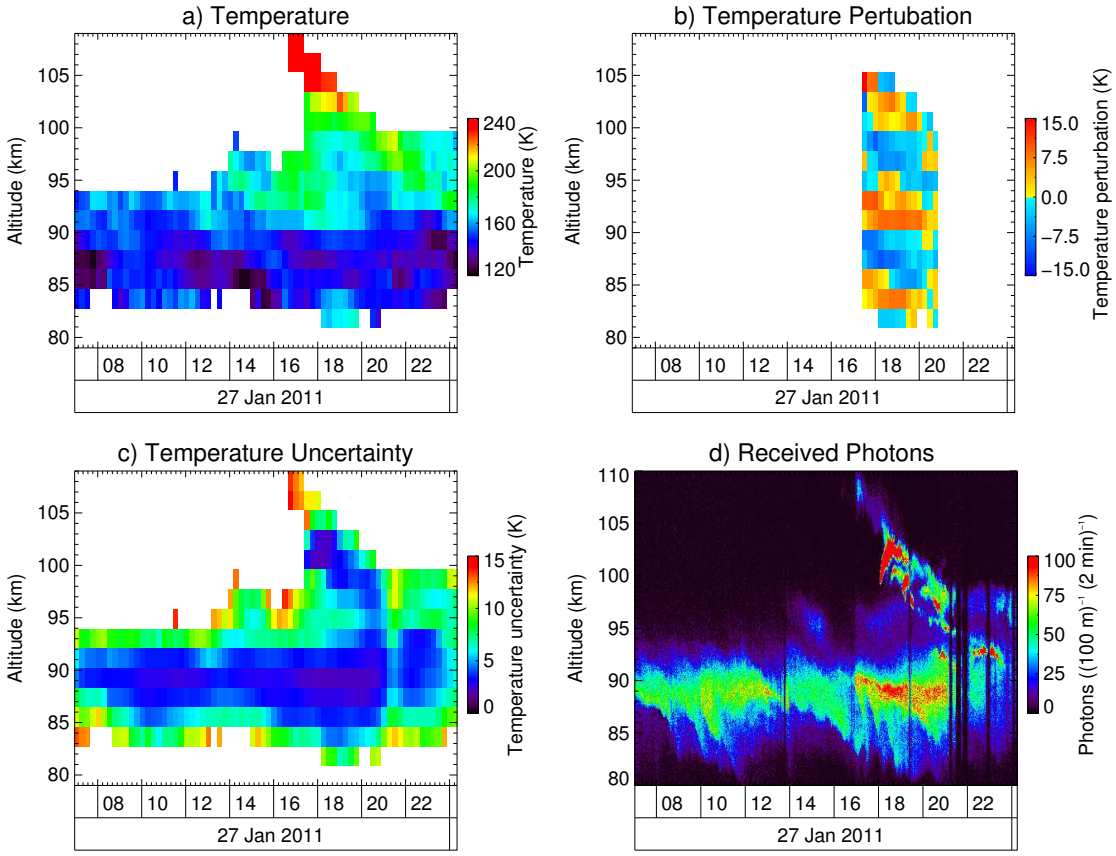


Figure 4.39.: 17 hours of temperature measurements on January 27, 2011. In summer low iron densities resulting in low photon counts (panel d) limit the vertical extent of the temperature profiles (panel a) to approximately 10 km. Longer profiles are restricted to periods with iron density enhancements caused by sporadic layers (around 1800 UT, panel d). Temperature perturbations (panel b) can be extracted if the length of the profile exceeds 20 km. Temperature uncertainties are shown in panel c).

layers. A typical example is shown in figure 4.39. For the largest part of the observation period the FWHM width of the iron layer estimated from the photon count profiles (panel d) remains below 10 km. Starting at around 1500 UT, a sporadic layer develops at 108 km altitude. This layer progresses downward over the next 8 hours and leads to a significant Fe density enhancement in the range 95 to 105 km, thus allowing temperature measurements above the main Fe layer. For a period of approximately 4 hours the length of continuous temperature profiles exceeds the 20 km threshold, and gravity wave-induced temperature perturbations can be extracted, as shown in panel b). Typical uncertainties for these temperature measurements range between 2 and 10 K (panel c).

In winter the width of the Fe layer is usually large enough to allow for temperature measurements covering more than the required 20 km in length, and temperature perturbation profiles can be extracted from lidar measurements for most of the observation period. In order to take into account the different statistical significance of long and short observations (typical observations periods range from few hours to several tens of hours, see section 4.1), long observations are split into segments of approximately 10 hours in length. Each of these segments is then separately analyzed using the gravity

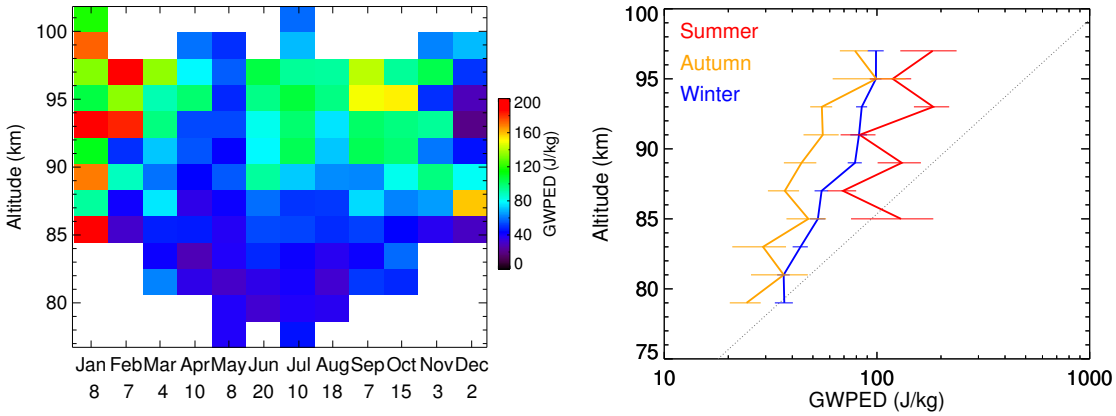


Figure 4.40.: Left: Monthly mean GWPED profiles. The number of profiles contributing to the means is written below the month. Right: Seasonal profiles: Summer (January and February), autumn (March and April), and winter (June to September). Also shown is the growth rate $\sim \exp(z/H)$ for conservative propagation (gray line). Horizontal lines mark standard errors.

wave extraction algorithm described in section 4.4.2. Ultimately, one GWPED profile is computed for each segment, and mean profiles for a given time frame are obtained by averaging all individual GWPED profiles within this time frame.

The limitation to sporadic layer events in summer has two important consequences for the gravity wave analysis. First of all, despite the vast amount of observation hours in summer (see section 4.1), only few observations capture sporadic layers large enough to allow for temperature measurements well above the main Fe layer. This severely limits the number of GWPED profiles, and averages are therefore less reliable compared to winter. Second, since sporadic layers are visible only few percent of the total observation time, the occurrence of these layers is indicative of a disturbed state of the atmosphere. Restricting gravity wave measurements to these events may lead to results which are not representative for the more common undisturbed state. For example, *Chu et al.* (2011) speculate that gravity waves play an important role in the formation and structure of sporadic Fe layers observed above the Antarctic station McMurdo. If the formation of sporadic layers is favoured by a certain type of gravity wave, then this wave would of course dominate the summer GWPED results presented in this thesis. These two issues concerning the summer data, the limited number of GWPED profiles and a possible bias, should be kept in mind when interpreting the results shown in the next section.

Results

Figure 4.40 shows monthly mean and seasonal mean GWPED profiles. In case of the monthly means the number of individual profiles contributing to the mean varies from 2 profiles in December to a maximum of 20 profiles in June. Mean profiles are truncated if, for a given altitude bin, the number of profiles with usable data falls below 80 percent of the maximum number of profiles for this month. For this reason, the length of the mean profile can be shorter than the minimum length of individual profiles (20 km), e.g. the mean profile in February covers only 14 km.

The monthly GWPED profiles in figure 4.40 show a general trend toward larger values with increasing altitude. This trend becomes more clear when looking at seasonal averages shown in the right panel of figure 4.40. In winter, the GWPED increases expo-

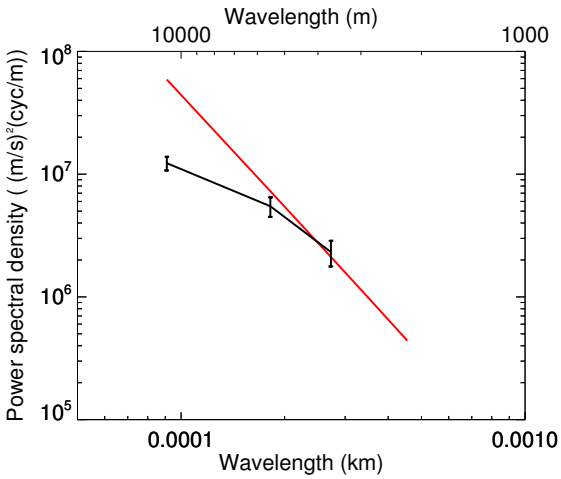


Figure 4.41.: Mean vertical wavenumber potential energy power spectral density in winter (months May-September) calculated for the 85–95 km range. The red line shows the saturation limit $N^2/10\,m^3$ (*Smith et al.*, 1987).

nentially from 36.7 J/kg at 79 km to 82.4 J/kg at 91 km altitude. This 2.3-fold increase is significantly smaller then the 7.4-fold growth expected for conservative propagation of gravity waves over the vertical distance of two atmospheric scale heights. In the conservative case the growth rate is $\sim \exp(z/H)$, where the density scale height H is approximately 6 km in the relevant altitude region (*Lübken*, 1999). The observation of GWPED values which grow slower than $\exp(z/H)$ is compatible with the assumption of breaking waves. As evident from figure 4.41, the vertical wavenumber power spectrum touches the saturation limit $N^2/10\,m^3$ (see section 4.4.4) at wavelengths shorter than 5.5 km. Note that the spectrum shown in figure 4.41 represents the atmospheric condition averaged over an altitude range of 12 km (the resolution of the temperature perturbation profiles is 2 km). Because spectral amplitudes can not grow beyond the saturation limit, however wave perturbation amplitudes increase steadily with altitude due to reasons of energy conservation, significant wave breaking is expected if spectral amplitudes approach the saturation level.

GWPED growth rates similar to the winter profile are also observed in autumn and summer (right panel in figure 4.40). Note that the summer profile is less reliable for reasons discussed in the previous section. This fact is also reflected in the magnitude of the error bars in figure 4.40. No vertical wavenumber spectrum could be calculated for the summer season due to large uncertainties in GWPED measurements. It is, however, likely that the gravity wave spectrum at short wavelengths is saturated throughout the year. This is different from stratospheric observations (see section 4.4.4) where breaking waves are observed in winter only.

The summer profile shown in 4.40 appears much more noisy compared to the other two profiles. As discussed in the previous section, the excessive variance results from sampled wave structures. As observation periods in summer are relatively short compared to winter, the variance of the summer GWPED profiles is dominated by the wave structure, while in winter the wave structure is largely smoothed out due to long averaging periods. Averaging a small number of GWPED profiles (15 profiles in summer) is not sufficient to suppress these wave structures in the seasonal mean.

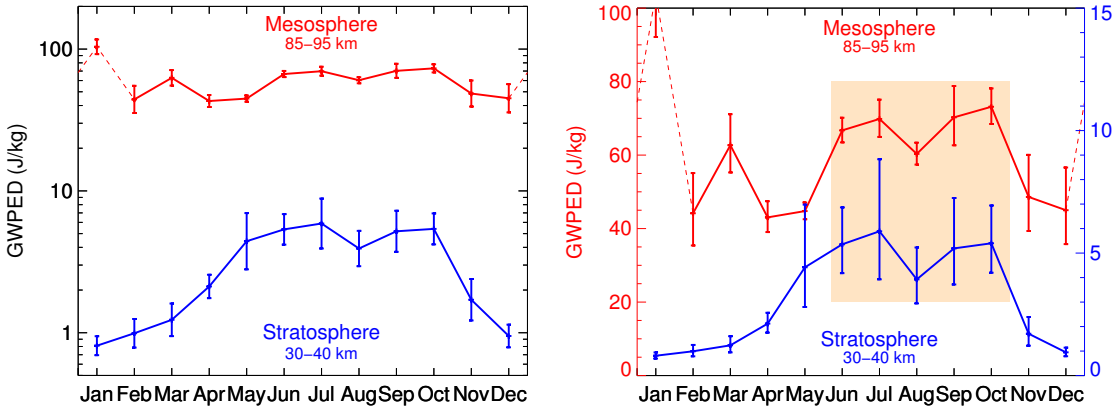


Figure 4.42.: Seasonal variation of the gravity wave potential energy density. The large increase in energy density from stratospheric heights to the mesosphere is evident from the logarithmic scale of in left panel. The linear representation shown in the right panel highlights the correlation observed in winter (shaded area).

Seasonal variation

Figure 4.42 shows the seasonal variation in observed GWPED for the stratosphere (30–40 km altitude) and the mesosphere (85–95 km altitude). The mesospheric GWPED follows approximately the same trend as the stratospheric GWPED. In particular, the double humped structure of the winter maximum with its characteristic dip in August (see section 4.4.3) is seen at both altitude ranges, followed by a minimum in December. The sharp peak in January is, however, not seen in the stratospheric data and may be questionable in light of its magnitude. It should be noted that the peak is statistically significant based on available data, but the data may be biased toward a certain condition favoring the formation of sporadic Fe layers, as discussed previously. The peak occurring in March is free from biases, but it is not statistically significant given that the monthly mean comprises 4 individual GWPED profiles only. The March peak may, however, very well be a robust feature of the annual cycle. Further evidence supporting this view yields the comparison between measured GWPED values and results of Kühlungsborn Mechanistic general Circulation Model (KMCM) shown in figure 4.43b. KMCM employs a dynamical core which solves the primitive equations by the spectral transform method (*Knöpfel and Becker*, 2011). The effect of gravity waves is included in this model by implementation of the Doppler Spread Parameterization (*Becker and McLandress*, 2009). A climatology was derived from the model by averaging five years of simulated data. As evident from figure 4.43, KMCM reproduces the March peak at the 0.001 mbar level with approximately the same magnitude as the observations.

Measured mesospheric GWPED values in winter are on average approximately twice as large as corresponding stratospheric values. Excluding the peak in January, maximum GWPED occurs in November (73.2 J/kg), and the minimum is observed in April (43.0 J/kg). Thus, monthly means in the mesosphere vary by a factor of 1.6. This factor is much smaller than the corresponding variation seen in stratospheric data, where the maximum in July is approximately 6.6 times larger than the minimum in January. Small seasonal variations at altitudes above 80 km are indeed expected if we assume a gravity wave spectrum which is predominantly saturated in this region. Even though the amplitudes of the gravity wave sources may vary, observed spectral amplitudes in

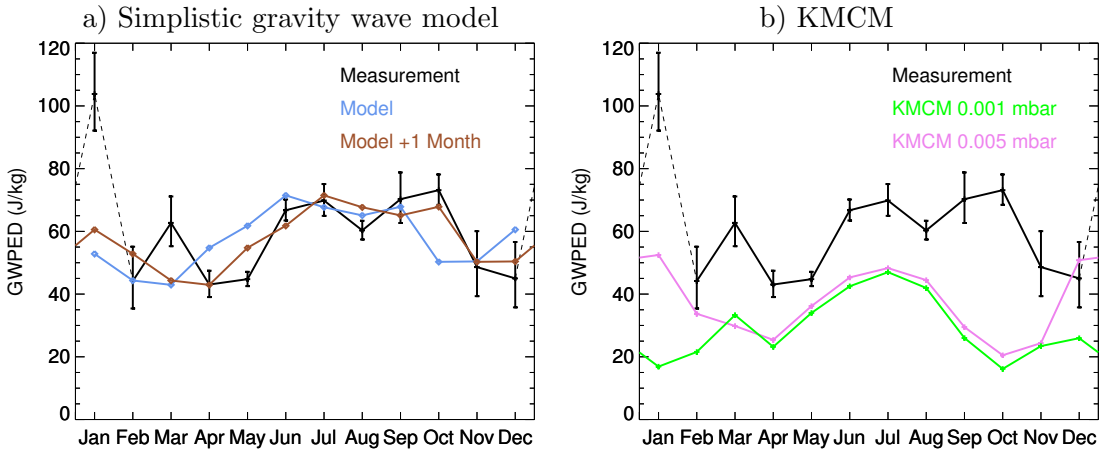


Figure 4.43.: a) Comparison between measured potential energy densities and model results. The winter peak in the model data precedes the measurements by approximately one month. b) Comparison between measurements and KMCM results for two pressure levels.

the saturated regime of the spectrum do not follow this variation because any excess energy above the saturation limit is dissipated. Following this argument, in case of a fully saturated wave spectrum the GWPED observed in the 85–95 km region should be constant throughout the year. The fact that a *small* seasonal trend is indeed observed – the peak in January is neglected here – suggest that the part of the wave spectrum the lidar is sensitive to is not completely saturated. Most likely, the non-saturated subrange results from the filtering of waves with low phase speeds at stratospheric altitudes. Further evidence in supporting this view yields the measured stratospheric horizontal wave spectrum shown in figure 4.14. In summer, the spectral amplitudes of low-frequency waves are significantly smaller than in winter. Low-frequency waves are associated with short vertical wavelengths according to the dispersion relation of gravity waves (equation 4.9). As the speed of the background mean flow approaches the phase speed of the wave, the vertical wavelength gets compressed and the wave becomes dynamically unstable. Waves with initially short wavelengths are thus more likely to encounter critical levels and being subsequently "removed" from the spectrum. Because most gravity wave sources are below the tropopause, amplitude minima produced by filtering of waves in the stratosphere are in general preserved in the middle atmosphere region. On the other hand, amplitude maxima may be reduced due to saturation of the gravity wave spectrum. These two factors taken together may explain why the seasonal variation of the GWPED observed in the mesosphere is similar to stratospheric observations albeit the peak-to-peak variation of monthly averages is much smaller. Hence, vertical propagation of gravity waves can be regarded as medium which couples the mesosphere to the stratosphere. For example, the wind anomaly in the lower stratosphere which causes the winter dip in the GWPED influences the dissipation of wave energy in the mesosphere, and thus also the momentum imparted on the background mean flow.

The semi-annual oscillation of the mesospheric gravity wave energy was previously reported by *Hoffmann et al.* (2010). *Hoffmann et al.* investigated the seasonal variation of the wave activity based on wind measurements with the meteor and MF radars at Andenes (69°N, 16°E) and Juliusruh (55°N, 13°E). According to their study, at both sites the gravity wave mean kinetic energy shows a broad maximum in winter and a smaller secondary maximum in summer. The seasonal variation of the GWPED

observed at Davis (this work) is consistent with the wave study by *Hoffmann et al.*.

To investigate the nature of the semi-annual oscillation in the Davis GWPED measurements, the simplistic gravity wave model which was used to simulate the stratospheric GWPED variation (section 4.4.6) was extended to the mesosphere. Two modifications were necessary: First of all, ECMWF model data is available up to the 0.01 mbar level ($\sim$ 75 km) only. In order to compare simulated GWPED variations with mesospheric lidar observations, zonal wind profiles ranging as high as 85 km are needed. The gap (70–85 km) was closed by extending the ECMWF profiles with MF radar measurements at Davis station[3]. Second, the effect of a saturating gravity wave spectrum must be included in the model as saturation becomes important in the mesosphere. In the model gravity waves are described in terms of phase speed and amplitude. A natural choice is therefore to limit the amplitude to a maximum value. It should be noted that the model does not discern between waves with different frequencies and vertical wavelengths, and saturation of the gravity wave spectrum can thus not be well represented. However, saturation reduces the GWPED as waves propagate upwards. Thus, saturation can be approximated by limiting the maximum GWPED per phase speed bin. The maximum value of 1.65 J/kg/bin was determined from the best fit to the measured GWPED.

Figure 4.43 shows the comparison between model results and mesospheric GWPED measurements. With exception of the peak in January the amplitude of the seasonal variation is surprisingly well reproduced. However, the phase of the semi-annual oscillation precedes the measurements by approximately one month. This suggests that wind reversal associated with the summer/winter and winter/summer transition is not captured at the right time. Shifting the model derived results by one month aligns both the broad winter maximum and the smaller summer maximum with measured data. The location of the minima in the model data is consistent with filtering of gravity waves in the stratosphere and lower mesosphere. Around the equinoxes the zonal wind speed is close to zero over a relatively wide range of heights, resulting in filtering of slow waves with both positive and negative phase speeds. The reduced GWPED in summer in comparison to the broad maximum in winter can be explained by filtering of most westward and slowly eastward propagating waves. In winter, the strong stratospheric jet makes sure that only the eastward propagating waves are filtered, while any (slow and fast) westward propagating wave can penetrate into the MLT region. In contrast to the study by *Hoffmann et al.* the winter maximum at Davis is larger in both amplitude and duration. As noted by *Hoffmann et al.*, the more dominant annual component of the gravity wave activity in the southern hemisphere is likely related to the much stronger polar night jet. A large annual component was also reported by *Hibbins et al.* (2007). *Hibbins et al.* studied the gravity wave activity at the Antarctic station Rothera (67°S, 68°W) based on MF radar wind measurements. The small maximum in March is not reproduced by the simplistic gravity wave model, but it does occur in the zonal mean climatology derived from KMCM data. This suggests that either the wind structure which is responsible for this maximum is missing in ECMWF data, or the maximum is produced by gravity waves that propagate horizontally over large distances. Because the simplistic model uses the local wind field only, any effects outside this domain can not be reproduced. The comparison with the zonal wind structure of KMCM reveals that the summer/winter transition happens roughly at the same time in the stratosphere as well as in the mesosphere, and there is a short period (2–3 weeks) with approximately zero zonal wind speed in the MLT region. During this time, westward propagating waves with low phase speeds can reach the mesopause region, while before and after the path for slow waves is blocked by the wind reversal in the mesosphere. This situation is very

[3]MF radar data courtesy of Damian Murphy, Australian Antarctic Division

much different from the winter/summer transition. The wind reversal associated with the breakdown of the polar vortex starts at high altitudes and progresses downward. This means that the path for waves with slow phase speeds is blocked during the whole transition time, and for this reason no additional GWPED maximum is observed.

4.4.9. Heat transport by gravity waves

The vertical heat transport caused by gravity waves is characterized by the cross-correlation between the wave-induced temperature perturbations, T', and the associated fluctuations observed in the vertical wind, w'. Although both temperature- and vertical wind fluctuations can be large in the MLT region, measurement of the heat flux associated with gravity waves is challenging because the covariance between temperature and vertical wind is typically small. In case of non-dissipating waves the polarization between temperature- and vertical wind fluctuations is 90 degrees, resulting in zero vertical heat flux. Nonzero heat fluxes can be obtained only if the phase relationships between T' and w' is altered. This is typically the case if waves dissipate. Thus, measurement of vertical heat transport can provide a measure of gravity wave dissipation (e.g. *Gardner et al.*, 2002).

The derivation of heat fluxes from high resolution resonance lidar measurements has already been demonstrated by *Gardner and Liu* (2007). *Gardner and Liu* used the 3.5-m-diameter telescope of the Starfire Optical Range (New Mexico) in connection with the transmitter of the University of Illinois Na wind/temperature lidar to obtain temperature- and wind profiles with 1.5 min temporal and 500 m vertical resolution. From these data they were able to extract wave-induced perturbations with observed periods longer than 3 min and vertical wavelengths in excess of 2 km. The root mean square uncertainties of the temperature and vertical wind measurements between 85 and 100 km altitude are about 1.0 m/s and 1.3 K, respectively. In contrast, the IAP Fe lidar is capable of measuring vertical wind and temperature with 60 min temporal and 2 km vertical resolution. Average uncertainties are 1.5 m/s and 4.9 K for vertical wind and temperature measurements. In comparing the performance of the two lidar instruments it is clear that the Fe lidar is less suited for heat flux studies. While the variance of the temperature field is dominated by waves with long periods (several hours), the largest contribution to the momentum and heat fluxes is expected from high-frequency waves with periods < 1 h. This important part of the wave spectrum can not be resolved by the Fe lidar. On the other hand, with more than 2600 observation hours the Fe lidar dataset is the largest dataset on record, exceeding the observations presented in *Gardner and Liu* (2007) (369.7 h) by a factor of 7. Because the noise in temperature and wind measurements resulting from photon noise is uncorrelated, the statistical significance of heat flux estimates is mainly determined by the length of the averaging window. The larger dataset obtained with the Fe lidar may thus allow the estimation of the heat flux in the low-frequency part of the wave spectrum in spite of the much smaller amplitudes. Although the magnitude of the heat flux estimates is expected to be different (due to the different spectral sensitivity), similar large-scale seasonal variations should be detectable in both lidar datasets, provided they do exist at both observation sites.

The seasonal variations and vertical structure in heatflux data derived from Fe lidar measurements agree qualitatively with the New Mexico data presented in *Gardner and Liu* (2007). However, as the Fe lidar is not sensitive enough to resolve the important high-frequency waves, the heatflux data are considered not reliable. Hence, no attempts have been made to draw conclusions based on these data. A detailed discussion of the data analysis as well as results can be found in appendix A.

5. Summary and Outlook

This work demonstrates that the Rayleigh measurements acquired by the IAP Fe lidar can be successfully applied for gravity wave analysis in the stratosphere. An overview of the technical aspects of the Fe lidar is given in chapter 2, followed by a step-by-step discussion of the data processing in chapter 3. Unlike conventional Rayleigh lidars, the Fe lidar employs a frequency scanning technique which results in spectral distortions in the Rayleigh signal. New algorithms were developed to correct these distortions, and their effect on the temperature retrieval was verified by comparing temperature profiles to reference profiles. Furthermore, a new technique to estimate the dead time of the detector based on actual lidar measurements was proposed and successfully implemented. If not corrected, dead time effects resulting from high photon counting rates are a major source of instrumental error. This applies in particular to the Fe lidar because a single detector is used for both mesospheric (low rate) and stratospheric (high rate) measurements. The effectiveness of the various corrections which are applied to the measured lidar return signal was demonstrated by comparing temperature profiles retrieved from lidar measurements to temperature profiles that have been independently acquired by radiosondes. Also, the impact of aerosols on stratospheric lidar measurements was briefly discussed, and seasonal changes of the stratospheric aerosol loading were characterized in terms of variation of the color ratio of the lidar return signal.

The stratospheric temperature dataset in conjunction with the mesospheric data was then used in chapter 4 to study variations in temperature on different temporal scales. This includes the seasonal variation of mean temperatures, spectra of temperature perturbation amplitudes in the range 2–12 h, and finally a more detailed study of gravity wave signatures. The seasonal variation of the stratospheric temperature at Davis was found to be similar to the northern hemisphere, however the stratopause in winter is on average approximately 10 K warmer. No stratospheric warming was detected at Davis. The comparison with ECMWF model temperature data revealed that most differences between model data and measurements in the stratosphere can be attributed to difficulties in reproducing the exact height of the stratopause. A small model bias towards low winter temperatures was found.

The spectral analysis of temperature perturbations in the 2–12 h band revealed a strong seasonal variation of the frequency dependence of spectral amplitudes. A $k^{-5/3}$ exponent was found to prevail at all altitudes in winter, while a clear altitude dependence of the exponent was observed in summer. Starting at 30 km altitude with an exponent of $-5/3$, the slope of the spectrum flattens continuously with increasing altitude, reaching -1 at 45 km. A similar flattening of the spectrum was found to occur at ALOMAR, and observed variations in the spectrum could be linked to selective filtering of waves through critical layers. Furthermore, the analysis of variances estimated from different spectral bands showed a clear annual cycle with a significant shift towards waves with short periods in summer.

The wave analysis was more refined in section 4.4 by adapting the wave extraction algorithm to retain gravity waves only. Gravity waves were then characterized in terms of the gravity wave potential energy density (GWPED). It is shown that conservative wave propagation at Davis occurs in winter as well as in summer throughout most of

the upper stratosphere, and significant wave breaking is only observed in winter above 43 km altitude. Latter results are consistent with previous studies of the Antarctic winter stratosphere. The fact that the Fe lidar GWPED winter profile does actually match the GWPED profile of a previous lidar based study suggest very stable winter conditions at Davis with very little year-to-year variation.

While previous lidar based studies of gravity wave activity are by majority limited to the winter season, the capabilities of the IAP Fe lidar in conjunction with the large number of observation hours allowed the detailed investigation of seasonal variations over the period of a full annual cycle for the first time. In the stratosphere, observational results show a strong GWPED maximum in winter and a minimum in summer, similar to results of previous radiosonde and lidar based studies at Rothera. The large seasonal variation is presumably linked to selective filtering of gravity waves in the stratosphere. In particular, a decrease in gravity wave energy was observed in vertical GWPED profiles at altitudes where the zonal wind speed is close to zero. A very simplistic model was developed to study selective filtering of waves by the background wind field. The observed seasonal variation in GWPED is surprisingly well reproduced by the model. This suggests that most of the observed variation is indeed caused by filtering of waves rather than through modulation of gravity wave sources.

In the mesosphere, observations show a semi-annual oscillation with a broad maximum in winter and a smaller secondary maximum in summer. A third maximum is observed in March. While the semi-annual oscillation is consistent with selective filtering of eastward and westward propagating waves, the third maximum most likely results from the zonal wind speed being close to zero throughout most of the mesosphere around autumn equinox. This unique wind structure permits slow westward propagating waves to penetrate into the MLT region.

Conclusions and outlook

The results presented in this thesis clearly demonstrate that the IAP Fe lidar is a very valuable tool for studies of atmosphere dynamics. However, potential availability of such a tool and its actual adoption appear to be two separate things. In the past discussions at IAP focused very much on what could *theoretically* be done with this lidar in optimum conditions rather than what is *realistically* achieved. A very capable and advanced lidar system is of little scientific value if the new capabilities can not be used for science studies due to lack of data processors and analysis algorithms. In order to maximize the science output of the Fe lidar it is important to focus on both instrument development and data processing. The latter has been neglected in recent years, and it would be desirable to increase efforts to create more extensive data products reflecting the capabilities of the instrument.

Regarding future development of the Fe lidar, increasing the time resolution would certainly open new possibilities, e.g. the determination of fluxes of heat and iron. Also, coordinated campaigns with horizontal wind resolving lidars (e.g. lidars at ALOMAR, the new Fe lidar under development at the Cooperative Institute for Research in Environmental Science) could provide a substantial contribution to our understanding of wave-wave interactions and the generation of secondary gravity waves. Finally, a detailed comparison between observations and model results could lead to improvements of gravity wave parameterization schemes in global circulation models.

A. Heat transport by gravity waves

A.1. Data analysis

In order to achieve a maximum of comparability, the Fe lidar dataset was analyzed in a similar way as described in *Gardner and Liu* (2007). However, some issues caused by technical differences between the two lidar systems needed to be addressed. While the Starfire Optical Range lidar could operate only at night, limiting continuous observations to approximately 10 h, no such limitation does exist for the Fe lidar. In fact, many Fe lidar observations span multiple days (see section 4.1), and the "daily" average may thus not be identical with the average of one observation as this was the case with the Starfire Optical Range lidar. Therefore long Fe lidar observations were split into segments of 9 to 18 h in length, each segment mimicking one "day" in the meaning of the Starfire Optical Range lidar. All segments were then analyzed separately.

First, data points with measurement uncertainties exceeding 10 m/s or 10 K were discarded. Temperature and vertical wind perturbations for each segment were computed by subtracting the linear trend in time at each altitude from temperature and vertical wind measurements. Unlike the analysis method presented in *Gardner and Liu* (2007), no values were discarded because of perturbations exceeding three standard deviations from the segment mean. A careful analysis of the variances shows that large perturbations are to be expected, and, more important, they do not correlate with large measurement uncertainties. Therefore, it can not be justified to classify these large perturbations as outliers and subsequently discard them. In fact, discarding these values would create "holes" near peaks of large-amplitude waves.

The lower and upper ends of measured temperature and vertical wind profiles may change over time because of fluctuations in the iron density. If the iron density becomes too low, measurement uncertainties increase beyond the threshold of 10 m/s or 10 K, and measured profiles are truncated accordingly. To avoid unreliable perturbation estimates resulting from the mean trend in time being determined from few measurements, altitude regions with less than 50% of usable data were excluded from the subsequent analysis. Finally, in order to remove biases resulting from waves with vertical wavelengths longer than approximately 30 km, the vertical mean was subtracted from each perturbation profile. The resulting perturbation profiles T' and w' were then multiplied with each other to form instantaneous heat flux profiles $T'w'$. These profiles were subsequently averaged in time to obtain the heat flux profile $\overline{T'w'}$. This data processing is carried out separately for each segment. As an example, major steps of the data analysis for an 18 h segment on 29/30 July 2011 are shown in figure A.1. The thermal structure of the atmosphere between 80 and 100 km altitude (figure A.1a) features large wave-induced perturbations, which become more pronounced after the linear trend in time and vertical means are removed (figure A.1b). Figure A.1c shows the result of the same data processing steps applied to measurements of vertical wind. While the temperature perturbations are dominated by low-frequency wave structures, fluctuations in the vertical wind appear at all resolvable frequencies with approximately equal amplitude. This qualitative classification agrees with typical spectra of horizontal temperature fluctuations with slopes in the order of $-5/3$ (figure 4.22) and vertical wind spectra which

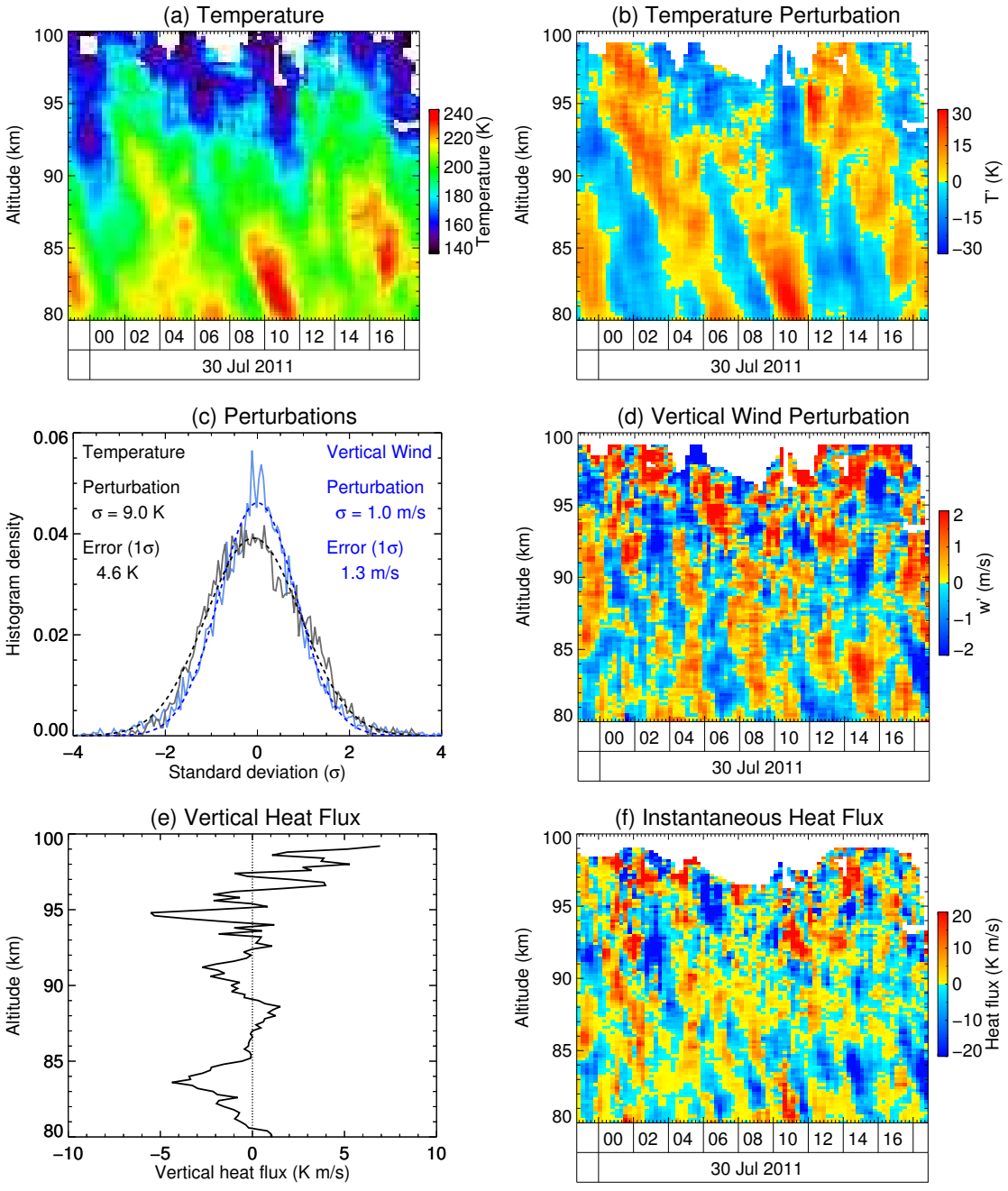


Figure A.1.: Illustration of data processing for temperature, vertical wind and heat flux on July 29/30, 2011: (a) measured temperature; (b) temperature perturbation after removing linear trend in time and vertical mean; (d) vertical wind after removing linear trend in time and vertical mean; (f) instantaneous heat flux $w'T'$; (e) heat flux profile $\overline{w'T'}$; (c) perturbations and error estimates for wind and temperature measurements. See text for details.

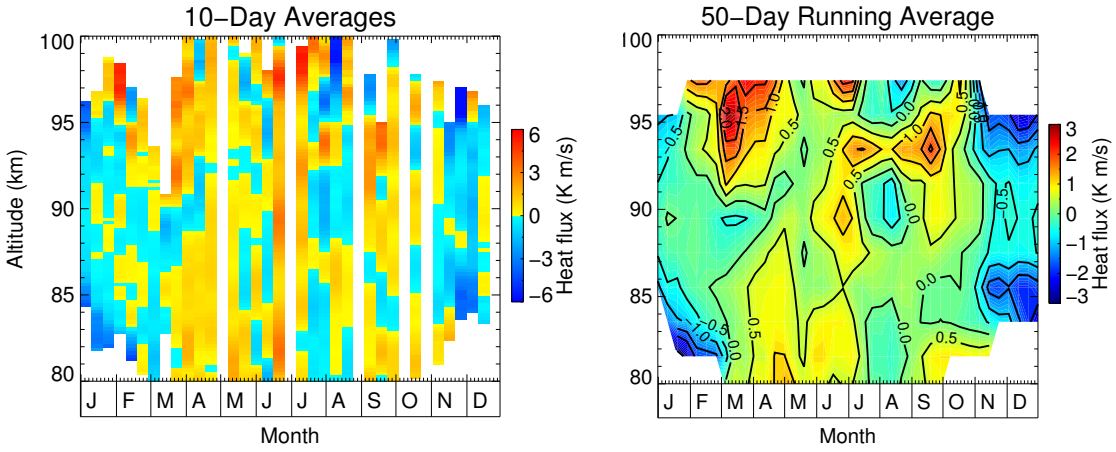


Figure A.2.: Seasonal variation of the measured vertical heat flux: 10-day averages (left) and 50-day running average (right). Note the different color coding which highlights the transition from positive to negative heat flux (left).

vary approximately as ω^0. Histograms of the observed temperature and vertical wind perturbations are shown in figure A.1c. For easy comparison, widths of both distributions are scaled to one standard deviation which amounts to 9.0 K and 1.0 m/s for temperature and vertical wind perturbations, respectively. As evident from the figure, the distributions form nearly perfect Gaussians (dashed lines). Typical uncertainties of the temperature and vertical wind measurements were estimated from the cumulative error distribution. Uncertainties of 68.5% of all measurements (1σ) are < 4.6 K or < 1.3 m/s. Consequently, for temperature measurements the typical perturbation amplitude is approximately twice as large as the typical measurement uncertainty. On the other hand, the typical uncertainty of vertical wind measurements is apparently larger than the corresponding perturbation amplitude. This is, however, not possible because if all variance is caused by instrumental noise (worst case), the width of the error distribution can be at most equal to the width of the observed perturbation distribution. Consequently, the error of vertical wind measurements appears to be overestimated. Based on the data alone it is thus not possible to tell how much of the variance is of geophysical origin as opposed to instrumental noise. The product of the temperature and vertical wind perturbation profiles, $T'w'$, is shown in figure A.1f. Between 80 and 85 km altitude, a stripe pattern with mainly negative values is visible in the second half of the observation segment. The location of this pattern agrees with a region of decaying temperature amplitudes (figure A.1b), which suggests dissipation of gravity waves. Large negative values at 84 km altitude are also observed in the heat flux $\overline{T'w'}$ (figure A.1e).

A.2. Results

10-day mean profiles and 50-day running averages of the vertical fluxes of heat are displayed in figure A.2. Negative values are predominantly observed during summer months, while positive heat fluxes mainly occur around spring and autumn equinoxes as well as midwinter. Except for the very first profile in January, all profiles show negative and positive values, with fractions varying from profile to profile. However, altitude regions with negative (positive) values may persist over multiple adjacent profiles, e.g.

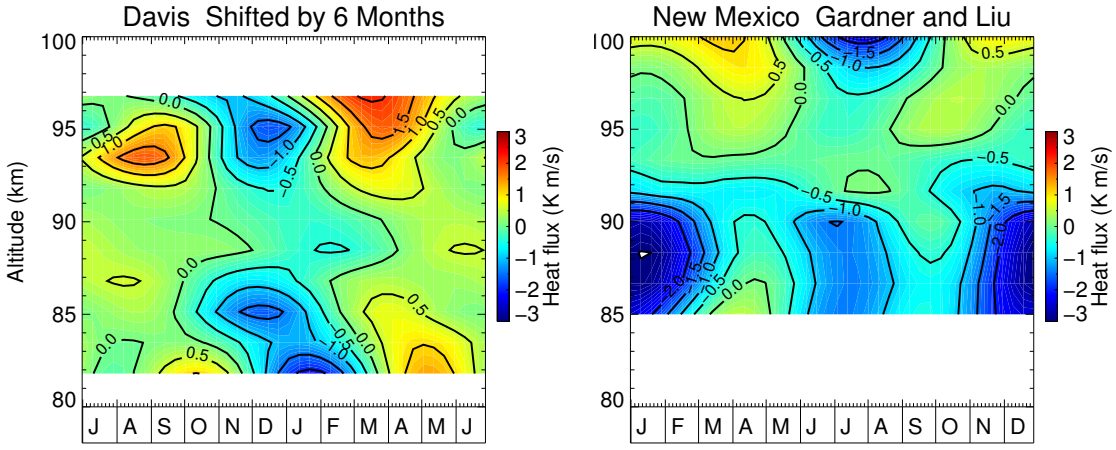


Figure A.3.: Harmonic model of the seasonal variation of the vertical heat flux measured at Davis (left) and Starfire Optical Range in New Mexico (*Gardner and Liu*, 2007) (right). The Davis dataset is shifted by six months in order to account for seasonal differences caused by Davis station being located in the southern hemisphere.

in July/August at 90 km (87 km). This suggests that regions of heights where wave dissipation occurs predominantly remain largely constant over a certain period of time, or at least progress slowly with time.

To illustrate seasonal variations, a harmonic model comprising the first two harmonics (6-month and 12-month oscillations) as well as the annual mean was fitted to the data. The result is plotted in contour format in the left panel of figure A.3. Note that the time axis is shifted in order to place the summer season in the center of the plot. The six month shift adjusts the phase of the annual cycle to match the cycle of the northern hemisphere, thus allowing easy comparison of the Davis observations with the Starfire Optical Range dataset (New Mexico) shown in the right panel of figure A.3. The fit parameters of the harmonic model which was used to generate the New Mexico plot were taken from Table 2 published in (*Gardner and Liu*, 2007). In the upper half of the plots (> 90 km for Davis, > 93 km for New Mexico) both datasets are qualitatively consistent. Maxima are observed around spring and autumn equinoxes, and a large minimum occurs in summer. In general, the Davis dataset appears to be biased toward positive values. Similar structures are also observed at lower altitudes. The location and magnitude of the summer minimum is in good agreement, and the maxima around the equinoxes occur approximately at the same time in both datasets. However, the New Mexico data show a very large minimum in winter below 90 km altitude which is not observed at Davis. This minimum overshadows the maximum occurring during the autumn equinox, resulting in negative heat fluxes throughout most of the year. For the Davis data the situation appears to be reversed. A large maximum (positive heat flux) occurs in autumn, and the winter minimun turns out much smaller with values in July/August barely becoming negative. In comparing the two plots in figure A.3, a vertical shift of approximately 4 km is evident not only from the different altitude ranges, but also from the vertical location of the maxima and minima. Because the Starfire Optical Range in New Mexico (35°N, 106.5°W) is closer to the equator than Davis Station (69°S, 78°E), pressure levels are shifted to larger geometric heights. This affects the height of the metal layers as well as the location of features which influence gravity wave breaking, e.g. the mesopause.

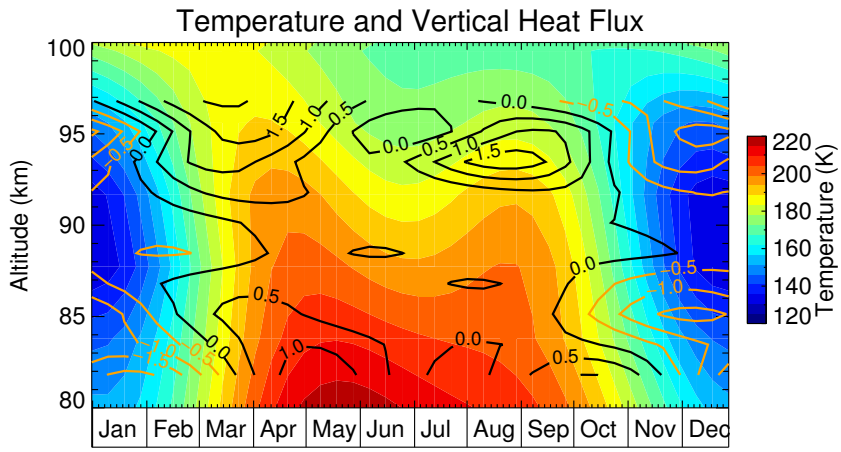


Figure A.4.: Harmonic model of the seasonal variation of the measured temperature (filled contours) and vertical heatflux (open contours).

Figure A.4 shows the seasonal variation of the measured temperature with the heat flux overlaid as open contours. Maximum downward heat fluxes are observed in summer below and above the temperature minimum where temperature gradients are largest. Because downward heat fluxes are produced by dissipating gravity waves, the coincidence of negative vertical temperature gradients below the temperature mininum (temperature decreases with altitude) and maximum downward heat fluxes suggests that convective instability may be the main driver for wave dissipation in the summer MLT region. As shown by *Gardner and Yang* (1998), large downward heat fluxes can result in substantial dynamical cooling of the atmosphere. The magnitude depends, however, on the observed heat flux which in turn depends strongly on the sampled part of the gravity wave spectrum. Because the IAP Fe lidar is not sensitive to the important high-frequency part of the spectrum which is thought to carry most of the fluxes (heat and momentum), the annual mean of the heat flux (figure A.5) is much smaller than values reported by *Gardner and Liu* (2007). It is therefore not surprising that cooling rates in the order of few K/d estimated from the Davis dataset are also small in comparison to results obtained from measurements with the Starfire Optical Range lidar. Based on observations with latter instrument *Gardner and Yang* (1998) report mean cooling rates as large as 60 K/d. This example illustrates that improper selection of the sampling window (e.g. not sampling the high-frequency part of the gravity wave spectrum) can introduce large biases in quantities such as heat fluxes and dynamical cooling rates. The IAP Fe lidar is, unfortunately, not powerful enough to resolve high-frequency waves. Heat fluxes and dynamical cooling rates derived from the Davis data are therefore not reliable although seasonal variations and vertical structure agree qualitatively with the New Mexico data (see figures A.3 and A.5).

The sampling bias is also evident from the heat flux profiles shown in figure A.6. Both profiles were derived from the same Fe lidar measurements taken between February 20 and February 23, 2011. The blue profile represents the mean heat flux computed from standard data products (temperature and vertical wind profiles with 60 min × 2 km resolution), while the red profile is based on data with increased temporal and vertical resolution. Increasing the temporal resolution from 60 min to 20 min results in on average larger downward heat fluxes over most of the altitude range. Smaller heat fluxes at 86 km and around 90 km are most likely caused by excessive noise in temperature and wind variances. As the resolution is increased, the signal-to-noise ratio decreases. This

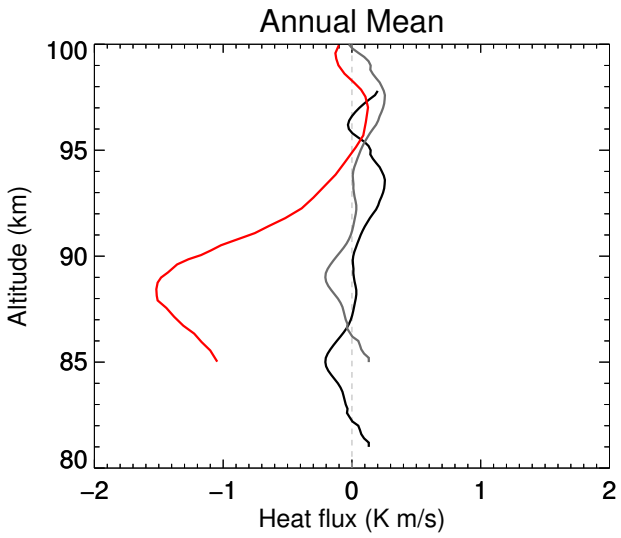


Figure A.5.: Annual mean heat flux profiles measured with the Starfire Optical Range lidar (red line) and the IAP Fe lidar (black line). In case of the Fe lidar profile, all values are close to zero. The gray line shows the Davis profile shifted by 4 km.

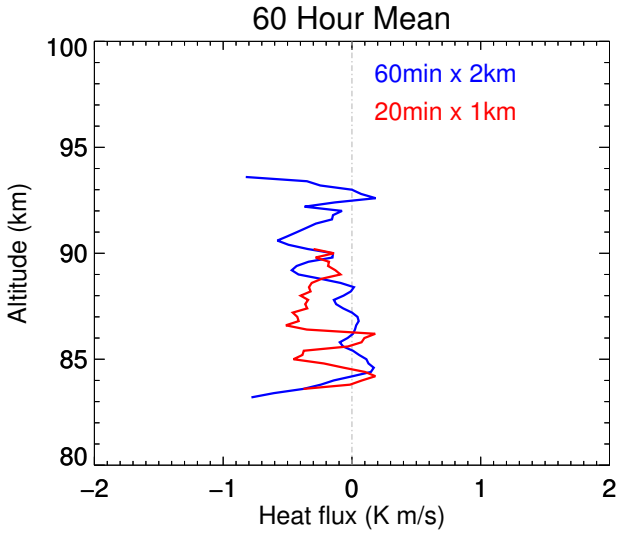


Figure A.6.: Mean vertical heat flux profiles derived from measurements taken between February 20 and February 23, 2011. Temperature and vertical wind data was processed in two resolutions: 60 min temporal and 2 km vertical (blue line) and 20 min temporal and 1 km vertical (red line). The mean profiles were smoothed vertically using a 3 km FWHM Hann window.

results not only in increased photon noise (which does not affect mean covariances), but also in buildup of systematic errors. Latter errors increase dramatically as estimated measurement uncertainties approach 10 m/s or 10 K. Consequently, computed heat fluxes become unreliable if the signal-to-noise ratio is too low. The temporal resolution of three samples per hour is therefore the final performance limit of the Fe lidar, and this limit is reached only under optimal conditions. In order to allow the determination of the heat flux in the high-frequency part of the gravity wave spectrum, the Fe lidar would have to be approximately ten times more powerful.

B. Filtering of gravity waves by tides

Correlations of gravity waves and tides have been previously reported in literature (e.g. *Thayaparan et al.*, 1995; *Manson et al.*, 1998; *Preusse et al.*, 2001). Based on numerical studies *Mayr et al.* (1998) conclude that due to absorption of gravity wave momentum in the mesopause region, the amplitude of the diurnal tide is amplified and its vertical wavelength is reduced. With the large dataset produced by the Fe lidar which covers most of the middle atmosphere, it should be possible to confirm this effect using observational data. However, obtaining meaningful results from correlations of tides and gravity waves, both derived from the Fe lidar temperature dataset, turned out to be very difficult. Main reasons are: 1) Properties of tides (vertical wavelength, phase) can not be determined to the required precision from sparse lidar observations. Depending on the length of the averaging window, tidal parameters show a high degree of variability. The question of whether this detected variability is inherent to tides or is caused by gravity waves leaking into the tide analysis, could not be satisfactorily answered. This makes the interpretation of correlations impossible. 2) The extent of the iron layer and therefore also the range of the iron Doppler temperature profiles, is highly variable. In order to obtain a sufficiently large vertical range for gravity wave and tide analysis, non-overlapping temperature profiles must be included in the analysis. The gaps in the resulting dataset are not randomly distributed, and also the precision of the temperature measurements varies strongly in the vicinity of the gaps. Therefore, a non-Gaussian statistical analysis of the temperature measurements is required to study local time effects on time scales of few days.

For reasons stated above, the correlation analysis of tides and gravity waves is not included in this thesis. The analysis of tidal signatures using long averaging windows is discussed in the following section and results are presented in section B.3 (stratosphere) and section B.4 (MLT region).

B.1. Tides

Atmospheric tides are global oscillations in the atmosphere whose periods are integral fractions of lunar or solar days. The term tide is usually limited to the first three harmonic components: Oscillations with a period of one day are called diurnal tides, with a period of half a day semidiurnal tides, and with a period of a third of a day terdiurnal tides. The detection of tides requires tracing the periodic signal for at least two cycles. In case of the diurnal lunar tide this amounts to 59 Earth days, a time frame comparable to the scale of seasonal variations in the atmosphere. The tidal signal is therefore hard to detect in sparse data such as lidar observations. Thus, the study of lunar tides is usually limited to analyzing model outputs rather than actual measurements (e.g. *Pedatella et al.*, 2012).

Solar tides, on the other hand, are detectable even in sparse lidar data provided there is adequate sampling of all local times. Latter requirement implies that the lidar must be capable of rejecting the solar background in order to be operational over a full diurnal cycle. Narrowband optical filters with sufficiently high transmission for use in mesospheric lidars have been available since the 1990's (*Chen et al.*, 1993; *McKay*,

1999). However, it took another decade to collect enough lidar data to permit systematic studies of thermal tides in the mesopause region. *She et al.* (2002a) published first results based on 18 sets of 24-hour continuous lidar observations. Since then a number of studies based on increasing datasets have been reported in literature (e.g *She et al.*, 2004; *Fricke-Begemann and Höffner*, 2005; *Yuan et al.*, 2008; *Lübken et al.*, 2011). With more than 2600 observation hours, the lidar dataset presented in this thesis offers a so far unprecedented opportunity to study tides in the stratosphere and the MLT region. As with any point observation, only migrating tides can be detected in the Fe lidar dataset.

B.2. Data reduction and analysis

The key challenge in isolating tidal signatures in lidar data is to resolve the ambiguity between atmospheric tides and gravity waves. Since gravity waves can have the same periods as tides, it is not sufficient to perform a harmonic analysis of measured time series and identify spectral amplitudes as tidal amplitudes. In doing so, tidal amplitudes are heavily overestimated because gravity wave-induced perturbations are most often much larger than the amplitude of tidal signals. A common method used to mitigate this ambiguity problem is coherent integration: all measurements which fall within the same hour of day are averaged. Because phases of gravity waves are random, averaging a large number of observations suppresses gravity wave-induced signals while coherent tidal signals are retained. The coherent signal may then be decomposed into Fourier components, where each component can be associated with a migrating tidal mode. The complex Fourier spectrum contains information about the amplitude as well as the phase of each mode.

In a first processing step the annual temperature oscillation (see section 4.2.2) was subtracted from the temperature dataset. Next, all individual data points with estimated uncertainties larger than 7 K in the MLT region and 5 K in the stratosphere were discarded. The reduced dataset was then split into monthly or seasonal intervals. Coherent integration was performed for each interval and each altitude separately, and complex spectra were computed from the resulting coherent signals using the Fast Fourier Transform (FFT). Finally, values for amplitude and phase were calculated from the spectra for the first three components: the diurnal, semidiurnal and terdiurnal tide. Error estimates for these results were obtained from Monte Carlo simulations using the error estimates of the observed temperature profiles.

B.3. Tides in the stratosphere and comparison with MERRA data

Figure B.1 shows seasonal tide amplitude profiles derived from lidar observations and the Modern Era-Retrospective Analysis for Research and Applications (MERRA) dataset. The amplitudes in both datasets generally increase with altitude, the amplitudes observed by lidar are, however, 3–6 times larger. One exception is the summer lidar profile of the diurnal tide which shows a minimum around 38 km altitude. The larger amplitudes below may be the result of systematic effects in the Rayleigh temperature retrieval (see section 4.1) and should therefore treated with caution. Interestingly, the local maximum of the diurnal tide in winter at 38 km altitude shows up in both lidar and MERRA data, although in lidar observations the maximum is much more pronounced and also covers a larger range of heights. Seasonal variations of tidal amplitudes are shown in

figure B.2.

The seasonal variation of the phase for the first three tidal components is plotted in panels a–c of figure B.3. In order to reduce noise, the phase was averaged over a 10 km altitude range centered at 45 km. A similar trend is observed in both datasets for most months. In case of the diurnal tide the largest difference occurs in months April and May. The phase determined from lidar observations jumps by approximately half a period between March and April. This jump leaves some ambiguity in answering the question of whether the phase advances or retards (see dashed line in figure B.3). Following the trend seen in the MERRA data, the phase of the diurnal tide is expected to advance between March and April. However, this requires a larger shift than would be necessary if the phase retards. Yet retardation is even less plausible given that the phase would have to jump 24 h between May and June in order to match the MERRA trend in June and July. A more careful analysis of the lidar data acquired in April and May did not show any abnormality. Considering that MERRA and lidar data agree within differences of less than 4 h for the remaining months, the nature of the jump seen in lidar data remains a mystery. No jumps or other peculiarities are evident from the traces of the semidiurnal or terdiurnal tide.

B.4. Tides in the MLT region

Thermal tides in the MLT region are characterized by coherent modes (similar phase and vertical wavelength of all three tidal components) in winter and a more complex structure in summer. The constructive interference of modes in winter results in large local time variations in temperature. As evident from figure B.4, the temperature at 96 km altitude changes by approximately 8 K between 6 and 10 local solar time. Decomposing the local time variation into tidal components yields amplitudes of 3.5 K, 2.0 K, and 1 K for the diurnal, semidiurnal, and terdiurnal tide, respectively. Phase progression of all three tidal components remains constant over more than 10 km (figure B.5). Since the vertical wavelength of the components are approximately equal, phases remain phase-locked, and constructive interference is thus observed over most of the altitude range. The situation is more complex in summer, however. As evident from figure B.7, the phases of the tidal components are much more variable than in winter. Moreover, depending on the particular selection of the altitude range, the vertical wavelength estimated from linear fits varies from 28 km (diurnal component) to 101 km (semidiurnal component). The non-phase-locked modes lead to a complex pattern in temperature (figure B.6). Also, amplitudes of the individual components are on average smaller compared to winter.

Figure B.8 shows the variation of temperature reconstructed from the first three tidal modes. The comparison with composites of lidar temperature observations (figures B.4 and B.6) reveals that all dominant features of the observed temperature variation can indeed be described using the first three modes only. Higher-order modes contribute marginally to the total variance.

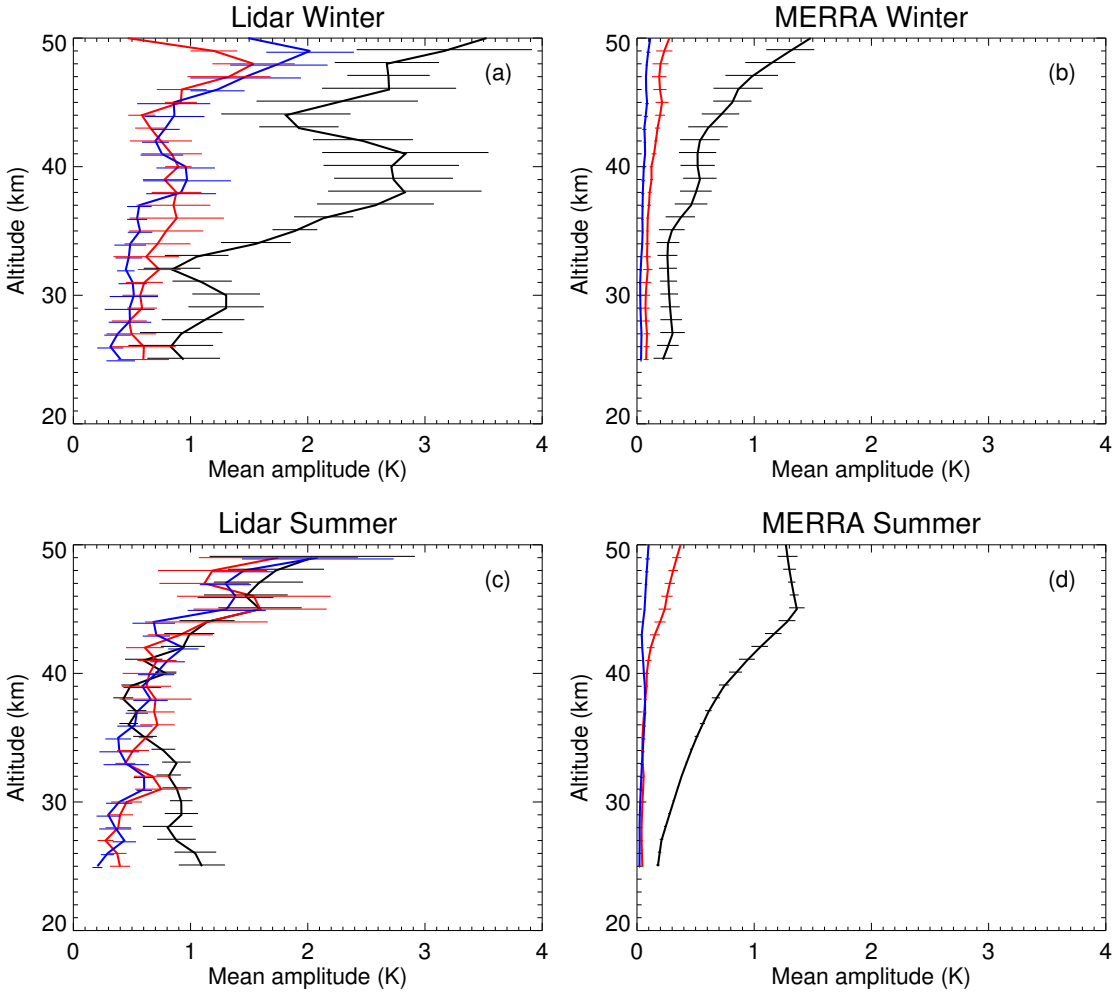


Figure B.1.: Thermal tide amplitude profiles determined from lidar observations (left column) and MERRA data (right column). The diurnal component is drawn in black, the semidiurnal component in red, and the terdiurnal component in blue. Horizontal lines indicate standard errors. The winter period comprises months May to September 2010, and the summer period months January, February and December in 2011 as well as January and February in 2012.

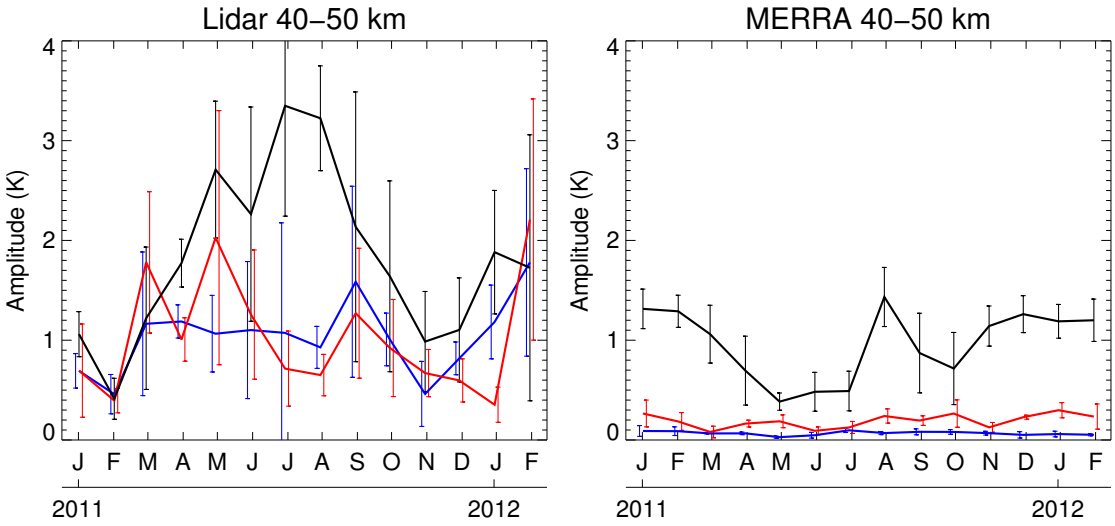


Figure B.2.: Seasonal variation of the tidal amplitudes (40–50 km altitude); black: diurnal component, red: semidiurnal component, blue: terdiurnal component. Error bars mark the standard deviation of amplitude estimates within the 10 km altitude range.

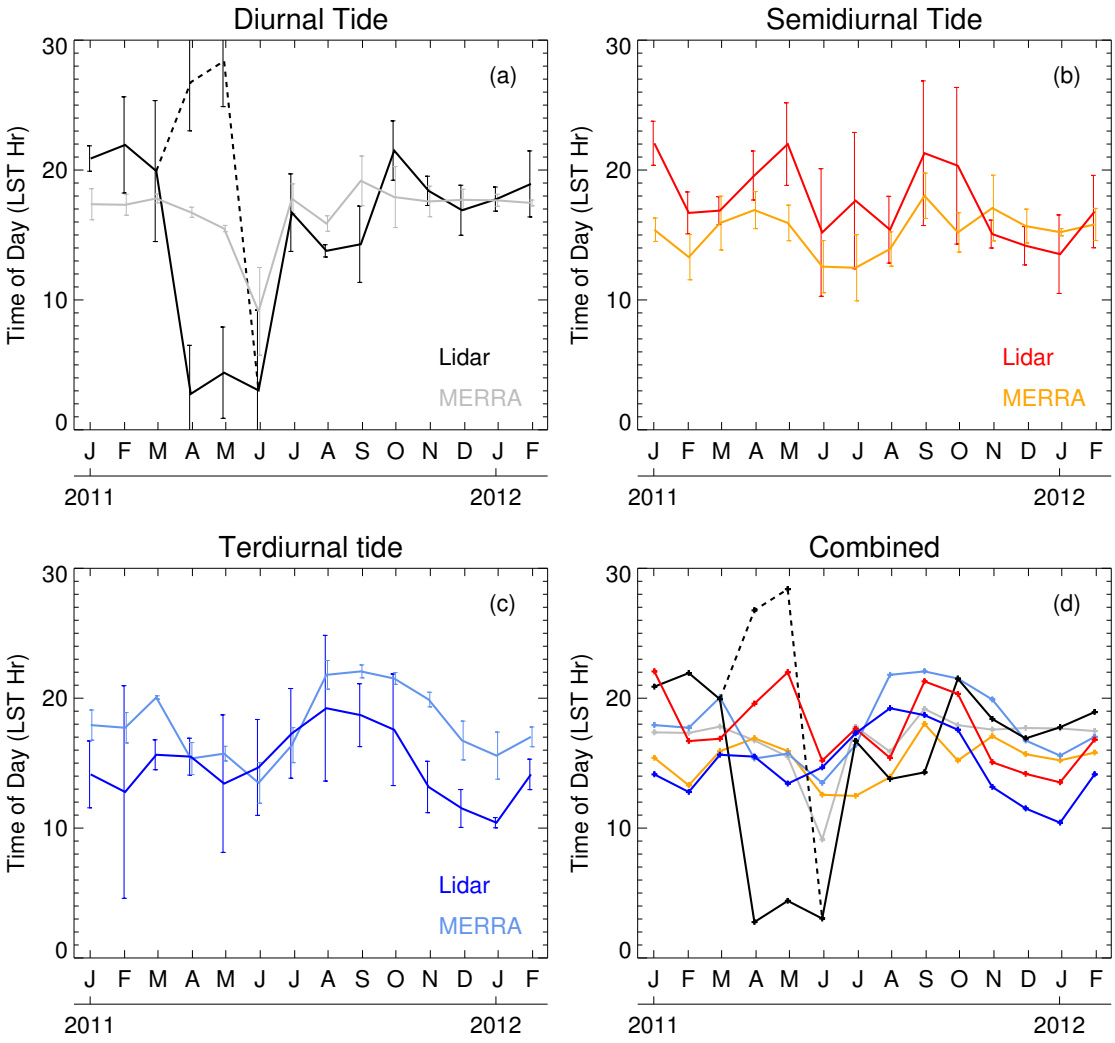


Figure B.3.: Seasonal variation of mean phases (40–50 km altitude). Panel a–c: Comparison between lidar results (dark colors) and phase values determined from the MERRA dataset (light colors). Error bars mark the standard deviation of phase estimates within the 10 km altitude range. Panel d: Combination of panels a–c.

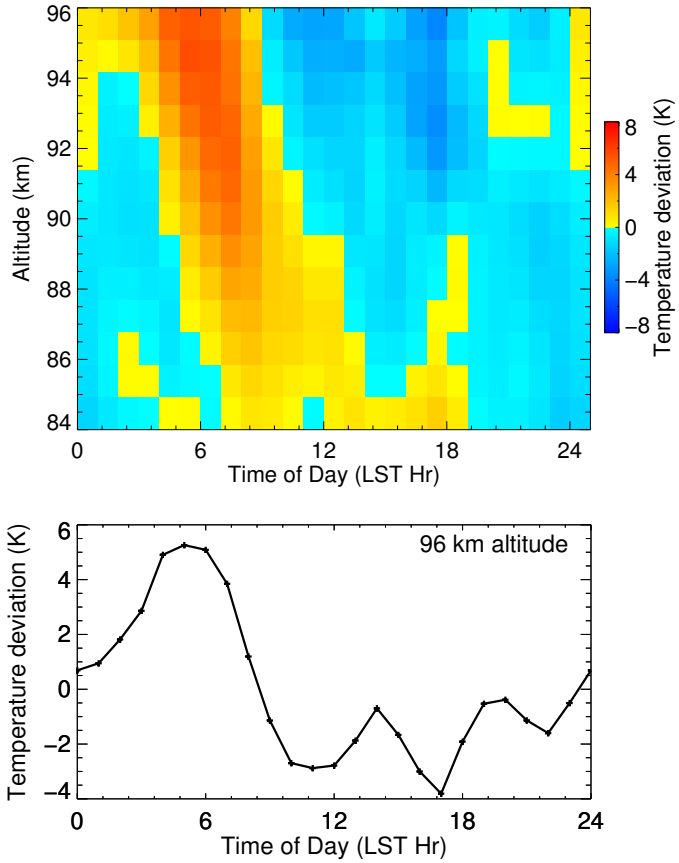


Figure B.4.: Local time variation of temperature in winter 2011 (months May to August). The composite comprises 726 hours of lidar observations.

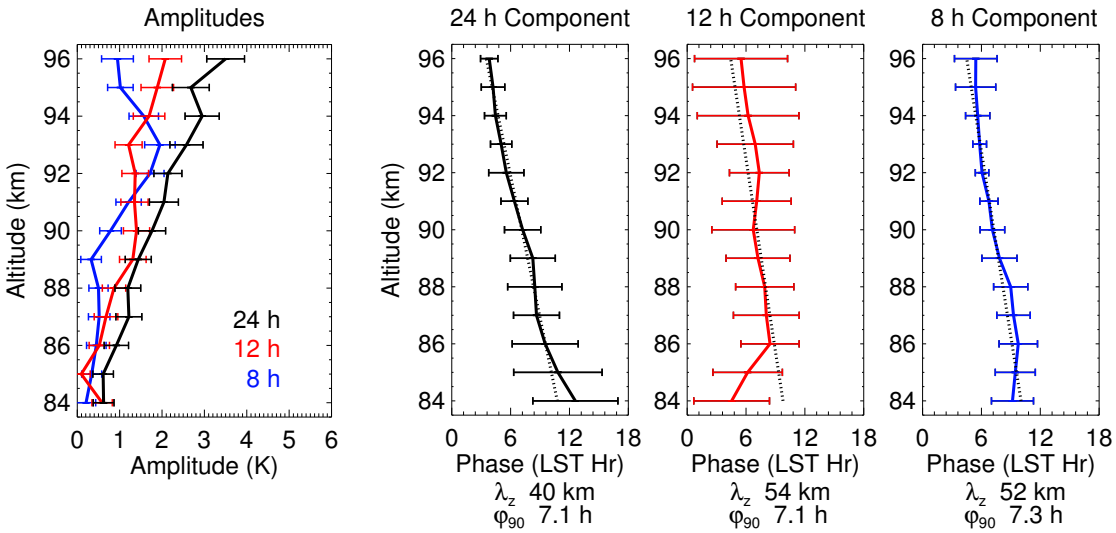


Figure B.5.: Amplitude and phase of the first three tidal components in winter 2011 (May to August). Also shown is the vertical wavelength λ_z and the phase at 90 km altitude φ_{90} determined by linear fit (dashed line).

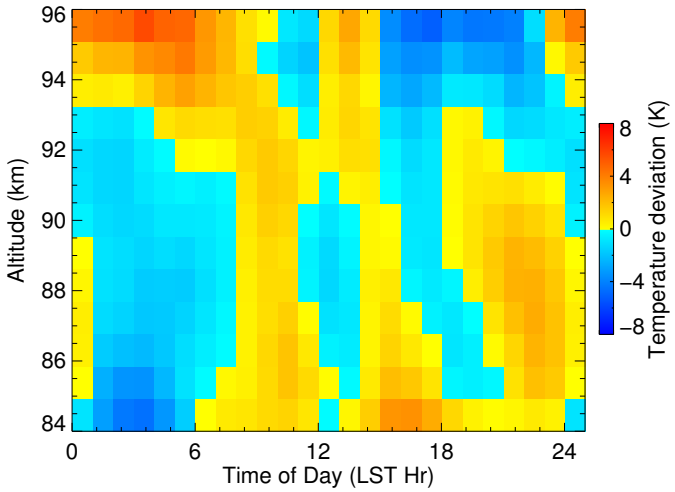


Figure B.6.: Local time variation of temperature in summer (January and February 2011, November to February in summer 2011/2012). The composite comprises 1120 hours of lidar observations.

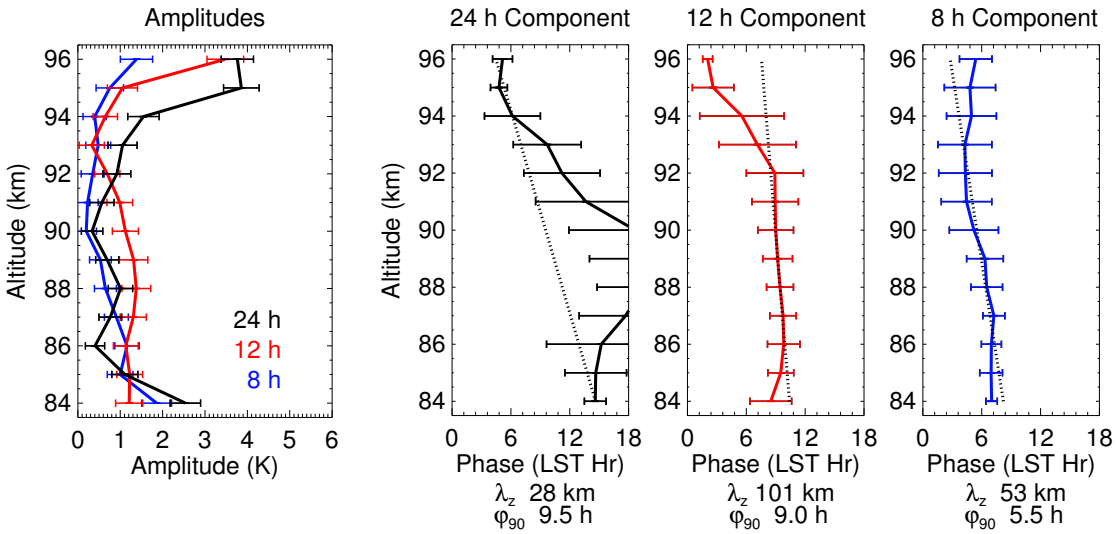


Figure B.7.: Amplitude and phase for the first three tidal components in summer (November to February). Also shown is the vertical wavelength λ_z and the phase at 90 km altitude φ_{90} determined by linear fit (dashed line). The fit range for the 24 h component is 84–85 km, and 94–96 km and 86–93 km in case of the 12 h and 8 h component.

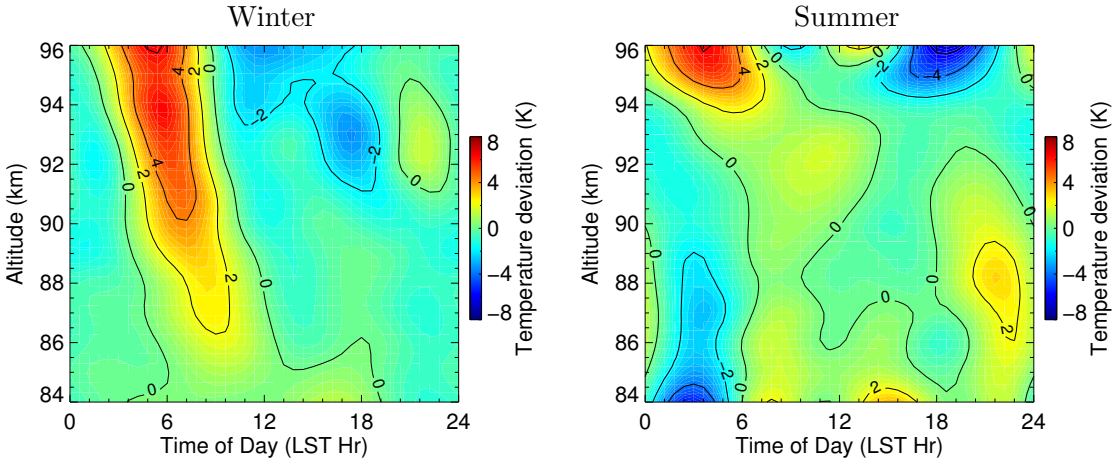


Figure B.8.: Temperature variation reconstructed from the first three tidal modes for winter (left) and summer (right).

C. Estimated fluorescence signal gain at 386 nm wavelength

The expected signal gain which results from probing a metal resonance line can be estimated by comparing the resonance scattering cross section to the molecular scattering cross section scaled by the relative densities of air molecules and metal atoms. In case of the fluorescence of iron at 386 nm the maximum scattering cross section at 200 K is $\sigma_{max} = 3.88 \times 10^{-18}\,\mathrm{m^2\,sr^{-1}}$ (*Lautenbach and Höffner*, 2004). The molecular backscattering cross section is expressed with the help of the phase function $q\,(\Omega)$ and the total molecular cross section σ_T as

$$\sigma_{\mathrm{Ray}}\,(180^\circ) = \sigma_T\,q\,(180^\circ)\,/4\pi \tag{C.1}$$

(*Fricke and von Zahn*, 1985). Evaluating the polynomial published by *Bucholtz* (1995) yields for the iron line at 386 nm $\sigma_T = 1.94 \times 10^{-30}\,\mathrm{m^2}$. Then equation C.1 reduces to $\sigma_{\mathrm{Ray}} = 2.28 \times 10^{-31}\,\mathrm{m^2\,sr^{-1}}$ assuming $q\,(180^\circ) = 300/203$ (*Fricke and von Zahn*, 1985). Thus, the resonance scattering cross section of iron σ_{max} is approximately 13 orders of magnitude larger than the molecular scattering cross section σ_{Ray}. Typical iron densities between 85 and 95 km range from 3×10^3 to $18 \times 10^3\,\mathrm{atoms\,cm^{-3}}$ (*Yu et al.*, 2012). As evident from figure 1.1, the density of the air at 90 km altitude is approximately $1 \times 10^{-9}\,\mathrm{g\,cm^{-3}}$. This value translates into $2 \times 10^{13}\,\mathrm{molecules\,cm^{-3}}$ using a mean molecular weight of 28.97. Although the number density of iron is about nine orders of magnitude lower than the air density, the larger resonance scattering cross section results still in a net signal gain of approximately four orders of magnitude.

Bibliography

Alexander, M. J., and R. A. Vincent (2000), Gravity waves in the tropical lower stratosphere: A model study of seasonal and interannual variability, *J. Geophys. Res.*, *105*(D14), 17,983–17,993, doi:10.1029/2000JD900197.

Alexander, S. P., A. R. Klekociuk, and D. J. Murphy (2011), Rayleigh lidar observations of gravity wave activity in the winter upper stratosphere and lower mesosphere above Davis, Antarctica (69°S, 78°E), *J. Geophys. Res.*, *116*(D13), D13,109, doi:10.1029/2010JD015164.

Alpers, M. (1993), Untersuchung der mesosphärischen Eisenschicht in polaren und niedrigen geographischen Breiten mittels Lidar, Ph.D. thesis, Universität Bonn, Bonn, Germany.

Alpers, M., R. Eixmann, C. Fricke-Begemann, M. Gerding, J. Höffner, et al. (2004), Temperature lidar measurements from 1 to 105 km altitude using resonance, Rayleigh, and Rotational Raman scattering, *Atmos. Chem. Phys.*, *4*(3), 793–800.

Andrews, D., J. Holton, and C. Leovy (1987), *Middle Atmosphere Dynamics*, Academic Press, Orlando.

Atmosphere, U. S. (1976), *US standard atmosphere*, National Oceanic and Atmospheric Administration.

Becker, E. (2012), Dynamical control of the middle atmosphere, *Space Sci. Rev.*, *168*, 283–314, doi:10.1007/s11214-011-9841-5.

Becker, E., and C. McLandress (2009), Consistent scale interaction of gravity waves in the Doppler spread parameterization, *J. Atmos. Sci.*, *66*(5), 1434–1449, doi:10.1175/2008JAS2810.1.

Beres, J. H., R. R. Garcia, B. A. Boville, and F. Sassi (2005), Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the Whole Atmosphere Community Climate Model (WACCM), *J. Geophys. Res.*, *110*(D10), doi:10.1029/2004JD005504.

Black, E. D. (2001), An introduction to Pound-Drever-Hall laser frequency stabilization, *Am. J. Phys.*, *69*, 79–87, doi:10.1119/1.1286663.

Blum, U., K. H. Fricke, G. Baumgarten, and A. Schöch (2004), Simultaneous lidar observations of temperatures and waves in the polar middle atmosphere on the east and west side of the Scandinavian mountains: a case study on 19/20 January 2003, *Atmospheric Chemistry and Physics*, *4*(3), 809–816, doi:10.5194/acp-4-809-2004.

Bucholtz, A. (1995), Rayleigh-scattering calculations for the terrestrial atmosphere, *Appl. Optics*, *34*(15), 2765–2773, doi:10.1364/AO.34.002765.

Chen, H., C. Y. She, P. Searcy, and E. Korevaar (1993), Sodium-vapor dispersive Faraday filter, *Opt. Lett.*, *18*(12), 1019–1021, doi:10.1364/OL.18.001019.

Chu, X., C. S. Gardner, and S. J. Franke (2005), Nocturnal thermal structure of the mesosphere and lower thermosphere region at Maui, Hawaii (20.7° N), and Starfire Optical Range, New Mexico (35° N), *J. Geophys. Res.*, *110*(D9), D09S03, doi:10.1029/2004JD004891.

Chu, X., Z. Yu, C. S. Gardner, C. Chen, and W. Fong (2011), Lidar observations of neutral Fe layers and fast gravity waves in the thermosphere (110–155 km) at McMurdo (77.8°S, 166.7°E), Antarctica, *Geophys. Res. Lett.*, *38*(23), doi:10.1029/2011GL050016.

Demtröder, W. (2008), *Laser Spectroscopy*, vol. 1, fourth ed., 138 pp., Springer.

Dikty, S., M. Weber, C. von Savigny, T. Sonkaew, A. Rozanov, and J. P. Burrows (2010), Modulations of the 27 day solar rotation signal in stratospheric ozone from Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (2003–2008), *J. Geophys. Res.*, *115*, D00I15, doi:10.1029/2009JD012379.

Dowdy, A., R. A. Vincent, K. Igarashi, Y. Murayama, and D. J. Murphy (2001), A comparison of mean winds and gravity wave activity in the northern and southern polar MLT, *Geophys. Res. Lett.*, *28*(8), 1475–1478, doi:10.1029/2000GL012576.

Drazin, P. G., and W. H. Reid (1981), *Hydrodynamic Stability*, Cambridge University Press.

Duck, T. J., J. A. Whiteway, and A. I. Carswell (2001), The Gravity Wave Arctic Stratospheric Vortex Interaction., *J. Atmos. Sci.*, *58*, 3581–3596, doi:10.1175/1520-0469(2001)058⟨3581:TGWASV⟩2.0.CO;2.

Farman, J. C., B. G. Gardiner, and J. D. Shanklin (1985), Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, *Nature*, *315*(6016), 207–210, doi:10.1038/315207a0.

Fernald, F. G., B. M. Herman, and J. A. Reagan (1972), Determination of aerosol height distributions by lidar, *J. Appl. Meteorol.*, *11*(3), 482–489, doi:10.1175/1520-0450(1972)011⟨0482:DOAHDB⟩2.0.CO;2.

Fiedler, J., G. Baumgarten, and F.-J. Lübken (2009), NLC observations during one solar cycle above ALOMAR, *J. Atmos. Solar Terr. Phys.*, pp. 424–433, doi:10.1016/j.jastp.2008.11.010.

Fleming, E. L., S. Chandra, J. Barnett, and M. Corney (1990), Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude, *Adv. Space Res.*, *10*(12), 11 – 59, doi:10.1016/0273-1177(90)90386-E.

Fricke, K. H., and U. von Zahn (1985), Mesopause temperatures derived from probing the hyperfine structure of the D_2 resonance line of sodium by lidar, *J. Atmos. Terr. Phys.*, *47*(5), 499–512, doi:10.1016/0021-9169(85)90116-3.

Fricke-Begemann, C. (2004), Lidar Investigations of the Mesopause Region: Temperature Structure and Variability, Ph.D. thesis, Rostock University.

Fricke-Begemann, C., and J. Höffner (2005), Temperature tides and waves near the mesopause from lidar observations at two latitudes, *J. Geophys. Res.*, *110*(D19), doi:10.1029/2005JD005770.

Fricke-Begemann, C., M. Alpers, and J. Höffner (2002a), Daylight rejection with a new receiver for potassium resonance temperature lidars, *Opt. Lett.*, *27*(21), 1932–1934, doi:10.1364/OL.27.001932.

Fricke-Begemann, C., J. Höffner, and U. von Zahn (2002b), The potassium density and temperature structure in the mesopause region (80-105 km) at a low latitude (28°N), *Geophys. Res. Lett.*, *29*(22), 2067, doi:10.1029/2002GL015578.

Fritts, D. C., and M. J. Alexander (2003), Gravity wave dynamics and effects in the middle atmosphere, *Rev. Geophys.*, *41*(1), 1003, doi:10.1029/2001RG000106.

Fritts, D. C., and L. Yuan (1989), An analysis of gravity wave ducting in the atmosphere: Eckart's resonances in thermal and doppler ducts, *Journal of Geophysical Research: Atmospheres*, *94*(D15), 18,455–18,466, doi:10.1029/JD094iD15p18455.

Gardner, C. S., and A. Z. Liu (2007), Seasonal variations of the vertical fluxes of heat and horizontal momentum in the mesopause region at Starfire Optical Range, New Mexico, *J. Geophys. Res.*, *112*(D9), doi:10.1029/2005JD006179.

Gardner, C. S., and W. Yang (1998), Measurements of the dynamical cooling rate associated with the vertical transport of heat by dissipating gravity waves in the mesopause region at the Starfire Optical Range, New Mexico, *J. Geophys. Res.*, *103*(D14), 16,909–16,926, doi:10.1029/98JD00683.

Gardner, C. S., G. Papen, X. Chu, and W. Pan (2001), First lidar observations of middle atmosphere temperatures, Fe densities, and polar mesospheric clouds over the North and South Poles, *Geophys. Res. Lett.*, *28*(7), 1199–1202, doi:10.1029/2000GL012622.

Gardner, C. S., Y. Zhao, and A. Z. Liu (2002), Atmospheric stability and gravity wave dissipation in the mesopause region, *J. Atmos. Solar Terr. Phys.*, *64*(8-11), 923 – 929, doi:10.1016/S1364-6826(02)00047-0.

Gardner, C. S., X. Chu, P. J. Espy, J. M. C. Plane, D. R. Marsh, and D. Janches (2011), Seasonal variations of the mesospheric Fe layer at Rothera, Antarctica (67.5°S, 68.0°W), *J. Geophys. Res.*, *116*(D2), D02,304, doi:10.1029/2010JD014655.

Gelbwachs, J. A. (1994), Iron Boltzmann factor LIDAR: proposed new remote-sensing technique for mesospheric temperature, *Appl. Optics*, *33*(30), 7151–7156, doi:10.1364/AO.33.007151.

Gerding, M., J. Höffner, J. Lautenbach, M. Rauthe, and F.-J. Lübken (2008), Seasonal variation of nocturnal temperatures between 1 and 105 km altitude at 54°N observed by lidar, *Atmos. Chem. Phys.*, *8*(24), 7465–7482, doi:10.5194/acp-8-7465-2008.

Gerrard, A. J., T. J. Kane, J. P. Thayer, and S. D. Eckermann (2004), Concerning the upper stratospheric gravity wave and mesospheric cloud relationship over Søndrestrøm, Greenland, *J. Atmos. Solar Terr. Phys.*, *66*(3-4), 229–240, doi:10.1016/j.jastp.2003.12.005.

Hagan, M. E. (1996), Comparative effects of migrating solar sources on tidal signatures in the middle and upper atmosphere, *J. Geophys. Res.*, *101*(D16), 21,213–21,222, doi:10.1029/96JD01374.

Hagan, M. E., and J. M. Forbes (2002), Migrating and nonmigrating diurnal tides in the middle and upper atmosphere excited by tropospheric latent heat release, *J. Geophys. Res.*, *107*(D24), 4754, doi:10.1029/2001JD001236.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*(4), doi:10.1029/2010RG000345.

Hauchecorne, A., and M.-L. Chanin (1980), Density and temperature profiles obtained by lidar between 35 and 70 km, *Geophys. Res. Lett.*, *7*, 565–568, doi:10.1029/GL007i008p00565.

Hedin, A. E. (1991), Extension of the MSIS thermosphere model into the middle and lower atmosphere, *J. Geophys. Res.*, *96*(A2), 1159–1172, doi:10.1029/90JA02125.

Hibbins, R., P. Espy, M. Jarvis, D. Riggin, and D. Fritts (2007), A climatology of tides and gravity wave variance in the MLT above Rothera, Antarctica obtained by MF radar, *J. Atmos. Solar Terr. Phys.*, *69*(45), 578 – 588, doi:10.1016/j.jastp.2006.10.009.

Hoffmann, P., E. Becker, W. Singer, and M. Placke (2010), Seasonal variation of mesospheric waves at northern middle and high latitudes, *J. Atmos. Solar Terr. Phys.*, *72*(1415), 1068 – 1079, doi:10.1016/j.jastp.2010.07.002.

Höffner, J. (1990), Messungen von Eisendichten in der polaren Hochatmosphäre, Master's thesis, Universität Bonn.

Höffner, J. (2010), Private communication.

Höffner, J. (2012), Private communication.

Höffner, J., and C. Fricke-Begemann (2005), Accurate lidar temperatures with narrowband filters, *Opt. Lett.*, *30*(8), 890–892, doi:10.1364/OL.30.000890.

Höffner, J., and F.-J. Lübken (2007), Potassium lidar temperatures and densities in the mesopause region at Spitsbergen (78°N), *J. Geophys. Res.*, *112*(D20), doi:10.1029/2007JD008612.

Hoppe, U.-P., and D. C. Fritts (1995), High-resolution measurements of vertical velocity with the European incoherent scatter VHF radar: 1. Motion field characteristics and measurement biases, *J. Geophys. Res.*, *100*(D8), 16,813–16,825, doi:10.1029/95JD01466.

Innis, J. L., and A. R. Klekociuk (2006), Planetary wave and gravity wave influence on the occurrence of polar stratospheric clouds over Davis Station, Antarctica, seen in lidar and radiosonde observations, *J. Geophys. Res.*, *111*(D22), doi:10.1029/2006JD007629.

Jäger, H. (2005), Long-term record of lidar observations of the stratospheric aerosol layer at Garmisch-Partenkirchen, *J. Geophys. Res.*, *110*(D8), doi:10.1029/2004JD005506.

Kaifler, B. (2009), Na Lidar at ALOMAR - electrooptic improvements, analysis algorithms, and selected atmospheric observations 80 to 100 km above Northern Norway, Master's thesis, University of Ulm.

Kaifler, N., and G. Baumgarten (2012), Private communication.

Kaifler, N., G. Baumgarten, A. Klekociuk, S. Alexander, J. Fiedler, and F.-J. Lübken (2013), Small scale structures of NLC observed by lidar at 69°N/69°S and their possible relation to gravity waves, *J. Atmos. Solar Terr. Phys.*, (0), doi:10.1016/j.jastp.2013.01.004.

Kaletta, D. (1969), Isotopenverschiebung im Eisen-I-Spektrum, Master's thesis, University of Hannover, Hannover, Germany.

Karlsson, B., C. McLandress, and T. G. Shepherd (2009), Inter-hemispheric mesospheric coupling in a comprehensive middle atmosphere model, *J. Atmos. Solar Terr. Phys.*, *71*(34), 518 – 530, doi:10.1016/j.jastp.2008.08.006.

Keller, P. (2006), Realisierung einer tageslichtfähigen optischen Nachweisbank für ein mobiles, tomographisches Eisen-Temperatur-LIDAR, Master's thesis, Universität Rostock.

Kent, G. S., and R. W. H. Wright (1970), A review of laser radar measurements of atmospheric properties, *J. Atmos. Terr. Phys.*, *32*(5), 917–943, doi:10.1016/0021-9169(70)90036-X.

Klekociuk, A. R., R. J. Morris, and J. L. Innis (2008), First Southern Hemisphere common-volume measurements of PMC and PMSE, *Geophys. Res. Lett.*, *35*, L24804, doi:10.1029/2008GL035988.

Knöpfel, R., and E. Becker (2011), An idealized radiative transfer scheme for use in a mechanistic general circulation model from the surface up to the mesopause region, *Journal of Quantitative Spectroscopy and Radiative Transfer*, *112*(9), 1460 – 1478, doi:10.1016/j.jqsrt.2011.02.014.

Kopp, M., and M. Gerding (2012), Private communication.

Körnich, H., and E. Becker (2010), A simple model for the interhemispheric coupling of the middle atmosphere circulation, *Adv. Space Res.*, *45*(5), 661 – 668, doi:10.1016/j.asr.2009.11.001.

Krueger, A. J., and R. A. Minzner (1976), A mid-latitude ozone model for the 1976 US Standard Atmosphere, *J. Geophys. Res.*, *81*(24), 4477–4481.

Kurucz, R. L., I. Furenlid, J. Brault, and L. Testerman (1984), *Solar flux atlas from 296 to 1300 nm*, National Solar Observatory Atlas, US. National Solar Observatory, Sunspot, NM.

Larsen, M. L., and A. B. Kostinski (2009), Simple dead-time corrections for discrete time series of non-Poisson data, *Meas. Sci. Technol.*, *20*(9), 095,101, doi:10.1088/0957-0233/20/9/095101.

Lautenbach, J. (2007), Experimentelle Untersuchungen mit einem Lidar zur thermischen Struktur der Mesopausen-Region bei polaren und mittleren Breiten, Ph.D. thesis, Universität Rostock, Rostock, Germany.

Lautenbach, J., and J. Höffner (2004), Scanning iron temperature lidar for mesopause temperature observation, *Appl. Optics*, *43*(23), 4559–4563, doi:10.1364/AO.43.004559.

Lindzen, R. S. (1981), Turbulence and stress owing to gravity wave and tidal breakdown, *J. Geophys. Res.*, *86*(C10), 9707–9714, doi:10.1029/JC086iC10p09707.

Lübken, F.-J. (1999), Thermal structure of the Arctic summer mesosphere, *J. Geophys. Res.*, *104*(D8), 9135–9149, doi:10.1029/1999JD900076.

Lübken, F.-J. (2012), Private communication.

Lübken, F.-J., and U. V. Zahn (1989), Simultaneous temperature measurements in the mesosphere and lower thermosphere during the Mac/Epsilon campaign, *Planet. Space Sci.*, *37*(10), 1303 – 1314, doi:10.1016/0032-0633(89)90020-2.

Lübken, F.-J., A. Müllemann, and M. J. Jarvis (2004), Temperatures and horizontal winds in the Antarctic summer mesosphere, *J. Geophys. Res.*, *109*(D24), D24112, doi:10.1029/2004JD005133.

Lübken, F.-J., J. Höffner, T. P. Viehl, B. Kaifler, and R. J. Morris (2011), First measurements of thermal tides in the summer mesopause region at Antarctic latitudes, *Geophys. Res. Lett.*, *38*, doi:10.1029/2011GL050045.

Manson, A., C. Meek, and G. Hall (1998), Correlations of gravity waves and tides in the mesosphere over Saskatoon, *J. Atmos. Solar Terr. Phys.*, *60*(11), 1089 – 1107, doi:10.1016/S1364-6826(98)00059-5.

Markwardt, C. B. (2009), Non-linear Least-squares Fitting in IDL with MPFIT, in *Astronomical Data Analysis Software and Systems XVIII*, *Astronomical Society of the Pacific Conference Series*, vol. 411, edited by D. A. Bohlender, D. Durand, and P. Dowler, p. 251.

Mayr, H. G., J. G. Mengel, K. L. Chan, and H. S. Porter (1998), Seasonal variations of the diurnal tide induced by gravity wave filtering, *Geophys. Res. Lett.*, *25*(7), 943–946, doi:10.1029/98GL00637.

McKay, J. A. (1999), Single and tandem Fabry-Perot etalons as solar background filters for lidar, *Appl. Optics*, *38*(27), 5851–5858, doi:10.1364/AO.38.005851.

Moffat-Griffin, T., R. E. Hibbins, M. J. Jarvis, and S. R. Colwell (2011), Seasonal variations of gravity wave activity in the lower stratosphere over an Antarctic Peninsula station, *J. Geophys. Res.*, *116*(D14), D14,111, doi:10.1029/2010JD015349.

Morris, R. (2012), Private communication.

Morris, R. J., J. Höffner, F.-J. Lübken, T. P. Viehl, B. Kaifler, and A. R. Klekociuk (2012), Experimental evidence of a stratospheric circulation influence on mesospheric temperatures and ice-particles during the 2010–2011 austral summer at 69°S, *J. Atmos. Solar Terr. Phys.*, *89*(0), 54 – 61, doi:10.1016/j.jastp.2012.08.007.

Müller, N. (2007), Assessment and Evaluation of ALOMAR Ozone Lidar Data 1995 to 2007. part ii: Background Estimation, Temperature Calculation, Chopper Analysis, Clound Detection and Receiver Diagnosis., project work.

Paul, J., F. Fortuin, and H. Kelder (1998), An ozone climatology based on ozonesonde and satellite measurements, *J. Geophys. Res.*, *103*(D24), 31,709–31,734, doi:10.1029/1998JD200008.

Pedatella, N. M., H.-L. Liu, and A. D. Richmond (2012), Atmospheric semidiurnal lunar tide climatology simulated by the Whole Atmosphere Community Climate Model, *J. Geophys. Res.*, *117*, A06327, doi:10.1029/2012JA017792.

Picone, J. M., A. E. Hedin, D. P. Drob, and A. C. Aikin (2002), NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, *J. Geophys. Res.*, *107*(A12), 1468, doi:10.1029/2002JA009430.

Plane, J. M. C. (1991), The chemistry of meteoric metals in the Earth's upper atmosphere, *International Reviews in Physical Chemistry*, *10*(1), 55–106, doi:10.1080/01442359109353254.

Plane, J. M. C. (2004), A time-resolved model of the mesospheric Na layer: constraints on the meteor input function, *Atmos. Chem. Phys.*, *4*(3), 627–638, doi:10.5194/acp-4-627-2004.

Preusse, P., S. Eckermann, J. Oberheide, M. Hagan, and D. Offermann (2001), Modulation of gravity waves by tides as seen in CRISTA temperatures, *Adv. Space Res.*, *27*(10), 1773 – 1778, doi:10.1016/S0273-1177(01)00336-2.

Rapp, M., I. Strelnikova, B. Strelnikov, M. Friedrich, J. Gumbel, U.-P. Hoppe, T. Blix, O. Havnes, P. Bracikowski, K. Lynch, and S. Knappmiller (2011), Microphysical properties of mesospheric aerosols: An overview of in situ-results from the ECOMA-project, in *Aeronomy of the Earth's Atmosphere and Ionosphere*, edited by M. A. Abdu, D. Pancheva, and A. Bhattacharyya, pp. 67–74, Springer Science+Business Media B. V., doi:10.1007/978-94-007-0326-1-4, IAGA Special Sopron Book Series.

Rauthe, M. (2008), Lidarmessungen von Temperaturen und Schwerewellen zwischen 1–105 km über Kühlungsborn (54° N, 12° O), Ph.D. thesis, Universität Rostock, Rostock, Germany.

Rauthe, M., M. Gerding, and F.-J. Lübken (2008), Seasonal changes in gravity wave activity measured by lidars at mid-latitudes, *Atmospheric Chemistry and Physics*, *8*(22), 6775–6787, doi:10.5194/acp-8-6775-2008.

Schöch, A. (2007), Thermal structure and gravity waves in the Arctic middle atmosphere above ALOMAR (69.3° N, 16.0° E), Ph.D. thesis, Universität Rostock, Rostock, Germany.

Schöch, A., G. Baumgarten, and J. Fiedler (2008), Polar middle atmosphere temperature climatology from Rayleigh lidar measurements at ALOMAR (69°N), *Ann. Geophys.*, *26*(7), 1681–1698, doi:10.5194/angeo-26-1681-2008.

She, C., S. Chen, B. Williams, Z. Hu, D. A. Krueger, and M. Hagan (2002a), Tides in the mesopause region over Fort Collins, Colorado (41°N, 105°W) based on lidar temperature observations covering full diurnal cycles, *J. Geophys. Res.*, *107*(D18), ACL–4, doi:10.1029/2001JD001189.

She, C., T. Li, R. L. Collins, T. Yuan, B. P. Williams, T. D. Kawahara, J. D. Vance, P. Acott, D. A. Krueger, H.-L. Liu, et al. (2004), Tidal perturbations and variability in the mesopause region over Fort Collins, CO (41°N, 105°W): Continuous multi-day temperature and wind lidar observations, *Geophys. Res. Lett.*, *31*(24), doi:10.1029/2004GL021165.

She, C. Y., and U. von Zahn (1998), Concept of a two-level mesopause: Support through new Lidar observations, *J. Geophys. Res.*, *103*(D5), 5855–5863, doi:10.1029/97JD03450.

She, C. Y., and J. R. Yu (1994), Simultaneous three-frequency Na lidar measurements of radial wind and temperature in the mesopause region, *Geophys. Res. Lett.*, *21*(17), 1771–1774, doi:10.1029/94GL01417.

She, C. Y., H. Latifi, J. R. Yu, R. J. Alvarez, R. E. Bills, and C. S. Gardner (1990), Two-frequency lidar technique for mesospheric Na temperature measurements, *Geophys. Res. Lett.*, *17*(7), 929–932, doi:10.1029/GL017i007p00929.

She, C.-Y., S. Chen, Z. Hu, J. Sherman, J. D. Vance, V. Vasoli, M. A. White, J. Yu, and D. A. Krueger (2000), Eight-year climatology of nocturnal temperature and sodium density in the mesopause region (80 to 105 km) over Fort Collins, CO (41° N, 105° W), *Geophys. Res. Lett.*, *27*(20), 3289–3292, doi:10.1029/2000GL003825.

She, C. Y., J. D. Vance, B. P. Williams, D. A. Krueger, H. Moosmüller, D. Gibson-Wilde, and D. C. Fritts (2002b), Lidar studies of atmospheric dynamics near polar mesopause, *Trans. Am. Geophys. Union (EOS)*, *83*(27), 289–293, doi:10.1029/2002EO000206.

Smith, S. A., D. C. Fritts, and T. E. Vanzandt (1987), Evidence for a saturated spectrum of atmospheric gravity waves, *J. Atmos. Sci.*, *44*(10), 1404–1410, doi:10.1175/1520-0469(1987)044⟨1404:EFASSO⟩2.0.CO;2.

Solomon, S. (2007), *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, vol. 4, Cambridge University Press.

Tabazadeh, A., R. P. Turco, K. Drdla, M. Z. Jacobson, and O. B. Toon (1994), A study of type I polar stratospheric cloud formation, *J. Geophys. Res.*, *21*, 1619–1622, doi:10.1029/94GL01368.

Thayaparan, T., W. K. Hocking, and J. MacDougall (1995), Observational evidence of tidal/gravity wave interactions using the UWO 2 MHz radar, *Geophys. Res. Lett.*, *22*(4), 373–376, doi:10.1029/94GL03270.

Thomason, L. W., G. S. Kent, C. R. Trepte, and L. R. Poole (1997), A comparison of the stratospheric aerosol background periods of 1979 and 1989–1991, *J. Geophys. Res.*, *102*(D3), 3611–3616, doi:10.1029/96JD02960.

Varotsos, C. (2004), The extraordinary events of the major, sudden stratospheric warming, the diminutive Antarctic ozone hole, and its split in 2002, *Environmental Science and Pollution Research*, *11*(6), 405–411, doi:10.1007/BF02979661.

Voigt, S., J. Orphal, K. Bogumil, and J. Burrows (2001), The temperature dependence (203–293 k) of the absorption cross sections of O_3 in the 230–850 nm region measured by Fourier-transform spectroscopy, *Journal of Photochemistry and Photobiology A: Chemistry*, *143*(1), 1 – 9, doi:10.1016/S1010-6030(01)00480-4.

von Zahn, U., and J. Höffner (1996), Mesopause temperature profiling by potassium lidar, *Geophys. Res. Lett.*, *23*(2), 141–144, doi:10.1029/95GL03688.

Wagner, R., C. Linke, K.-H. Naumann, M. Schnaiter, M. Vragel, M. Gangl, and H. Horvath (2009), A review of optical measurements at the aerosol and cloud chamber AIDA, *J. Quant. Spectrosc. Radiat. Transfer*, *110*(11), 930 – 949, doi:10.1016/j.jqsrt.2009.01.026.

Whiteway, J. A., T. J. Duck, D. P. Donovan, J. C. Bird, S. R. Pal, and A. I. Carswell (1997), Measurements of gravity wave activity within and around the Arctic stratospheric vortex, *Geophys. Res. Lett.*, *24*(11), 1387–1390, doi:10.1029/97GL01322.

Wright, C. J., and J. C. Gille (2013), Detecting overlapping gravity waves using the S-Transform, *Geophys. Res. Lett.*, doi:10.1002/grl.50378.

Xu, J., A. K. Smith, W. Yuan, H.-L. Liu, Q. Wu, M. G. Mlynczak, and J. M. Russell (2007), Global structure and long-term variations of zonal mean temperature observed by TIMED/SABER, *J. Geophys. Res.*, *112*(D24), doi:10.1029/2007JD008546.

Yamashita, C., X. Chu, H.-L. Liu, P. J. Espy, G. J. Nott, and W. Huang (2009), Stratospheric gravity wave characteristics and seasonal variations observed by lidar at the South Pole and Rothera, Antarctica, *J. Geophys. Res.*, *114*(D12), D12,101, doi:10.1029/2008JD011472.

Young, A. T. (1981), Rayleigh scattering, *Appl. Optics*, *20*(4), 533–535, doi:10.1364/AO.20.000533.

Yu, Z., X. Chu, W. Huang, W. Fong, and B. R. Roberts (2012), Diurnal variations of the Fe layer in the mesosphere and lower thermosphere: Four season variability and solar effects on the layer bottomside at McMurdo (77.8°S, 166.7°E), Antarctica, *J. Geophys. Res.*, *117*(D22), D22,303, doi:10.1029/2012JD018079.

Yuan, T., H. Schmidt, C. She, D. A. Krueger, and S. Reising (2008), Seasonal variations of semidiurnal tidal perturbations in mesopause region temperature and zonal and meridional winds above Fort Collins, Colorado (40.6° N, 105.1° W), *J. Geophys. Res.*, *113*(D20), doi:10.1029/2007JD009687.

Zahn, U., G. Cossart, J. Fiedler, K. Fricke, G. Nelke, G. Baumgarten, D. Rees, A. Hauchecorne, and K. Adolfsen (2000), The ALOMAR Rayleigh/Mie/Raman lidar: objectives, configuration, and performance, *Ann. Geophys.*, *18*(7), 815–833, doi:10.1007/s00585-000-0815-2.

List of Figures

1.1. Typical temperature and air density profiles at high latitudes 2
1.2. Illustration of the working principle of a lidar 4
1.3. Doppler-broadening of the iron resonance line at 386 nm 5
1.4. Lidar return signal acquired on April 19, 2011 6
1.5. Map of the Antarctic continent 8
1.6. Areal photography of Davis station 9
1.7. Timeline of the project 10

2.1. Sketch of the transmitter system 13
2.2. Sketch of the optical bench of the receiver 16
2.3. Picture of the Fe lidar container 17

3.1. Effect of detector saturation on temperature profiles 22
3.2. Time series of detected photons at 20 km altitude 25
3.3. Signal ratio as function of count rate 26
3.4. Estimated dead time values 27
3.5. Optical depth of the atmosphere at 386 nm wavelength 29
3.6. Example of a measured backscatter spectrum 31
3.7. Temperature profiles with and without spectral correction 34
3.8. Spectrum of the laser line, Doppler-broadening and filter transmission . 35
3.9. Measured filter transmission spectrum and drift of the filter 37
3.10. Seasonal variation of the aerosol loading 38
3.11. The lidar- and radiosonde temperature datasets 40
3.12. Comparison between lidar and radiosonde temperature data 41
3.13. Seasonal variation of the temperature deviation and correlation with the color ratio 42

4.1. Lidar observations as function of time of day and season 44
4.2. Histogram and cumulative distribution of observation hours 45
4.3. Number of observation hours per month. 46
4.4. Local time distribution of lidar observations 46
4.5. Monthly mean temperature profiles 48
4.6. Fe lidar temperature dataset 49
4.7. Seasonal variation of the mean temperature at Davis 50
4.8. Comparison of seasonal mean temperatures with ECMWF model data . 52
4.9. Seasonal variation of the height of the stratopause 53
4.10. Seasonal variation of the mean temperature in the southern and northern hemisphere 54
4.11. Time series of mean temperatures 55
4.12. Amplitude spectrum of seasonal temperature variations 56
4.13. Illustration of the temperature perturbation analysis 59
4.14. Temperature perturbation spectra 61
4.15. Seasonal temperature perturbation amplitude profiles 61

4.16. Relative variability of mean temperature perturbation amplitudes 62
4.17. Seasonal temperature perturbation amplitude profiles for different sites . 64
4.18. Temperature perturbation spectrum observed at ALOMAR 65
4.19. Seasonal variation of the ratio of temperature variances estimated from two different spectral bands 66
4.20. Seasonal variation of temperature perturbation amplitudes 66
4.21. Mean temperature perturbation amplitudes versus variance ratio 67
4.22. Mesospheric temperature perturbation spectra 68
4.23. Mesospheric mean temperature perturbation amplitude profiles 69
4.24. Illustration of the gravity wave analysis process 74
4.25. Seasonal variation of the gravity wave potential energy density 77
4.26. Correlation lengths of gravity wave potential energy density measurements 78
4.27. Monthly mean gravity wave potential energy densities 79
4.28. Seasonal gravity wave potential energy density profiles 80
4.29. Vertical wavenumber spectra 82
4.30. Illustration of the filtering of gravity waves 84
4.31. Zonal and meridional wind extracted from the ECMWF model 85
4.32. Seasonal mean ECMWF zonal wind profiles 86
4.33. Monthly gravity wave potential energy density profiles 88
4.34. ECMWF mean zonal wind profiles 89
4.35. Interaction of gravity waves with the background wind field 90
4.36. Illustration of the gravity wave model 92
4.37. Comparison between measured gravity wave potential energy densities and model results 94
4.38. Monthly mean gravity wave potential energy densities for different sites 96
4.39. Temperature profiles acquired on January 27, 2011 98
4.40. Mesospheric seasonal mean and monthly mean gravity wave potential energy density profiles 99
4.41. Mesospheric vertical wave number spectrum 100
4.42. Seasonal variation of the gravity wave potential energy density 101
4.43. Comparison between measured mesospheric gravity wave potential energy densities and model results 102

A.1. Illustration of the data processing for temperature, vertical wind, and heat flux 108
A.2. Seasonal variation of the measured vertical heat flux 109
A.3. Seasonal variation of the vertical heat flux at Davis and Starfire Optical Range 110
A.4. Seasonal variation of temperature and vertical heat flux 111
A.5. Annual mean heat flux profiles at Davis and Starfire Optical Range . . . 112
A.6. Mean vertical heat flux profiles 112

B.1. Seasonal tide amplitude profiles: observations and MERRA data 118
B.2. Seasonal variation of tidal amplitudes: observations and MERRA data . 119
B.3. Seasonal variation of mean phases: observations and MERRA data . . . 120
B.4. Variation of mesospheric temperature in winter caused by tides 121
B.5. Amplitude and phase of the first three tidal components in winter . . . 121
B.6. Variation of mesospheric temperature in summer caused by tides 122
B.7. Amplitude and phase of the first three tidal components in summer . . . 122
B.8. Variation of mesospheric temperature caused by tides 123

List of Tables

2.1. Technical data of the Fe lidar system 12

4.1. Summary of the datasets used in the comparison of temperature perturbation amplitudes observed at different sites 63
4.2. Mean relative variability of gravity wave potential energy densities . . . 76
4.3. Multiplication factors used in the gravity model 93

Abbreviations

AAD	Australian Antarctic Division
ALOMAR	Arctic Lidar Observatory for Middle Atmosphere Research
APD	Avalanche Photo Diode
BAS	British Antarctic Survey
BoM	Australian Bureau of Meteorology
COCOS	Command & Control System
ECMWF	European Center for Medium-Range Weather Forecasts
FFT	Fast Fourier Transform
FSR	Free Spectral Range
GCM	Global Circulation Model
GWPED	Gravity Wave Potential Energy Density
IAP	Institute of Atmospheric Physics
KMCM	Kühlungsborn Mechanisic general Circulation Model
LBO	Lithium TriBOrate
LOS wind	line-of-sight wind
MERRA	Modern Era-Retrospective Analysis for Research and Applications
MLT region	Mesosphere-Lower Thermosphere region
PDH	Pound-Drever-Hall
PMC	Polar Mesospheric Cloud
PMT	Photo Multiplier Tube
PSC	Polar Stratospheric Cloud
PSD	Power Spectral Density
RMR lidar	Rayleigh-/Mie-/Raman lidar
SHG	Second Harmonic Generation
WACCM	Whole Atmosphere Community Climate Model

Acknowledgements

The "Fe Lidar at Davis" project was a collaborative effort with many people involved. I would like to thank all previous colleagues at IAP for their support during the painstaking months of hard work to get the lidar instrument ready for Antarctica, and for their support during the eleven month period I have been at Davis, in transit to or from Davis, or in training at Hobart.

My time at Davis was made successful and enjoyable in large part due to the friendship and helpfulness of the wintering crew 2011: Andrew Hooper, Andrew Stanley, Brad Collins, Brian McKechnie, Cliff Simpson-Davis, Corey Brazendale, Craig Ingrames, David Hawley, David Tulloh, David Hosken, Doug McVeigh, Graham Cook, Jenn McGhee, Jenny Feast, Jim Dunnett, Justin Chambers, Kerry Steinberner, Lloyd Fletcher, Ray Morris, Rick Besso, Rob Deerson, Shane Mitchell, Tony Bojkovski. Every single one contributed in one way or another to the success of the project. I am especially grateful to Ray Morris and David Hosken who assisted in operating the lidar, and I thank Josef Höffner and Timo Viehl for continuing the observation program after my return to Kühlungsborn.

I gratefully acknowledge the role of the Australian Antarctic Division (AAD). AAD provided transport to and from Davis Station, as well as the infrastructure for operating the lidar. In regard to exchange of scientific expertise I thank Damian Murphy, Andrew Klekociuk, and Simon Alexander for valuable discussions while I was at Hobart and at conferences, as well as for providing radar and lidar data.

I wish to thank Gerd Baumgarten for his support, encouragement, and helpfulness. The discussions were extremly valuable for refocusing the scientific aim of this thesis. I am also grateful to Erich Becker for discussing theoretical aspects of gravity waves and tides. Comparisons between lidar measurements and KMCM results provided insights into atmospheric processes on a much broader level, thus making interpretation of observational data easier.

In looking back at the past years, I realize that there are some individuals who had a strong influence on my scientific career and thus knowingly or unknowingly contributed to the success of my PhD project. The first person is certainly Rudolf Lehn, who spent an innumerable number of hours with groups of students interested in physics. After school hours we built many experiments in the physics lab at the secondary school in Bad Saulgau. It soon became clear that I was more into doing experiments than working out theories on paper. I credit him for becoming an experimenter. Then there is Paul Ziemann. During my undergraduate studies he offered me work at the ion beam line. I spent many hours in the lab doing Rutherford backscatter measurements, and I spent even more hours repairing the vintage ion accelerator. Finding out how this really big machine worked was good fun and I learned a lot. Ulf-Peter Hoppe first introduced me to atmospheric physics and lidars during the two semester I was at the University of Tromsø. Natalie and I asked him if we could do a student project in this field, and he immediately said "yes". His enthusiasm is largely responsible for my change in plans to continue with atmospheric physics and not go back to the field of solid state physics. During the time I spent at ALOMAR I met Biff Williams. He showed me all the intricacies of the Na lidar and, most important, allowed me to try many things

myself, e.g. align the various components of the laser system. This hands-on experience proved to be very valuable during the overwinter at Davis. Finally, there is Franz-Josef Lübken who sent me to Antarctica with a 20 foot container full of lidar hardware and 9 months time to get some lidar measurements done. The result is this thesis.

Lastly, I would like to thank my family for all their encouragement. For my parents who raised me with a curious attitude towards nature and supported me in all possible ways. And most of all for my wife Natalie, who shares my passion. Thank you for being so patient.