# $O_2$ and OH airglow in the mesosphere through the lens of the Odin/OSIRIS Infrared Imager

ANQI LI

*Department of Space, Earth and Environment*
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

# O$_2$ and OH airglow in the mesosphere through the lens of the Odin/OSIRIS Infrared Imager

Anqi Li

# Abstract

Observing airglow emissions allows us to explore the chemical composition and dynamics of the upper mesosphere. The Odin satellite has routinely measured $O_2(a^1\Delta_g)$ and OH Meinel band airglow emissions from 2001 until 2015. In this thesis, the Odin/OSIRIS infrared imager data have been studied in some depth for the first time. Numerical inversions of the observed radiances have been carried out to retrieve the volume emission rates and, thereafter, the mesospheric ozone density. This resulted in a new, long-term high-resolution airglow and ozone datasets of the middle atmosphere.

The photolysis reactions are affected by periodic changes in solar irradiance. Thus, the $OH_v$ emission should vary with the solar cycle. In terms of the 11-year solar cycle, as expected, we observed that the vertically integrated intensity of $OH_v$ correlates positively with the Lyman-$\alpha$ flux and that the emission height correlates negatively at most latitudes except near the equator. Employing a time-dependent photochemical model, we showed that the changing local time sampling of the Odin satellite was the cause of the observed distortion of the solar cycle signature near the equator.

We also observed the episodic signatures in the two airglow emissions of sudden stratospheric warming (SSW) events. With the aid of the temperature and $H_2O$ measured from Odin-SMR, we qualitatively assessed the events that occurred in 2009, 2012, and 2013. Using analytical expressions, we derived proxy O and $OH_v$ number densities. A significant amount of atomic oxygen-rich air descends into the mesosphere a few days after the SSW onsets, resulting in unusually intense airglow emissions at a much lower altitude than average. The modelled $OH_v$ largely resembles the temporal evolution of the observed $OH_v$. The synchronous structure of the two airglow emissions indicates that the vertical transport of O plays a dominant role in the observed changes.

This thesis work has set a valuable foundation as part of the preparations for the future MATS mission.

**Keywords:** Satellite limb observation, mesosphere, photochemistry, atmospheric waves, airglow, mesospheric ozone.

# Acknowledgments

This thesis work could not have been completed without many individuals who have assisted me along the journey. First of all, Donal Murtagh, Kristell Pérot, and Ole Martin Christensen have provided me with excellent academic guidance and beyond by being my mentors as well as role models. Their in-depth knowledge of the scientific fields and acceptance encouraged me to grow both as a researcher and person. Also, thanks to Patrick Eriksson and Francesco Grieco for proofreading this thesis and providing valuable input.

Secondly, Chris Roth, Adam Bourassa, and Doug Degenstein in Saskatoon have shown continuous interest and provided comprehensive input to this work. I am incredibly honoured to collaborate with you on the OSIRIS project. Also, Linda and Jörg in Stockholm have given me their trust, in particular, the 'wavy part,' and included me in the MATS science team.

Finally, my fellow doctoral students and other colleagues in the department have given me a great environment to develop many other skills. Special thanks to dear Alexander, who has become my home-office co-worker throughout the second half of my doctoral journey (due to the pandemic) and provided me with a safety net in all aspects. And, of course, without my family's support, I would not have ended up where I am now.

Anqi Li
Göteborg, April 2022

# List of Publications

This thesis is based on the following appended papers:

**Paper 1.** Anqi Li, Chris Roth, Kristell Pérot, Ole Martin Christensen, Adam Bourassa, Doug Degenstein and Donal Murtagh. *Retrieval of daytime mesospheric ozone using OSIRIS observation of $O_2(a^1\Delta_g)$ emission.* Atmos. Meas. Tech., 13, 6215–6236, https://doi.org/10.5194/amt-13-6215-2020, 2020.

**Paper 2.** Anqi Li, Chris Roth, Kristell Pérot, Ole Martin Christensen, Adam Bourassa, Doug Degenstein and Donal Murtagh. *The OH (3-1) nightglow volume emission rate retrieved from OSIRIS measurements: 2001 to 2015.* Earth Syst. Sci. Data, 13, 5115–5126, https://doi.org/10.5194/essd-13-5115-2021, 2021.

**Paper 3.** Anqi Li, Chris Roth, Adam Bourassa, Doug Degenstein and Donal Murtagh. *11-year solar cycle influence on OH (3-1) nightglow observed by OSIRIS* Published (2022). Journal of Atmospheric and Solar-Terrestrial Physics, 229, 105831, 1364-6826, https://doi.org/10.1016/j.jastp.2022.105831, 2022.

**Paper 4.** Anqi Li, Kristell Pérot and Donal Murtagh. *Case studies of the OH and $O_2$ airglow response to sudden stratospheric warmings with elevated stratopause* Manuscript in preparation.

Contributions of the authors: Anqi Li is the principal author of all appended papers. She has taken full responsibility for the data processing, modelling work and data analysis of $O_2(a^1\Delta_g)$ volume emission rate and ozone in Paper 1 and 4, and OH (3-1) volume emission rate in Paper 2, 3 and 4. Professor Donal Murtagh is the principal academic supervisor and has contributed with discussion, guidance and editing of all papers. He wrote Sect. 2.2 in Paper 4. Associate Professor Kristell Pérot is the assistant academic supervisor and has contributed with discussion and guidance of Paper 1, 2 and 4, and editing of all papers. Ole Martin Christensen has contributed with guidance regarding retrieval methods, discussion and editing of Paper 1 and 2. Chris Roth wrote Sect. 2.1 in Paper 1. Professor Adam Bourassa, Professor Doug Degenstein and Chris Roth have produced the calibrated IRI data and contributed with discussion and editing of all papers.

Other relevant publication co-authored by Anqi Li but have not been included in the thesis:

Jörg Gumbel, Linda Megner, Ole Martin Christensen, Nickolay Ivchenko, Donal P. Murtagh, Seunghyuk Chang, Joachim Dillner, Terese Ekebrand, Gabriel Giono, Arvid Hammar, Jonas Hedin, Bodil Karlsson, Mikael Krus, **Anqi Li**, Steven McCallion, Georgi Olentšenko, Soojong Pak, Woojin Park, Jordan Rouse, Jacek Stegman, Georg Witt. *The MATS satellite mission-Gravity wave studies by Mesospheric Airglow/Aerosol Tomography and Spectroscopy* Atmospheric Chemistry and Physics 2020 vol: 20 (1) pp: 431-455.

# Contents

# Part I

# Introductory chapters

# Chapter 1

# Welcome to the mesosphere

The Earth's atmosphere can be divided into different layers based on their characteristics. One of the most common divisions of the atmosphere is based on its vertical temperature gradient, which is either positive or negative, as shown in Fig. 1.1. The mesosphere[1] is one of the 'spheres' among others, where the temperature gradient is negative, and it ranges approximately from $50\,\mathrm{km}$ to $100\,\mathrm{km}$ altitude. In this thesis, other 'spheres' such as the troposphere, stratosphere, and thermosphere are not studied explicitly, although some of the considered phenomena reach slightly above the boundary of the mesosphere. In such cases, the considered region is commonly referred to as the MLT (Mesosphere and Lower Thermosphere).

The examination of mesospheric processes requires us to address several aspects simultaneously. Especially, the coupled effects of the dynamical and photochemical processes should be carefully considered. In addition, inversion theory must be addressed while handling data collected from remote sensing techniques and attempting to convert the measured signals to the desired quantities. This thesis focuses on the following three main areas: dynamics that influence the motion in the mesosphere, oxygen and hydroxyl photochemistry, and retrieval techniques using the optimal estimation method.

## 1.1 Why is the mesosphere so interesting?

The top of the mesosphere, the mesopause, can be considered as a region between $80\,\mathrm{km}$ and $100\,\mathrm{km}$ where the coldest temperatures of Earth's atmosphere can be found. The coldest place is located at the summer pole, not the winter pole. In turn, the extremely cold temperatures at the summer mesopause help to create noctilucent clouds, despite the fact that the humidity is very low there. So why is it so cold at the summer pole even though the solar heating rate there is more significant than in the winter hemisphere? A simple explanation is that the extremely cold temperatures are due to the net upward rising motion and the associated adiabatic cooling process. This is one of the examples showing that dynamical aspects modify the thermal balance in the atmosphere. In addition to the adiabatic cooling and
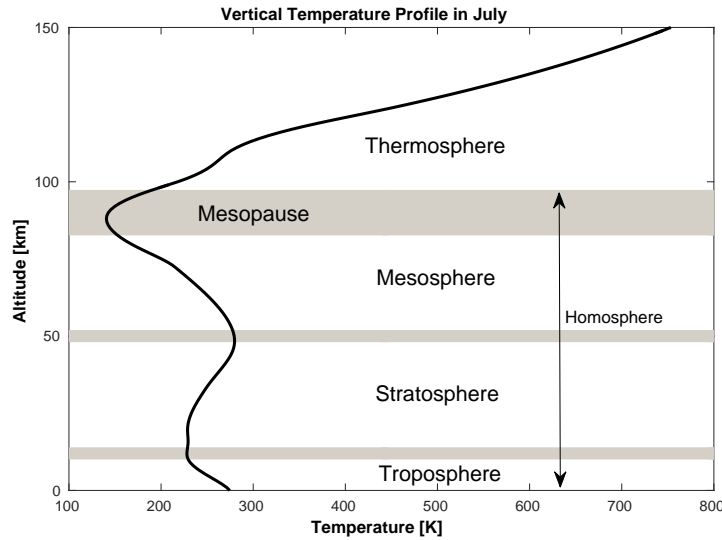
---

[1]from Greek 'mesos', middle

Figure 1.1: Divisions of the atmosphere based on vertical temperature gradient. The vertical temperature profile is obtained from MSISE90 model (Hedin (1991)) in July at the latitude of $80\,°N$

heating mechanisms, processes such as the transport of thermal energy by turbulent diffusion, heating due to chemical reactions as well as radiative cooling by carbon dioxide are all playing their crucial roles in the energy budget of this altitude region.

Another distinct feature in the mesosphere, besides the extreme temperature at the top, is the existence of airglow resulting from the active chemical interactions with the solar light in the ultraviolet and visible spectrum. Airglow is similar to aurora, both appear in the same altitude region and are linked to the chemical activity of a similar group of gases. However, in contrast to aurora, airglow corresponds to the light emitted by gases produced by photochemical reactions and is distributed around the globe rather than being confined to the polar regions. Moreover, the uneven structure of the airglow reflects variations of local temperature and chemical species. Thus, the analysis of the airglow distribution helps us understand the internal and external sources of disturbances such as atmospheric waves, seasonal and solar cycles.

The mesosphere is a challenging altitude region to probe by in-situ measurements. The air is too thin at that altitude and cannot provide enough buoyancy to lift an aircraft or even a balloon (commonly used to study the troposphere and stratosphere). On the other hand, the atmospheric friction is too significant for an orbital spacecraft to fly around. The only feasible way to directly access the altitude region is by sending sounding rockets, which allows taking measurements for a few minutes per mission. Thus, remote sensing plays an essential role in studying the mesosphere. The Swedish National Space Agency (SNSA) is involved in two satellite missions that support the scientific investigation of the mesosphere, namely Odin and MATS. The following sections provide a summary of the features of these two satellites. This thesis is centred around the data collected by one of the instruments onboard Odin as part of the preparatory work for the MATS mission.

## 1.2   Current observations with Odin

Since 2001 the Odin satellite has been orbiting Earth at ca. 600 km altitude, around 15 times per day (D. Murtagh et al., 2002) and is fully active to date. The main scientific objectives of Odin include middle atmospheric ozone and noctilucent clouds science and the coupling between the upper and lower atmosphere. The two main instruments aboard Odin are SMR (Sub-Millimetre Radiometer, measuring at 486-580 GHz and at 119 GHz) and OSIRIS (Optical Spectrograph and Infrared Imaging System, measuring at 274-810 nm, 1255-1275 nm and 1510-1550 nm, respectively). They measure the concentration of various species closely related to ozone chemistry such as NOx, CO, $H_2O$, ClO, $N_2O$, $HNO_3$ by observing the Earth's limb. Furthermore, OSIRIS consists of two optically independent instruments: the optical spectrograph (hereafter OS) and the infrared imager (IRI). IRI has three vertical channels. Two of them measure the oxygen infrared atmospheric (0-0)[2] band (IRA-band) emissions centred at 1.27 μm and one measures the OH Meinel (3-1) band emission centred at 1.53 μm. Two example orbits of the measured limb radiance by channels 1 and 3 are shown in Fig 1.2. Unfortunately, the data coverage of IRI has significantly decreased after 2015 due to a power supply unit problem of OSIRIS; thus the airglow measurements are highly limited, particularly after 2018.

Odin follows a near-terminator sun-synchronous orbit, in which the satellite passes over any given location at the same local time. For example, the local time is ca. 18h or 6h whenever Odin passes over the equator (i.e. ascending/descending node). Thus, data from such satellites are collected in relatively consistent illumination conditions. Moreover, daytime and nighttime measurements are separated by hemispheres and seasons. For instance, if one wants to study the daytime ozone concentration, only the data taken in the summer hemisphere are available for analysis, i.e. October - March in the Southern hemisphere and April - September in the Northern hemisphere. And since the local time of the ascending node (LTAN) of Odin is 18h (descending node at 6h), most of the data at low latitudes are collected near sunset or sunrise.

## 1.3   Future observations with MATS

Mesospheric Airglow/Aerosol Tomography and Spectroscopy (MATS) is a future satellite mission (Gumbel et al., 2020). The research satellite is scheduled to be launched in 2022 into a 600 km sun-synchronous orbit[3]. The main scientific objective of MATS is to determine the global distribution of gravity waves and other dynamical structures in the MLT region. Two instruments, a limb imager and a nadir imager, will be on board. Observations will primarily be done by the limb imager in six spectral channels, with two in the ultra-violet (UV) spectral region centred at 270 nm and 305 nm and four in the infrared spectral region centred at around 762 nm. These spectral channels are designed to target NLCs and oxygen airglow to study the dynamical structures, as the atmospheric waves can alter the homogeneity of NLCs

---

[2]refers to the vibrational levels

[3]However, due to the current geopolitical situation, the launch date and site are very uncertain.
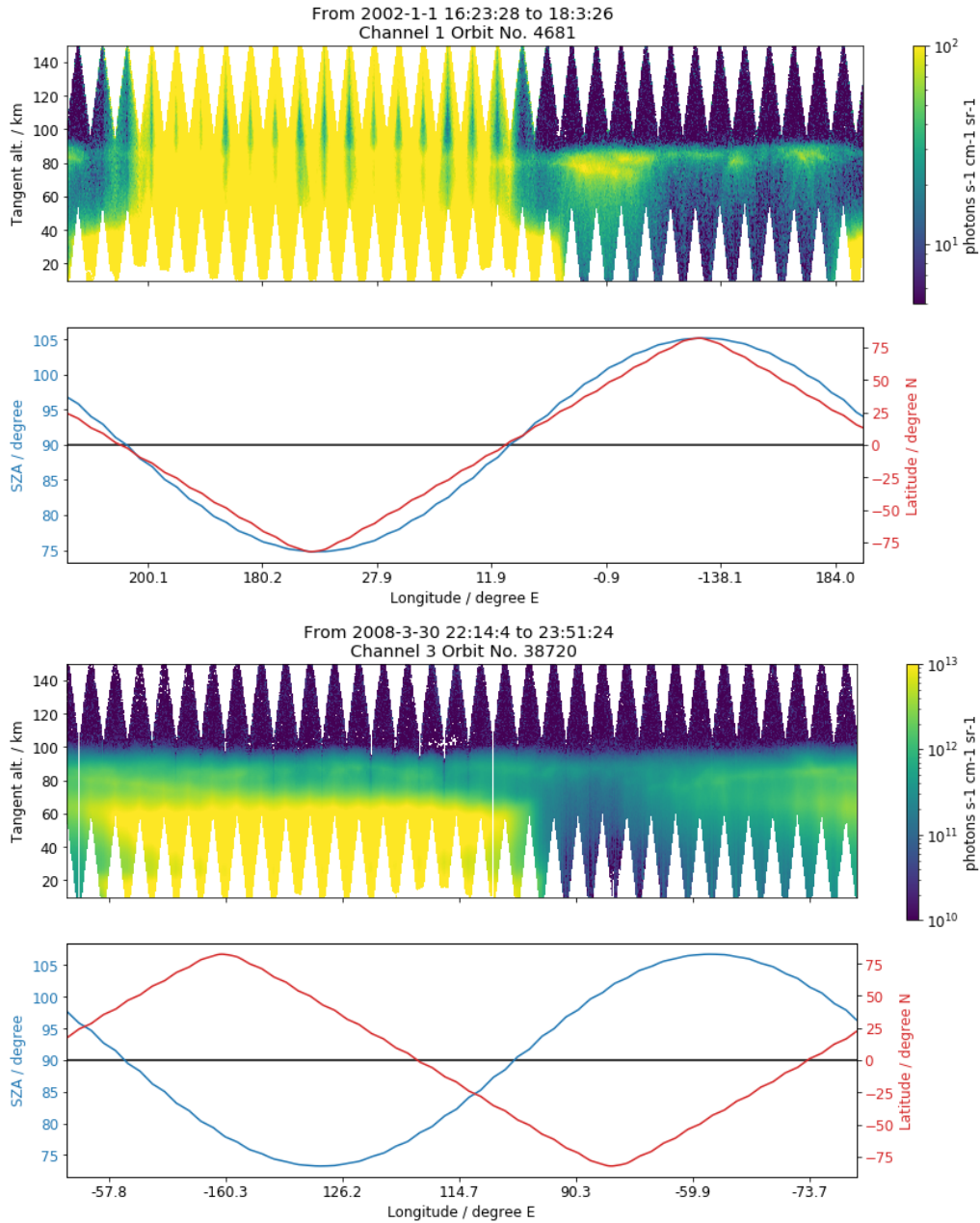
Figure 1.2: Two example orbits of the limb radiance measured at different times by Odin/OSIRIS infrared imager channel 1 (upper) and 3 (lower), the spectral band centred at $1.27\,\mu m$ and $1.53\,\mu m$, respectively. The red curves indicate the latitude coordinate at the tangent point, whereas the longitude coordinate is shown as the horizontal axes. The blue curves indicate the tangent point's solar zenith angle (SZA). The day part of the orbit corresponds to an SZA lower than 90°, and vice versa.
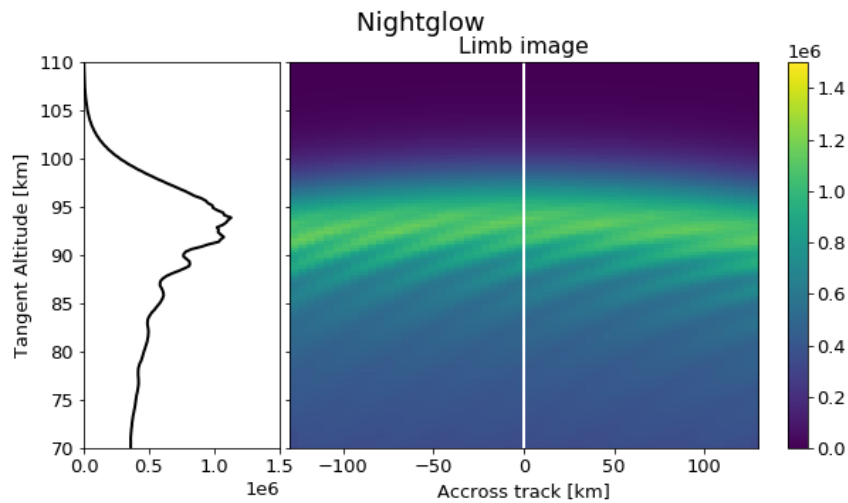
Figure 1.3: A virtual limb image taken from the MATS satellite. The simulated limb image is based on the viewing geometry of MATS measuring an atmospheric volume with a monochromatic wave presented under a background wind shear. The left panel shows the vertical profile of the limb radiance that corresponds to the white vertical line indicated on the limb image to the right. Values of the limb radiance in this figure are in an arbitrary unit. The model couples the $O_2(b^1\Sigma_g^+)$ photochemistry and gravity waves, and is described in Li (2017).

and airglow in A-band emission. Figure 1.3 shows a virtual limb image of MATS simulated by a model that couples the oxygen airglow photochemistry and gravity wave.

# Chapter 2

# Mesospheric dynamics

## 2.1 Introduction

As discussed in Sect. 1.1, dynamical processes can substantially modulate the thermal balance, which results in the extremely cold temperature in the summer mesopause. Dynamical patterns such as the zonal and meridional circulations are induced by atmospheric waves interacting with the mean flow. In many modelling studies, a 'Rayleigh friction coefficient' has to be introduced to account for the wave effect in order to produce a realistic thermal structure (Brasseur & Solomon, 2005). In turn, dynamical processes influence the observed tracers such as ozone and hydroxyl radicals.

Atmospheric waves are dynamic features initiated by a displacement from an equilibrium state caused by, among other things, the forcing of air parcels passing over obstacles such as mountains, frontal systems, etc. The oscillations are driven by a restoring force, like a spring force, that acts in opposition to the displacements. Once waves are excited near the surface of the Earth, they may propagate upwards and act as a coupling agent between the lower and higher atmosphere. Their amplitudes grow substantially in the mesosphere due to the decreasing air density (due to conservation of kinetic energy). When the amplitudes are so large that the restoring force can not be maintained, these waves eventually break and deposit their energy into the mean flow. The influence on the mean atmospheric state due to the breaking of these waves becomes significant at higher altitudes. In fact, the global circulation in the mesosphere is rather dynamically than thermally driven (Holton, 1982; Karlsson & Shepherd, 2018).

In this chapter, we will start by discussing the basic linear theory of gravity wave characterisation. Then, we will look at how the gravity waves play such an essential row in atmospheric dynamics in terms of zonal and meridional circulation.

## 2.2   Fundamental linear wave theory

Let us begin with a generic description of a sinusoidal wave form to represent a displacement field $\xi$

$$\xi = A \cos \phi \tag{2.1}$$

or in exponential notation

$$\xi = \Re(A e^{i\phi}), \tag{2.2}$$

where $\Re$ denotes the real part of a complex number, $A$ the amplitude of the wave and $\phi$ the phase angle which is a function of space and time

$$\phi(\vec{r}, t) = \vec{k} \cdot \vec{r} - \Omega t, \tag{2.3}$$

where $\vec{r}$ is the position vector, $t$ the time and $\vec{k}$ the wave vector, which points in the direction of the travelling wave and $\Omega$ is the wave frequency. Both $\vec{k}$ and $\Omega$ are fundamental properties of a wave. If we look at the space in the Cartesian coordinate system, i.e., $\vec{r} = (x, y, z)$, the wave vector can be decomposed into three components along each direction, i.e.,

$$\vec{k} = (k_x, k_y, k_z). \tag{2.4}$$

Conventionally, we often use $x$ and $y$ to denote the two directions on the horizontal plane and $z$ to denote the vertical direction. $k_x$ and $k_y$ sometimes can be composed into the 'total' horizontal component, $k_h$, if we focus on a 2 dimensional problem. Each component is termed the wavenumber, which can be considered $2\pi$ times the number of wave oscillations per unit length or wavelength per unit radian. Likewise, the wave frequency is $2\pi$ times the number of wave oscillations per unit time or radians per second. In other words,

$$(k_x, k_y, k_z, \Omega) = (\frac{2\pi}{\lambda_x}, \frac{2\pi}{\lambda_y}, \frac{2\pi}{\lambda_z}, \frac{2\pi}{\tau}), \tag{2.5}$$

where $\lambda_x$, $\lambda_y$ and $\lambda_z$ are wavelengths in each direction and $\tau$ the wave period. The phase speed of the wave, $c$, the speed at which constant phase moves in the direction of the travelling wave, can also be seen (and is measurable) along each direction, i.e.,

$$(c_x, c_y, c_z) = (\frac{\Omega}{k_x}, \frac{\Omega}{k_y}, \frac{\Omega}{k_z}) = (\frac{\lambda_x}{\tau}, \frac{\lambda_y}{\tau}, \frac{\lambda_z}{\tau}). \tag{2.6}$$

However, it is interesting to note that the phase speed of a wave is not a vector, but that the total phase speed follows the following relation

$$\frac{1}{c^2} = \frac{1}{c_x^2} + \frac{1}{c_y^2} + \frac{1}{c_z^2}. \tag{2.7}$$

For a given wave period, waves with long wavelength travel faster than those with short wavelength which leads to the wave *dispersion*. The *dispersion relation* connects the wave's spatial characteristics (wavenumbers) and the wave frequency. For internal gravity waves in the atmosphere, the dispersion relation reads

$$k_z^2 = \frac{(k_x^2 + k_y^2)(N^2 - \Omega^2)}{(\Omega^2 - f^2)} - \frac{1}{4H^2}, \tag{2.8}$$

or alternatively

$$\Omega^2 = N^2 \frac{k_x^2 + k_y^2}{k_x^2 + k_y^2 + k_z^2 + 1/(4H^2)} + f^2 \frac{k_z^2 + 1/(4H^2)}{k_x^2 + k_y^2 + k_z^2 + 1/(4H^2)}, \qquad (2.9)$$

where $N$ is the buoyancy frequency, also known as the Brunt-Väisälä frequency, $f$ the Coriolis parameter, and $H$ the scale height. This dispersion relation is derived from the fundamental fluid equations, which takes into account the buoyancy, the gravity, and the Coriolis force, but eliminating the viscosity and the acoustic property of the fluid (i.e., the air). Note that although the acoustic wave is eliminated, the compressibility term related to the atmospheric density gradient, the so-called '4H term', is retained in the expression. The detailed derivation can be found in e.g., Fritts and Alexander (2003), Holton (1982), and Nappo (2002). If the Coriolis parameter, $f$, is negligible compared to $\Omega$ and $N$, and the compressibility term is small, the dispersion relation can be further simplified to

$$k_z^2 = (k_x^2 + k_y^2)\frac{(N^2 - \Omega^2)}{\Omega^2} \qquad (2.10)$$

or alternatively

$$\Omega^2 = N^2 \frac{k_x^2 + k_y^2}{k_x^2 + k_y^2 + k_z^2}. \qquad (2.11)$$

The Brunt-Väisälä frequency, $N$, characterises the background atmosphere properties, specifically the vertical temperature gradient. This frequency is related to the difference between the *atmospheric lapse rate*, $\gamma_a$ and the *adiabatic lapse rate*, $\Gamma$

$$\begin{aligned} N^2 &= \frac{g}{T_a}\left(\frac{g}{c_p} - \frac{\partial T_a}{\partial z}\right) \\ &= \frac{g}{T_a}(\Gamma - \gamma_a), \end{aligned} \qquad (2.12)$$

where $T_a$ is the atmospheric temperature, $c_p$ the specific heat capacity at constant pressure and $g$ is the gravitational acceleration. The Brunt-Väisälä frequency represents the maximum frequency for vertically propagating gravity waves, that is when the air parcel is purely vertically displaced. In reality, a propagating gravity wave always has a direction at an angle to the vertical, so the air parcel is displaced at an angle to the vertical. Thus the frequency of the gravity wave is expressed as

$$\Omega = N \cos\beta. \qquad (2.13)$$

If one compares 2.13 to Eq. 2.11, one can easily find that $\beta$ is the angle between the wave vector and the horizontal plane.

Of course, the oscillation of the air parcel that has been mentioned above is under the assumption that $N$ is a real number, which is referred to as a *stably-stratified* environment. On the other hand, if $N$ is an imaginary number, i.e., $\gamma_a > \Gamma$, it is referred to as a *convective instability* in the atmosphere, which causes the air parcel to exhibit unbounded growth in response to the vertical displacement. Such a

condition may occur primarily when the wave amplitude has grown so large at high altitudes that the local temperature eventually exhibits a gradient greater than the adiabatic lapse rate, i.e., a *superadiabatic lapse rate.* At this point, wave breaking occurs and transfers the wave energy into turbulent kinetic energy. Likewise, the amplitude of a water surface wave on a coastline reaches a point where the crest of the wave overturns. Around the wave breaking region, the linear wave theory often becomes invalid to describe the dynamical behaviour and thus will not be further discussed in this thesis.

### 2.2.1   Interaction with the mean flow

Our discussions so far characterise the behaviour of internal gravity waves within the mean flow, as these waves are essentially the perturbations of the mean. We intuitively describe the airflow above the surface as 'moving air', or more commonly, wind. The so-called 'wind speed' and 'wind direction' are always measured relative to the surface ground. Therefore, the characterisation of gravity waves should also be studied in the frame of reference that we often refer to, i.e., the ground. Classically, we tend to think of how the background wind affects the behaviour of internal gravity waves. Though, these waves can not solely exist without the flow field and are bonded together with the background flow regardless of how we observe them (e.g., fixed on the ground, on a balloon, on an aircraft, etc.). In fact, the heading of this section could have been 'From Lagrangian[1] to Eulerian[2] frame of reference' instead.

   Let us begin the discussion with a constant background wind that is independent of space and time, i.e., a spatially uniform steady wind. A modification needs to be made on how we describe the phase angle as seen on the ground, thus Eq. 2.3 becomes

$$\phi(\vec{r}, t) = \vec{k} \cdot (\vec{r} - \vec{U}^w t) - \Omega t$$
$$= \vec{k} \cdot \vec{r} - (\vec{k} \cdot \vec{U}^w + \Omega)t \tag{2.14}$$

where $\vec{U}^w$ is the wind velocity. Similar to the wave vector $\vec{k}$, $\vec{U}^w$ can be decomposed into three components, $x$, $y$ and $z$ in the Cartesian coordinates

$$\vec{U}^w = (U_x^w, U_y^w, U_z^w). \tag{2.15}$$

From Eq. 2.14, we can replace $\Omega + \vec{U}^w \cdot \vec{k}$ with $\omega$, introducing the concept of the *ground-based frequency*, which is the frequency observed in the reference frame relative to the ground. The *intrinsic frequency* $\Omega$ is the frequency observed within the medium. The terminologies and symbols of these two frequencies are not always consistent between literature and should not be confused. Note that the intrinsic frequency of a wave shall not be interpreted as a constant, unchangeable property, even though the word intrinsic might imply this, or confused with the source frequency (an

---

[1]Lagrangian description is a description of the fluid motion by following an individual parcel, e.g., a wave packet in our case, as it travels through space and time.

[2]Eulerian description is a description of the fluid motion that fixes with locations in space through time. As in our case, we fixed to the ground and observed the wave.

example argument would be 'the intrinsic frequency is constant as long as the wave is generated by the same source'), which will be discussed further in the next section. The intrinsic frequency shall only be understood as the frequency we would observe as if we were moving with the mean flow (e.g., a balloon measurement observes $\Omega$).

$\Omega$ and $\omega$ are related by the Doppler relationship (Lighthill, 1967)

$$\Omega = \omega - \vec{U}^w \cdot \vec{k}, \tag{2.16}$$

where $\vec{U}^w \cdot \vec{k}$ is referred to the Doppler shift term (Buhler, 2009). Since it is a dot product between the wind vector and the wave vector, the actual wind component that affects the wave properties seen from the ground is the wind vector projected on the wave vector,

$$U_e^w = \vec{U}^w \cdot \vec{k}/|\vec{k}|, \tag{2.17}$$

where $U_e^w$ is termed the effective wind speed in this thesis.

In the upper atmosphere, the vertical component of the wind velocity can be safely assumed to be negligible compared to the horizontal component, i.e., the horizontal wind component is dominant. Thus, in the rest of this thesis, we will simplify the equations by retaining only the horizontal components of the background wind to study the mean flow interaction with waves. With this assumption, the Doppler relationship is simplified to

$$\begin{aligned} \Omega &= \omega - U_x^w k_x - U_y^w k_y \\ &= \omega - U_e^w k_h, \end{aligned} \tag{2.18}$$

where $U_e^w$ here is the wind component projected on the direction of the horizontal component of the wave vector, $k_h$.

It is worth emphasising that in the Doppler relationship, the ground-based frequency $\omega$, sometimes called the observed frequency, should be more precisely defined as the frequency observed relative to **the same reference frame in which $\vec{U}^w$ is measured**. This is crucial to keep in mind as we may encounter wave observations in a different reference frame than the ground (e.g., a moving spacecraft), and we are required to adequately account for the Doppler shift effect. The analysis becomes even more complicated when we also characterise the wavefield in the reference frame fixed to a moving source (e.g., a convection system), which was briefly summarised in the licentiate thesis (Li, 2020, Sect. 2.2.2).

So far, we have discussed the idealistic condition of constant background wind. The background wind is neither spatially uniform nor steady in the real world. To study the effect of the mean wind, one of the options, though more computationally expensive, is to use a first principle computational fluid dynamics model to simulate the waves in a complicated background situation. Another less expensive option is to apply the ray trace equations to analyse the wavenumbers and the ground-based frequency given the initial conditions (Heale & Snively, 2015; Marks & Eckermann,

1995). In accordance with Heale and Snively (2015), the ray trace equations read

$$
\begin{aligned}
\frac{d\vec{r}}{dt} &= \vec{C}_g \\
\frac{dk_h}{dt} &= -k_h \frac{\partial U_e^w}{\partial x} \\
\frac{dk_z}{dt} &= -k_h \frac{\partial U_e^w}{\partial z} \\
\frac{d\omega}{dt} &= k_h \frac{\partial U_e^w}{\partial t}
\end{aligned}
\tag{2.19}
$$

where $\vec{C}_g$ is the group velocity which indicates the velocity of a wave packet. An immediate interpretation of the ray trace equations above is that the characteristics of a gravity wave depend on the spatial gradients or rate of change in the effective background wind speed. For instance, a purely vertically varying background flow field modifies the $k_z$ of a wave as a function of $z$ and $t$, but the horizontal wavenumber and the ground-based frequency keep constant, and thus $\Omega$ should be modified due to the dispersion relation.

One of the most important effects of the spatially inhomogeneous mean flow is the channelling of the wave propagation path. Let us still focus on the case of vertically varying horizontal mean flow (or vertical shear flow), If the magnitude of the mean flow is equal to the ground-relative phase speed, i.e., $U_e^w = \frac{\omega}{k_h}$, the intrinsic phase speed $\frac{\Omega}{k_h}$ becomes zero due to the Doppler-shift effect (see Eq. 2.18) and the vertical wavelength $\lambda_z$ approaches zero (i.e., $k_z \to \infty$) due to the wave dispersion (see Eq. 2.10). When this happens, the gravity wave approaches the *critical level* where the wave can not propagate further up anymore. In contrast, when the intrinsic frequency is Doppler shifted by the mean wind to become close to the local buoyancy frequency of the medium (i.e., $\Omega \to N$), the vertical wavelength $\lambda_z$ approaches infinity (i.e., $k_z \to 0$) due to the wave dispersion (see Eq. 2.11). This layer is called the *reflection level* or the *evanescent level*, because the wave will be reflected and maybe partially transmitted through this layer. The analysis is similar in the case of a horizontal shear background, and we can refer to the modelling study done by Heale and Snively (2015).

The phenomenon of critical level can be clearly examined by a tank experiment shown in Fig. 2.1, where an internal gravity wave was generated by a corrugated wall moving under a tank of stratified fluid. The shadows visualise where the phase lines are within the fluid. The background flow field has a vertical shear which can be verified by the velocity profile superimposed on the photograph (note: measured in the coordinate system moving with the corrugated wall). We can see that the wave train can not penetrate through the critical level as the vertical wavelength becomes infinitely small. Since the vertical wavelength approaches zero $\lambda_z \to 0$, the vertical shear in the perturbation of horizontal flow speed will become so large that the Richardson number[3] becomes small, indicating turbulence production. The increasingly rapid oscillations near the critical region result in *dynamical instability*

---

[3]Richardson number is the dimensionless number that expresses the ratio of the buoyancy to flow shear.

that leads to wave breaking before reaching the theoretical critical level. Another similar tank experiment is shown in Fig. 2.2 in which an oscillating cylinder initiates the internal waves. The waves are generated on both sides of the cylinder, one propagating to the left and the other to the right in a two-dimensional flow field. Again, a critical level occurred on the left side for the same reason as in Fig. 2.1. A reflection level occurred for the wave on the right side due to its intrinsic frequency matched up with the background buoyancy frequency. The reflected wave train interferes with the original wave train, which can be studied in the shadow graph.
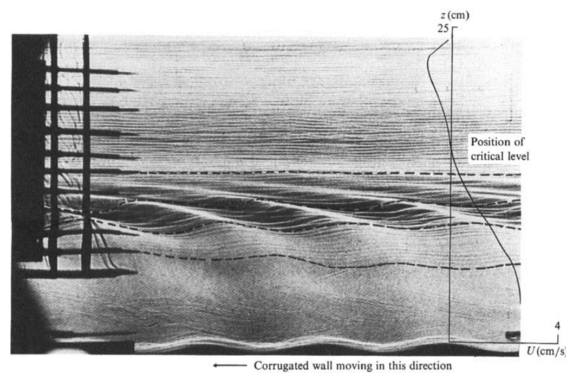


Figure 2.1: A shadowgraph taken from experiments in which an internal wave is generated by leftward-moving set of sinusoidal hills in a stratified shear flow, superimposed a velocity profile of the shear flow measured relative to the velocity of the corrugated wall. [Reproduced, by permission of Cambridge University Press, from Figure 5 in Koop and McGee (1986)]
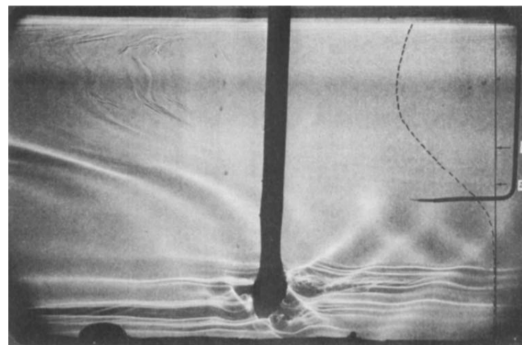


Figure 2.2: A shadowgraph taken from experiments in which internal waves generated by an oscillating cylinder in a leftward-shear flow, superimposed a velocity profile of the shear flow. The leftward-propagating wave encounters a critical level; the rightward-propagating wave encounters a reflection level. [Reproduced, by permission of Cambridge University Press, from Figure 10 in Koop (1981)]

The implications of the critical level and reflection level are significant in the context of atmospheric dynamics. The in-homogeneity of the background wind's zonal- and meridional components may lead to wave ducting, wave filtering, and many other exciting mechanisms. For example, when gravity waves are trapped between two reflection levels, they are prohibited from travelling vertically and may

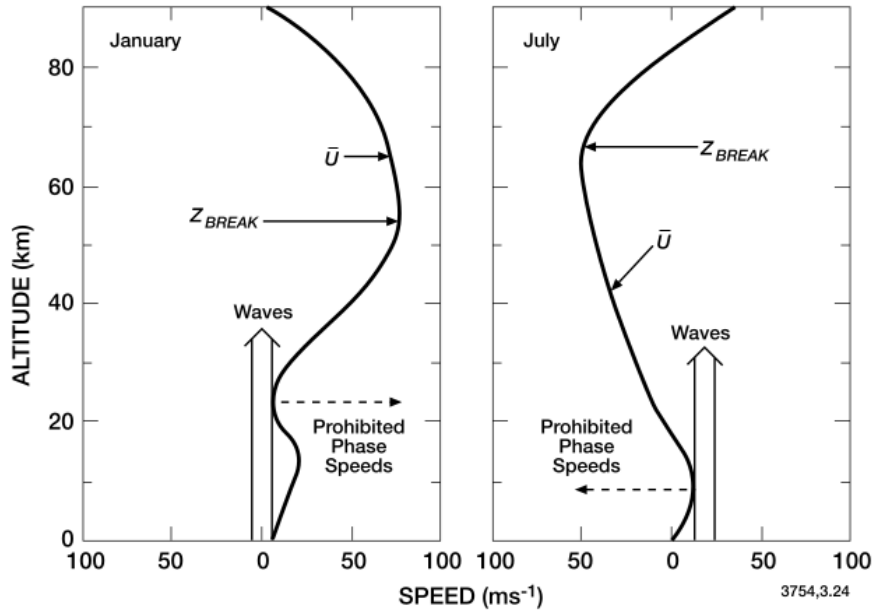Figure 2.3: A schematic illustration of the wave filtering effect by vertical shear wind in the atmosphere. Left: approximated vertical profile of the mean zonal winds in winter. Right: same as left but in summer. The prohibited and permitted phase speeds of the upward propagating gravity waves and their breaking levels are shown. [Reproduced, by permission of Springer Nature, from Figure 3.11 in Brasseur and Solomon (2005)]

propagate at a large horizontal distance before they break. This wave tunnelling mechanism may act as an 'information link' between non-contiguous geographic regions. Similarly, the critical level prohibits gravity waves from propagating upwards at a certain altitude.

The stratospheric zonal winds act like filters to absorb certain waves with the equivalent phase speeds, as illustrated in Fig. 2.3. A seasonal variation also incorporates the pattern of filtering since the stratospheric zonal wind changes its direction with seasons. The remaining waves that propagate higher generate 'wave drag' in the mesosphere, and that in turn produces a mesospheric mean flow pattern, which can be seen in various observations (e.g., Holton, 1982; Lindzen, 1981). The theoretical studies of wave filtering explain well the observed mesospheric wind pattern that could not be answered previously by pure thermodynamic studies.

To give a brief summary, the propagation path of the atmospheric internal wave depends on the spatial distributions of the background wind due to the Doppler effect and wave dispersion. In turn, the results of wave ducting and filtering influence where the wave dissipation occurs in the atmosphere, especially in the mesosphere. However, the origin and magnitude of the wave dissipation still remain an active topic for debate in recent studies (Fritts & Alexander, 2003; B. Kaifler et al., 2015; N. Kaifler et al., 2017; Vadas & Becker, 2018; Vadas et al., 2018).
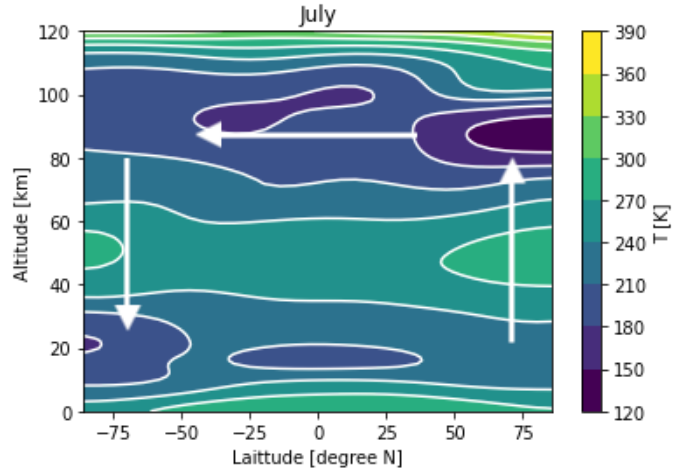
Figure 2.4: Altitude-latitude distribution of the zonal mean temperature in July. Data from the MSIS model. The arrows illustrate the pole-to-pole circulation.

## 2.3 Atmospheric circulation

In this section, we will discuss the meridional and zonal components of the global mesospheric circulation. In other words, the residual transport in the latitudinal and longitudinal directions, respectively. Moreover, we will look at their implications in the seasonal variations manifested in the observed variables such as temperature, ozone and hydroxyl radical distributions.

Fig. 2.4 displays a two-dimensional view of the zonal mean temperature distributions in July, as a function of latitude and altitude. We can clearly see that around 80-100 km, the summer pole (in this case, the north pole) is characterised by a much colder temperature than the winter pole, even though there is a positive net radiative heating in summer and a net cooling in winter. An air parcel displaced upward (decreasing pressure) will expand and cool down based on simple thermodynamic considerations. This mechanism is referred to as adiabatic cooling. Such low observed temperatures are therefore suggesting a large scale upward motion above the summer pole. In fact, the rising motion above the summer pole is part of the residual meridional transport. The pole-to-pole circulation up to the mesopause is driven by gravity waves breaking.

Since the pole-to-pole circulation is always from summer to winter, the seasonal variability occurs at high latitudes and oscillates annually. The signature of this annual oscillation is seen in many tracer studies such as the $O_2(a^1\Delta_g)$ and $OH_v$ airglows observed by IRI (see Paper 1 and 2), and CO, $H_2O$ and temperature observed by SMR (Grieco et al., 2020; Grieco et al., 2021).

The mean zonal wind structure in the stratosphere is eastward in the winter hemisphere and westward in the summer hemisphere due to the meridional temperature gradient (driven by radiative heating) balanced by Coriolis force (i.e. geostrophic balance). For example, when it is winter in the northern hemisphere, the pressure gradient points from south to north. The Coriolis effect forces the flow to turn right in the northern hemisphere and left in the southern hemisphere. Hence, the
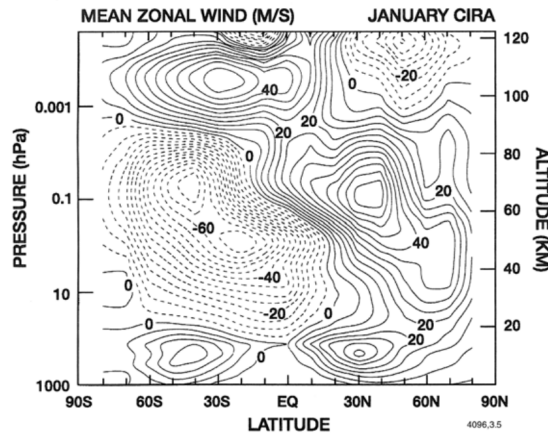
Figure 2.5: A typical averaged zonal wind distribution in January from CIRA. [Reproduced, by permission of Springer Nature, from Figure 3.5 in Brasseur and Solomon (2005)]

prevailing wind is eastward in the northern and westward in the southern hemisphere. A similar analysis can be made for the opposite case, corresponding to the northern summer season. A typical mean zonal wind distribution over latitude and altitude is displayed in Fig. 2.5.

As discussed in Sect. 2.2, the mesospheric prevailing zonal wind is weakened or in the opposite direction to the wind in the stratosphere, as a result of the gravity wave critical level filtering and breaking mechanism. The stratospheric zonal wind only allows the gravity waves with the opposite direction in their phase velocities to penetrate a higher altitude level. When these waves break in the mesosphere, momentum is dissipated into the mean flow resulting in a 'wave drag' decelerating the stratospheric jets and even reversing the prevailing directions above the mesopause. Such 'wave drag' is then balanced by the Coriolis force resulting in the pole-to-pole meridional transport as discussed earlier in this section. In effect, the wave filtering mechanism explains well the observed zonal wind and the thermal structure in the mesosphere, which a pure radiative balance assumption can otherwise not explain.

Besides the annual variation of the wind in the upper atmosphere, gravity wave filtering also plays a crucial role in the semi-annual oscillation (SAO) in the tropical region. Such an oscillation can be readily observed in the distributions of chemical compounds such as the OH airglow shown in Paper 2. In addition, Smith et al. (2017) showed the observed mesospheric zonal wind at the equator. Above the stratopause, the westerly (lower mesosphere) and easterly (upper mesosphere) wind maximise at the two equinoxes.

In Fig. 2.5, an eastward wind near the winter pole in the stratosphere is noticeable, which is known as the polar night jet or the polar vortex. Occasionally, large planetary-scale wave breaking decelerates or even reverses the vortex's direction. When such an event occurs, the stratosphere becomes warmer than usual (sudden stratospheric warming, or SSW), accompanied by mesospheric cooling. The implication of such dynamical change in the stratosphere is significant even up to the mesopause region

where the airglow emissions are observed (see Paper 4). However, planetary wave activity is relatively lower and the polar vortex appears to be more robust in the southern hemisphere. Thus, the ozone hole often appears in the early springtime in Antarctica as a result of the strong polar vortex acting as a barrier preventing horizontal transport.

To conclude this section, the large scale circulation in the stratosphere is mainly determined by the thermal structure, whereas it is induced by the gravity wave filtering mechanism in the mesosphere. The wave-mean flow interaction is so crucial that radiative balance alone can not otherwise explain the observed wind and temperature structures in the mesosphere. When, occasionally, the strong polar vortex is perturbed by planetary-scale waves, a sudden stratospheric warming occurs and can have large impacts on the middle atmosphere.

# Chapter 3

# Mesospheric photochemistry

## 3.1 Introduction

Almost all atoms and molecules in the mesosphere participate in chemical processes, which result in their spatial and temporal distributions being closely linked to one another. Processes that involve the interaction with light (photons) are called photochemical processes. In the field of aeronomy, the source of light we are typically concerned with is the sun. The ultra-violet photons in the solar spectrum have sufficient energy to dissociate molecules (i.e., photo-dissociation) or may excite molecules from a ground state to an energetic state (i.e., photo-excitation). These excited molecules are often unstable and spontaneously release their energy as photons (i.e., fluorescence) or by colliding with other molecules to transfer this energy (e.g., intermolecular energy transfer, quenching). Ozone, atomic- and molecular oxygen are typical examples of constituents in the mesosphere that are constantly dealing with these types of energy transfer. Thanks to these photochemical processes that act as a protecting layer for life on Earth, the most powerful solar wavelengths are mostly attenuated before reaching the lower atmosphere.

Among the aforementioned photochemical processes, the light that is spontaneously emitted by the electronically excited chemical species in the atmosphere is called airglow. For instance, Fig. 3.1 shows a thin greenish layer on the night side of Earth's atmosphere that human eyes can observe from space. Depending on whether the solar energy is available or not when observed, airglow is called either dayglow or nightglow. Common airglow emissions originate from excited OH, NO, Na, Li and $O_2$. This thesis deals with airglow emissions from $O_2$ at wavelengths around $1.27\,\mu m$ and from OH at $1.53\,\mu m$ in the mesopause region.

One of the objectives of measuring airglow emissions is to reveal the many dynamical features present in the atmosphere. Because the kinetics of the airglow chemistry (described in Sect. 3.3) involves numerous reaction rates that are sensitive to temperature and to the density of various species, any fluctuation in these parameters will affect the distributions of the excited species. In turn, the observed airglow emissions often exhibit a 'wavy' pattern, such as the observation shown in Fig. 3.2. Therefore, observing the airglow layer can be an essential tool for studying the morphology and distribution of gravity waves in the mesopause region.

Figure 3.1: Airglow above the horizon captured from the International Space Station. Image from NASA.



Figure 3.2: Airglow exhibits in a stripped pattern observed from the ground over Maine, USA. [Permitted by Mike Taylor and Sonia MacNeil]

Many other dynamical events also affect the airglow distribution both on short and long time scales, such as the sudden stratospheric warming (SSW) events and the pole-to-pole residual circulation.

Moreover, the oxygen airglow is closely connected to the photodissociation of ozone in the mesosphere. Typically, the ozone concentration is measured by observing thermal emission, such as in the spectral region around $9.6\,\mu m$ and in the microwave region, or by measuring absorption in the UV-visible region, such as limb-scatter technique and solar/stellar occultation techniques. However, ozone density in the mesosphere changes drastically from nighttime to daytime. During the day, mesospheric ozone is as little as only $10\,\%$ of the maximum values observed in the night. In addition, the total density in the mesosphere is much lower than in the stratosphere. In turn, assessing daytime ozone density in the mesosphere requires instruments with high sensitivity. Fortunately, as previously mentioned, oxygen airglow serves as a valuable tool to evaluate ozone density indirectly.

This chapter will start by listing the basic chemical kinetics concepts and some of the terminology in photochemistry. After that, we will discuss the kinetic models that describe the oxygen and hydroxyl airglow photochemistry. The kinetic models are the foundation for uncovering the critical linkage between ozone, atomic oxygen, and the observed airglow species.

## 3.2 Fundamental chemical kinetics

Almost all of the minor constituents in the atmosphere are created and destroyed by chemical processes. The concentration of each chemical species highly depends on the rates at which they are produced and removed. These reaction rates describing the kinetics must be determined in the laboratory, typically by recording the concentration of the reactants or products as a function of time in controlled background conditions. A discussion about the measurement techniques and the associated uncertainties of these measured parameters is out of the scope of this thesis. This section aims at providing necessary information to understand some of the important nomenclature, as well as how a photochemical model is implemented such as in Paper 1 and 4.

### 3.2.1 The steady state assumption

Let us begin with a general expression of a reaction between species A, B, C,...,

$$a\,\mathrm{A} + b\,\mathrm{B} + c\,\mathrm{C} + ... \xrightarrow{k} e\,\mathrm{E} + f\,\mathrm{F} + g\,\mathrm{G} + ... \tag{3.1}$$
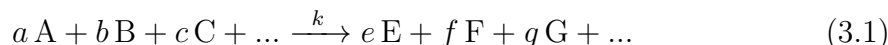
where lower case letter represent integers that indicate the stoichiometry and $k$ is the *rate coefficient* or *rate constant* for this particular reaction, which often has a dependency on the local temperature.

The rate of change of a particular reactant or of a particular product is expressed as
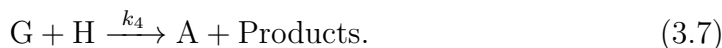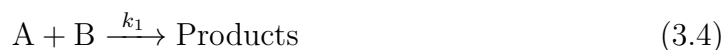
$$R = -\frac{1}{a}\frac{\mathrm{d[A]}}{\mathrm{d}t} = -\frac{1}{b}\frac{\mathrm{d[B]}}{\mathrm{d}t} = ... = \frac{1}{e}\frac{\mathrm{d[E]}}{\mathrm{d}t} = \frac{1}{f}\frac{\mathrm{d[F]}}{\mathrm{d}t} = ... \tag{3.2}$$

The reaction rate depends on the concentrations of reactants so the rate equation often, but not always, takes the form of

$$R = k[\mathrm{A}]^a[\mathrm{B}]^b[\mathrm{C}]^c... \tag{3.3}$$

where square brackets denote the concentration of species.

In an environment like the atmosphere, where many species are reacting with each other simultaneously, chemical reactions often occur through an interconnected network called *reaction mechanism*. In such case, the rate of change of a certain species must consider all the possible production and loss processes. For example, a reaction mechanism represented by

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_1} \mathrm{Products} \tag{3.4}$$

$$\mathrm{A} + \mathrm{C} + \mathrm{M} \xrightarrow{k_2} \mathrm{Products} \tag{3.5}$$

$$\mathrm{A} + \mathrm{F} \xrightarrow{k_3} \mathrm{Products} \tag{3.6}$$

$$\mathrm{G} + \mathrm{H} \xrightarrow{k_4} \mathrm{A} + \mathrm{Products}. \tag{3.7}$$

If we focus on the species A, the sum of all the production terms (also called sources) $P_i$ is

$$\sum_i P_i = k_4[\mathrm{G}][\mathrm{H}], \tag{3.8}$$

and the sum of all the losses (also called sinks) of species A, noted $L_i[A]$, is

$$\sum_i L_i[A] = (k_1[B] + k_2[C][M] + k_3[F])[A]. \tag{3.9}$$

The rate of change of species A results in

$$\frac{d[A]}{dt} = \sum_i P_i - \sum_i L_i[A] \tag{3.10}$$

Specifically, the lifetime of A is the inverse of the losses sum $\tau \equiv 1/\sum_i L_i$.

When the density of A is not changing over time, i.e., $\frac{d[A]}{dt} = 0$ , the constituent is said to be in *steady state* or *equilibrium*. This means that whenever A is newly produced, a similar amount will be removed by other reactions of the network so that A will neither accumulate or dissipate in the system. Thus the concentration of A under the steady state assumption can be derived as

$$[A]_{equi} = \frac{\sum_i P_i}{\sum_i L_i} = \frac{k_4[G][H]}{k_1[B] + k_2[C][M] + k_3[F]} \tag{3.11}$$

In a chemical model which includes multiple reaction mechanisms, such as the one described in the next two sections, the steady state assumption is particularly useful to derive the concentrations of all species considered in the model. However, if the time scale of the dynamics in a certain location is comparable to the lifetime of a certain species, the effect of transport must be taken into account in the balance equation (i.e., the continuity equation)

$$\frac{d[A]}{dt} = \sum_i P_i - \sum_i L_i[A] - \nabla \cdot ([A]\vec{v}) \tag{3.12}$$

where the last term added represents the advection effect on the species A by the mass flow which has a velocity $\vec{v}$. We will not discuss further the details about the effect of transport in this thesis, but shall be aware of the bias introduced by considering the chemical steady state assumption only.

When integrating the differential equation Eq. 3.10, the concentration of A in a given time is expressed as

$$[A](t) = (1 - \exp(-\frac{t}{\tau}))[A]_{equi}, \tag{3.13}$$

where $[A]_{equi}$ is the concentration of A in equilibrium (i.e. Eq. 3.11). In the expression above, we understand that the ratio $\frac{t}{\tau}$ reflects how far the current $[A]$ has reached to the final equilibrium level, i.e. $[A](t)/[A]_{equi}$, in response to an instantaneous change in the system. For instance, when $[A]$ has reached 80%, 90%, 95%, 98% and 99% of the equilibrium level, $\frac{t}{\tau}$ takes the values of 1.6, 2.3, 3, 4 or 4.6. In Paper 1, we incorporate the factor $(1 - \exp(-\frac{t}{\tau}))$ as an equilibrium index to the retrieval scheme and take $\frac{t}{\tau} \geq 3$ as a recommended screening criterion to address the issue associated with long lifetime species $O_2(a^1\Delta_g)$.

## 3.2.2 Photochemical processes

A photochemical reaction is defined as a chemical reaction involving the interaction with electromagnetic waves. The reactants and products of the processes are often excited states (which was previously referred to having "excess energy" in the introductory section of the chapter). Depending on the amount of energy that is absorbed, the excited state can be electronically, vibrationally and rotationally excited. In this thesis, we will mainly consider electronically and, to a lesser extend, vibrationally excited states.

As briefly mentioned earlier in the introduction section, photo-excitation can be the first step of a chain of photochemical processes. This is also termed *resonance absorption*, or in the context of the atmosphere, *solar-excitation* since the sun is considered as the only energy source. As an example, the reaction of the solar-excitation of molecule XY takes the form of

$$XY + h\nu \xrightarrow{g} XY^* \tag{3.14}$$

where the superscript $^*$ denotes that the molecule is in an excited state, $g$ is the excitation rate which has a unit of $s^{-1}$. If the associated photon energy $h\nu$ (Planck's constant times the photon frequency) is sufficiently high, the molecule XY may fall apart to form multiple constituents which is referred to as *photo-dissociation*, or *photolysis* process, represented by,

$$XY + h\nu \xrightarrow{J} X + Y^* \tag{3.15}$$

note that either, both or neither of the products can be in an excited state. Here, $J$ is termed photolysis rate. Since the reaction only involves one reactant, the photolysis rate has a unit of $s^{-1}$.

Both excitation rate $g$ and photolysis rate $J$ are determined by how many photons are available to interact with the gas, the ability of the particular species to absorb these photons and the probability that the photon leads to the specific products of the photochemical reaction. This means that to compute these rates, we need radiative transfer modelling along the path from the light source to the location of interest for each wavelength. In other words, the solar irradiance, the absorption cross-sections of the species and the quantum yield for the particular photo-dissociation are needed to proceed with the computation of $g$ and $J$.

After the initiation of an excited molecule, it may transfer its excess energy to another molecule by collision, which is called the *inter-molecular energy transfer* process, i.e.,

$$XY^* + AB \xrightarrow{k} XY + AB^* \tag{3.16}$$
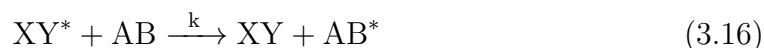
or may collide with another molecule, losing its energy as heat/kinetic energy, and simply returns back to its basic ground state, which is referred to as the *quenching* process, i.e.,

$$XY^* + M \xrightarrow{k} XY + M. \tag{3.17}$$

Another way of returning back to the basic ground state without collision is the *fluorescence* or *chemiluminescence* process (naming depends on how the molecule

was excited). That consists in the spontaneous release of the photon energy at a specific spectral characteristic, i.e.,

$$XY^* \xrightarrow{\ A\ } XY + h\nu, \tag{3.18}$$

where $A$ is termed the Einstein A coefficient, as known as the inverse of radiative lifetime, i.e., $A = 1/\tau_{XY^*}$. Similarly to the reaction coefficients $k$, the Einstein A coefficient must be obtained by laboratory measurements. This spontaneous relaxation of the excited state is exactly the reason why we can observe the airglow by measuring the released photon energy $h\nu$ in a specific spectral region. This emitted photon energy may be re-absorbed to repeat the excitation process, which is commonly referred to as *self-absorption*.

In summary, we have mentioned some fundamental chemical kinetic concepts that are important for implementing a photochemical model, as well as introduced most of the nomenclature that will be used in Sect. 3.3 where we discuss the photochemical model that describes the relationship between ozone and several excited states of oxygen in the mesosphere.

## 3.3   Oxygen airglow

As mentioned in Chapter 1, the Odin satellite includes the OSIRIS instrument onboard. The optical spectrograph (hereafter OS) routinely measures the oxygen atmospheric band (A-band) emission centred at 762 nm, and the infrared imager (hereafter IRI) measures the oxygen infrared atmospheric band (IRA-band) intensity centred at 1.27 µm. In addition, the future satellite MATS will also capture the A-band emission. Oxygen A-band and IRA-band emissions are the conventional names for the (0-0) band of airglow emissions by the two electronically excited states of the oxygen molecule, $O_2(b^1\Sigma_g{}^+)$ and $O_2(a^1\Delta_g)$, respectively. Both of them are closely connected to the available ozone amount in the mesosphere under sunlit conditions, since they are the 'by-products' of the ozone photolysis in the Hartley band (around 310 nm to 350 nm, see Fig. 3.4). For this reason, the observed intensities of these two oxygen band emissions during the day can often be utilised as proxies for the ozone density in the mesosphere (Mlynczak et al., 2001; Mlynczak et al., 2007; P. Sheese, 2009; Thomas et al., 1984; Yankovsky et al., 2016). In this section, we will investigate the most important reactions that connect ozone and the several electronically excited states of the oxygen atom and molecule. The addition of their vibrational-rotational sub-levels can be referred to the model in Yankovsky and Manuilova (2006) and Yankovsky et al. (2016) and in Paper 1. The kinetic scheme of the photochemical model considered here is illustrated in Fig. 3.3.

The reaction chain starts by ozone molecules being photo-dissociated by the solar energy in the Hartley band (see Fig. 3.4 for the spectral range)

$$O_3 + h\nu \xrightarrow{\ J_H\ } O(^1D) + O_2(a^1\Delta_g), \tag{3.19}$$
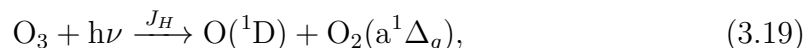
where $J_H$ is the photolysis rate in this spectral range. While the theoretical combination of the quantum levels of the products (oxygen atom and molecular oxygen)
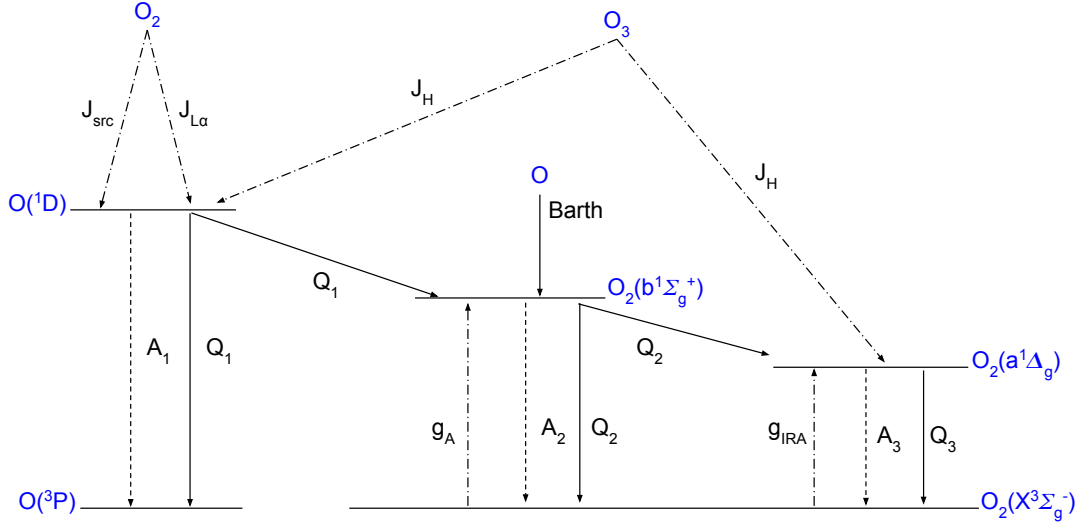
Figure 3.3: Scheme of kinetics that is considered in this section. Dashed-dotted lines represent either photolysis or resonance absorption, dashed lines represent the spontaneous de-excitation processes, and solid lines represent quenching processes.

of this particular photochemical reaction depends on the wavelength of the incident photon energy, $O(^1D)$ and $O_2(a^1\Delta_g)$ are the most probable products and have a quantum yield of $90\%$ based on a laboratory experiment. Besides the Hartley band, other absorption bands of ozone are in longer wavelength regions, such as Huggins bands and Chappuis bands, as shown in Fig. 3.4. Although the Huggins bands and Chappuis bands are not in the focus on our airglow photochemistry, it is worth mentioning that they are routinely measured by the Odin mission (more specifically, the optical spectrograph OSIRIS).

Another important source of the excited oxygen atom $O(^1D)$, besides the photolysis of ozone, comes from the photo-dissociation of molecular oxygen

$$O_2 + h\nu \xrightarrow{J_{src}, J_{L\alpha}} O + O(^1D), \tag{3.20}$$
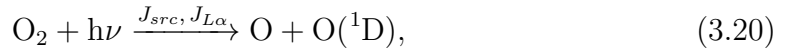
where $J_{src}$ and $J_{L\alpha}$ are the photolysis rates corresponding to the Schumann-Runge Continuum and Lyman $\alpha$ line, respectively (see Fig.3.5 for spectral range). Here, chemical species without brackets denote their ground states, for instance $O_2$ is equivalent to $O_2(X^3\Sigma_g^-)$ and O is $O(^3P)$. The absorption cross section of molecular oxygen is shown in Fig. 3.5.

The excited atomic oxygen $O(^1D)$ then either spontaneously releases its energy or transfer its energy by quenching processes

$$O(^1D) \xrightarrow{A_1} O + h\nu \tag{3.21}$$

$$O(^1D) + N_2, O_2 \xrightarrow{k_1^{N_2}, k_1^{O_2}} O + N_2, O_2 \tag{3.22}$$
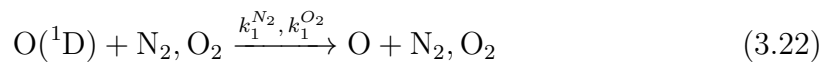
where subscripts of the reaction rate coefficients corresponds to the quenching series represented in Fig. 3.3, likewise for the Einstein $A$ coefficient. The superscripts
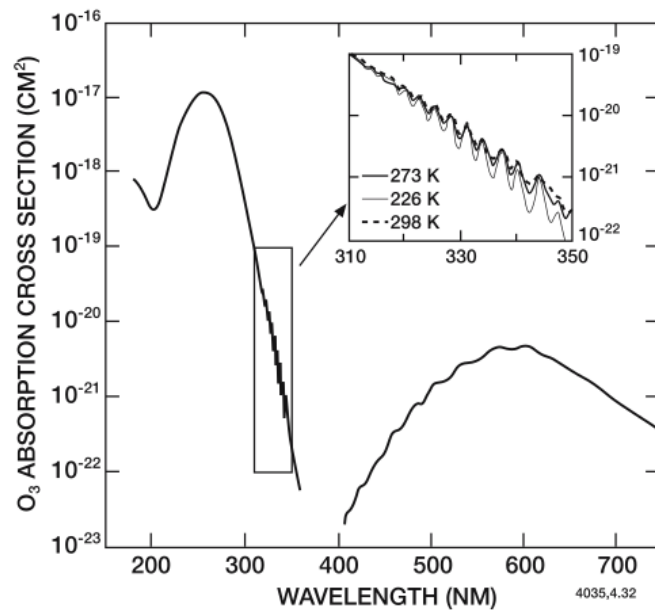
Figure 3.4: The ozone cross section in the Hartley band (200 nm to 300 nm), Huggins bands (310 nm to 350 nm) and Chappuis bands (410 nm to 750 nm). [Reproduced, by permission of Springer Nature, from Figure 4.35 in Brasseur and Solomon (2005)]
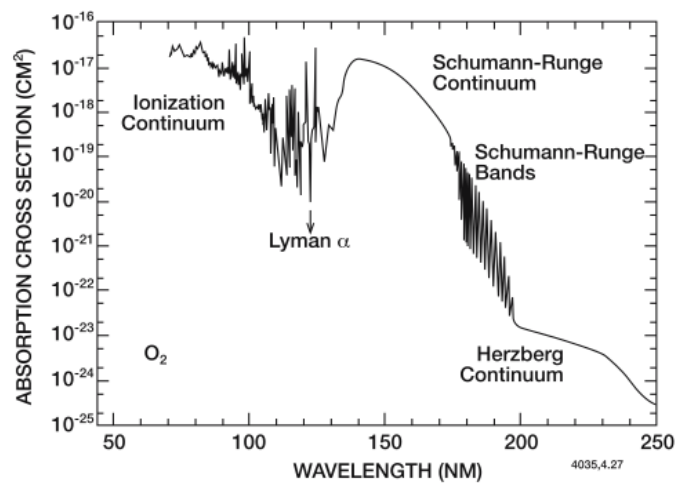


Figure 3.5: The molecular oxygen cross section. [Reproduced, by permission of Springer Nature, from Figure 4.30 in Brasseur and Solomon (2005)]

denote the rate coefficients corresponding to the quenching by the species. Hence, the concentration of O($^1$D) can be derived by assuming the sources and the sinks are balanced (i.e. photochemical steady-state), i.e.,

$$[O(^1D)] = \frac{(\phi_{src}J_{src} + \phi_{L\alpha}J_{L\alpha})[O_2] + \phi_H J_H[O_3]}{A_1 + Q_1}, \tag{3.23}$$

where $\phi_{src}$, $\phi_{L\alpha}$ and $\phi_H$ denote the quantum yields of each of the photolysis processes, and

$$Q_1 = k_1^{N_2}[N_2] + k_1^{O2}[O_2]. \tag{3.24}$$

Among the two quenching processes represented in reaction 3.22, the most critical reaction in our kinetic scheme is the collision with $O_2$ that partially forms the excited oxygen molecule $O_2(b^1\Sigma_g^+)$, i.e.,
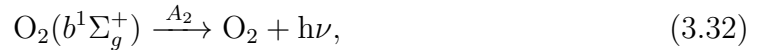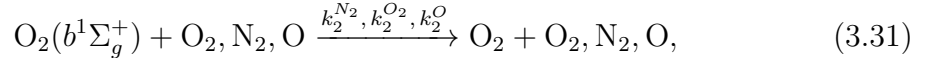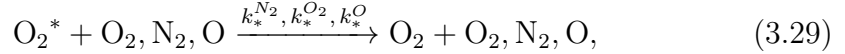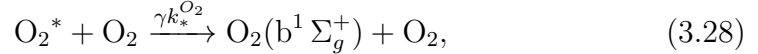
$$O(^1D) + O_2 \xrightarrow{\beta k_1^{O_2}} O + O_2(b^1\Sigma_g^+), \tag{3.25}$$
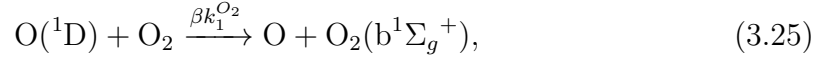
where $\beta$ represents the fractional efficiency of this product path. Another important source of $O_2(b^1\Sigma_g^+)$ is the solar-excitation at the A-band wavelengths of the oxygen molecule itself, i.e.,

$$O_2 + h\nu \xrightarrow{g_A} O_2(b^1\Sigma_g^+), \tag{3.26}$$

where $g_A$ is the A-band excitation rate.

A small contribution to the production of $O_2(b^1\Sigma_g^+)$ is by the two-step transfer mechanism, as known as the Barth-type mechanism, represented by (after McDade et al. (1986))

$$O + O + M \xrightarrow{\alpha k^O} O_2^* + M \tag{3.27}$$

$$O_2^* + O_2 \xrightarrow{\gamma k_*^{O_2}} O_2(b^1\Sigma_g^+) + O_2, \tag{3.28}$$

$$O_2^* + O_2, N_2, O \xrightarrow{k_*^{N_2}, k_*^{O_2}, k_*^O} O_2 + O_2, N_2, O, \tag{3.29}$$

$$O_2^* \xrightarrow{A_*} O_2 + h\nu, \tag{3.30}$$

$$O_2(b^1\Sigma_g^+) + O_2, N_2, O \xrightarrow{k_2^{N_2}, k_2^{O_2}, k_2^O} O_2 + O_2, N_2, O, \tag{3.31}$$

$$O_2(b^1\Sigma_g^+) \xrightarrow{A_2} O_2 + h\nu, \tag{3.32}$$

where $\alpha$ and $\gamma$ represent the fractional efficiency of the three-body recombination into the unspecified excited state $O_2^*$ and from this state to $O_2(b^1\Sigma_g^+)$, respectively. The subscript $_*$ denotes the Einstein A coefficient and the quenching rate coefficients of the unspecified excited state $O_2^*$. In photochemical equilibrium, the concentration of $O_2(b^1\Sigma_g^+)$ solely resulting from the Barth-type mechanism can be written as

$$[O_2(b^1\Sigma_g^+)]_{Barth} = \frac{\alpha k^O[O]^2[M]\gamma k_*^{O_2}[O_2]}{(A_2 + Q_2) \cdot (A_* + Q_*)}, \tag{3.33}$$

where

$$\begin{aligned} Q_2 &= k_2^{O_2}[O_2] + k_2^{N_2}[N_2] + k_2^O[O], \\ Q_* &= k_*^{O_2}[O_2] + k_*^{N_2}[N_2] + k_*^O[O]. \end{aligned} \tag{3.34}$$

However, $\alpha, \gamma, A_*$ and $k_*$ are quantities that are often unknown and difficult to measure, thus introducing empirical quenching coefficients can simplify the problem as (D. P. Murtagh et al., 1990)

$$[O_2(b^1\Sigma_g^+)]_{Barth} = \frac{k_1[O]^2\,[M]\,[O_2]}{(Q_2 + A_2)\cdot(C^{O_2}[O_2] + C^O[O])}, \qquad (3.35)$$

where

$$\begin{aligned} C^{O_2} &= \frac{1 + Rk_*^{N_2}/k_*^{O_2}}{\alpha\gamma}, \\ C^O &= \frac{k_*^O/k_*^{O_2}}{\alpha\gamma}. \end{aligned} \qquad (3.36)$$

These newly introduced empirical quenching coefficients $C^{O_2}$ and $C^O$ were evaluated by rocket measurements of nightglow emissions (D. P. Murtagh et al., 1990). Although the Barth-type mechanism is not a significant source compared to other sources to form $O_2(b^1\Sigma_g^+)$, it is special in a way that it does not involve absorption of solar radiation, and thus becomes the only source in the nighttime. Nevertheless, summing up all the sources, the concentration of $O_2(b^1\Sigma_g^+)$ can be derived as
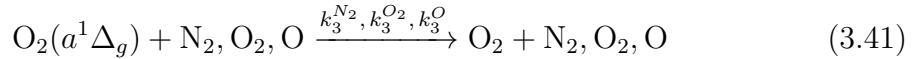
$$[O_2(b^1\Sigma_g^+)] = \frac{\beta k_1^{O_2}[O(^1D)][O_2]}{A_2 + Q_2} + \frac{g_A[O_2]}{A_2 + Q_2} + [O_2(b^1\Sigma_g^+)]_{Barth} \qquad (3.37)$$

The quenching processes of $O_2(b^1\Sigma_g^+)$ upon collisions (in reaction 3.31) result in the production of $O_2(a^1\Delta_g)$, e.g.,

$$O_2(b^1\Sigma_g^+) + N_2, O_2, O \xrightarrow{\epsilon k_2^{N_2}, \epsilon k_2^{O_2}, \epsilon k_2^O} O_2(a^1\Delta_g) + N_2, O_2, O, \qquad (3.38)$$
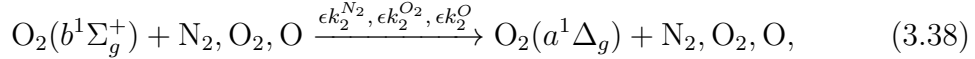
where $\epsilon$ is the fractional efficiency of the quenching processes that produce $O_2(a^1\Delta_g)$.

Similar to $O_2(b^1\Sigma_g^+)$, $O_2(a^1\Delta_g)$ can also be produced by resonant absorption from the ground state, removed by spontaneous emission and quenching processes, i.e.,

$$O_2 + h\nu \xrightarrow{g_{IRA}} O_2(a^1\Delta_g), \qquad (3.39)$$

$$O_2(a^1\Delta_g) \xrightarrow{A_3} O_2 + h\nu, \qquad (3.40)$$

$$O_2(a^1\Delta_g) + N_2, O_2, O \xrightarrow{k_3^{N_2}, k_3^{O_2}, k_3^O} O_2 + N_2, O_2, O \qquad (3.41)$$

Together with the direct production from ozone photolysis in the Hartley band, the concentration of $O_2(a^1\Delta_g)$ under photochemical steady state assumption can be derived as

$$[O_2(a^1\Delta_g)] = \frac{g_{IRA}[O_2] + \epsilon Q_2[O_2(b^1\Sigma_g^+)] + \phi_H J_H[O_3]}{A_3 + Q_3}, \qquad (3.42)$$

where

$$Q_3 = k_3^{O_2}[O_2] + k_3^{N_2}[N_2] + k_3^O[O]. \qquad (3.43)$$

In a photochemical model such as described above, the concentrations of $O(^1D)$, $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$ can be calculated, given the prescribed number density
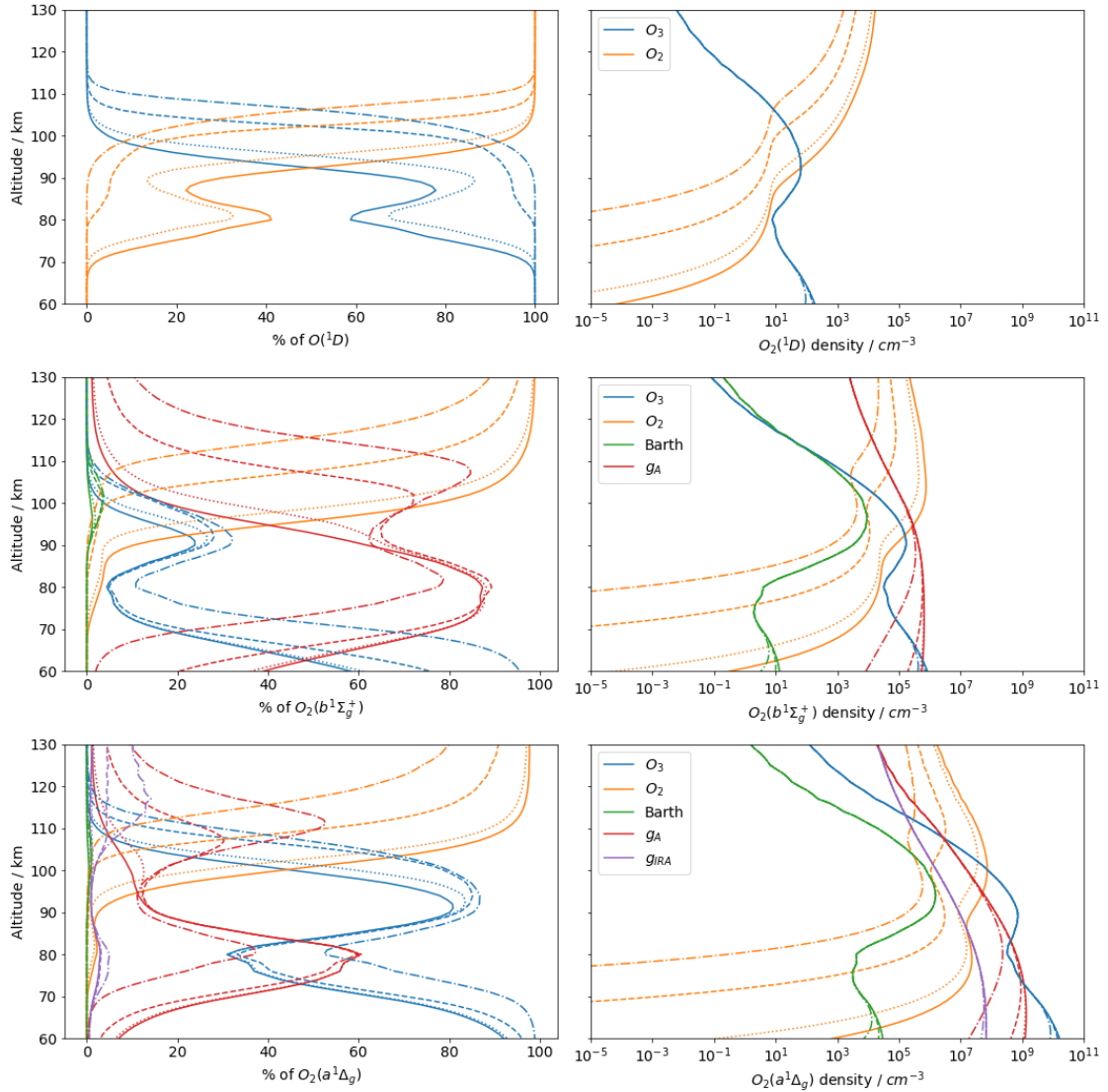
Figure 3.6: Altitude profiles of contributions from different sources (represented by colours, see text) to the total productions of $O(^1D)$ (first row), $O_2(b^1\Sigma_g^+)$ (second row) and $O_2(a^1\Delta_g)$(third row), while panels on the left show the relative contributions in percentage and panels on the right show the absolute contributions in number density of the species. Note that the horizontal axes of the right column are in logarithmic scale. Different line styles correspond to solar zenith angles of 30° (solid lines), 60° (dotted lines), 85° (dashed lines) and 90° (dash-dotted lines). Background density profiles of $N_2$, $O_2$ and temperature profile are taken from MSIS90 and $O_3$ taken from CMAM climatology in January at 15° N latitude.

profiles of $O_2$, $N_2$ and $O_3$, as well as the temperature profile for some rate coefficients that have dependency on the background temperature. Figure 3.6 displays the results of such a photochemical model in the altitude range of $60\,\text{km}$ to $130\,\text{km}$, calculated for four different solar zenith angles (SZAs) in order to illustrate that the effect from the height of the sun varies between sources and in altitude. The label '$O_3$' in Fig. 3.6 indicates productions through the ozone photolysis in the Hartley band, '$O_2$' the molecular oxygen photolysis in the Schumann-Runge Continuum and Lyman $\alpha$ line, 'Barth' the three-body recombination mechanism and '$g_A$' and '$g_{IRA}$' the solar excitation of the oxygen in the A-band and the IRA-band, respectively. It can be seen that for the production of $O(^1D)$, ozone photolysis is more important below $90\,\text{km}$ and oxygen photolysis above $90\,\text{km}$. For the production of $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$, the general picture remains similar to $O(^1D)$, with the additional important sources from the solar excitation from the ground state $O_2$. The Barth-type mechanism contributes to the excited oxygen molecules to a lesser extent than other sources.

Note that the calculation above (Fig. 3.6) was made assuming photochemical equilibrium (i.e. steady state, see Sect. 3.2) for all species (Eq. 3.23, 3.37 and 3.42). Such an assumption should be taken cautiously for $O_2(a^1\Delta_g)$ whose radiative lifetime is relatively long (about an hour), dominating the total photochemical lifetime at higher altitudes. Problems associated with long lifetime species are the slow response to those relevant changes of the basic states. For example, when $O_3$ concentration changes drastically around sunset and sunrise in the mesosphere, $O_2(a^1\Delta_g)$ responds at a much slower rate to reach the new balance. Consequently, if $O_2(a^1\Delta_g)$ measured not long after the sunrise is used as a proxy to inversely derive $O_3$ under the equilibrium assumption, the retrieved $O_3$ is most likely underestimated (the significance depends on the altitude concerned). More details concerning the issue of the long $O_2(a^1\Delta_g)$ photochemical lifetime is described in Paper 1. One may consider using a time-dependent photochemical model to retrieve ozone in such conditions.

To conclude this section, we described the general kinetic scheme of the oxygen airglow photochemistry. Of course, variations can be made by adding vibrational sub-levels or having different assumptions on the fractional efficiencies of each path. Nevertheless, it can be shown in all variations that ozone photolysis in the Hartley band is one of the most significant contributors to the production of both $O_2(b^1\Sigma_g^+)$ and $O_2(a^1\Delta_g)$ in their (secondary) peak region in the mesosphere. For this reason, these two airglow emissions can be used as proxies to infer the ozone density in the mesosphere during daytime, which is challenging to measure by using other types of remote sensing techniques. However, special treatment should be considered when dealing with long lifetime species such as $O_2(a^1\Delta_g)$ to avoid underestimating the ozone density.

## 3.4  OH airglow

The vibrationally excited hydroxyl radical $OH_v$ ($v = 1, ..., 9$) emits infrared radiation, as known as the Meinel band emission, between 80 and $100\,\text{km}$. The layer's inten-
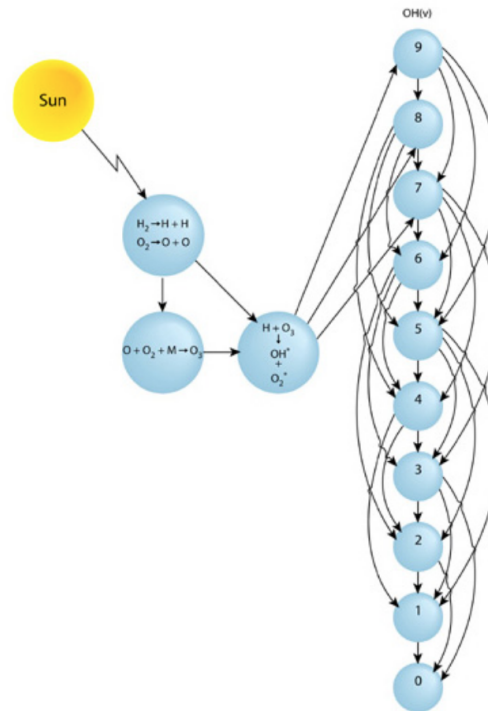
Figure 3.7: The $OH_v$ kinetic scheme. [Reproduced, by permission of Physica Scripta, from Figure 7 in Baker et al. (2007)]

sity and emission height depend on the two vibrational levels under consideration. Odin/OSIRIS continuously measures emissions corresponding to the (5-1), (8-3), (9-4) transitions with the optical spectrograph, and to the transition (3-1) with the infrared imager. The latter is the main focus of this thesis.

Unlike the oxygen airglow, the reaction chain of the hydroxyl airglow does not directly start by a photolysis reaction but rather by recombination of atomic hydrogen and ozone

$$H + O_3 \longrightarrow OH_v + O_2, \tag{3.44}$$
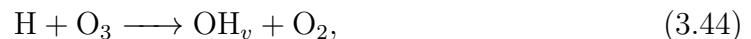
and a minor contribution from the reaction

$$O + HO_2 \longrightarrow OH_v + O_2, \tag{3.45}$$

which is mainly important for the lower vibrational levels (Grygalashvyly, 2015).

Its losses include quenching by another molecule M ($N_2$, $O_2$ or O)

$$OH_v + M \longrightarrow OH + M \tag{3.46}$$

and emission

$$OH_{(v=v_1)} \longrightarrow OH_{(v=v_2)} + h\nu, \tag{3.47}$$
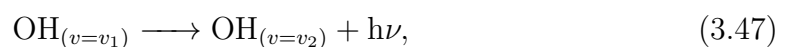
where the vibrational levels $v_1 > v_2$.

Figure 3.7 illustrates some of the $OH_v$ production and loss processes involved in the reaction chain. Note that not all processes mentioned above are included in

this figure. Depending on the level of details of the OH layer in question, there are different ways to model the production and losses of the nine vibrationally excited levels. The well-known ones are the so-called 'sudden death' and 'multi-quantum quenching' models (see Grygalashvyly, 2015; Llewellyn et al., 1978; McDade & Llewellyn, 1987). This thesis focuses on a qualitative assessment on the variations that occur in the $OH_v$ observations, rather than the quantitative assessment, e.g. comparing emissions from different vibrational levels or atomic oxygen retrieval. Thus, the following provides the simplest model, which serves merely as a first-order approximation.

If we omit the minor contributions and only consider the primary one, the production of $OH_v$ is

$$P_{OH_v} \approx k_{H+O_3}[H][O_3]. \tag{3.48}$$

Assuming that $O_3$ is in steady state, and if we, once again, only consider the primary source and loss term of $O_3$ around the $OH_v$ layer altitude,

$$k_{O+O_2+M}[O][O_2][M] \approx k_{H+O_3}[H][O_3], \tag{3.49}$$

we obtain

$$P_{OH_v} = k_{O+O_2+M}[O][O_2][M]. \tag{3.50}$$

If we follow the 'sudden death' model (i.e. a given vibrational level is quenched to ground level in one step) and assume that the quenching process is dominated by $O_2$, the loss of $OH_v$ can be described by

$$L_{OH_v} = Q_v[O_2], \tag{3.51}$$

where $Q_v$ is the 'sudden death' quenching rate for $O_2$ (Llewellyn et al., 1978). With the explicitly written $k_{O+O_2+M}$ and the ideal gas law, the balance of the $OH_v$ is finally expressed as

$$[OH_v] = BpT^{-3.4}[O], \tag{3.52}$$

where $B$ is a vibrationally dependent constant, $p$ the pressure and $T$ the temperature (the detailed derivation refers to Grygalashvyly et al. (2014)). From this expression, we can see that the concentration of $OH_v$ is proportional to pressure and O, and inversely proportional to $T^{3.4}$.

Figure 3.8 illustrates the typical structure of the hydroxyl layer modelled from Eq. 3.52 and the background atmosphere taken from the MSIS climatology. The pole to pole circulation pushes the layer downward in winter. In turn, the layer becomes more intense due to a higher pressure at a lower altitude. The semi-annual oscillation around the equator is typically observed in the mesosphere, such as the seasonal variation manifested in the $OH_v$ observation (see Paper 2).

Moreover, because of its strong ties with atomic oxygen, the hydroxyl layer is often used as a proxy to retrieve the concentration of O. Since the atomic oxygen at the altitude of the $OH_v$ layer is mostly controlled by vertical fluxes (changes in meteorological conditions), the $OH_v$ layer exhibits short term variation such as during the sudden stratospheric warming (SSW) events when the winter polar cap experiences unusual vertical movements (see Chapter 2). Correlation with the solar
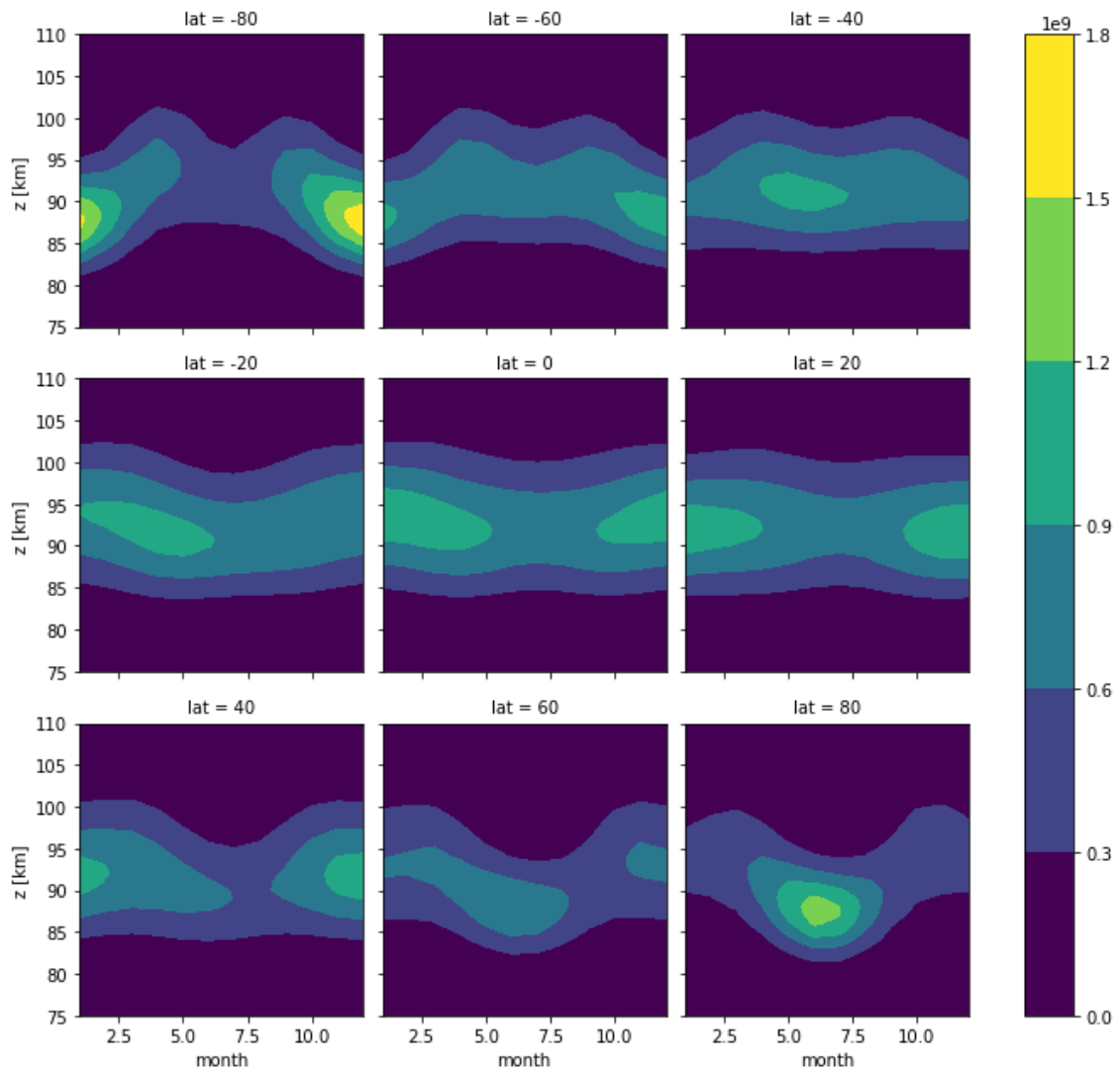
Figure 3.8: The $OH_v$ layer calculated based on Eq. 3.52
. Pressure, temperature and atomic oxygen are taken from MSIS climatology. Note that the contour scale is unitless as the vibrationally dependent constant $B$ is unknown.

flux (changes in external conditions) of the $OH_v$ layer, such as those signatures following the solar cycle, are also expected. Variations in the solar flux affect the photo-disassociation rate of the molecular oxygen to produce atomic oxygen at the mesopause and in the thermosphere. Thus, even though the production and losses of $OH_v$ are not directly linked to any photolysis process, the close connection with O makes $OH_v$ follow changes in the solar fluxes. Paper 3 describes such signature manifested in the $OH_v$ observations from OSIRIS.

In conclusion, we described the general production and loss mechanisms of the vibrationally excited hydroxyl radical $OH_v$. The close relationship between $OH_v$ and atomic oxygen and pressure is demonstrated by a simple expression by considering only the most important sources and sinks. Of course, a multi-quantum level quenching model has to be considered if we want to assess the precise quantity. However, this thesis addresses the $OH_v$ layer variations in time observed by Odin-OSIRIS, and, in that context, a first-order approximation is sufficient to aid our interpretations.

# Chapter 4

# The inversion problem

## 4.1 Introduction

Remote sensing is one of the most effective ways to collect data over large areas, allowing us to study the global atmosphere. Sensing the mesosphere would otherwise need e.g. rocket sondes which can hardly cover a large geographic area and a long-term perspective. Also, space-borne limb sounders are convenient for capturing the small variability in the middle and upper atmosphere since the majority of the air mass is concentrated in the lower atmosphere that will otherwise dominate the signal. However, the quantity actually being measured is usually an indirect parameter that somehow is linked to the desired geophysical property. A typical example in atmospheric remote sensing is that the measured quantity is an electromagnetic signal, while the desired property is actually the temperature or the density of the air. Thus, an inverse problem arises when we attempt to find the best representation of the desired parameter.

Another example of the inverse problem arises from the viewing geometry when we attempt to study the vertical structure of various properties of the atmosphere. Such atmospheric sounding can be made by looking up to the sky from the ground, looking down to the ground from space (i.e., nadir sounding) or pointing to the side of Earth through a vertical range (i.e., limb sounding). The aforementioned sounding methods collect the integrated signal along the line-of-sight (LOS). However, the desired quantity is often the local signal at a given position, as if we would have sent a rocket sonde to the high atmosphere. In this thesis, limb sounding is the viewing geometry in focus, as we are sensing the mesosphere with the satellite Odin now and with MATS in the future.

The integrated electromagnetic signal that reaches the instrument is conventionally called the *limb radiance*, and the desired quantity to retrieve could, for example, be the volume emission rate (see Fig. 4.1 for an illustration). In such a case, the inversion problem is how to 'unravel' the local volume emission given the total integrated limb radiance (Paper 1 and 2). Additionally, in Paper 1, we also discuss a different inversion problem which is to estimate the ozone density from the airglow volume emission rate.
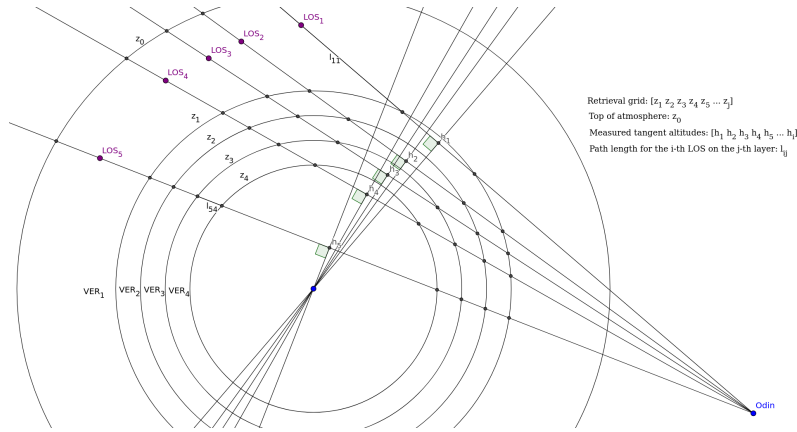
Figure 4.1: A graphical illustration of the limb geometry. VER denotes the volume emission rate, and LOS is line-of-sight.

In this chapter, we will address a model-based statistical approach to solve our inverse problems. The theory of how we can find the best estimates is based on a mathematical foundation that can be applied to different types of inversion problems. Although different mathematical approaches can be used to estimate the parameters, the physical relationship between the measured and desired quantities must be prescribed. In this thesis, the relationships we will consider include radiative transfer physics (from limb radiance to volume emission rate) and the airglow photochemistry (from IRA-band emissions to ozone). This chapter aims to provide the basic mathematical foundation on the optimal estimation method and some highlights of the practical implementation in our application.

## 4.2 Theory of optimal estimation in Bayesian philosophy

Let us begin by defining the *measurement vector* $\mathbf{y}$ and the *state vector* $\mathbf{x}$, which correspond to the measured and the desired quantity that we attempt to estimate, such as limb radiance and ozone number density, respectively. The aforementioned physical relationship, that maps from the the state vector to the measurement vector, is the *forward model* $\mathbf{F}$, and the relationship can be expressed as

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon, \tag{4.1}$$

where $\epsilon$ is the random experimental error, e.g., that may come from the instrument's noise, which cannot be precisely described by $\mathbf{F}$.

Even though we do not know the exact value of $\epsilon$ (otherwise, we could have corrected it in $\mathbf{y}$), we can assume the error term follows a particular probability distribution. The choice of the probability density function (pdf) depends on the system that we are modelling, but usually, it is assumed to be a Gaussian distribution (i.e., normal distribution), so we will do the rest of the thesis. The assumption of Gaussian distribution is justified by the need for a mathematically tractable model

and is consistent with most problems in nature. The unknown variable $\mathbf{x}$ that we attempt to estimate is viewed as a pdf, i.e., as a realisation of a random variable, instead of a deterministic constant. In other words, the representation of the optimal solution of an inversion tells us the most likely $\mathbf{x}$, $\hat{\mathbf{x}}$, with an uncertainty characterised by a pdf.

Furthermore, some prior information about the state $\mathbf{x}_a$ can help to regularise the solution to be physically reasonable. This particularly applies to an ill-posed problem, which often is the case. Such a condition occurs when we attempt to retrieve more parameters than we actually can based on the amount of information given by the measurement. This can lead to multiple solutions, or having too much information that leads to inconsistency (no solution), or the uncertainty of the given information is too large that leads to a solution sensitive to noise (e.g., overfitting). This prior knowledge that we have before the actual measurement is made is termed the a priori (means 'from the earlier' in Latin), which is also described by a pdf. Bayes' theorem helps us to mathematically combine the a priori pdf and the measurement pdf to find the desired a posteriori (means 'from the later' in Latin) pdf. In other words, the process of inversion can be seen as updating the a priori information with the given measurements and uncertainties mapped to the state space. This is known as the Bayesian approach since its implementation is based on Bayes' theorem.

A pdf of the Gaussian distribution is associated with two parameters: the mean (or expected value) $\mu$ and the standard deviation $\sigma$. The mean is the random variable most likely to be. The covariance is expressed in the form of a matrix which has elements of $S_{ij} = \rho_{ij}\sigma_i\sigma_j$ where the correlation coefficient $\rho_{ij}$ has value of between -1 and 1. In this chapter, we use $\mathbf{S}_y$, $\mathbf{S}_x$ and $\mathbf{S}_a$ to denote the covariance matrix of the measurement vector, the estimated state vector and the a priori state vector, respectively.

In the subsequent sections, two inversion schemes implemented in this thesis will be presented. One applies to a linear forward model and another to a non-linear forward model. The formalisation of the solutions is adopted from Rodgers (2000).

## 4.2.1 Linear optimal estimation

If the measurement vector $\mathbf{y}$ is linearly related to the state vector $\mathbf{x}$, the forward model $\mathbf{F}$ can be expressed in a linear form

$$\begin{aligned} \mathbf{y} &= \mathbf{F}(\mathbf{x}) + \epsilon \\ &= \mathbf{K}\mathbf{x} + \epsilon, \end{aligned} \tag{4.2}$$

where $\mathbf{K}$ is called the weighting function, or the Jacobian, which is a matrix comprised of all the first order partial derivatives of $\mathbf{F}(\mathbf{x})$ upon each element of $\mathbf{x}$, i.e., $\mathbf{K} = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$.

If we neglect the error term, an immediate solution (only if $\mathbf{K}$ is invertible, see text below) of a pure linear inversion problem, may seem to be

$$\hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{y} = \mathbf{G}\mathbf{y} \tag{4.3}$$

where $\mathbf{G}$ compiles all the partial derivatives of the estimated state vector $\hat{\mathbf{x}}$ upon each element of the measurement vector $\mathbf{y}$, i.e., $\mathbf{G} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}}$. However, as we can expect, this simple inverse of the matrix $\mathbf{K}$ may be impossible to compute due to e.g., rank deficiency, thus it is not practical for solving many physical problems. However, it can be used to illustrate the overall idea of the matrix $\mathbf{G}$ as a generalised inverse of $\mathbf{K}$, as later we will expand its formulation to more complex formats. $\mathbf{G}$ is called the *gain matrix* or *contribution function* in atmospheric remote sensing literature.

As mentioned earlier, the probabilistic approach is to make use of the full pdf of the measurement, which includes both the mean and the uncertainty.Moreover, we want to incorporate the prior knowledge to regularise the retrieved quantity. Thus, the pdf of the a priori is also utilised in the contribution function. In such case, the inversion can be viewed as an update from the a priori. Hence, the gain matrix $\mathbf{G}$ maps the difference between the $\mathbf{y}$ vector and the modelled a priori in measurement space to $(\hat{\mathbf{x}} - \mathbf{x}_a)$. The equations become

$$\mathbf{G} = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1}$$
$$\hat{\mathbf{x}} - \mathbf{x}_a = \mathbf{G}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) \tag{4.4}$$

alternatively

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{G}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) \tag{4.5}$$

This solution is known as the optimal estimation method (OEM) or maximum a posteriori method (MAP) (Rodgers, 2000). For this method, the procedure minimises the cost consisting of both the measurement noise and the a priori assumption, i.e.,

$$\chi^2 = (\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_y (\mathbf{y} - \mathbf{K}\mathbf{x}) + (\hat{\mathbf{x}} - \mathbf{x}_a)^T \mathbf{S}_a (\hat{\mathbf{x}} - \mathbf{x}_a). \tag{4.6}$$

Thanks to the mathematical tractability property of the Gaussian pdf, we can evaluate the covariance of the retrieval uncertainty based on the measurement noise (termed *retrieval noise* in Rodgers (2000)) following the form

$$\mathbf{S}_m = \mathbf{G} \mathbf{S}_y \mathbf{G}^T. \tag{4.7}$$

However, $\mathbf{S}_m$ is not the only component that contributes to the total error of the inversion. If OEM/MAP is used, another important source comes from the error in the a priori uncertainty, called the *smoothing error* evaluated by

$$\mathbf{S}_s = (\mathbf{A} - \mathbf{I}) \mathbf{S}_a (\mathbf{A} - \mathbf{I})^T, \tag{4.8}$$

where $\mathbf{I}$ is the identity matrix which has the same size as $\mathbf{S}_a$, and $\mathbf{A}$ is named the averaging kernel,

$$\mathbf{A} = \mathbf{G}\mathbf{K} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}. \tag{4.9}$$

The $\mathbf{A}$ matrix tells us how sensitive the estimation is to the true state. In an ideal case, $\mathbf{A}$ would be an identity matrix $\mathbf{I}$ with the same size. This represents, in a ratio of one-to-one, the estimated state changes with the true state. However, in practice, as we have modified the formulation of $\mathbf{G}$ not to be the exact inverse of $\mathbf{K}$, the diagonal elements in $\mathbf{A}$ may be 'spread' to the off-diagonal elements. In fact, the spreading aspect reflects the true retrieval resolution being lower than the grid spaces predefined by the state vector. Thus, the actual (e.g., spatial) resolution does not necessarily equal to the predefined grid cell size.

### 4.2.2 Non-linear optimal estimation

In the case of a non-linear inversion problem, the general idea of finding the solution is to linearise the forward model at a given state condition $\mathbf{F}(\mathbf{x}_n)$, e.g., initially $\mathbf{F}(\mathbf{x}_0)$. Following the previous formulation of the MAP method, the solution found at each iteration is

$$
\begin{aligned}
\mathbf{G}_n &= (\mathbf{K}_n^T \mathbf{S}_y^{-1} \mathbf{K}_n + \mathbf{S}_a^{-1})^{-1} \mathbf{K}_n^T \mathbf{S}_y^{-1} \\
\mathbf{x}_{n+1} &= \mathbf{x}_a + \mathbf{G}_n [\mathbf{y} - \mathbf{F}(\mathbf{x}_n) + \mathbf{K}_n(\mathbf{x}_n - \mathbf{x}_a)].
\end{aligned}
\tag{4.10}
$$

When the iteration number $n \to \infty$, the solution found is $\mathbf{x}_n \to \hat{\mathbf{x}}$. This method is called the Gauss-Newton method. When convergence is reached, the error analysis and characterisation are essentially the same for the linear case based on the Jacobian and the gain matrix at the last iteration.

To conclude this section, the probabilistic approach helps determine the desired parameters as pdf rather than a deterministic constant. The mathematical formulations of several well-developed methods, including OEM/MAP and Gauss-Newton, are described here to treat linear and non-linear inversion problems.

## 4.3 Practical implementation

In practice, numerous aspects are needed to be treated with caution when we carry out the implementations of the statistical method presented in the previous section. In this section, we discuss some of these considerations in the inversion problems included in this thesis. In particular, the retrieval problem deals with the vertical profile of chemical species number density, which exhibit high gradients. More details are provided in Paper 1.

### A priori assumption

If the retrieval is based on a high response from the information given by the measurement, the resulting estimate of the state vector shall be, to a negligible extent, influenced by an arbitrarily assumed a priori state $\mathbf{x}_a$. Typically, the a priori state is taken from a mean value of independently measured results or model outputs to approximate the prior knowledge about the state vector.

Once the expected value of the prior knowledge, $\mathbf{x}_a$, is defined, the approximation for its standard deviation $\sigma_a$ shall also be made in an appropriate manner. In our application, the state vector elements are expected to have several orders of magnitude differences (e.g., the atmospheric number density follows approximately an exponential function in altitude). If $\mathbf{x}_a$ is taken from a model output that resembles the true state reasonably well, a convenient way of defining $\sigma_a$ is to choose values that are relative to the mean. For instance, $\sigma_a(i) = r x_a(i)$ for the element $i$ where $r$ is a factor scalar to tighten or loosen the constraints on the a priori state. In other words, the diagonal elements of the a priori covariance matrix

$$
\mathrm{diag}(\mathbf{S}_a) = (r\mathbf{x}_a)^2.
\tag{4.11}
$$

Another type of constraint is to tackle the correlation between adjacent levels for the a priori assumption. As it is true for most of the atmospheric variables, closely spaced elements are less likely to have large differences than elements that are located far from each other. Thus, the off-diagonal elements of the a priori covariance matrix can be defined by an ad-hoc constraint, for example

$$S_a(i,j) = \sigma_a(i)\sigma_a(j)\exp(-|i-j|\frac{\delta z}{h}), \tag{4.12}$$

where $i$, $j$ are the element indices, $\delta z$ is the physical spacing represented by the retrieval grid and $h$ is the length scale which has the same unit with $\delta z$. This process prevents large oscillation in the retrieved solutions and thus is often termed the smoothing process.

## Relative averaging kernel

As we previously proposed that the covariance of the a priori can be represented by a scaling factor multiplies the mean state $\mathbf{x}_a$, the averaging kernel can also be viewed in a relative term to the a priori state, $\mathbf{A}^{rel}$. This can be implemented by the transformation from the ordinary averaging kernel $\mathbf{A}$

$$A^{rel}(i,j) = A(i,j)\frac{x_a(j)}{x_a(i)}. \tag{4.13}$$

The relative averaging kernel is more convenient to interpret when the state vector is expressed in a ratio depart from the a priori state.

To summarise this chapter, we discussed the mathematical formulations of several inversion methods, including OEM/MAP and Gauss-Newton. The essence of the Bayesian statistical approach is to incorporate prior knowledge about the state for regularisation. In a practical implementation of OEM/MAP, there are several important considerations for the retrieval of atmospheric constituents' concentrations which roughly follow an exponential function in altitude.

# Chapter 5

# Summary of appended publications

## Paper 1

Odin-OSIRIS consists of two optically independent instruments, the optical spectrograph (OS) and the infrared imager (IRI). The latter routinely collects data on the oxygen infrared atmospheric band (IRA-band) emissions observed by one of its channels. The measurement principle and geometry are very similar to the future mission of MATS. Thus, the exploration of this IRI data set certainly contributes to the preparatory process of the MATS mission, which includes testing the retrieval algorithms, understanding the photochemical processes and possibly finding small scale dynamical features in the mesosphere.

In Paper 1 we presented a retrieval scheme to derive the mesospheric daytime ozone profiles from the IRI limb measurements of the IRA-band emissions ($1.27\,\mu m$). First, we briefly described the updated calibration scheme of the limb radiance data product. Then, we applied the optimal estimation method (OEM) to the two inversion problems, the retrievals of volume emission rate (VER) of the $O_2(a^1\Delta_g)$ and the ozone number density profiles. We used a linear forward model for the VER retrieval, which includes the absorption of the IRA-band along the line-of-sight in the radiative transfer modelling. For the $O_3$ retrieval, the forward model is non-linear since the photochemistry model includes coefficients that depend on ozone concentrations. Hence, we applied the iterative Levenberg-Marquardt method to derive $O_3$ concentrations from VER of oxygen IRA band emissions. In addition, we addressed the problem of the $O_3$ retrieval associated with the long lifetime of $O_2(a^1\Delta_g)$.

We demonstrated the fidelity of the retrieval scheme by testing on a small but significant amount of data (5% in 2008). The IRI ozone profiles were shown to be consistent with the two other ozone products obtained from the same spacecraft Odin, namely SMR[1] and OS, even though their measurement principles are intrinsically different. It was also shown that the IRI ozone data have the advantage of the high along-track resolution. The zonal mean monthly average profiles of IRI ozone were then compared with other independent ozone data measured from a different

---

[1]SMR: Sub-Millimetre Radiometer

satellite, MIPAS[2]. We found that IRI appears to have a positive bias of up to 25% below 75 km, and up to 50% in some regions above. We attributed these differences to uncertainties in the IRI calibration and in the photochemical constants. These results indicate that the retrieval technique can be applied to process all the data collected throughout the Odin mission, leading to a long term mesospheric ozone data set that can be used to study mesospheric dynamics and photochemistry.

In fact, the method has been applied to retrieve $O_2(a^1\Delta_g)$ VER and $O_3$ profiles for the entire Odin mission after the publication of Paper 1.

# Paper 2

Besides the IRA-band emission (Paper 1), the Odin-OSIRIS infrared imager (IRI) also collects data of the hydroxyl Meinel-band emissions (around $1.53\,\mu m$) with one of its channels. The hydroxyl airglow is an important feature to study in the MLT because of its close relationship to ozone, atomic hydrogen, atomic oxygen and temperature. The variability that appears in the airglow layer is thus a useful tool to investigate how different mechanisms, for example the sudden stratospheric warming (SSW) or solar activity change, influence the MLT.

In Paper 2, we retrieved the night-time OH (3-1) volume emission rate (VER) using the optimal estimation method (OEM), similar to the work done in Paper 1 for the IRA-band but this time for the entire 15 years of data set. Subsequently, we used each VER profile to fit a Gaussian profile to approximate the layer height, thickness, peak intensity, zenith intensity, and their corresponding error estimates. The paper describes the necessary steps of the retrieval of these data products and the screening recommendations. The entire data set is registered on Zenodo.org and is publicly accessible.

We showed the fidelity of the retrieval scheme by investigating the monthly zonal averages of the VER profiles, layer height, thickness and peak intensity. They all depict the well-known annual oscillation and semi-annual oscillation signatures, which are consistent with previous studies. The new IRI OH (3-1) data set is unique owing to the coverage over the high latitudes and the long lifespan of the Odin satellite. The data set makes possible the investigation of MLT variability over a complete solar cycle and multiple major SSW events.

# Paper 3

After the completion of Paper 2, we proceed to the first investigation about the long-term variability in the MLT based on this particular OH (3-1) data set.

The OH nightglow is sensitive to changes in incoming solar flux. $OH_v$ is mainly produced by the $O_3$ + H reaction. Furthermore, $O_3$ is primarily produced by O + $O_2$ + M reaction and destroyed by the $O_3$ + H reaction to create $OH_v$. Ultimately, $OH_v$ production rate is controlled by O, temperature and pressure and all of them

---

[2]MIPAS: Michelson Interferometer for Passive Atmospheric Sounding

are affected by the incoming solar radiation. The 11-year solar cycle is one of the well-known cycles where the incoming radiation is changing periodically. However, space-borne measurements rarely last such a long period of time, and thus many studies relied on data obtained in a short period or on ground-based measurements, for which a good geographical coverage is lacking. As described in Paper 2, the IRI OH (3-1) data set is well suited to this investigation as it covers more than a full solar cycle.

Based on the previous studies, we expect that the $OH_v$ emission rate to be positively correlated and the emission height to be negatively correlated to the solar flux. The IRI data set does confirm that the $OH_v$ layer has the expected signature at most latitudes, except for the observations near the equator. Suspecting that the drift in the local time of the orbit was interfering with our interpretation, we used a time-dependent model to simulate the $OH_v$ diurnal variation at different latitudes and re-sampled the modelled data in line with the Odin sampling schedule. We demonstrated that the changing local time sampling of the Odin satellite during the mission was the cause of the observed distortion of the solar cycle signature near the equator.

# Paper 4

As shown in Paper 2, the new IRI data set is characterised by an excellent temporal coverage over the high latitudes, which is very well suited to the investigation of the impact of sudden stratospheric warming (SSW) events on the middle atmospheric composition. Occasionally, the stratopause collapses and reforms at a higher altitude, known as SSW with an elevated stratopause (SSW-ES) event. When such an event occurs, both the $OH_v$ and $O_2(a^1\Delta_g)$ airglow layers are greatly perturbed. The IRI data have nicely recorded these episodic events throughout the Odin mission.

In Paper 4, we analyse the daily average zonal mean profiles at 70-degree latitude pole-ward during the winter and early spring. In addition, we incorporate the simultaneously measured temperature and $H_2O$ profiles obtained from SMR to aid our interpretation. IRI has captured SSW-ES events in 2001/2002, 2003/2004, 2005/2006, 2008/2009, 2009/2010, 2011/2012 and 2012/2013 (identified by Limpasuvan et al. (2016)). However, we focus only on the three latest events because the temporal coverage of the SMR profiles was not sufficient in the earlier years. The operational activity of IRI has significantly decreased after 2015, thus we do not cover the event in 2019 which is among the most important SSW-ES during the Odin operational time.

The $O_2(a^1\Delta_g)$ nighttime emission results from the $O + O + M$ recombination. Thus, its concentration is directly related to the atomic oxygen number density. On the other hand, the $OH_v$ emission is primarily produced from the recombination of ozone and atomic hydrogen. Moreover, if we assume that ozone is in photochemical equilibrium, the dependency on atomic hydrogen is gone, and $OH_v$ concentration is proportional to atomic oxygen concentration. We used analytical expressions (following Grygalashvyly (2015)) to derive proxy [O] and [$OH_v$]. We could therefore

model the $OH_v$ layer by utilising the observed $O_2(a^1\Delta_g)$ and temperature and compare that with the observed $OH_v$.

The temporal evolution of the modelled $[OH_v]$ resembles well the recorded data. Right after the onset of the SSW events, the $OH_v$ layer is first lifted and weakened in its intensity. When the vortex starts to recover and the stratopause to reform, the mesosphere experiences a warming and a strong downwelling seen in the $H_2O$ data. The temporal structure that appears in the two airglows is synchronous. Such signature indicates that the downward transport of the atomic oxygen from the thermosphere causes the episodic emission seen in the $OH_v$ at a much lower altitude. The analytically derived model serves as a tool to perform a qualitative assessment of the events. However, the low vertical resolution of the measured temperature profiles may be the limitation leading to the thicker appearance seen in the modelled $[OH_v]$ profiles.

# Chapter 6

# Outlooks

This thesis work has contributed to a better understanding of the chemical and dynamical processes controlling the mesosphere. We analysed the airglow observation taken from Odin/OSIRIS infrared imager. The work also includes preparing the complete data set, which serves as a foundation for further investigations. This chapter will discuss several possible research ideas and their challenges. They include expanding the current retrieval methods and atmospheric chemistry and dynamics studies by using the readily retrieved datasets.

## 6.1 Time-dependent retrieval

In Paper 1, a method has been developed to retrieve the daytime $O_3$ in the mesosphere based on a photochemical model, which led to a valuable $O_3$ dataset to help us understand various physical processes in the middle atmosphere. However, the main problem associated with this retrieval method is that it heavily relies on the equilibrium assumption, while $O_2(a^1\Delta_g)$ has a relatively long radiative lifetime (ca. 74 min). Therefore, a significant amount of IRI observations cannot be used for ozone retrieval due to the measured $O_2(a^1\Delta_g)$ being far from photochemical equilibrium. In particular, such an issue applies to the profiles taken near the equinoxes and close to the equator when the local time sampling is most likely not long after the sunrise and the $O_3$ concentration changes quickly. Thus, these data are treated as invalid for ozone retrieval in Paper 1 to avoid the underestimation of ozone by incorporating the equilibrium index into the underlying assumption.

An idea to 'rescue' the missing part of the dataset is to develop a time-dependent model of ozone and $O_2(a^1\Delta_g)$ photochemistry. Such a model would simulate the diurnal behaviour of these species, which would allow us to have the flexibility to estimate the mesospheric $O_3$ and $O_2(a^1\Delta_g)$ at any local time. Furthermore, since the mesospheric $O_3$ has a diurnal variation, the comparison of our $O_3$ dataset to another dataset obtained from other satellites at a different local time sampling, for example, SABER, could be made in a relatively more straightforward way with such time-dependent retrieval.

## 6.2   Tomographic method

As demonstrated, the OSIRIS-IRI imager has the advantage of having a high sampling rate. A two-dimensional tomographic retrieval technique can therefore be applied to refine the resolution in the along-track horizontal direction one step further. The main difference to the previous 1D inversion is that a sequence of image exposures is used together for one inversion to retrieve a 2D atmospheric field, contrary to the retrieval of image by image. For example, the measurement vector $\mathbf{y}$ contains $a$ pixels per image and $b$ images, and the atmospheric state vector $\mathbf{x}$ contains the volume emission rate (VER) represented in a 2D grid space in the orbital plane, with $n$ and $m$ numbers of grid points in the vertical and horizontal directions, respectively. Thus the inversion problem to pose can be described by the following expression

$$
\begin{bmatrix} y_{11} \\ \vdots \\ y_{1a} \\ \vdots \\ y_{b1} \\ \vdots \\ y_{ba} \end{bmatrix} = \mathbf{F}(\begin{bmatrix} x_{11} \\ \vdots \\ x_{1n} \\ \vdots \\ x_{m1} \\ \vdots \\ x_{mn} \end{bmatrix}) + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1a} \\ \vdots \\ \epsilon_{b1} \\ \vdots \\ \epsilon_{ba} \end{bmatrix} \tag{6.1}
$$

where $\mathbf{F}$ is the forward model which describes the relationship between the measured radiance and the VER.

The advantage of the tomographic technique is to more correctly allocate the origin of the atmospheric signal in the along-track direction and thus improve the spatial resolution. However, there are several difficulties with the tomographic retrieval, one of which is computationally costly. As we can tell from the above equation, when the forward model is linearised, the Jacobian matrix $\mathbf{K}$ will contain $ab \times mn$ elements, and the covariance matrices $\mathbf{S}_a$ and $\mathbf{S}_y$ will have dimensions of $mn \times mn$ and $ab \times ab$ , respectively. Even though the non-zero elements of the matrices can be stored wisely, such as in the form of sparse matrices, the computation is still substantially increased compared to a 1D inversion.

Another difficulty for the tomographic retrieval method, perhaps not so trivial, comes from the scanning schedule of the Odin satellite compromised by the needs of the different instruments onboard. Shown as an example orbit in Paper 1 Fig. 4, the so-called mesospheric scanning mode creates data gaps up to $90\,\text{km}$ penetrating through the important airglow layer. These data gaps eliminate many of the images in the orbit that can be useful for tomographic retrieval. Thus, unique treatments are needed for orbits affected by such data gaps, such as a strategy for sectioning the orbit.

Otherwise, IRI data obtained from the normal scanning mode and, even better, the staring mode can be readily used to implement tomographic retrieval of the VERs. The resulting data would be beneficial to the study of small-scale structures that IRI is capable of capturing.
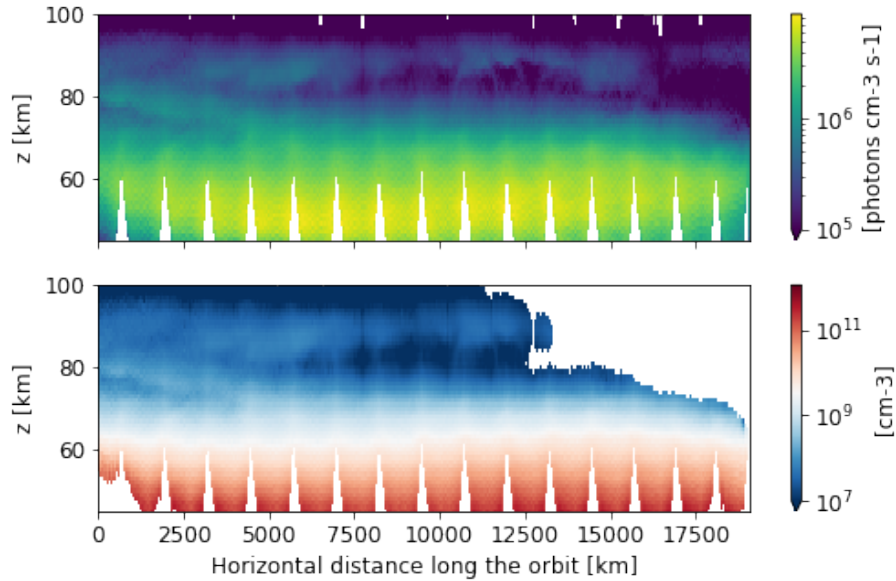
Figure 6.1: The volume emission rate of the oxygen IRA-band emission (top) and the ozone number density (bottom) measured in the day part of an example orbit (orbit No. 38720). This measurement was collected on March 20th, 2008. A large-scale feature appears at the beginning of the orbit around 70 to 80 km.

## 6.3 Tertiary ozone peak

The previously observed tertiary ozone peak at around 72 km is explained by the shortage of odd-hydrogen for the catalytic destruction of odd-oxygen, while the production of the odd-oxygen species is unchanged (Marsh et al., 2001). Unlike the primary and secondary ozone layers, the occurrence of the tertiary ozone maximum highly depends on the season, latitude and altitude, according to multiple observations and modelling studies (e.g., Kaufmann et al., 2003; Marsh et al., 2003). Moreover, Degenstein et al. (2005) demonstrated that the IRI measurements of oxygen IRA-band and OH Meinel (3-1) band were capable of capturing the signature of the tertiary maximum. As an example, the tertiary peak appears to be a finger-like, large-scale feature in a two-dimensional orbital scan shown in Fig. 6.1.

The daytime mesospheric ozone data product presented in Paper 1 can readily serve as a tool for further investigation on the behaviour of the tertiary ozone peak.

## 6.4 The future mission MATS

As mentioned in Chapter 1, this thesis work is part of the preparation for the future satellite MATS, which is planned to be launched at the end of 2022[1]. Numerous two-dimensional images measuring the oxygen atmospheric band (emission from $O_2(b^1\Sigma_g^+)$) will be available for analysis. After the calibration process, the limb radiance can be inverted to VER using a retrieval scheme similar to the one described

---

[1]However, due to the current geopolitical situation, the launch date and site are very uncertain.

in Paper 1. Since the photochemical lifetime of $O_2(b^1\Sigma_g^+)$ is shorter than $O_2(a^1\Delta_g)$, more refined structures due to small scale dynamics, such as gravity waves, are expected to be captured by MATS.

As the production of $O_2(b^1\Sigma_g^+)$ is related to the available ozone in the mesosphere under sunlit conditions, mesospheric ozone can be retrieved by using the same forward model as in Paper 1. During the nighttime, the reaction O+O+M is the only source to produce $O_2(b^1\Sigma_g^+)$. Hence, atomic oxygen density can be retrieved. In addition, the two spectral passbands in the infrared region will allow us to derive mesospheric temperature as the spectral filters are selected to be temperature sensitive. The retrieval scheme is similar to the one described in P. E. Sheese et al. (2010) where the temperature is derived from the OSIRIS optical spectrograph measurement on the A-band emission.

This thesis work has set a foundation as part of the many valuable preparations for the MATS mission.

# Bibliography

Baker, D. J., Thurgood, B. K., Harrison, W. K., Mlynczak, M. G., & Russell, J. M. (2007). Equatorial enhancement of the nighttime OH mesospheric infrared airglow. *Physica Scripta*, *75*(5), 615–619. https://doi.org/10.1088/0031-8949/75/5/004 (cit. on p. 33)

Brasseur, G., & Solomon, S. (2005). *Aeronomy of the Middle Atmosphere* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004. (Cit. on pp. 9, 16, 18, 28)

Buhler, O. (2009). *Waves and Mean Flows.* Cambridge University Press. https://doi.org/10.1017/CBO9780511605499. (Cit. on p. 13)

Degenstein, D. A., Gattinger, R. L., Lloyd, N. D., Bourassa, A. E., Wiensz, J. T., & Llewellyn, E. J. (2005). Observations of an extended mesospheric tertiary ozone peak. *Journal of Atmospheric and Solar-Terrestrial Physics*, *67*(15), 1395–1402. https://doi.org/10.1016/j.jastp.2005.06.019 (cit. on p. 49)

Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of Geophysics.* https://doi.org/10.1029/2001RG000106 (cit. on pp. 11, 16)

Grieco, F., Pérot, K., Murtagh, D., Eriksson, P., Forkman, P., Rydberg, B., Funke, B., Walker, K. A., & Pumphrey, H. C. (2020). Recovery and validation of Odin/SMR long-term measurements of mesospheric carbon monoxide. *Atmospheric Measurement Techniques*, *13*(9), 5013–5031. https://doi.org/10.5194/AMT-13-5013-2020 (cit. on p. 17)

Grieco, F., Pérot, K., Murtagh, D., Eriksson, P., Rydberg, B., Kiefer, M., Garcia-Comas, M., Lambert, A., & Walker, K. A. (2021). Improvement of Odin/SMR water vapour and temperature measurements and validation of the obtained data sets. *Atmospheric Measurement Techniques*, *14*, 5823–5857. https://doi.org/10.5194/amt-14-5823-2021 (cit. on p. 17)

Grygalashvyly, M. (2015). Several notes on the OH* Layer. *Annales Geophysicae*, *33*(7), 923–930. https://doi.org/10.5194/angeo-33-923-2015 (cit. on pp. 33, 34, 45)

Grygalashvyly, M., Sonnemann, G. R., Lübken, F. J., Hartogh, P., & Berger, U. (2014). Hydroxyl layer: Mean state and trends at midlatitudes. *Journal of Geophysical Research*, *119*(21), 391–12. https://doi.org/10.1002/2014JD022094 (cit. on p. 34)

Gumbel, J., Megner, L., Christensen, O. M., Ivchenko, N., Murtagh, D. P., Chang, S., Dillner, J., Ekebrand, T., Giono, G., Hammar, A., Hedin, J., Karlsson, B., Krus, M., Li, A., McCallion, S., Olentšenko, G., Pak, S., Park, W., Rouse, J., . . . Witt, G. (2020). The MATS satellite mission-Gravity wave studies by

Mesospheric Airglow/Aerosol Tomography and Spectroscopy. *Atmospheric Chemistry and Physics*, *20*(1), 431–455. https://doi.org/10.5194/acp-20-431-2020 (cit. on p. 5)

Heale, C. J., & Snively, J. B. (2015). Gravity wave propagation through a vertically and horizontally inhomogeneous background wind. *Journal of Geophysical Research.* https://doi.org/10.1002/2015JD023505 (cit. on pp. 13, 14)

Hedin, A. E. (1991). Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere. *Journal of Geophysical Research*, *96*(A2), 1159–1172. https://doi.org/10.1029/90JA02125 (cit. on p. 4)

Holton, J. R. (1982). The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of the Mesosphere. *Journal of the Atmospheric Sciences.* https://doi.org/10.1175/1520-0469(1982)039<0791:TROGWI>2.0.CO;2 (cit. on pp. 9, 11, 16)

Kaifler, B., Lübken, F. J., Höffner, J., Morris, R. J., & Viehl, T. P. (2015). Lidar observations of gravity wave activity in the middle atmosphere over Davis (69°S, 78°E), Antarctica. *Journal of Geophysical Research.* https://doi.org/10.1002/2014JD022879 (cit. on p. 16)

Kaifler, N., Kaifler, B., Ehard, B., Gisinger, S., Dörnbrack, A., Rapp, M., Kivi, R., Kozlovsky, A., Lester, M., & Liley, B. (2017). Observational indications of downward-propagating gravity waves in middle atmosphere lidar data. *Journal of Atmospheric and Solar-Terrestrial Physics.* https://doi.org/10.1016/j.jastp.2017.03.003 (cit. on p. 16)

Karlsson, B., & Shepherd, T. G. (2018). The improbable clouds at the edge of the atmosphere. *Physics Today*, *71*(6), 30–36. https://doi.org/10.1063/PT.3.3946 (cit. on p. 9)

Kaufmann, M., Gusev, O. A., Grossmann, K. U., Martín-Torres, F. J., Marsh, D. R., & Kutepov, A. A. (2003). Satellite observations of daytime and nighttime ozone in the mesosphere and lower thermosphere. *Journal of Geophysical Research D: Atmospheres*, *108*(9). https://doi.org/10.1029/2002jd002800 (cit. on p. 49)

Koop, C. G. (1981). A Preliminary Investigation Of The Interaction Of Internal Gravity Waves With A Steady Shearing Motion. *Journal of Fluid Mechanics.* https://doi.org/10.1017/S0022112081003546 (cit. on p. 15)

Koop, C. G., & McGee, B. (1986). Measurements of internal gravity waves in a continuously stratified shear flow. *Journal of Fluid Mechanics.* https://doi.org/10.1017/S0022112086001817 (cit. on p. 15)

Li, A. (2017). *A 3D-model for O2 airglow perturbations induced by gravity waves in the upper mesosphere (M.Sc. Thesis).* (Cit. on p. 7).

Li, A. (2020). *New perspectives in mesospheric wave dynamics and oxygen photochemistry (Lic. Thesis).* (Cit. on p. 13).

Lighthill, M. J. (1967). Waves in fluids. *Communications on Pure and Applied Mathematics.* https://doi.org/10.1002/cpa.3160200204 (cit. on p. 13)

Limpasuvan, V., Orsolini, Y. J., Chandran, A., Garcia, R. R., & Smith, A. K. (2016). On the composite response of the MLT to major sudden stratospheric warming

events with elevated stratopause. *Journal of Geophysical Research, 121*(9), 4518–4537. https://doi.org/10.1002/2015JD024401 (cit. on p. 45)

Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical Research, 86*(C10), 9707. https://doi.org/10.1029/jc086ic10p09707 (cit. on p. 16)

Llewellyn, E. J., Long, B. H., & Solheim, B. H. (1978). The quenching of OH{*} in the atmosphere. *Planetary and Space Science, 26*(6), 525–531. https://doi.org/10.1016/0032-0633(78)90043-0 (cit. on p. 34)

Marks, C. J., & Eckermann, S. D. (1995). A Three-Dimensional Nonhydrostatic Ray-Tracing Model for Gravity Waves: Formulation and Preliminary Results for the Middle Atmosphere. *Journal of the Atmospheric Sciences.* https://doi.org/10.1175/1520-0469(1995)052<1959:ATDNRT>2.0.CO;2 (cit. on p. 13)

Marsh, D., Smith, A., Brasseur, G., Kaufmann, M., & Grossmann, K. (2001). The existence of a tertiary ozone maximum in the high-latitude middle mesosphere. *Geophysical Research Letters, 28*(24), 4531–4534. https://doi.org/10.1029/2001GL013791 (cit. on p. 49)

Marsh, D., Smith, A., & Noble, E. (2003). Mesospheric ozone response to changes in water vapor. *Journal of Geophysical Research D: Atmospheres, 108*(3). https://doi.org/10.1029/2002jd002705 (cit. on p. 49)

McDade, I. C., & Llewellyn, E. J. (1987). Kinetic parameters related to sources and sinks of vibrationally excited OH in the nightglow. *Journal of Geophysical Research, 92*(A7), 7643. https://doi.org/10.1029/ja092ia07p07643 (cit. on p. 34)

McDade, I., Murtagh, D., Greer, R., Dickinson, P., Witt, G., Stegman, J., Llewellyn, E., Thomas, L., & Jenkins, D. (1986). ETON 2: Quenching parameters for the proposed precursors of O2(b1Σg+) and O(1S) in the terrestrial nightglow. *Planetary and Space Science, 34*(9), 789–800. https://doi.org/10.1016/0032-0633(86)90075-9 (cit. on p. 29)

Mlynczak, M. G., Morgan, F., Yee, J. H., Espy, P., Murtagh, D., Marshall, B., & Schmidlin, F. (2001). Simultaneous measurements of the O2(1Δ) and O2(1Σ) airglows and ozone in the daytime mesosphere. *Geophysical Research Letters, 28*(6), 999–1002. https://doi.org/10.1029/2000GL012423 (cit. on p. 26)

Mlynczak, M. G., Marshall, B. T., Martin-Torres, F. J., Russell, J. M., Thompson, R. E., Remsberg, E. E., & Gordley, L. L. (2007). Sounding of the Atmosphere using Broadband Emission Radiometry observations of daytime mesospheric O2 (1Δ) 1.27 $\mu$m emission and derivation of ozone, atomic oxygen, and solar and chemical energy deposition rates. *Journal of Geophysical Research Atmospheres, 112*(D15306). https://doi.org/10.1029/2006JD008355 (cit. on p. 26)

Murtagh, D. P., Witt, G., Stegman, J., McDade, I. C., Llewellyn, E. J., Harris, F., & Greer, R. G. H. (1990). An assessment of proposed O(1S) and O2 (b 1Σg+)nightglow excitation parameters. *Planetary and Space Science, 38*(1), 43–53. https://doi.org/10.1016/0032-0633(90)90004-A (cit. on p. 30)

Murtagh, D., Frisk, U., Merino, F., Ridal, M., Jonsson, A., Stegman, J., Witt, G., Eriksson, P., Jiménez, C., Megie, G., De la Noë, J., Ricaud, P., Baron, P., Pardo, J. R., Hauchcorne, A., Llewellyn, E. J., Degenstein, D. A., Gattinger, R. L., Lloyd, N. D., . . . Oikarinen, L. (2002). An overview of the Odin atmospheric mission. *Canadian Journal of Physics*, *80*(4), 309–319. https://doi.org/10.1139/P01-157 (cit. on p. 5)

Nappo, C. J. (2002). *An Introduction to Atmospheric Gravity Waves*. Academic Press. (Cit. on p. 11).

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding : theory and practice*. World Scientific. (Cit. on pp. 39, 40).

Sheese, P. E., Llewellyn, E. J., Gattinger, R. L., Bourassa, A. E., Degenstein, D. A., Lloyd, N. D., & McDade, I. C. (2010). Temperatures in the upper mesosphere and lower thermosphere from OSIRIS observations of O2 A-band emission spectra. *Canadian Journal of Physics*, *88*(12), 919–925. https://doi.org/10.1139/p10-093 (cit. on p. 50)

Sheese, P. (2009). *Mesospheric ozone densities retrieved from OSIRIS observations of the O2 A-band dayglow* (Doctoral dissertation). (Cit. on p. 26).

Smith, A. K., Garcia, R. R., Moss, A. C., & Mitchell, N. J. (2017). The Semiannual Oscillation of the Tropical Zonal Wind in the Middle Atmosphere Derived from Satellite Geopotential Height Retrievals. *Journal of the Atmospheric Sciences*, *74*(8), 2413–2425. https://doi.org/10.1175/JAS-D-17-0067.1 (cit. on p. 18)

Thomas, R. J., Barth, C. A., Rusch, D. W., & Sanders, R. W. (1984). Solar Mesosphere Explorer near-infrared spectrometer: measurements of 1.27 micrometer radiances and the inference of mesospheric ozone. *Journal of Geophysical Research*, *89*(D6), 9569–9580. https://doi.org/10.1029/JD089iD06p09569 (cit. on p. 26)

Vadas, S. L., & Becker, E. (2018). Numerical Modeling of the Excitation, Propagation, and Dissipation of Primary and Secondary Gravity Waves during Wintertime at McMurdo Station in the Antarctic. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2017JD027974 (cit. on p. 16)

Vadas, S. L., Zhao, J., Chu, X., & Becker, E. (2018). The Excitation of Secondary Gravity Waves From Local Body Forces: Theory and Observation. *Journal of Geophysical Research: Atmospheres*, *123*(17), 9296–9325. https://doi.org/10.1029/2017JD027970 (cit. on p. 16)

Yankovsky, V. A., & Manuilova, R. O. (2006). Model of daytime emissions of electronically-vibrationally excited products of O_3 and O_2 photolysis: application to ozone retrieval. *Annales Geophysicae*, *24*(11), 2823–2839. https://doi.org/10.5194/angeo-24-2823-2006 (cit. on p. 26)

Yankovsky, V. A., Martyshenko, K. V., Manuilova, R. O., & Feofilov, A. G. (2016). Oxygen dayglow emissions as proxies for atomic oxygen and ozone in the mesosphere and lower thermosphere. *Journal of Molecular Spectroscopy*, *327*, 209–231. https://doi.org/10.1016/j.jms.2016.03.006 (cit. on p. 26)