UNIVERSITY OF SOUTHAMPTON

Faculty of Environmental and Life Sciences
School of Ocean and Earth Science

Observations of thermospheric heating signatures associated with sub-kilometer scale auroral electrodynamics

DOI: 10.1002/0470841559.ch1

by

David John Price
MPhys
ORCiD: 0000-0002-1825-0097

A thesis for the degree of
Doctor of Philosophy

November 2021

University of Southampton

Abstract

Faculty of Environmental and Life Sciences
School of Ocean and Earth Science

Doctor of Philosophy

Observations of thermospheric heating signatures associated with sub-kilometer scale auroral electrodynamics

by David John Price

This thesis presents analysis of thermospheric neutral temperature-altitude profiles derived from ground-based observations during intervals of auroral emission. Neutral temperature-altitude profiles are obtained in the magnetic zenith at unprecedented spatial ($<$10 km) and temporal resolution ($<$5 s). Electrodynamic processes associated with the formation mechanisms of auroral arcs are inferred via enhancements in the neutral temperature. The observations agree with leading arc-formation theories and provide further evidence supporting the existing literature. New evidence for the existence of filamentary field-aligned current (FAC) sheets, embedded in auroral arcs and directly linked to the formation of auroral curls, is presented.

This thesis also provides a detailed outline of the development of a novel observational technique for determining neutral temperature-altitude profiles during intervals of auroral emission. This technique combines spectral observations from the High Throughput Imaging Echelle Spectrograph (HiTIES) and images from Auroral Structure and Kinetics (ASK) with radar observations from the EISCAT Svalbard Radar and the Southampton Ionospheric Model. A spectral inversion technique is utilised. Trial temperature profiles are converted into a synthetic emission spectrum via an integration in altitude across modelled N_2 volume emission rate profiles. A least squares residual fitting process is used to evaluate the goodness-of-fit of the synthetic spectra that correspond to each random trial altitude-temperature profile, by determining its ability to recreate the HiTIES observed spectrum. Analysis of the outcome of multiple random trial profiles is used to estimate a final true thermospheric temperature profile. Neutral temperature variations are associated with specific auroral dynamics by comparing the temperature profiles with concurrent imaging observations from the ASK instrument.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
   DJ Price, DK Whiter, JM Chadney, and BS Lanchester. High-resolution optical observations of neutral heating associated with the electrodynamics of an auroral arc. *Journal of Geophysical Research: Space Physics*, 124(11):9577–9591, 2019

Signed:..... Date:.....

Acknowledgements

I've had the opportunity to meet lots of great people during my PhD who have played no small part in making it such a enjoyable and interesting experience.

Primarily I would like to thank Dan Whiter - I really couldn't have asked for a better supervisor. I never really felt lost or stressed during my PhD, something that I am sure was mostly your doing thanks to your advice and help. A massive thanks for being so cool about everything, from introducing me to academia and exposing me to so many interesting things and places, to messing around on google maps and pushing the limits of sleep deprived problem solving on campaign.

Thanks to Rob, Caitriona, Imogen and Betty. The space environment physics group is a really great and friendly place to work and I've thoroughly enjoyed my time there. A special thanks to Betty Lanchester - the master of constructing sentences and a font of helpful advice and knowledge for all things regarding the aurora.

I am grateful for the countless friends I've made during my PhD - at Southampton and otherwise. Meeting new people and exploring new places at conferences and on campaign is one of the best things you can do during a PhD and I can't overstate how much I loved it. Thanks Noora and Mikko for supporting me on campaign and showing me what Svalbard has to offer (including extremely cold morning swims and claustrophobic ice caves). A shout-out to the Southampton group (Dale, Jade, John, James, Josh, Laura, Pip, Ed and Micheal) for the Pub trips, snooker games and putting up with my incessant competitiveness.

Thanks to all my friends from home - who always reminded me it could be worse, I could be filling in excel spreadsheets all day. Thanks for all past, present and future laughs! Let it forever be known that I was the first.

Finally I'd like to thank my family, for their love, and for always being there and supporting my every decision. Thanks for making me who I am.

Some people are always trying to ice skate uphill - Blade, 1998 (paraphrased)

Chapter 1

Introduction

The primary subject of this thesis is the observation and analysis of the aurora borealis, a polar atmospheric phenomenon that results from interactions between the Earth's magnetic field and the solar wind; a hot stream of plasma streaming from the Sun. The spectacular optical light shows produced by the aurora have been a subject of fascination and awe for as long as there has been humans to observe them. Scientifically they provide an excellent and relatively accessible testing ground for a fascinating mixture of atmospheric and plasma research. The aurora typically occurs in the upper atmosphere or thermosphere, at heights between approximately 100 and 500 km, occupying a unique region in altitude which is too high for weather balloons, and too low for satellites. As a result, short of one-time (and incredibly expensive) rocket flights, studying the aurora provides the only glimpse into an altitude layer that is proving to be increasingly important in building a complete picture of the highly dynamic global circulation of the atmosphere.

The work presented here concerns the analysis of the variability in the temperature of the local neutral atmosphere during intervals of auroral emission, with an aim to develop a complete picture of the electrodynamical structures responsible for auroral formation. To facilitate this research, I developed and employed a novel observational technique that produces neutral atmospheric temperature profiles at high temporal and spatial resolution. A variety of instruments and methods have been used in this thesis, the most notable of which are optical observations from two Southampton owned and operated instruments: Auroral Structure and Kinetics (ASK) and the High Throughput Imaging Echelle Spectrograph (HiTIES). These observations are supported by radar measurements from coincident EISCAT (European Incoherent Scatter) campaigns and an ion chemistry and transport model.

The chapters following provide an overview of the relevant scientific background, a detailed outline of the instrumentation used, the full development process behind the production of the observational technique, results from a number of auroral case studies and a discussion of potential avenues for future work. Specifically the chapters can be summarised as: Chapter 2 gives an overview of the relevant theoretical background for the following research, including a simple summary of the currently understood

electrodynamical structure of auroral arcs; Chapter 3 outlines the roles and technical specifications for each of the instruments employed in the research, as well as the corresponding ion chemistry modelling techniques and applications; Chapter 4 gives a detailed chronology of the development of the novel observational technique at the centre of this research, discussing how the process evolved throughout the course of the PhD; Chapters 5, 6 & 7 present case studies of auroral events wherein the neutral atmospheric temperature response to auroral structures was observed. Neutral temperature changes are used to develop a picture of the underlying electrodynamical structure responsible for the aurora; finally Chapter 8 concludes the thesis and suggests a range of potential future work and projects.

Chapter 2

Background

2.1 Solar-Magnetosphere Interactions

2.1.1 The Solar Wind

Ultimately, the aurora results from a chain of physical processes that begins 150 million kilometers away at the Sun, and as such the dynamics and variations of the aurora are intrinsically linked to those of the Sun. The Sun is a large ball of incandescent hot plasma; a highly electrically conductive gas of charged particles with equal numbers of free positive and negative charge carriers, such that the net-charge is zero. Solar plasma is primarily composed of hydrogen and helium atoms which are heated to ionisation by radioactive processes in the core. The complex rotational and convectional motion of solar plasma generates an intricate, highly twisted and variable stellar magnetic field, which extends well beyond the solar surface and out into interplanetary space. The region of influence of the solar magnetic field is known as the 'heliosphere'. The heliosphere is an roughly spherical sector of space which terminates at the 'heliopause' - marking the transition into interstellar space. The heliopause is located at approximately 100-200 AU (astronomical units) from the Sun, meaning that the heliosphere easily encompasses all of the solar system's planetary bodies. The magnetic field that defines and permeates the heliosphere is therefore known as the interplanetary magnetic field (IMF).

Above the Sun's photosphere (the visible 'surface' of the Sun) is a large and super-hot atmosphere of plasma known as the stellar corona. The corona is observed to be significantly hotter than the solar surface, and the exact heating mechanism responsible for this is not clearly understood. However, the enhanced temperatures in the corona play a crucial role in the production of the 'solar wind'. Due to the extreme temperatures, a small fraction of the coronal plasma has sufficient thermal energy to escape the Sun's gravitational pull and propagate radially outwards into the solar system. The constant stream of charged plasma particles flowing from the Sun as a result of this process is known as the solar wind. The solar wind varies in density, temperature and speed, both with time and with regard to solar latitude and longitude. It primarily consists of

electrons, protons and alpha particles and typically flows at velocities between 250 and 750 kms^{-1}.

The solar wind is inherently linked to the interplanetary magnetic field. In certain conditions, such as those found in the solar wind, plasma and magnetic fields are coupled by the 'frozen-in' theorem. Under frozen-in conditions, the motion of the plasma drags the superposed magnetic field along with it, keeping each particle's centre of gyration fixed on a particular field line. Thus, the flow of the solar wind carries the IMF outwards and into the heliosphere. As the activity of the Sun fluctuates, so does the outflow of the solar wind and the corresponding structure and orientation of the coupled IMF, which results in ever-changing magnetic and plasma conditions in the heliosphere. The relative size of the Earth with respect to the heliosphere means that random fluctuations (or noise) in the solar wind appear as coherent structures, increasing the perceived variability of the IMF at earth scales. The result of these factors is a highly variable plasma and magnetic environment in the region of space directly surrounding the Earth.

2.1.2 The Earth's Magnetosphere and the Dungey Cycle

The Earth's magnetic field, and those of other planetary bodies in the solar system, represent islands in the stream of the solar wind. Two magnetic fields cannot occupy the same space, so when the expanding solar wind and the frozen-in IMF encounters the geomagnetic field generated by the Earth's core, it is slowed down and deflected around it. A simple diagram of the interaction between the magnetosphere (Earth's magnetic field) and the solar wind is shown in Figure 2.1. The kinetic pressure of the solar wind plasma distorts the magnetosphere significantly: compressing it at the nose, and elongating it into a long thin tail on the night-side. A bow shock (left hand side of Figure 2.1) is formed due to the rapid deceleration of solar wind particles, which also thermalises them; converting the majority of their kinetic energy into thermal energy. The region of space containing these particles, and directly succeeding the bow shock, is called the 'magnetosheath' (white jagged field lines in Figure 2.1) and the plasma in this region is much denser and hotter than the surrounding solar wind. The magnetosheath surrounds the magnetosphere and ends at the boundary between the two magnetic fields: the magnetopause (indicated by the transition from white to red field lines in Figure 2.1). Inside the magnetopause the geomagnetic field dominates, and the plasma contained within is separate from the solar wind, only able to mix with the solar wind plasma under special circumstances.

The Earth's magnetosphere is approximately dipolar in structure, with magnetic poles in the northern and southern hemisphere that are slightly displaced from the geographic poles. Plasma particles in the magnetosphere are trapped on the field lines, gyrating around them due to the Lorentz force, and moving parallel along the field lines in a spiral motion. The convergence of field lines at the magnetic poles increases the magnetic field strength in those regions, introducing a magnetic gradient and adding a

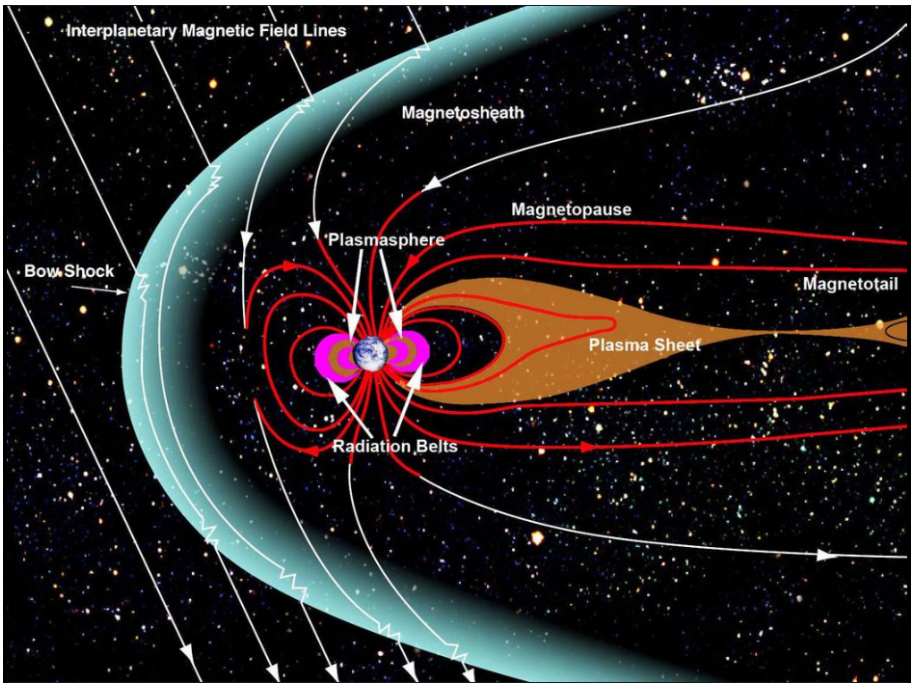


FIGURE 2.1: A simple depiction of the interaction between the Earth's magnetosphere and the Solar wind. The magnetopause is represented by the transition from white (IMF) to red (geomagnetic) field lines. Credit: NASA

component of the Lorentz force that is directed in opposition to approaching particles (away from the region of increased field strength). The magnetic gradient force slows and eventually reverses the parallel velocity of the particles, creating a 'magnetic bottle' and trapping them so that they bounce repeatedly between the two poles. The depth at which each particle is reflected depends solely on its pitch angle: the angle between the particle's velocity and the magnetic field lines in the equatorial plane. A pitch angle of 90° represents a gyrating particle with a parallel velocity component of zero. Particles with low enough pitch angles are able to penetrate deep into the magnetic bottle and precipitate into the Earth's polar atmosphere, creating the aurora. For an isotropic distribution of particles in the magnetosphere, a small fraction of them have sufficiently low pitch angles to escape the magnetic bottle in this way, leaving a gap in the final distribution. The range of angles that allows this is known as the 'loss cone' - so named for its conical shape consisting of the small angles centered around 0°.

The primary source of the plasma contained in the magnetosphere is the Earth's atmosphere; thermal particles escape from the Earth's gravitational attraction and become trapped in the terrestrial magnetic field. However a small portion of the magnetospheric plasma also comes from the solar wind. Transfer of plasma (and energy) between the two regimes occurs during a process called magnetic reconnection. Magnetic reconnection is a physical process during which the topologies of two interacting magnetic fields are re-arranged. Field lines belonging to each field 'break' and are 're-connected' with a field line from the other field. The result is that field lines that were

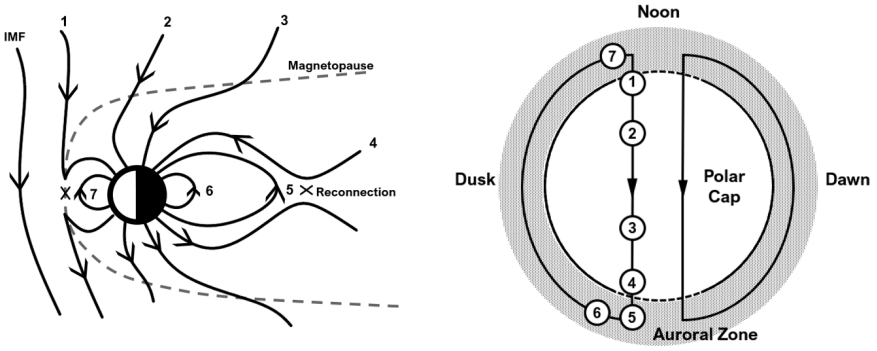


FIGURE 2.2: Magnetic reconnection in the magnetosphere during steady-state convection. Left: A dusk-side view (Sun to the left) of the interaction between the terrestrial magnetic field and the IMF. Closed magnetic field lines, oriented 'northward' undergo reconnection with the 'southward' IMF, opening them (1), where they then convect into and across the polar cap (2-4). The open field lines are compressed in the tail and reconnect again, forming a closed field line (5), which contracts Earthward (6). Newly closed field lines then convect around the dusk or dawn flanks of the earth, their footprints forming the auroral oval, eventually restarting the process in the nose (7). Right: a top down view of the polar cap, showing the footprints of the magnetic field lines in the left hand panel (labelled 1-7). The boundary between the footprints of open and closed field lines is shown as a solid and dashed black semicircle, and the general 2-cell convection pattern of field lines during steady state convection is displayed as black stream-lines. The auroral zone is denoted by the shaded region.

once exclusive to each field are joined together into one - releasing energy and allowing the plasma that was trapped on each of them to mix. Magnetic reconnection is most efficient when the field lines of the interacting magnetic fields are anti-parallel. The predominant locations where this can occur in the magnetosphere are: at the nose of the magnetosphere, when the orientation of the variable IMF is oppositely directed to the (relatively) static terrestrial field; and in the magnetotail, when the looped field lines are compressed into an elongated, flat structure. The rate of reconnection at both locations is highly variable, resulting in a dynamic magnetosphere. However, magnetic flux cannot be added or subtracted from the magnetosphere indefinitely, leading to an approximate balance between the two reconnection rates over long enough timescales. The large-scale general convection of magnetic flux through the magnetosphere is described by a process known as the Dungey Cycle, which is outlined by the following steps:

1. Magnetic reconnection at the nose of the magnetosphere connects an IMF field line and a 'closed' (meaning both ends of the field line originate at the earth) geomagnetic field line.

2. The two newly 'opened' magnetic field lines, each attached to the polar regions of their respective hemisphere, are now directly connected with the IMF.

3. A combination of magnetic tension and the flow of the solar wind (and IMF) drags the open field lines over the poles of the Earth, eventually stretching them

tailward. The regions of the Earth's atmosphere that contain the footprints of these open field lines are known as the polar caps.

4. The compression of the two stretched open field lines in the tail causes them to magnetically reconnect, creating a closed geomagnetic field line and an IMF field line.

5. The closed field line, no longer under tension from the solar wind, relaxes, moving Earthward and eventually convecting around the Earth's flanks to reach the nose and repeat the process again. The IMF field line continues to propagate deeper into the heliosphere.

A schematic outline displaying this process is shown in Figure 2.2. During steady-state convection, the footprints of the magnetic field lines follow a two-cell convection pattern in the polar cap (shown as black stream-lines in the right-hand sub figure). Newly opened field lines begin by flowing directly tail-ward across the polar cap. When these field lines are closed in the tail their footprints convect Sun-ward, through the dawn- and dusk-side atmosphere, forming the auroral oval. The auroral oval is generally defined as a 5-10° wide ring centered on approximately 65° latitude. The movement of the field lines during the Dungey cycle also drags frozen-in plasma along the same convection pattern.

The motion of the frozen-in plasma particles ($\mathbf{v}$) through the Earth's magnetic field ($\mathbf{B}$) generates an electric field ($\mathbf{E}$) defined by:

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} \tag{2.1}$$

The result of this relationship is a polar-convection driven electric field, the direction of which is dependent on the velocity of local plasma flow. In the northern polar cap the electric field is orientated towards dusk, and in the northern auroral oval the electric field is oriented pole-ward on the dusk-side and equator-ward on the dawn-side. A schematic overview of the convection electric field can be seen in Figure 2.3.

2.2 The Aurora

2.2.1 The Ionosphere

The ionosphere is the ionised region of the Earth's upper atmosphere. Incoming solar radiation regularly ionises atmospheric constituents, and the ionosphere is defined as the altitude region at which the atmospheric density is sufficiently low that the rate of plasma recombination is smaller than the ionisation rate, allowing a stable population of plasma to be maintained. The lowest altitude that this balance is typically achieved is between approximately 80 and 100 km. It is important to note that the atmosphere is not fully ionised until extreme altitudes, and much of the ionosphere remains in a partially ionised state, coexisting with a significant neutral density. Nonetheless, the

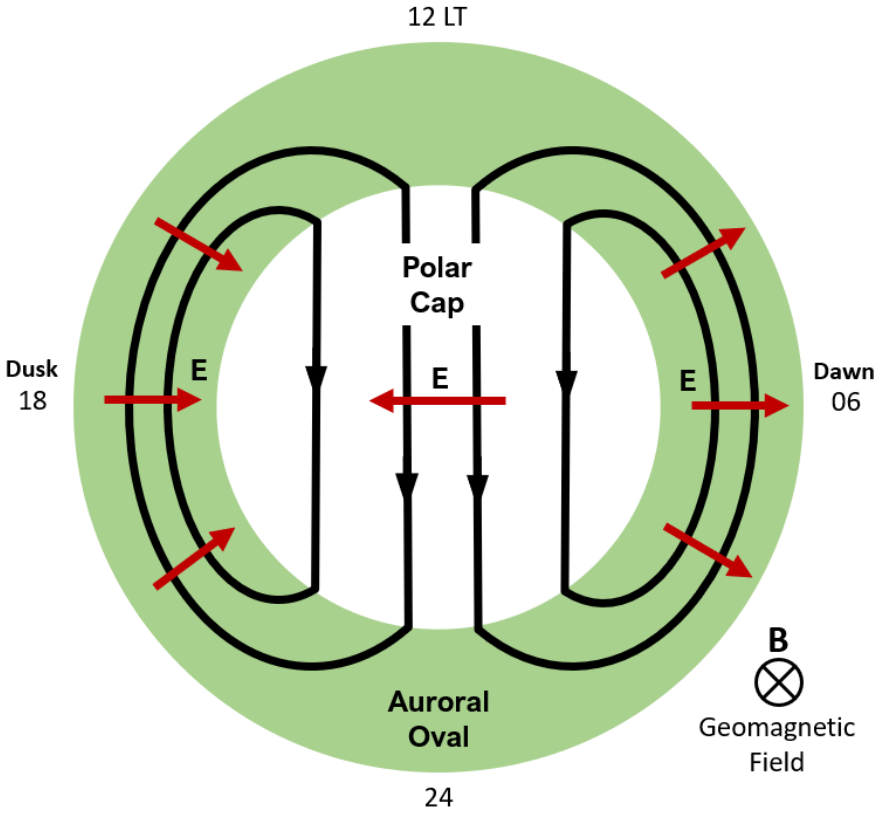


FIGURE 2.3: Schematic of the two-cell convection pattern and resultant electric field orientation in the Northern hemisphere's high latitude ionosphere during steady state Magnetospheric convection. Black streamlines show the motion of the magnetic field lines, and the plasma carried by them. Red arrows display the orientation of the corresponding electric fields. Adapted from Baumjohann and Nakamura [2007].

plasma population in these regions allows for the flow of numerous electric currents, and the ionosphere is an important component and the prime medium for many magnetospheric current systems. The ionisation is typically separated into three broad altitude regions, each with a distinct set of ionisation mechanisms: the D-region, E-region and F-region. The aurora is generally observed in the E and F-region ionosphere, at altitudes between approximately 90 and 500 km.

The ionosphere is composed of ions from a variety of atmospheric atoms and molecules. The neutral component of atmosphere in this region is known as the thermosphere, and its most abundant constituents are molecular Nitrogen (N_2), molecular Oxygen (O_2), and atomic Oxygen (O). Figure 2.4 displays atmospheric parameters produced by the Mass Spectrometer Incoherent Scatter (MSIS-E-90) model [Hedin, 1991]. The left-hand panel displays modelled concentrations of the primary neutral species above Svalbard, as a function of altitude. The right-hand panel shows the modelled neutral temperature profile. At low altitudes, stronger gravitational constraints mean that the heavier molecular structures (N_2 and O_2) dominate the relative concentrations, with their abundances dropping off quickly with increasing altitude. The atomic Oxygen in the atmosphere is generated by the photo-disassociation of O_2 by solar UV radiation,

and its concentration peaks at approximately 100 km. However, O only becomes the most abundant species at altitudes above 200 km, where its lighter mass allows it to maintain a larger population relative to the heavier molecular competitors.

Neutral temperatures (right hand panel of Figure 2.4) rapidly increase with altitude above the mesopause (the boundary between the mesosphere and the thermosphere where the temperature stops decreasing with altitude), before plateauing to ∼1000 K in the upper thermosphere and exosphere. The neutral temperature profile modelled here represents a general case; in reality it is highly dependent on a large range of extremely variable atmospheric dynamics and processes, including the aurora, and the true profile likely displays significant local-scale variations in temperature.

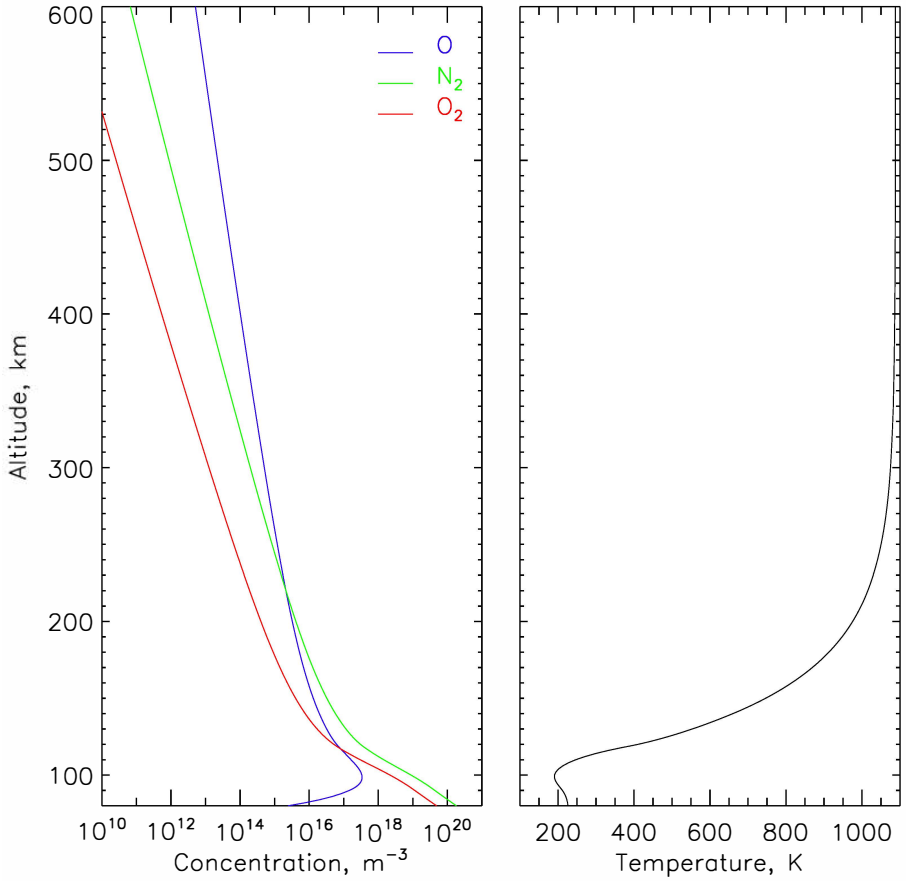


FIGURE 2.4: Ionospheric number densities for atmospheric O, N_2 and O_2 (left). Atmospheric neutral temperature profile (right).

2.2.2 Ionospheric Conductivity

The unique conditions present in the high latitude ionosphere result in the manifestation of plasma conductivities which can drive important electrodynamical processes related to the formation and evolution of auroral features. As mentioned, the ionosphere is partially ionised plasma that is permeated by the Earth's magnetic field. At

low altitudes, the geomagnetic field lines in the polar regions are approximately radial and the magnetic field can be assumed to be spatially uniform and consisting of only a vertical component. Furthermore, in the polar cap and auroral zones there is often an additional and variable convection-driven electric field, which is superposed with the geomagnetic field. The interaction between the partially ionised plasma in the ionosphere and these magnetic and electric fields is an important aspect of high-latitude ionospheric dynamics. In any plasma, the response of the ion and electron populations to external magnetic and electric fields is governed by the Lorentz force ($\mathbf{F} = q[\mathbf{E} + \mathbf{V} \times \mathbf{B}]$). Thus, in a collisionless plasma, electrons and ions undergo an $\mathbf{E} \times \mathbf{B}$ drift in the same direction (perpendicular to both the $\mathbf{B}$ and $\mathbf{E}$ fields) and at the same velocity. As a result, there is no net flow of charge, no anomalous conductivity and no current flow. However, in a collisional (partially ionised) regime, the motion of the ions and electrons are impacted by the collisions, both with other charged particles and with neutrals, which disrupts the $\mathbf{E} \times \mathbf{B}$ drift. Crucially, if the local neutral density is significant enough that the ion or electron-neutral collision frequency approaches the respective cyclotron frequency, the collisions can significantly decouple the charged particles from their usual Lorentz force governed motion. Furthermore, at high enough neutral densities the charged particles collide so often with the neutral particles that their motion is completely coupled to, and dictated by the motion of the neutrals.

Ions and electrons have different cyclotron frequencies, and as such, the range of neutral densities required to impact their Lorentzian motion also differs. Furthermore, since the neutral density of the thermosphere depends significantly on the altitude, the motion of the ion and electron populations will be impacted differently at different altitudes. However, the collision frequencies generally approach the respective cyclotron frequencies of both particles in around the E-region ionosphere (between $\sim$75–150 km), with the ratio between the two frequencies approaching unity at altitudes of $\sim$140 km and $\sim$80 km for ions and electrons respectively. If either, or both of the ion or electron populations are impacted in this way, the resulting divergence from their joint motion under the Lorentz force introduces currents in the ionosphere. A simple schematic of this mechanism is show in Figure 2.5. In the absence of collisions, charged particles exhibit a semi-circular motion: starting from rest, the particle is initially accelerated by the $\mathbf{E}$ field; once it has a non-zero velocity, the perpendicular $\mathbf{V} \times \mathbf{B}$ component of the Lorentz force becomes non-zero and swings the particle motion into a circular path. After the point of maximum displacement (the peak of the semi-circle), the particle now begins to move against the $\mathbf{E}$ field, reducing its velocity, until it eventually comes to rest with a net displacement in the $\mathbf{E}$ direction of zero. This cycle repeats, forming a semi-circular motion. Introducing collisions (indicated by black dots in Figure 2.5) complicates this picture; working under the simplified assumption that a collision brings the

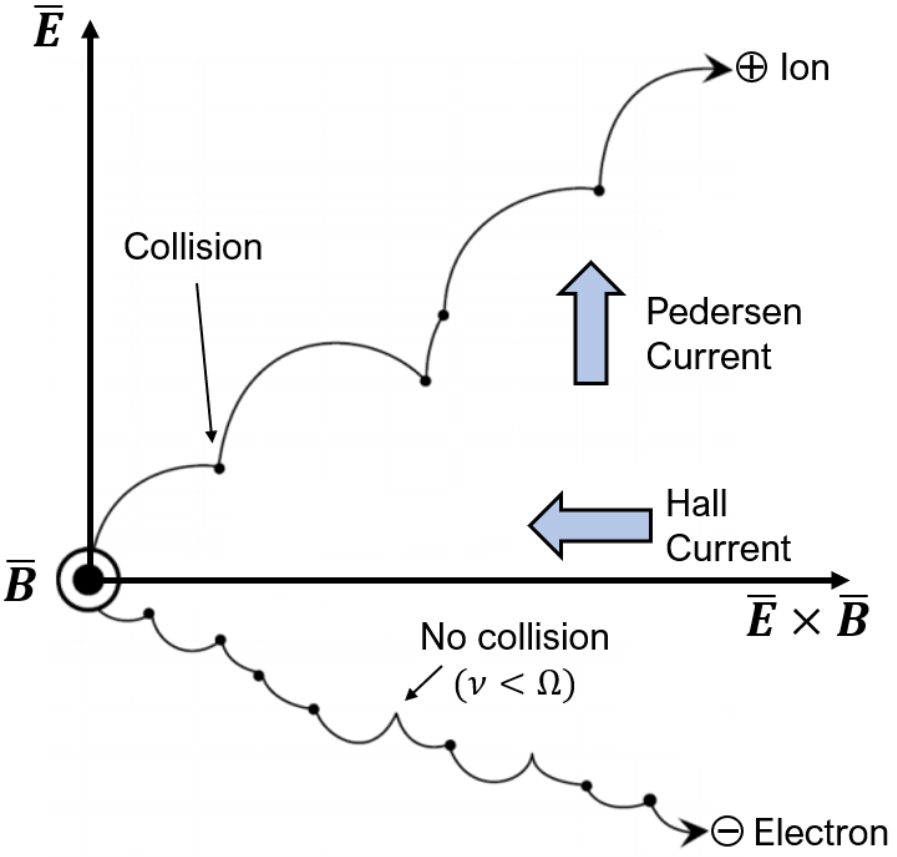


FIGURE 2.5: Schematic example displaying how plasma collisions in the presence of perpendicular magnetic and electric fields lead to Pedersen and Hall currents.

charged particle to rest, each collision resets the semi-circular cycle, breaking the positional equilibrium in the **E** direction and introducing a net drift into the particle's motion. Since electrons and ions are oppositely charged, this collision-driven component of their motion is mirrored: ions drift in the positive **E** direction and electrons drift in the negative **E** direction - indicated by the two separate particle paths in Figure 2.5.Furthermore, due to their different cyclotron and collision frequencies, the electron and ion populations exhibit a difference in their velocities in the $\mathbf{E} \times \mathbf{B}$ direction because one is being deflected from the usual $\mathbf{E} \times \mathbf{B}$ more significantly that the other.

To summarise, the introduction of collisions in the plasma derails the coupled Lorentzian motion of the two charge populations, in turn resulting in a net motion of charge in a particular direction - which forms a current.

The decoupled ion and electron motion in the **E** direction results in the Pedersen current, and that in the $\mathbf{E} \times \mathbf{B}$ direction results in the Hall current. The magnitudes of the Pedersen and Hall currents are parameterised by their respective conductivities, which incorporate the relevant ionisation fractions, collision frequencies and cyclotron frequencies to determine how easily each of the currents can flow. The Pedersen conductivity peaks at both 140 and 80 km where the motions of either ions or electrons are significantly impacted by collisions. The Hall conductivity peaks near 100 km, where

both of the populations are affected by collisions.

2.2.3 Auroral Emission

Charged particles from the magnetosphere, accelerated by mechanisms that are still unclear, enter the ionosphere by escaping the magnetic bottle and precipitating down magnetic field lines. Upon arrival they collide and exchange energy with atmospheric constituents of both the neutral thermosphere, and the ionosphere, creating the aurora. The precipitating particles lose some kinetic energy with each collision, and are eventually stopped. More energetic particles are therefore able to penetrate further into the atmosphere before losing all their energy, and have an increased likelihood to interact with the heavier atoms and molecules present at lower altitudes. There are two primary outcomes of these collisions: ionisation, in which an atom or ion loses one of its bound electrons, increasing its charge and releasing a free electron; and excitation, wherein an atom's or molecule's electron(s) enter a higher energy level, but remain bound. The free or 'secondary' electrons produced by ionisation continue to propagate through the atmosphere, and can also interact further with particles. An extremely large number of secondary electrons are produced, contributing the majority of the ionisation and excitation that occurs during this process. Once excited, atoms and molecules can return to their non-excited states either via quenching or by emitting photons which are observed as aurora. Quenching is the de-excitation of excited particles via collision with other particles. It inhibits auroral emission if the mean free path of the excited particle is small enough that it is likely to collide with another before the lifetime (time before de-excitation via emission) of the emission passes. Quenching is most efficient at low altitudes where the atmospheric density is increased, or for emissions with long lifetimes where there is a increased time window for a quenching collision to occur.

The wavelengths of photons emitted by excited molecules or particles are directly linked to the change in energy level of the bound electrons, and as a result they are highly dependent on the species of the emitting atom or molecule. Atomic emissions (e.g. from O) produce emission lines at specific wavelengths; the distinctive green colour associated with the aurora originates from the atomic oxygen emission at 557.7 nm (the 'auroral green line'). Molecular emissions (e.g. from N_2 or O_2) produce more complex and structured 'band' emissions, covering a range of wavelengths, that arise due to the large variety of energy levels made available by the rotational and vibrational freedom of molecules.

A particular precipitating particle with a unique characteristic energy will deposit that energy via primary and secondary electron collisions at a range of altitudes. The ratio of molecular concentrations at the altitudes at which this energy is deposited determines the relative intensities of the resulting emission. Low energy particles require few interactions before exhausting all their kinetic energy, and therefore produce a limited number of secondary electrons, restricting their energy deposition to high altitudes. Conversely, high energy particles can interact many times and thus produce

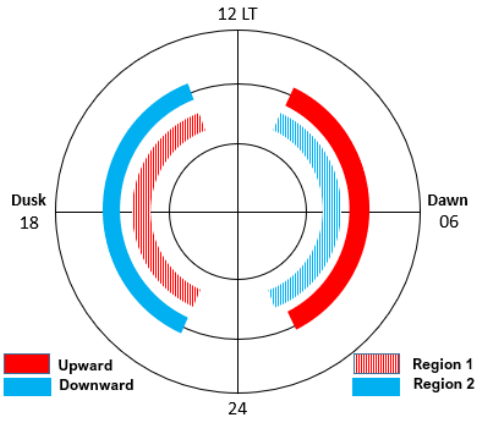
large numbers of secondary electrons, increasing the total amount of energy deposited and distributing it over a large range of ionospheric altitudes. The altitudes at which the energy is deposited impacts the relative concentrations of atmospheric constituents that are available to produce auroral emission (Figure 2.4) and thus different altitudes will result in a different distribution of emission.

Understanding of the relationship between altitude and emission characteristics allows for the determination of useful analytic quantities. By observing the ratio of emission intensities, the altitude of emission and an estimate of the precipitation energy can be determined. The specific auroral emissions used in this thesis are given in Chapter 4.

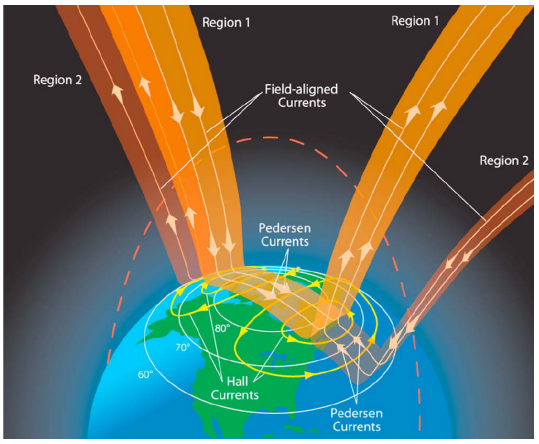
2.2.4 Auroral Electrodynamics and Joule Heating

The net motion of charge, such as that carried by the charged particles flowing down magnetic field lines, constitutes an electrical current. Thus, the quantification and analysis of the resulting current systems associated with the aurora is a key part of understanding auroral formation mechanisms and dynamics. It has long been understood that the auroral displays present at high latitudes are linked with a system of extensive magnetic field-aligned currents (FAC), which flow between the magnetosphere and the ionosphere [Iijima and Potemra, 1976, 1978]. The general picture is of two large rings of field-aligned current, approximately aligned with the auroral oval, and separated into 'Region 1' currents at high latitudes ($\sim$66–76°) and 'Region 2' currents at lower latitudes ($\sim$62–72°), with some crossover depending on the magnetic local time. These regions display opposite polarities at a given MLT and are closed through the ionosphere via meridional Pedersen currents, some exceptions to the separation of the two regions can be found near noon and midnight MLT where region 1 and 2 currents of the same polarity may overlap. A further ionospheric current system is also observed in the form of significant Hall currents flowing horizontally sunward across the centre of the polar cap, and anti-sunward along the oval in the morning and evening sectors. The latter of these Hall currents are known as the westward and eastward auroral electrojets respectively [Boström, 1964]. An idealised schematic of the described current system is shown in Figure 2.6. The overarching picture of upward and downward flowing region 1 & 2 currents is complicated by a variety of smaller-scale currents which are embedded within the large scale pattern. FAC systems in auroral regions have been observed to display a variety of sizes, ranging from between a few hundreds of meters to up to $\sim$500 km in horizontal extent [Lühr and Warnecke, 1994, Stasiewicz and Potemra, 1998, Wu et al., 2017], suggesting that the net flow of the region 1 & 2 currents is comprised of the sum of both upward and downward smaller-scale FACs. From these observations, it is a simple assumption to associate these small-scale current systems with the widely varied and extraordinarily dynamic structures observed across similar scales in the optical aurora. However, these small and often intense FAC structures are difficult

to accurately quantify from satellite observations, due to the spatial and temporal ambiguity inherent to the relative velocity of the satellite with respect to the size of the observable features.



(A) A simple top-down cartoon displaying the generalised locations of the Region 1 & 2 current systems in the Northern Hemisphere.



(B) Schematic overview of the current systems observed in the polar-cap. The field aligned Region 1 and 2 currents a connected through the ionosphere by meridional Pedersen currents. Hall currents flow in a two-cell pattern and form the auroral electrojets in the auroral oval. Figure from Wikimedia Commons

FIGURE 2.6

The electric currents flowing through the ionosphere - in the form of the aforementioned field-aligned, Pedersen and Hall currents - have important consequences for the energy balance of the entire magnetosphere-ionosphere (M-I) system. It is generally believed that there are two primary components through which magnetospheric energy is deposited into the ionosphere: Joule heating and particle heating. Ionospheric Joule heating results from thermal collisions between atmospheric neutrals and ions due to a net difference in their bulk velocities, and is typically associated with Pedersen currents, and hence ionospheric electric fields (see Section 2.2.2). Joule heating can occur on many scales, large scale ionospheric Joule heating is driven by variability in the polar convection field, and smaller more intense Joule heating can arise from short lived and localised electric fields associated with aurora. Particle heating is produced from the energy degradation of precipitating particles down magnetic field lines. These magnetospheric particles interact and exchange thermal energy with the neutral population of the thermosphere. Of these two mechanisms, large-scale ionospheric Joule heating is understood to be the most significant factor in the dissipation of the total solar energy entering the magnetosphere. Sharber et al. [1998] report that, for a particular storm, the northern hemispheric Joule heating rate (JHR) can be as high as 290 GW, whereas the corresponding power deposited via direct particle participation is approximately 50 GW. This result, among others, suggests that Joule heating is responsible for dissipating approximately 55-65% of the total energy budget of the magnetosphere-ionosphere system [Fujii et al., 1999, Østgaard et al., 2002]. It is clear then, that between

both Joule heating and particle precipitation heating, a large amount of magnetospheric energy is deposited directly into the polar atmosphere, with much variability in both time and space. Accordingly, the accurate inclusion of these processes in both local and global atmospheric models is essential to developing a complete understanding of their impact [Ridley et al., 2006]. However despite their significance, the spatial distribution and magnitude of these mechanisms have historically carried a large amount of uncertainty, and many models resort to computation using mean or averaged convection and precipitation patterns, smoothing out small-scale variability [e.g. Chun et al., 2002, Foster et al., 1986]. For example, when calculating the local or global ionospheric Joule heating, studies generally start with an expression for the Joule heating rate, typically given as the following:

$$Q_j(h) = \sigma_p(h)[\mathbf{E}_\perp + \mathbf{u}(h) \times \mathbf{B}]^2 \tag{2.2}$$

where $\sigma_p(h)$ is the Pedersen conductivity, $\mathbf{E}_\perp$ is the perpendicular (with respect to the magnetic field) electric field, $\mathbf{u}(h)$ is the thermospheric neutral wind velocity, $\mathbf{B}$ is the ambient geomagnetic field and h is the altitude. The aforementioned difficulties in determining the relevant ionospheric electric field patterns and conductivities often result in relying on a number of simplifying assumptions to produce an approximation of the total Joule heating. One such approach is the identification and utilisation of relevant magnetic indices, which are then used as proxies for the global Joule heating rate [Chun et al., 1999, 2002, Kosch and Nielsen, 1995]. Another approach is to solve either a height integrated version of Equation 2.2, in which the neutral wind speed is often assumed to be negligible [Kosch et al., 2011], or to solve the equation for fixed altitude points at which the required parameters are measurable [Anderson et al., 2013]. For larger-scale hemispheric attempts, in which full coverage of the required conductivities and electric fields cannot be realistically measured, researchers have employed a combination of data sources and statistical models; estimated maps of the relevant electrodynamic properties in the polar regions are produced, which can then be used to determine the desired Joule heating rates, among other distributions of interest [Richmond, 1992, Richmond et al., 1990]. Occasionally, ionospheric measurements are circumvented entirely and Joule heating rates are derived using empirical models based on satellite acquired solar-wind data [Weimer, 2005]. However, the majority of estimates are highly reliant on measurements and observations to inform the relevant models and equations. It stands to reason that enhanced quality of observation will therefore result in higher quality Joule heating estimates. A large part of the uncertainty arises due to lack of high-resolution (both spatially and temporally) observations of Joule heating on scale sizes below what is easily observable with current techniques [Codrescu et al., 1995, Rodger et al., 2001]. Since we know the heating mechanisms are associated with auroral structures, and extremely variable auroral forms are readily observable on scale-sizes as small as 100 m, events of this type are a logical place to start when attempting to quantify local thermospheric heating effects.

2.2.5 Auroral Arcs

Perhaps the simplest sub-oval auroral structure is the auroral arc; an east-west aligned structure with variable thickness [Partamies et al., 2010], that is relatively steady on timescales between 10 – 100 s [Marghitu, 2012]. The electrodynamic structure of auroral arcs has been extensively studied for many years using radar and rocket-borne instrumentation. Marklund [1984] combined personal observations with further results from the literature to develop an arc classification scheme that is based on observed electric field features. He defined three main categories of arc: polarisation arcs, Birkeland current arcs and combination arcs (a transitional case). The categories and their names are determined by the relative contribution of polarisation electric fields and Birkeland (field-aligned) currents to maintaining the required current continuity in the ionosphere. In the polarisation regime, the background polar convection field drives a charge separation process in the region of heightened conductivity associated with the arc, 'closing' the current system. However, once the resulting polarisation field produced by the charge separation is of equal magnitude to the background field, resulting in a net-zero field within the arc, this mechanism can no longer provide current continuity. In this regime, a field-aligned current system is required to ensure closure of the current loop, with the upward field-aligned current associated with the aurora being closed by an equivalent downward field-aligned current adjacent to the arc. In this 'Birkeland' regime, the two field-aligned currents are connected via a horizontal Pedersen current in the same way as the larger-scale Region 1 & 2 systems are (see Section 2.2.4) meaning that a horizontal electric field is still required to produce the requisite Pedersen current. Indeed, a common but not universal feature observed in multiple case studies of auroral arcs is an enhanced electric field perpendicular to the arc's alignment (e.g. north or south directed) [Aikio et al., 1993, Lanchester et al., 1996, 1998, Marklund et al., 1982]. These electric fields are observed adjacent to the visual arc structure but are often reported to be asymmetrically distributed; they are generally found to be on the equatorward edge of the arc in the evening sector, and the poleward edge in the morning sector. This spatial configuration is thought to relate to the direction of the background ionospheric convection field [Aikio et al., 1993, Marklund et al., 1982, Opgenoorth et al., 1990, Timofeev et al., 1987]. The mechanism behind this asymmetry is discussed more thoroughly in Chapter 5. Crucially, the presence of electric fields and the resulting Pedersen currents creates the potential for Joule heating processes on the same scale and, correspondingly, multiple radar and rocket studies of discrete auroral arcs have reported localised observations of enhanced ion or electron temperatures adjacent to the emission region [Han et al., 2019, Kosch et al., 2011, Lanchester et al., 2001, Opgenoorth et al., 1990].

The magnitude of ion heating can be determined using the approximation that during quiescent periods the ion temperature is equivalent to the neutral temperature [Nozawa et al., 2006] and then comparing the enhanced ion temperatures to the estimates of the neutral temperature. For example, Maeda et al. [2005] reported that

during periods of intense electric field enhancements ($>$ 100 mVm^{-1}) Joule heating can produce ion temperatures that exceed that of the modelled neutral temperature by upwards of 300 K [Maeda et al., 2005]. Whilst ion temperatures are relatively easy to measure remotely via incoherent scatter radar, neutral temperatures are comparatively much more difficult to determine - usually requiring an expensive and one-off rocket campaign. As a result, the magnitude of the response of the atmospheric neutral temperature to local scale auroral heating mechanisms is not well understood. Kurihara et al. [2009] recorded N_2 rotational temperatures from an instrument on-board a Japanese sounding rocket during an auroral event; their measurements showed a uniform neutral temperature enhancement of 70–140 K in the 110–140 km altitude range, with the MSIS-E-90 atmospheric model used as the baseline. They attributed the majority of the heating to a 'passive energy deposition' rate, which they describe as an alternative to Joule heating wherein the neutral wind effect is neglected, and reported a peak heating rate of 0.5 μWm^{-3} at 120 km. Measuring atmospheric neutral temperatures via N_2 rotational spectra is an often employed diagnostic tool (see Jokiaho et al. [2008] and Section 2.3) but has previously been limited to measurements of altitude averaged emission profiles, or one-time rocket trajectories (as above), making it hard to infer the presence of any local heating [Henriksen et al., 1987].

The focus of the research contained in this thesis is the measurement of neutral temperature altitude profiles in the altitude region of auroral emission, at both high temporal and spatial resolution. The novel observational technique used to measure thermospheric temperatures is presented in Chapter 4.

2.3 Neutral atmospheric temperature measurements

This section gives an overview on how various neutral atmospheric temperature measurements are obtained via ground based observations. To facilitate clarity during the discussion of auroral emissions here and in the remainder of the thesis, a brief discussion on atomic and molecular spectroscopic notation is also included.

2.3.1 Spectroscopic Notation

Atomic and molecular bonding, reactivity and excitation are dominated by the behaviour of electrons residing in the outermost orbital or energy level of the corresponding atoms. These electrons are known as 'valence' electrons. The characteristics of each electron can be described by various quantum numbers - each denoting the value of a conserved quantised property, such as spin. The 'configuration' of an orbital is expressed using these quantum numbers as:

$$n(l)^x \tag{2.3}$$

where n is the principle quantum number, l is the orbital quantum number and x is the number of electrons. Each of an atom's electrons fills the available orbitals following a strict set of placement rules. As a result, an atom can be represented by a full set of configurations (as defined in Equation 2.3), detailing the location of each of the electrons, for example:

$$1s^2 2s^2 2p^3 \tag{2.4}$$

is the full electron configuration for Nitrogen - containing two electrons in the 1s-orbital, two electrons in the 2s-orbital and three electrons in the 2p-orbital, for a total of seven electrons. The sum of properties from all of the electron configurations in a particular atom can be expressed as its 'state', defined as:

$$^{M_s}L_J \qquad \text{or} \qquad ^{2S+1}L_{L+S} \tag{2.5}$$

where S is the sum of individual electron spins ($\sum_i s_i$), M_s is the electron spin multiplicity, which is equal to $2S+1$, L is the orbital quantum number, and J is the total angular momentum ($L+S$). Each electron contributes either $+\frac{1}{2}$ or $-\frac{1}{2}$ to the total spin, thus, paired electrons and full orbital shells contribute a net spin of zero to the total. A 'singlet' quantum state is defined as a state with no unpaired electrons ($S=0$), a 'doublet' has a single unpaired electron ($S=\frac{1}{2}$) and a triplet has two unpaired electrons ($S=1$).

For an atom, the 'term' is a full representation of the relevant electron properties; it includes the configuration of the valence electrons and the state of the atom, as they are defined above. 'A multiplet' includes all possible transitions between two terms, for example:

$$3s(^5S) - 3p(^5P) \qquad\qquad Configuration(State) \tag{2.6}$$

and a 'line' is a specific transition between two terms, subject to appropriate selection rules on the total angular momentum J.

Molecular structure has a separate term symbol, defined as:

$$^{M_s}\Lambda^{+/-}_{\Omega,g/u} \qquad \text{or} \qquad ^{2S+1}\Lambda^{+/-}_{\Omega,g/u} \tag{2.7}$$

where M_s and S are the spin multiplicity and total spin quantum number, respectively, Λ is the orbital angular momentum along the inter-nuclear axis and Ω is the total angular momentum along the inter-nuclear axis. The $(+/-)$ and (u/g) sub and super-scripts refer to symmetrical and reflectional properties of the molecule. Molecular states with the same spin multiplicity form a 'family'. The ground state of a family is denoted by a preceding $'X'$ before the term symbol, with the following increasing energy levels labelled alphabetically (A, B, C, ...etc.).

A molecular band system consists of all possible transitions between two molecular electronic states, and is represented as:

$$B(^3\Pi) \rightarrow A(^3\Sigma) \tag{2.8}$$

A molecular state has further vibrational and rotational degrees of freedom, afforded to them by their multi-atomic structure, described by the quantum numbers ν and J, respectively. A molecular band consists of all possible transitions between a pair of two specific vibrational levels:

$$B^3\Pi(\nu' = 3) \rightarrow A^3\Sigma(\nu'' = 1) \tag{2.9}$$

where ν' and ν'' denote high and low vibrational energy levels. A rotational line is a transition between rotational energy levels, for a fixed vibrational band, for example:

$$B^3\Pi(\nu' = 3, J' = 3) \rightarrow A^3\Sigma(\nu'' = 1, J'' = 4) \tag{2.10}$$

A branch is composed of all lines with the same ΔJ value; the Q-branch has a ΔJ of 0, the R-branch has a ΔJ of -1 and the P-branch has a ΔJ of 1.

2.3.2 Neutral Temperatures

The observation and analysis of key auroral emissions provides a unique opportunity to determine neutral atmospheric temperatures at altitudes which are difficult to measure otherwise. The fundamental principle behind these measurements is that the spectral characteristics of certain auroral emissions are dependent on the thermodynamical temperature of the emitting species. Chemical theory and lab based experiments are used to investigate the relationship between emission characteristics (such as relative line intensities and spectral shapes) and thermodynamic temperature. This knowledge is then used to retrieve an estimate of the temperature from spectral observations taken in the field. Temperature measurements of this kind typically rely on two assumptions relating to the emitting species:

1. The population in the low energy state is thermalised through frequent collisions with the neutral atmosphere - they are in local thermodynamic equilibrium (LTE) with the neutrals.

2. The emitting species is limited to a sufficiently thin atmospheric layer such that there are no significant temperature gradients across the extent of the emission region. The emission can be considered to correspond to a single uniform temperature.

The first of these assumptions is generally well accepted at auroral altitudes. For example, the rotational temperatures of neutral atmospheric molecules such as N_2 are

expected to be approximately equal to the neutral temperature in the lower thermosphere [Kurihara et al., 2006].

The second assumption requires careful consideration; its validity is highly dependent on both the species responsible for the emission and the method of observation. As discussed in Section 2.2.3, emissions from oxygen and nitrogen molecules can originate from a wide range of altitudes. Therefore, ground based observations are the sum of the emission from all the altitudes along the observed atmospheric column. The result of this height-integration effect is that observations are dominated by the brightest section of the altitude profile, and any deduced atmospheric characteristics will be heavily weighted by that region. OH airglow emission (non-auroral) provides an exception to this rule, since it arises from a relatively narrow layer of excited hydroxyl molecules near the mesopause, over which the temperature can be assumed to remain approximately constant. The aurora almost never occurs in a thin altitude sheet like this, and as a result there will always be significant temperature gradients throughout the emission region. In-situ measurements from rocket-borne instruments can overcome this limitation by observing emission as they pass directly through the altitudes of interest, but are limited otherwise by their one-off nature. The specifics of the techniques used to derive atmospheric temperature measurements from both OH and N_2 emission are discussed in the following sections.

2.3.3 OH Temperature

Measuring the temperature of the excited hydroxyl molecules found in the 'OH layer' is a well understood and commonly applied atmospheric temperature measurement. The OH layer is the source of OH airglow emission in the Meinel band [Meinel, 1950], a consistent and non-auroral atmospheric emission from excited OH radicals produced via two body collisions between hydrogen atoms and ozone [Bates and Nicolet, 1950]. The peak emission altitude is dependent on the vibrational energy level of the radical [Savigny et al., 2012], and is located between approximately 80 and 90 km. The range of allowable vibrational levels translates to a total mean thickness of approximately 8 km [Baker and Stair Jr, 1988], which is a relatively narrow region in comparison to auroral emission. Crucially, as mentioned above, the thin altitude extent of the emission region means that the neutral temperature profile is approximately constant such that all of the OH is emitting at the same temperature. Thus, the ground based drawback of observing the entire atmospheric column is minimised, and a single temperature can be reliably assigned to the emission. The discovery of the OH layer, which is located at a consistent altitude and in a previously difficult to measure portion of the atmosphere, unlocked an extensive field of further research [Sivjee, 1992].

It is possible to derive the OH temperature by determining the ratio of intensity between various emission lines. Assuming that the population in each of the energy levels is described by a Boltzmann distribution, the intensity (I) of a particular Meinel band line can be calculated as [Mies, 1974]:

$$I = N_{v'} A \frac{2(2J'+1)}{Q_{v'}(T)} exp\left[\frac{-hcF}{kT}\right] \tag{2.11}$$

where $N_{v'}$ is the total number of molecules in that vibrational level (v'), A is the transition probability (Einstein coefficient), J' is the total angular momentum quantum number of the upper state of the transition, Q'_v is the partition function, F is the energy level of the initial rotational level, T is the rotational temperature, c is the speed of light and h & k are the Planck and Boltzmann constants, respectively.

Defining the emission intensities of two OH lines with different upper rotational states, but the same initial state, as I_a and I_b, and assuming that the rotational temperature of the two states is the same ($T_a = T_b$), an expression for that rotational temperature can be obtained by deriving the quotient of the two:

$$T = \frac{hc(F_b - F_a)}{k} \left\{ \ln\left[\frac{I_a A_b (2J'_b + 1)}{I_b A_a (2J'_a + 1)}\right] \right\}^{-1} \tag{2.12}$$

Therefore, with observations of I_a and I_b and values for the remaining constants from the literature, an OH rotational temperature can be easily calculated. However, a more robust method for calculating the OH temperature is to use a number of measured OH line intensities to create a Boltzmann plot [e.g. Chadney et al., 2017, Sigernes et al., 2003]. A Boltzmann plot is a plot of $\ln\{I/[A(2J+1)]\}$ as a function of $(hc/k)F$, which produces a straight line with a gradient equal to the negative reciprocal of rotational temperature $-1/T$ (see Equation 2.11). A linear fit can then be calculated to give a final value of the gradient and therefore temperature. Furthermore, other products of the analysis such as the variance of the fit give a quantitative measure of how valid the assumptions inherent to the method are. A high variance could indicate the breakdown of LTE in the OH layer, or it could be a symptom of the measured OH line intensities being affected by external factors such as auroral contamination or water vapour absorption. Because of this, the variance is a convenient metric to help with identifying and discarding unsuitable observation intervals.

During intervals of weak auroral emission, HiTIES is able to retrieve accurate observations of OH(8-3) emission lines in the 728-740 nm wavelength region. The intensities of four strong P-branch lines ($P_1(2)$, $P_1(3)$, $P_1(4)$ & $P_1(5)$) are determined using a least squares fitting technique (discussed in detail in Section 4.7) and are used to produce a Boltzmann plot and subsequently calculate an estimate of the OH temperature.

2.3.4 N_2 Temperature

A particularly relevant and powerful outcome of the past century's developments in molecular theory is the ability to generate synthetic band emission spectra from fundamental molecular characteristics. A molecule holds quantised potential energy in

its electronic structure and in molecular vibrational and rotational terms. Steady state solutions of the Schrödinger equation exist for the vibrational and rotational energy levels of the molecule, which give the distribution of emission wavelengths for all possible transitions, and quantum selection rules and Einstein coefficients provide the relative intensity of each corresponding line. A full consideration of these factors results in the ability to produce a synthetic emission spectra, given the appropriate inputs. An extensive discussion on the topic of molecular spectra is provided in Herzberg [1950], and will not be repeated here. The most relevant aspect of molecular theory is the relationship between a molecule's rotational temperature and the shape of its emission bands. Figure 2.7 displays an example of this; the top panel shows a single synthetic emission spectra (black curve) in the HiTIES wavelength range, which is derived from the convolution of the full selection of theoretical emission lines (grey lines) that correspond to a rotational temperature of 200 K. The convolved spectra for a range of rotational temperatures are plotted in the bottom panel. It is evident that the large scale structure of the synthetic spectra depends heavily on the rotational temperature. Thus, by obtaining spectral measurements of rotational molecular bands and analysing and comparing their structure to the synthetic equivalents, estimates of the neutral temperature of the local atmosphere can be determined - subject to the assumptions described in Section 2.3.2.

Koehler et al. [1981] presented some of the first results of atmospheric rotational temperatures deduced this way, using emission from N_2^+ ions. They employed narrow band interference filters on photometers to measure the P and R branches of the $N_2^+ 1N(0,1)$ band emission. The ratio of intensities of these two branches provided an appropriate proxy for the variation in the total shape of that rotational band, which in turn allowed for an estimate of the rotational temperature to be calculated. This type of observation was widely used when adequately high resolution spectral measurements of entire band structures were not easily obtainable. However, improvements in imaging instrumentation have opened the way to more accurate assessments of spectral rotational temperatures. Scanning spectrometers or imaging spectrographs can produce high resolution emission spectra over wavelength regions that contain large sections of relevant spectral bands. These emission spectra can then be fit with synthetic spectra to determine more accurate rotational temperatures.

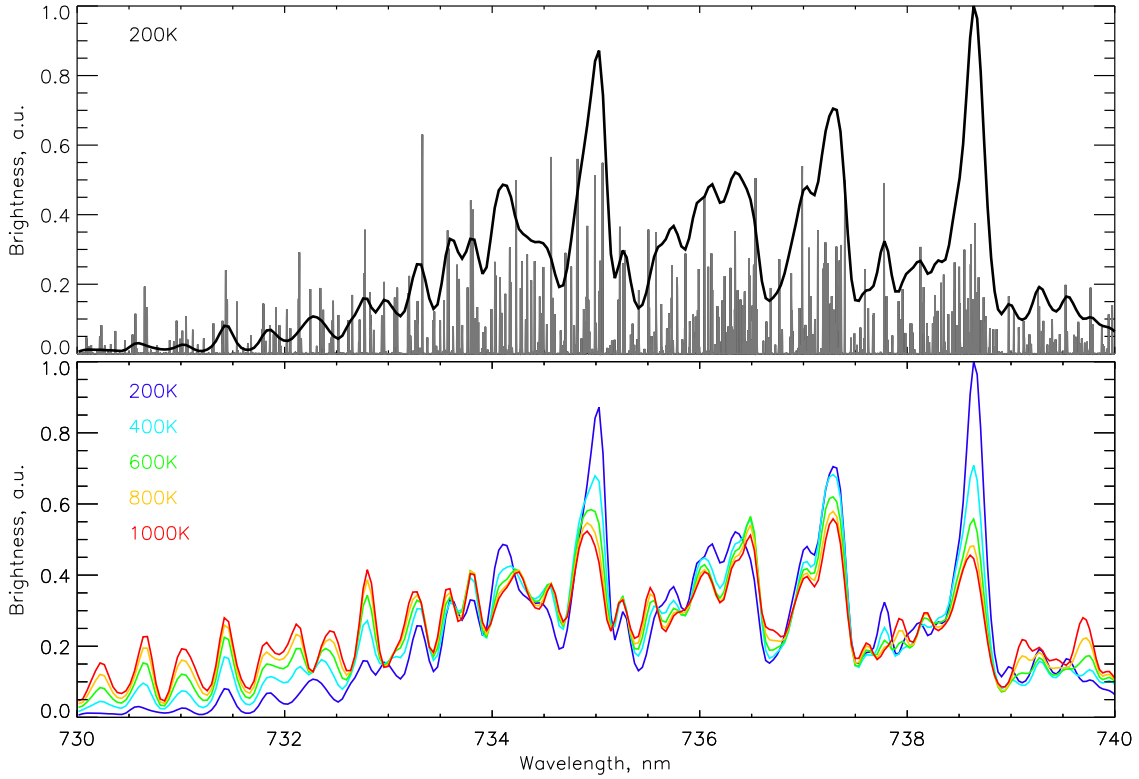


FIGURE 2.7: Example N_2 emission spectra. Top panel: component line spectra (grey lines) and corresponding convolved spectra (solid black line) for N_2 emission at 200 K. Bottom panel: example convolved N_2 emission spectra for a range of emission temperatures.

Chapter 3

Instrumentation and Models

The research conducted and outlined in this thesis relied heavily on the application and use of data from a variety of instruments and sources. The primary source of data were observations of the aurora by two Southampton owned and operated optical instruments: HiTIES (High Throughput Imaging Echelle Spectrograph) and ASK (Auroral Kinetics and Structure). HiTIES and ASK are approximately co-located on the Arctic archipelago Svalbard. Optical observations were supported by simultaneous ESR (EISCAT Svalbard Radar) campaigns and model runs performed by Southampton's Ion chemistry and electron transport model. Throughout the course of the PhD, I participated in six fieldwork campaigns to Svalbard where I assisted in the running and maintenance of the optical instruments. During these campaigns I was also at times responsible for the running and operation of the EISCAT Svalbard radar.

3.1 HiTIES

The High Throughput Imaging Echelle Spectrograph (HiTIES) is the primary instrument on the Spectrographic Imaging Facility (SIF, University of Southampton and University College London). SIF (and HiTIES) is located at the Kjell Henrikson Observatory (KHO) on Svalbard, at geographic coordinates 78.15 N, 16.04 E. HiTIES is an imaging spectrograph that is capable of measuring a selection of different wavelength regions simultaneously and at high spectral resolution [Chakrabarti et al., 2001]. The advantage provided by HiTIES is that it allows for the design and implementation of a specialised mosaic filter that is capable of separating overlapping diffraction orders to produce simultaneous high spectral resolution observations from multiple distinct wavelength windows (discussed in Section 3.1.1). HiTIES can therefore accurately observe specific wavelengths of interest, regardless of the total wavelength separation between them, and without the need for either an extremely large detector or a process involving stepping through the wavelength region and sacrificing temporal resolution.

Observations are made in the local magnetic zenith via a north-south aligned slit with an angular size of 8° along the slit and 0.05° across it. Incoming light is collimated, diffracted by a reflective echelon grating and imaged onto the plane of the mosaic filter,

which sorts and separates the diffraction orders. Finally, the light is re-imaged onto a position sensitive Electron Multiplying Charge Coupled Device (EMCCD). The dimensional axes of the final image plane are wavelength of emission and position along the observational slit. As a result, observed emissions are also separated spatially, with the centre of the image plane corresponding to the magnetic zenith (observed by the centre of the slit) and the remainder of the axis imaging a range of 4° to the north and south of the zenith. A schematic of the layout of the HiTIES instrument is displayed in Figure 3.1.

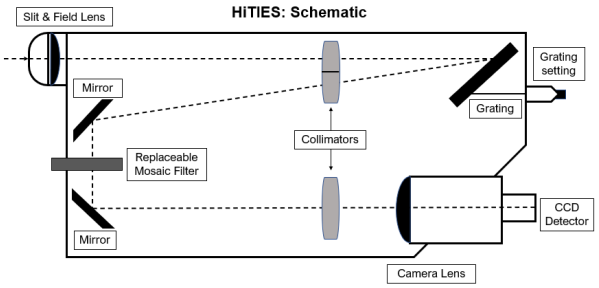


FIGURE 3.1: A simple schematic of the optical layout of the HiTIES instrument. The echelle grating is adjustable using the grating setting dial on the outside of the instrument (right). The mosaic filter is placed in the filter plane between two reflective mirrors (left). Adapted from Chakrabarti et al. [2001].

The fundamental principle of a diffraction grating is described by the grating equation:

$$m\lambda = d(sin(\alpha) + sin(\beta)) \tag{3.1}$$

where m is the diffraction order, λ is the wavelength, d is the groove spacing, α is the incident angle, and β is the diffracted angle. α is determined by the angle of the echelle grating, known as the 'grating setting', and is controlled using a micrometer on the outside of the instrument. Altering the grating setting within the range accommodated by the layout of the detector allows for the wavelengths present in each diffraction order to be adjusted. HiTIES operates at high diffraction orders (20-50), such that for a given grating setting Equation 3.1 can be satisfied for a range of wavelengths and orders, resulting in a significant amount of order overlap.

3.1.1 Mosaic Filter

The separation of overlapping diffraction orders using a mosaic filter provides HiTIES with the advantage of being able to simultaneously image a number of distinct wavelength regions, without losing any temporal or spatial resolution. The mosaic filter is a 50 mm by 50 mm set of interference filters which is placed at the first image plane of the detector. The filters are positioned alongside each other in a patchwork or mosaic pattern, and are joined together with an extremely small joint width ($<$0.5 mm). Each filter has a specifically designed transmission curve, allowing the diffraction orders of interest to be transmitted, whilst blocking the others. Thus, by carefully designing and

selecting a number of individual filters to comprise the mosaic, HiTIES can target wavelength regions and emissions of interest. The mosaic utilised in this work is shown in Figure 3.2.

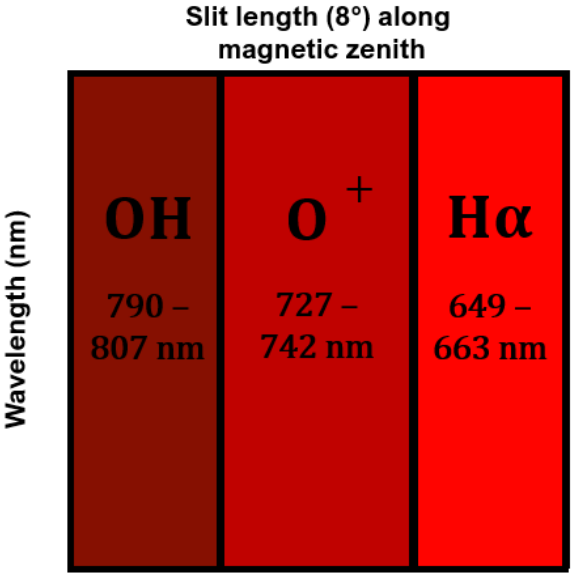


FIGURE 3.2: Schematic layout of the HiTIES mosaic filter in use during the observations discussed in this thesis. The central panel of the mosaic filter (728-740 nm), labelled as O^+ as this was the initial emission of interest, is predominantly used during the research presented in this thesis.

The primary mosaic panel used during the observations discussed in this thesis utilises the central 2° of the slit and has a wavelength range between 727 and 742 nm, with a resolution of 0.12 nm. This panel contains a few emissions of interest; the most notable and relevant for this work is the N_2 1PG vibrational band ($B^3\Pi$ - $A^3\Sigma$). The N_2 1PG band is the prevailing structure in the observed spectra during intervals of bright aurora. The specific bands observed in this region are the (4,1) and (5,2) bands, the latter of which has a distinct band head at approximately 738 nm. This panel also contains emissions from two auroral O^+ doublets (λ = 732.01/731.90 nm & λ = 733.08/732.97 nm), the OH(8-3) airglow band and a minor contribution at the short wavelengths from the O_2^+ 1N auroral emission ($b^4\Sigma_g^-$ - $a^4\Pi_u$).

A second panel from the same mosaic was also utilised in a contribution towards Reidy et al. [2020]. This panel also contains contributions from N_2, O_2^+, N^+ and OH emission, but notably it also includes Hα emission. The Hα line is a key diagnostic tool for detecting the presence of proton aurora and the isolation of this emission provided supporting evidence for the conclusions of Reidy et al. [2020], which is discussed in more detail in Chapter 7.

3.1.2 Observations

HiTIES observes the raw auroral spectrum at half-second resolution, but to ensure an adequately high signal-to-noise ratio, the emissions are typically post-integrated up to

a two minute resolution. However, during intervals of extremely bright aurora the emission spectra can be sufficiently resolved at integration intervals of the order of seconds. Thus, during intense events, HiTIES is able to provide spectral observations at an approximately comparable time resolution to the corresponding small-scale auroral dynamics.

3.2 Auroral Structure and Kinetics

Auroral Structure and Kinetics (ASK) is a narrow angle optical imager composed of three cameras with identical fields of view (ASK1, ASK2 and ASK3), and two photometers. Each camera is fitted with a distinct spectral interference filter which permits the transmission of a narrow band of wavelength, centered on a specific auroral emission of interest. The filters utilised in this thesis are discussed in Section 3.2.2. Each camera has a square field of view (FOV) centered on the local magnetic zenith, with an angular size of 6.2° by 6.2°; this FOV corresponds to a spatial extent of approximately 10 km by 10 km in the E-region ionosphere, where the N_2 emission commonly originates. Each imager has a resolution of 512 $\times$ 512 pixels, which are binned into 256 $\times$ 256 'super' pixels to increase time resolution. Each super pixel therefore corresponds to a scale of approximately 40 m. The ASK cameras are time synchronised together, and operate at 20Hz during the standard operation mode used throughout the optical season, and at 32Hz during intervals when EISCAT Svalbard Radar (see Section 3.3) observations are coordinated. The careful temporal and spatial alignment of the three cameras allows for direct comparisons between each of the emissions. Images taken from a ground based instrument provide certain advantages over satellite-based or radar equivalents. Single-point measurements from a radar beam or passing satellite come with an inherent space-time ambiguity, wherein it can be difficult to distinguish whether observed dynamics are a result of variations in space or time. For example, lateral movement of ionospheric plasma into the radar beam could be confused with intrinsic changes to the ionosphere itself. By taking continuous two-dimensional images, ASK is able to disentangle these ambiguities and provide a clearer picture of the full range of the aurora's variability in both time and space. The resulting high resolution measurements of the spatial distribution and temporal evolution of selected auroral emissions unlocks a number of powerful analytical techniques [e.g. Ashrafi, 2007, Dahlgren et al., 2008, Sandahl et al., 2008].

ASK's specific contribution to the research presented in this thesis was twofold. Firstly, it provided the crucial first step in this works central observation method (discussed in Chapter 4); emission intensities from ASK, when combined with simple model runs, allow for the determination of the characteristic energy and flux of the precipitating particles responsible for the aurora's formation. Secondly, ASK also contributed high resolution observations of the dynamic motion of the optical auroral

forms, providing essential spatial context for the final neutral temperature measurements.

3.2.1 Operations

The ASK instrument is located at the EISCAT Svalbard Radar (ESR) building (78.15°N, 16.03°E), where it is separated from KHO and HiTIES by approximately half a kilometer. It resides on a shelf at the back of the building, and is consequently designed to operate outside and withstand the extreme Arctic weather conditions. A photograph of the ASK instrument and a schematic of ASK3 (which is mirrored for ASK1 & 2) are displayed in the left and right panels of Figure 3.3, respectively. The ASK cameras are mounted on an optical bench, contained within the 'bucket': the outer shroud of the directional portion of ASK. The bucket is split into two sections, one of which contains ASK3 whilst the other contains both ASK1 and ASK2, mounted side-by-side. Each half can be accessed to make minor adjustments via a small removable hatch, or they can be entirely exposed via the complete removal of the shell to perform more extensive maintenance. In Figure 3.3(B) the camera is highlighted in yellow, with the two photometers on either side of it highlighted in red. Fine-scale adjustments to the pointing direction of the cameras, to ensure alignment between all three, is manually achieved using reference stars and small screws located on the side of each camera. The lenses and interference filters are fitted directly on top of the cameras below the optional Galilean converters (telescopes). The discretional inclusion of the telescopes (highlighted in blue) provides a 2× magnification effect, reducing the ASK FOV to 3.1° by 3.1° if desired. However, the telescopes were absent during all of the observations used in this thesis, retaining the original 6.2° by 6.2° field of view. The direction of the entire bucket is controlled using hand-winches at its base, allowing for independent adjustments of the azimuth and elevation to ensure the instrument is observing along local magnetic zenith.

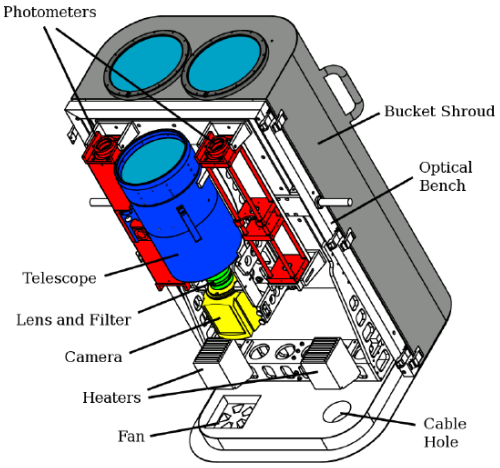
3.2.2 Observed Emissions

The installation of an interference filter allows an ASK camera to target a specific emission of interest, contained within the filter-selected wavelength region. A number of different filters and filter combinations have been used by the ASK instrument since its initial deployment in December 2006. The ASK observations presented in this thesis took place primarily during the 2016/17 observational season and utilised the corresponding interference filters attached to ASK1 and ASK3. The central wavelengths and full width half maximums (FWHM) of the filters are shown in Table 3.1.

TABLE 3.1: Properties of the relevant ASK cameras

Camera	Central wavelength (nm)	FWHM (nm)	Target Emission
ASK 1	673.0	15.0	N_2 (1PG)
ASK 3	777.4	1.3	O ($3s^5S - 3p^5P$)

(A) Auroral Structure and Kinetics during the aurora. Located on the shelf behind the ESR building. Photo credit: Dan Whiter



(B) A schematic of the components of ASK 3, located within the 'bucket'. Credit: Department of Physics Workshop, University of Southampton

FIGURE 3.3

The 673 nm filter fitted to ASK1 measures emission that originates from two N_2 1PG vibrational band transitions: (4,1) and (5,2). The same emission is observed by the HiTIES mosaic panel discussed above (see Section 3.1.1). N_2 density is concentrated at lower altitudes (Figure 2.4), and as a result, emission from these bands is responsive to high energy particle precipitation that is able to penetrate the lower thermosphere. The transmission curve of the 673 nm filter has a relatively wide full width half maximum (with respect to the other filters) which affords it an increased signal-to-noise ratio. Significant N_2 emission is, by requirement, present during all the observations discussed in this thesis and the ASK1 camera is used when images of the aurora are required for analysis of its shape and evolution.

The 777. 4nm filter fitted to ASK3 isolates atomic oxygen emission arising from the $3s^5S - 3p^5P$ multiplet. This emission arises from a mixture of direct and dissociative excitation. Direct excitation of atomic O is the dominant process for low energy electron precipitation, and dissacociative excitation of molecular oxygen (O_2) becomes more prevalent at higher precipitation energies. Consequently, the 777.4 nm emission is sensitive to a range of precipitation energies, but is most efficient for low energies.

3.2.3 Intensity Calibration

The ASK cameras measure intensity in counts/s, which is then converted into more useful units such as Rayleighs via an absolute intensity calibration process using catalogued reference stars. Multiple images of the background star field are taken during a dark clear interval that is close to the observation time and with no aurora, no cloud cover, and secondary sources of light are minimised (e.g. midnight with a new Moon,

or whilst the Moon is below the local horizon). The observed brightnesses of a number of stars in ASK are obtained by individually integrating the pixels containing them and removing a background value from each, taken from the surrounding dark region. The expected spectra of the observed stars are then retrieved from publicly available star catalogues, convolved with the ASK filter transmission curve, and integrated. This process produces an expected value of intensity in photons/s that each particular star should produce in each ASK camera. A scaling factor that reconciles the star's observed brightness with those expected from the catalogued properties is then calculated and the process is repeated at multiple times to determine an accurate absolute intensity calibration. This scaling factor is then applied to any auroral observations taken from intervals close to the calibration data in time. The sensitivity of the ASK cameras can fluctuate due to external factors such as weathering on the lens windows and general ageing of the instrument. As a result, the total calibration process is repeated often to ensure up-to-date values are used.

3.3 EISCAT Svalbard Radar

The EISCAT Svalbard Radar (ESR) is an ultra high frequency (500 MHz) incoherent scatter radar station owned by the EISCAT (European Incoherent Scatter) radar scientific association. Incoherent scatter radars (ISRs) are ground-based instruments which utilise incoherent scatter theory to observe parameters of the ionosphere. ESR is one of three radar systems operated by EISCAT in northern Scandinavia, and is the only one located on the Svalbard Archipelago. All of the radar data obtained and used for the research presented in this thesis was done so from the EISCAT Svalbard Radar. ESR consists of a fixed field aligned 42 m dish and a steerable 32 m dish which can be used in fast scanning observation modes to cover larger regions of the ionosphere, at the cost of temporal resolution. However, for all of the observations discussed in this thesis, the 32 m dish was aligned with the 42 m dish and consequently the local magnetic zenith. A photograph of the fixed dish is shown in Figure 3.4.

ISRs function by illuminating a column of the atmosphere with an extremely powerful radar beam at a fixed frequency. The frequencies used are typically between 50 MHz and 2 GHz, so chosen because they are well above the plasma frequency, (the resonance frequency of ionospheric plasma) making the radar waves almost completely un-attenuated by the ionosphere. An extremely small fraction of the radiated energy is absorbed and re-emitted (at approximately the same frequency) by the electron population in the ionosphere. High power transmitters are required in order to amplify the small echoed signal to readable levels. The strength of the returned signal is dependent on microscopic fluctuations and gradients in the atmospheric electron density - and it is these fluctuations that ISRs are measuring. However, when the ionospheric neutral density is significantly higher than the electron density, the motion of the much heavier

FIGURE 3.4: The EISCAT Svalbard Radar Fixed 42 m Dish. Photo taken by the author.

and slower moving ions exerts a heavy influence on the bulk motion of the electrons, allowing for ion dynamics to also be inferred from ISR measurements.

Assuming that the ionospheric plasma is thermalised, the shape of the returned incoherent scatter spectrum (Figure 3.5) depends on, and can be fit to retrieve, a number of useful plasma parameters:

- **Electron Density:** each illuminated electron radiates a small amount of energy so that the total power of the returned spectrum is proportional the the density of electrons in the irradiated volume
- **Line of sight plasma velocity:** the average frequency of the returned spectrum is Doppler shifted by the bulk motion of the plasma. Thus, the line of sight plasma velocity can be calculated using the shift in frequency with respect to operating frequency of the radar.
- **Ion and Electron Temperature:** the plasma constituents exhibit random thermal motion, and as a result, the frequency of the individual signal received from each electron is also Doppler shifted by a small amount. The incoherent spectrum therefore is composed of a range of frequencies surrounding the transmitted frequency. The wider the received spectrum, the greater the thermal velocities of the emitting plasma, which corresponds to higher temperature. The ion and electron

temperatures can be separated by examining the intensity of the 'wings' of the spectrum.

The altitude resolution of ISR measurements is determined using simple distance calculations for a electro-magnetic wave with a constant velocity. The return time of a radar pulse is dependent on the total distance travelled. The first received signal originates from the lowest altitude (shortest distance) and the trailing edge corresponds to the highest altitude (furthest distance). Combining this fact with a careful implementation of a time gating process on the received signal allows for the segmentation of the received spectra into distance bins, such that altitude profiles of the aforementioned plasma parameters can be created. However, ISR measurements made using a simple pulse (known as 'long pulse') measurements are limited to measuring altitudes above the most significant and interesting auroral emissions. A variety of pulse coding techniques have been developed to help lower the altitude floor of ISR measurements. The specific technique relevant to this thesis is the 'alternating-code' pulse, developed by Lehtinen and Häggström [1987]. The development of these pulse coding techniques not only allowed for measurements in the E and lower F-region ionosphere, but also greatly increased the spatial and time resolution of ISR observations.

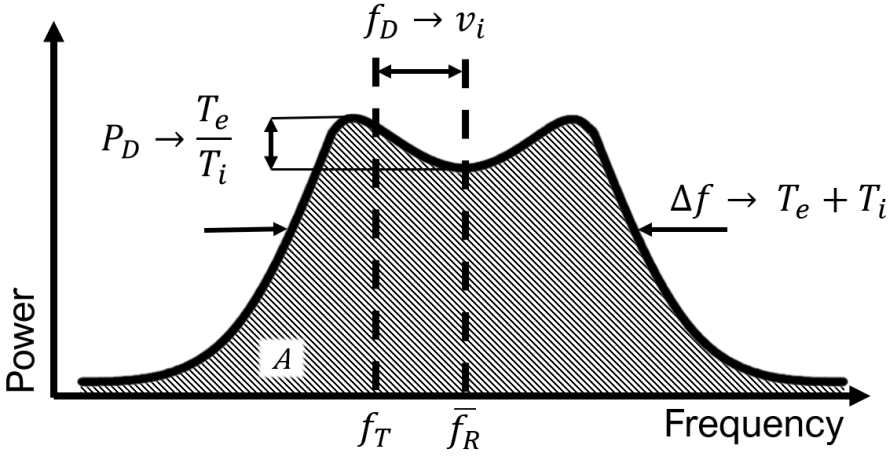


FIGURE 3.5: A schematic showing a simplified ISR returned power spectrum. The total returned power (A) is proportional to the electron density. The frequency shift (f_D) between the transmitted frequency (f_T) and the mean returned frequency $(\overline{f_R})$ gives the ion velocity (v_i). The total frequency width (Δf) gives the sum of the electron and ion temperatures. The prominence of the frequency wings (P_D) gives the ratio of electron to ion temperature.

The stochastic nature of the fluctuations in the ionospheric plasma means that a single back-scattered incoherent spectrum is not of particular use, and the signal-to-noise ratio (SNR) is too low to reliably fit most of the plasma parameters. To counteract this and increase the SNR, returned spectra are typically integrated in time up to resolutions varying from seconds to minutes. This can pose an issue for studies of small-scale and rapidly changing ionospheric conditions which can change on timescales significantly shorter than the EISCAT integration intervals, essentially 'smoothing out' interesting

features. An alternative work-around to this time resolution issue is to use EISCAT 'power profiles'; power profiles are a scaled measure of the total returned power, the simplest component of the returned spectrum, which can be used to determine an estimate of the electron density profile. Electron densities retrieved from power profiles are not as accurate as their fitted counterparts, but they are returned at a significantly improved time-resolution of approximately 0.5 seconds. Furthermore, power profiles only contain an estimate of the electron density, and the other plasma parameters cannot be retrieved at the same resolution.

For all of the observations discussed in this thesis, the radar experiment 'arc_slice' was run, which uses ESR's fixed field-aligned 42 m antennae. This experiment utilises a 64 bit alternating code that allows EISCAT to return plasma parameters at 5 second resolution and at altitude steps of 0.9 km between 85 and 481 km. Power profiles of the electron density can also be retrieved at 0.5 second resolution over the same altitude bins. The finely spaced altitude ranges and high time resolution of the arc_slice experiment make it ideal for studies of the E and F-region ionosphere.

3.4 Southampton Ion Chemistry and Transport Model

Instrumental observations used in this work were supported by the University of Southampton's Ionospheric model. The model is a time-dependent, one-dimensional ionospheric model which utilises a two-step combination of an electron transport model [Lummerzheim, 1987] and ion chemistry model [Lanchester et al., 2001, Palmer, 1995, appendix] to solve coupled continuity equations for a number of important ionospheric ions and neutral species. The electron transport code is run first, followed by the ion chemistry model. The electron transport code calculates ionisation and excitation rates at each descending altitude step and updates the newly degraded electron spectrum accordingly. Back-scattered (upward deflected) electrons and secondary electrons are also included. The ionisation and excitation rates produced at this stage are used to calculate auroral volume emission rates and intensities. The ion chemistry code solves an extensive selection of time-dependent chemical equations for all of the relevant positive ion species and produces density-altitude profiles for each. Assuming quasi-neutrality the electron density is then calculated as the total of the positive ion density. At the end of each ion chemistry step these profiles are used as the new starting ionosphere for the next step, and the process is repeated.

The model's two primary inputs are an electron energy spectrum, and MSIS-E-90 neutral concentration profiles for the predominant ionospheric species (N_2, O_2 and O). The electron energy spectrum at each time-step of the model can be specifically selected by the user and is typically in the form of either: a simple default combined Gaussian and/or Maxwellian curve centered at the desired energy and flux (with optional low energy tail), or a manually generated arbitrary electron energy flux spectrum of any form. For the high resolution and dynamic observations used in this thesis, fine

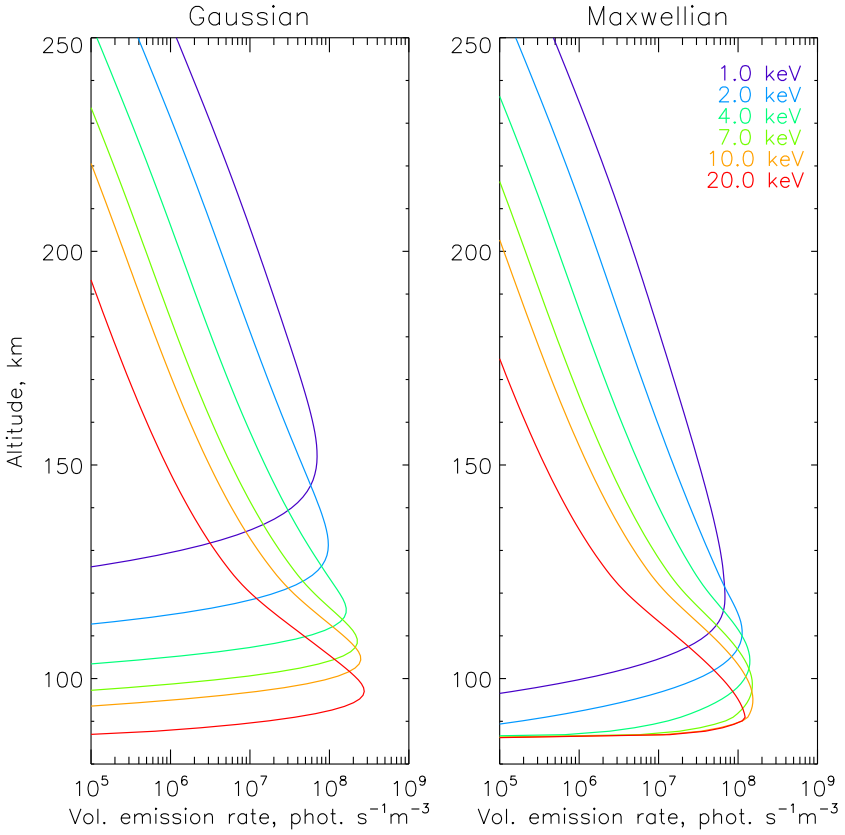


FIGURE 3.6: Modelled N2 (673 nm) volume emission rate profiles for both Gaussian and Maxwellian input spectrum over a range of energies.

control over the ratio of Gaussian to Maxwellian input spectra, and the ability to carefully adjust the individual properties of each, such as spectral width and peak energy, was desired. Since the first of the two input methods relies on pre-determined spectral widths, the runs used for this thesis almost exclusively employed the latter choice (arbitrary spectra). These spectra are generated in the energy flux vs energy domain using the standard form of each distribution for specified parameters, normalised to 1 mWm^{-2} total flux, and then scaled to match the desired input flux for the time-step in question. The MSIS-E-90 concentration profiles are generated for the specified latitude and longitude, as well as for the time of day and day of the year, and F10.7 radio flux and AP indices (magnetic activity proxies).

Volume emission rates are calculated in the model by multiplying the rate of electron impact excitation by a numerical factor which encapsulates the branching ratio, and other required coefficients, of that species. These factors and their values are determined from the relevant literature and are discussed fully in Palmer [1995]. The altitude distribution of the modelled volume emission rate profiles is highly dependent on both the emitting species and the shape of the electron energy spectrum. Emission rates naturally depend on the relative concentration of emitting species and on the altitudes at

which the magnetospheric energy is deposited in the ionosphere by the primary and (most significantly) secondary electrons. Adjustments to the peak energy, electron flux, spectral width, and overall shape of the electron energy spectrum can have a profound impact of the resulting volume emission rate profiles. Examples of the energy sensitivity for both Gaussian and Maxwellian input spectra are displayed in Figure 3.6. For a Maxwellian input with a high characteristic energy (∼20keV +), the population of electrons at extremely high energies becomes non-negligible. This high-energy population can reach to very low altitudes, resulting in a fraction of the total electron flux being 'lost' below the model's altitude floor at 80 km. This effect can be observed in the red Maxwellian curve, which experiences an anomalous decrease in volume emission rate just below its peak emission altitude due to this effect. The effect has a minimal effect on the accuracy of the model, as the rapidly increasing neutral density in the mesopause quickly inhibits further precipitation and quenches even the most short-lived excited states before they have a chance to emit.

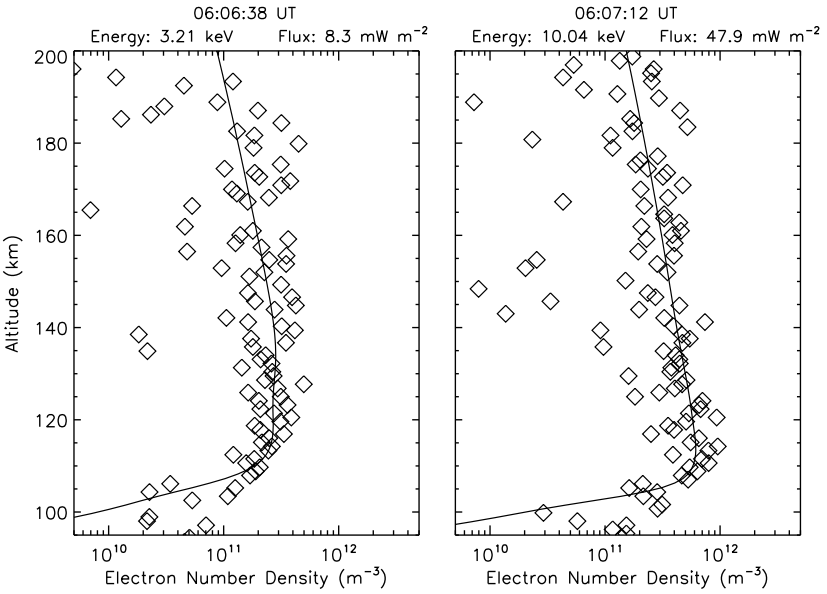


FIGURE 3.7: Modelled electron density altitude profiles (solid line) and coincident EISCAT power profiles (diamonds) during an auroral event. Time of observation and modelled energy and flux values are displayed above each respective panel.

The model's outputs for a particular auroral event can be easily compared with observations from ground-based instrumentation such as ASK. Integrating a modelled volume emission rate profile in altitude produces a value for the total column emission intensity - the intensity as it would be viewed from the ground. Then, applying the filter transmission curve from one of the ASK cameras allows for a prediction of the 'absolute brightness' of the emission, as observed by that camera. Thus, a comparison between the 'expected' brightness from a known model run, and the directly observed brightness from an optical instrument can be made. By relating the model's parameters to observed differences between the two methods, properties of the auroral emission can be inferred. As was mentioned in Section 3.2, this principle can be applied to ASK

observations to determine the precipitation energy and flux of auroral electrons and a full description of the process is given in Section 4.3.

The model has a number of sources of potential error. The restriction to one-dimension limits the ability to reproduce mechanisms of horizontal plasma transport in the ionosphere, and as such, processes like thermal expansion, electron diffusion and neutral winds are not accounted for. The model's chemical reaction rates, interaction cross-sections and other externally provided coefficients are generally well constrained by the literature, and are updated to recent values, but none-the-less still contain an aspect of uncertainty. However, once the above factors are acknowledged and their impact is minimised as much as is reasonably possible, the primary remaining source of model error is the inputs. More specifically, errors are expected to arise from the extremely important MSIS-E-90 neutral density profiles. Changes in the neutral density profiles have a non-negligible impact on predicted emissions; the total neutral density, as well as the altitude dependent concentrations of the core neutral species in the ionosphere play crucial roles in the resulting intensities of the respective volume emission rate profiles. Furthermore, verifying the MSIS-E-90 model, and quantifying the magnitude of the uncertainties of these parameters is difficult. However, model runs of EISCAT observed auroral events can be easily verified via comparison between model outputted electron-density-altitude profiles with the independently determined EISCAT electron densities or power profiles. Electron density fits do not directly verify the accuracy of the MSIS-E-90 inputs to the model, but a well fit power profile lends credibility to the validity of the remaining outputs, particularly considering the principle role electrons play in the relevant chemistry. This verification is discussed in more detail in Section 4.4 and an example comparison is shown in Figure 3.7 where ESR power profile measurements (diamonds) have been fit with time-congruent model produced electron-density-altitude profiles (solid line).

Chapter 4

Retrieval of Neutral Temperature Profiles in Aurora

A significant portion of the research presented in this PhD was spent in the design and development of a novel observational technique that measures neutral atmospheric temperatures. This section discusses the motivation behind the need for such a development, provides a detailed description of each step of the method and how it was implemented and includes chronological documentation of how the method progressed over the course of the PhD.

4.1 Introduction

Typically, optical measurements of atmospheric temperatures, such as those discussed in Section 2.3, are partnered with modelled atmospheric temperature profiles to estimate the altitude of emission and derive the corresponding auroral precipitation energies. However, it seems a natural evolution of these methods to extend their application to investigate atmospheric heating processes associated with the aurora. Observing neutral temperatures in and around auroral structures can help to further scientific understanding of the atmospheric energy budget and magnetosphere-ionosphere coupling processes. Unfortunately, emissions from the primary auroral molecular species, such as N_2 and O_2, provide complications due to the extended altitude distributions of their emission and the column-integrated nature of ground based observations.

Henriksen et al. [1987] succinctly addressed this problem in their investigation into neutral temperature and emission height changes in the aurora. They compared two independent measurements to determine if observed neutral temperature changes were the result of heating, or of a changing emission altitude. First, they used the intensity ratio of $N_2^+1N(0,3)$ to $O_2^+1N(1,0)$ emission to determine one estimate of altitude. The ratio between these two emissions is dependent on the relative concentration of the corresponding neutral populations, which in turn is altitude dependent. Thus, by using the MSIS atmospheric model to determine concentrations, they can relate the

observed intensity ratio to an auroral emission altitude. Second, they calculated a rotational temperature of the emitting molecules by fitting observed rotational emission bands using the method described above. This temperature is then converted into an alternative, but independent estimate of emission altitude using the same MSIS model temperature profile. Throughout the auroral event in question, the derived altitude from variations in both the neutral temperature and the ratio of emission intensities showed close correlation, implying that any observed temperature changes were likely a result of a changing altitude of emission, and not of local heating. They posited that "the evidence in the literature points to the conclusion that heating effects are not easily studied using simple spectroscopic observations of auroral arcs".

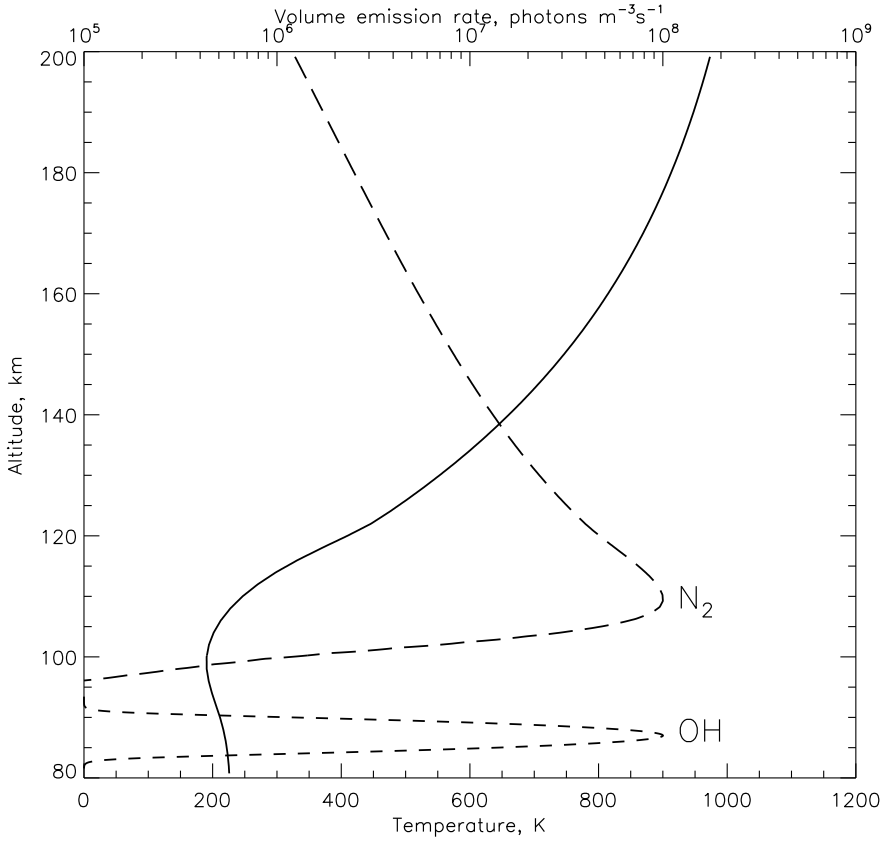


FIGURE 4.1: Example N_2 and OH volume emission rate profiles (dashed lines) plotted alongside an MSIS-E-90 neutral temperature profile (solid line). Note that the magnitude of the OH layer emissions peak value has been arbitrarily scaled to allow comparison with the modelled N_2 volume emission rate profile.

The Henriksen et al. [1987] result highlights a fundamental issue behind ground based observations of neutral temperatures: the temperature-altitude ambiguity. As mentioned, N_2 emission originates from a wide range of altitudes; the problem resides in the atmospheric temperature profile, which displays significant changes in temperature over the same altitude regions. An illustration of this issue with respect to OH and N_2 emission is shown in Figure 4.1. The relatively narrow OH emission layer covers

a region where there is an approximately constant neutral temperature. In contrast, the significantly wider N_2 volume emission rate profile extends over neutral temperatures ranging from 200 to 800 K. Therefore, the structure and shape of an N_2 emission spectrum observed from the ground is the result of the weighted contributions from all altitudes where significant emission is present, each with a different temperature. Thus, observed N_2 spectra are not well represented by a single neutral temperature, and a retrieved temperature would most likely correspond to the temperature at the altitude of the most intense emission. This ambiguity between altitude changes and temperature changes needs to be resolved before any potential neutral heating processes associated with the aurora can be accurately investigated. Instead of a single temperature, the emission spectra would need to be decomposed such that the entire temperature profile within the region of emission is retrieved.

A substantial part of my PhD was spent developing a process to achieve this goal. The following chapters outline the pre-requisite observations and the analytical steps required to estimate the neutral atmospheric temperature profile during intervals of auroral emission.

4.2 Overview

The core aspect of the method presented here is an inversion process: converting an observed N_2 emission spectrum into an estimate of the corresponding neutral temperature profile.

This is accomplished by combining a large selection of trial temperature profiles with modelled N_2 volume emission rate profiles to generate an extensive library of synthetic N_2 emission spectra. These synthetic emission spectra are fit using least squares residual analysis to an observed spectrum from HiTIES which was taken over the same time interval as the modelled volume emission. An estimate of the true neutral temperature profile is obtained using a weighted average of the trial temperature profiles, with weights set according to how well the corresponding synthetic spectra fit to the observed spectrum.

To reiterate, the method can be separated into the following 5 steps, each of which is discussed in detail in the sections below:

1. Determination of the characteristic particle precipitation energy and flux of auroral electrons using the ASK instrument.

2. Retrieval of a corresponding estimated N_2 volume emission rate profile, by using the energy and flux from step 1 as an input to an ion chemistry and electron transport model.

3. Generation of a synthetic N_2 volume emission spectrum via a combination of a trial temperature profile and the modelled N_2 volume emission rate profile.

4. Fitting of the synthetic N_2 emission spectrum to the spectrum observed by HiTIES.

5. Repeat steps 3 & 4 for many trial temperature profiles

6. Calculate an estimated true neutral profile using a weighted average of the trialled profiles.

4.3 Precipitation Energy and Flux

An accurate estimation of the shape and magnitude of the N_2 volume emission rate profile during auroral emission is crucial to the application of the method. Without an idea of the height distribution of the emission, it would be impossible to decode the individual contributions at each altitude and determine the temperature profile. The shape and variation of an N_2 volume emission rate profile depends primarily on both the atmospheric N_2 mixing ratio and on the characteristic energy and flux of the precipitating particles responsible for the excitation of the N_2 molecules. Therefore, an estimate of both must be obtained before any volume emission rate profiles can be accurately modelled. Atmospheric models such as the MSIS-E-90 model provide the first of these requirements, and the second can be determined using a well established observational technique.

As was briefly discussed in Section 2.2.3 and Section 3.4, the relative intensity of different auroral emissions is dependent on the relative concentration of the responsible atmospheric neutral species. For example, if the concentration of a particular neutral species is comparably large with respect to another, emissions from the denser of the two species will be more intense. Intuitively, this is because a larger fraction of the precipitating and secondary electrons will interact with the more abundant species, which in turn leads to more excitations or ionisations of that species, and thus more resulting emission. The concentration of neutrals is altitude dependent (Figure 2.4), such that different emissions will be more or less intense at different altitudes. The altitude at which a precipitating particle deposits its energy is dependent on a number of factors, but the prime component is the particle's characteristic energy. More energetic particles are able to penetrate further into the atmosphere than their lower energy counterparts and produce relatively more emission from the heavier molecules found there. The ASK instrument was designed to exploit this relationship and facilitate the determination of auroral precipitation energies. ASK1 (673 nm) measures molecular N_2 emission that originates from two 1st positive (1P) vibrational band transitions (4,1) and (5,2). N_2 density is concentrated at lower altitudes and thus this emission is responsive to high energy precipitation. Conversely, ASK3 (777.4 nm) isolates atomic oxygen emission from the $3s^5S$ – $3p^5P$ multiplet. This atomic O emission is a mixture of direct and dissociative excitation and is sensitive to a large range of precipitation energies. Since

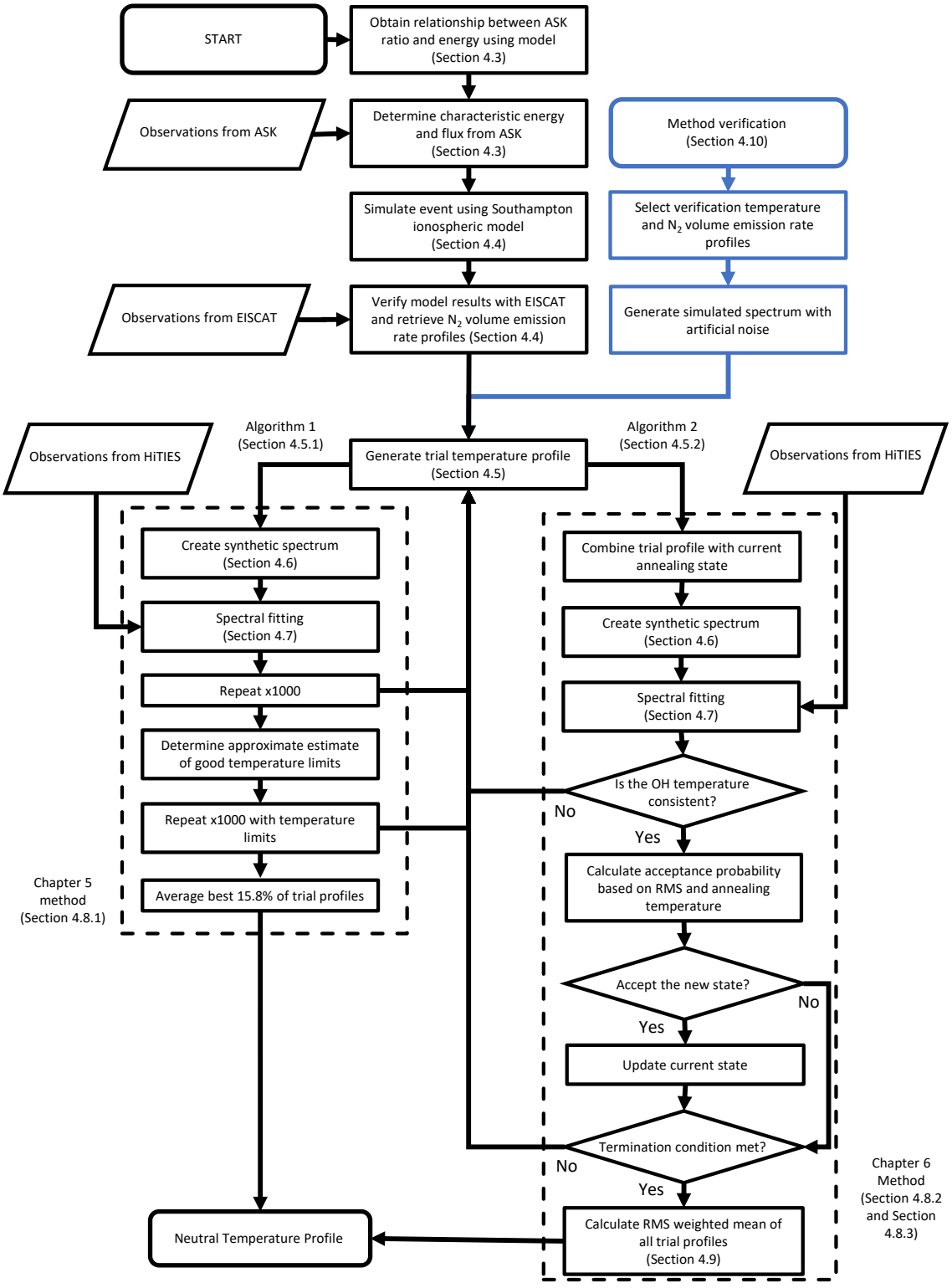


FIGURE 4.2: A flowchart describing the various aspects and versions of the method.

the two emissions observed in ASK1 and ASK3 have different energy responses they can be compared in order to determine an estimate of the precipitation energy.

First, the model is initially run with a simple Gaussian or Maxwellian input spectrum in steps of energy ranging from low to high (20 eV to 40 keV), and with a constant flux of 1 mWm^{-2}. The local time, geographic location, and geomagnetic conditions (the Ap index and the 81-day average F10.7 solar radio flux) for the event under investigation are also included. The model run outputs volume emission rate profiles for both of the ASK observed emissions for each of the energy steps, and a total emission brightness as seen from the ground can be calculated by integrating each of these profiles in altitude. A plot of this output as a function of the input energy is shown in the top panel of Figure 4.3. The N_2 (673 nm) brightness is plotted in blue, and shows the pre-described sensitivity to higher energies. The total O (777.4 nm) brightness is plotted in green and the dashed and dotted lines represent the two separate generation mechanisms responsible for this emission: direct excitation and excitation via O_2 dissociation, respectively. The combination of these two mechanisms results in a generally decreasing energy-brightness curve, indicating the Oxygen emission's preference for lower precipitation energies. The bottom panel of Figure 4.3 shows the modelled brightness ratio O/N_2. The ratio displays a clear dependence on energy, and as such the observed ratio between the same auroral emissions can be directly converted into the energy of the responsible electron precipitation.

The second important quantity needed for the eventual modelling of auroral volume emission rate profiles is the energy flux of the precipitating electrons. The energy flux, hereafter referred to as 'flux', is measured in mWm^{-2} and represents the total power entering the atmosphere per unit area. The flux can be represented a number of particles at a given energy such that, in some ways, the flux is a measure of the rate of precipitating particles. Flux has a significant impact on auroral emission rates; a higher flux generally means more total emission as there are more particles entering the atmosphere and interacting with the neutrals there. Notably, the flux has a negligible impact on the altitude distribution of emission, and variations in the flux do not affect the calculation of precipitation energy.

The model run shown in Figure 4.3 was produced using a constant flux equal to 1 mWm^{-2}. Therefore, the plotted brightnesses are the brightnesses that would be observed given this flux ($1mWm^{-2}$). Thus once the energy has been determined via the ratio, the top panel can be used to read off the expected brightness of each emission, at that energy, and with a flux of $1mWm^{-2}$. Then, to determine the flux of the observed emission, you simply calculate the ratio between the brightness seen in ASK and that produced by the model, and make the assumption that all of the increased brightness is the result of a linear increase in flux. For example, a precipitation energy of 1 keV (determined from the ratio) would produce a modelled N_2 (673 nm) brightness of $\sim$200 R. If the observed brightness of the N_2 emission was 400 R, the determined flux would be $2mWm^{-2}$, which accounts for the missing intensity without changing the energy.

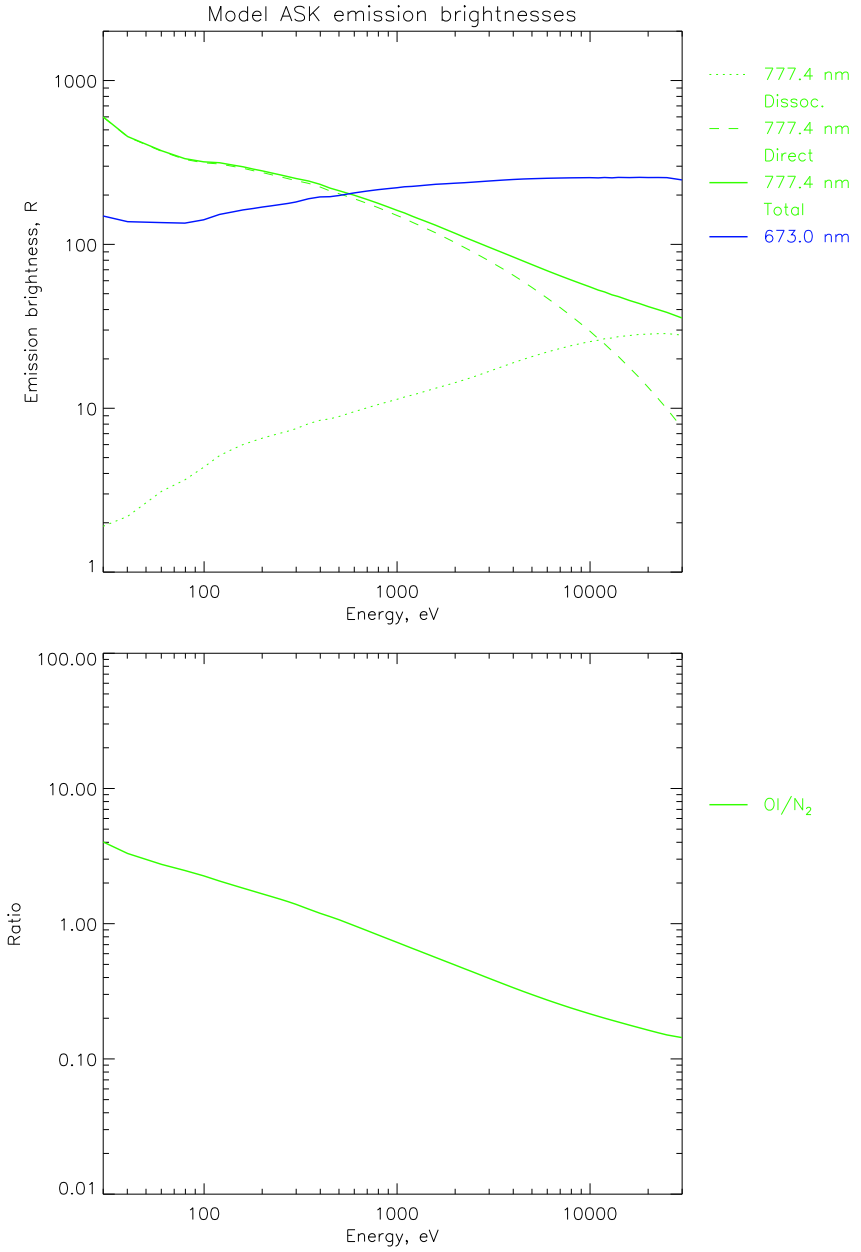


FIGURE 4.3: Modelled ASK brightnesses and corresponding ratio for N_2 (673 nm) and O (777.4 nm) emissions as a function of precipitation energy. Gaussian input electron spectra with a width of 10% were used.

Once the model runs for an increasing energy and constant flux have been produced they can be repeatedly utilised to calculate the values of the characteristic energy and flux of the precipitation responsible for observed emission in ASK. Energy and flux can be determined in this way at the time resolution of the ASK instrument, which is up to 32 Hz. It is worth mentioning that this technique only works in the magnetic zenith, or for small angles surrounding the zenith, since the N_2 and O emission needs to be observed along the same field line. Moving away from the zenith results in column observations that cross a number of field lines and the corresponding ratio of emission

gets contaminated. Determining energy off-zenith is possible but is a more complicated process [Tuttle et al., 2014] and was not required for the work presented in this thesis.

4.4 N_2 Volume Emission Rate Profiles

Once the energy and flux of the precipitating particles have been estimated via ASK observations they can be supplied as inputs to the Southampton ionospheric model in order to investigate the auroral event in question. The eventual goal is the combination of the model results with spectral observations from the HiTIES instrument, so the model is run at a matching time resolution of 0.5 s. To this end, the energies and fluxes from the ASK instrument are averaged over half second intervals before they are used to generate the input spectrum for the corresponding step of the model. The initial model runs utilise a combined spectrum that is comprised of an 80-20 ratio of flux between Gaussian and Maxwellian curves, respectively, with a Gaussian width of 10 %. This setup was found to be a reasonable starting point for most of the high-energy precipitation and intense N_2 emission observed in this work, but the ratio of the flux and the width of the Gaussian component can be adjusted as necessary.

A combination of the ratio of flux budget between Gaussian and Maxwellian input spectra, small uncertainties in the ASK acquired intensities (and hence energies) from errors in calibration, and uncertainties in the MSIS-E-90 neutral density profiles, means that the model's initial output is potentially erroneous. To account for and reduce the impact of these inaccuracies, it is useful to have a verification method which utilises independent observations to help substantiate the modelled results. Co-located incoherent scatter radar measurements provide an excellent opportunity to do this verification. ESR power profiles provide completely separate but simultaneous half second resolution measurements of electron density profiles in the auroral ionosphere. These power profiles are derived from the total returned radar power, and as such are completely independent of the model output electron density profiles. Electron densities provide a useful and reasonable verification check as they are an important aspect of auroral chemistry. Therefore, if modelled electron densities are well matched by the independently obtained ESR power-profiles, confidence in the rest of the more difficult to verify model outputs is substantially increased.

To ensure as accurate an output as possible, a comparison between modelled electron densities and ESR power profiles was employed on every model run used in this thesis. Typically, an initial model run was produced using the verbatim values of energy and flux from the ASK instrument and the aforementioned Gaussian and Maxwellian flux ratio, which gave an appropriate starting point for further adjustments. The output electron density profile at each time-step was individually examined and compared with the simultaneous ESR power profile, and the characteristics of the corresponding input spectrum were tweaked in order to ensure as accurate a match between the two as possible. The primary adjustment techniques used were the variation of the ratio

between Gaussian and Maxwellian input spectra and small changes to the width of the Gaussian spectrum. This ratio and the width of the Gaussian component of it were fine-tuned over multiple model runs to produce as closely matching an electron density profile as possible. Further adjustments were sometimes required in the form of small alterations to the characteristic energy of precipitation, which were usually no more than a 5-10% change from the ASK determined value. This careful process ensured the best possible match between modelled and observed electron densities, in turn resulting in increased confidence in the corresponding modelled N_2 volume emission rate profiles. Once this verification process has concluded, the required N_2 volume emission rate profiles were retrieved.

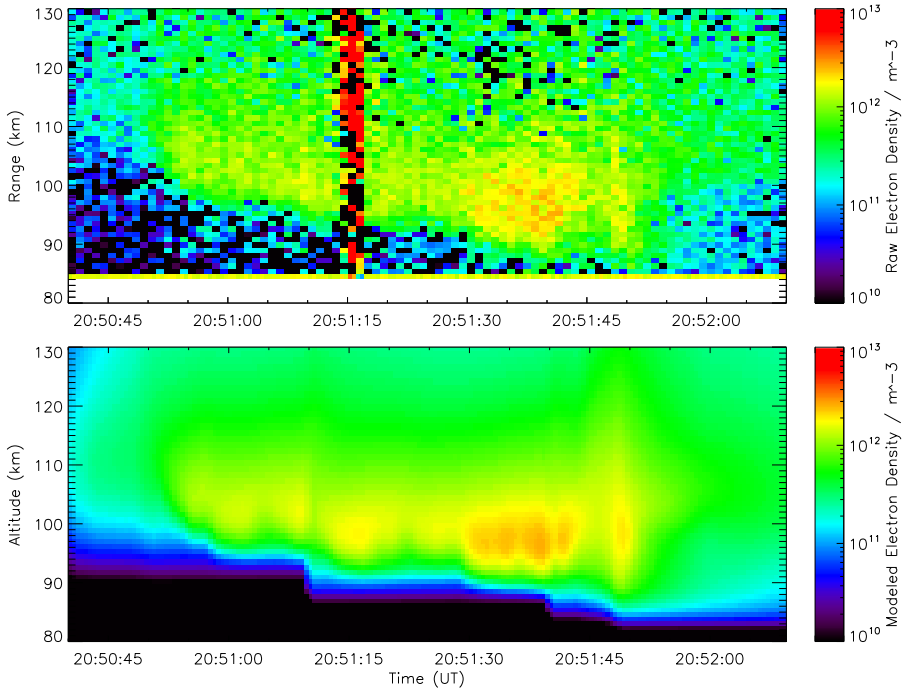


FIGURE 4.4: Comparison between the observed raw electron density from the EISCAT Svalbard Radar power profiles (top panel) and the modelled electron density from the University of Southampton ion chemistry and electron transport model over the same time interval (bottom panel). The large red artefact visible in the ESR power profiles is a result of a poor fit in the ESR analysis software and was ignored in the analysis.

Figure 4.4 displays an example comparison between observed and modelled electron densities at half second resolution, showing good agreement between the two data sets. The model successfully reproduces the increase in electron density at the arrival of the aurora in the radar beam (20:50:50 UT), and correctly responds as the peak density descends to altitudes as low as 95 km during the interval of most intense emission (20:51:30 – 20:51:52 UT). Fluctuations in electron density on time scales of the order of seconds are also correctly reproduced, most notably at 20:51:45 UT when a drop-off and subsequent resurgence in electron density is seen in ESR, and mirrored in the model.

4.5 Trial Temperature Profiles

The next stage of the method is to generate trial thermospheric temperature profiles for combination with the modelled N_2 volume emission rate profiles. This process proved to be extremely important to the overall accuracy of the method. The trial temperature profiles have to be realistic enough that they do not deviate too far from the expected thermospheric temperature profile and do not over-fit the observed spectrum using unrealistic shapes. However, they also need to be variable enough in large-scale structure that they cover a range of possibilities including profiles that are generally very cold and those that are very hot. It should be possible for the algorithm to create profiles as cold or colder than 'flat' MSIS profile discussed in Section 4.10 whilst also being able to create as hot or hotter profiles than the 'steep' MSIS profile discussed in Section 4.11. On the other hand, the profiles also need to show sufficient individual variability that smaller scale temperature features, such as localised enhancements over narrow altitude ranges, can also be resolved - features smaller than the allowed curvature will effectively be smoothed over. However, too 'wiggly' a profile will result in an ambiguous synthetic spectrum with many degenerate temperature components from multiple altitudes which can lead to over-fitting. Since any potential auroral heating signatures are unknown in extent and magnitude, careful consideration of all these factors must be included in any attempt to generate a trial temperature profile. As a result, multiple iterations of the temperature profile generation algorithm have been tested over the course of the research presented here. The following discussion will focus on two stages: that in use during the publication of the research discussed in Chapter 5 and Chapter 7, and the final version utilised for Chapter 6.

4.5.1 Trial Temperature Profiles: Algorithm 1

First, a temperature profile from the time and location of the observation was generated using the NRLMSISE–00 empirical model of the atmosphere [Picone et al., 2002]. Then a uniform random sample of temperatures and altitudes - centred around the NRLMSISE–00 defined mesopause - were chosen as initial conditions for two random walks. These random walks were performed in both decreasing and increasing altitude from the initial starting point, with minimum and maximum values of altitude at 80 km and 200 km respectively. The altitude grid used for the walk was taken from the ionospheric model discussed in Section 3.4, which has step sizes that vary between 0.07 and 0.5 km from the minimum altitude of 80 km to the maximum at 500 km. The largest allowable change in temperature per step was set to be approximately 10 K, which was determined via trial and error to be a reasonable upper limit. This was achieved by using a scaled normal distribution to generate a random temperature gradient at each altitude, centered on the current gradient, and with a maximum allowable gradient corresponding to a 10 K change. The result is a temperature profile that favours maintaining an approximately constant temperature gradient on a step-by-step basis,

but can also change significantly over the course of the walk. Finally, the initial temperature gradient during the random walk process was weakly constrained to favour a positive gradient above the mesopause, and vice versa below the mesopause, by generating the initial gradients using a slightly displaced normal distribution, such that the average gradient was slightly above or below 0, as required:

$$A \sim \mathcal{N}(\mu, \sigma^2) \tag{4.1}$$

where A is the new temperature gradient, μ is the mean and σ is the standard deviation. For the initial gradients, μ was chosen such that $\mu \times dh = 2$ K; the mean initial gradients were positive or negative, for above and below the mesopause respectively, and corresponded to an absolute temperature change of 2 K. There were no initial constraints on the allowable temperature range, other than being above absolute zero - a condition that was almost never needed. This algorithm was chosen to avoid as many biases as possible in the generation process. No strong assumptions were made about the overall shape of the profile and a large range of temperatures were possible at any altitude, aside from the initial conditions located near the mesopause. The profiles displayed the generally desired structure of increasing temperature with altitude above the mesopause, and a careful scaling via trial and error of the allowed gradient distribution function significantly reduced the chance of unrealistic profiles being generated. The result is a large collection of random and extremely variable trial temperature profiles which provided significant coverage of an enormous range of potential temperatures and gradients within the chosen altitude range.

4.5.2 Trial Temperature Profiles: Algorithm 2

The second temperature profile generation algorithm was built on the same foundation as that described in the previous section: namely a random walk algorithm loosely based on an initial mesopause location. However, a number of further constraints to the allowable shapes and temperatures were added in order to address specific problems that arose in the analysis.

The primary issue was a potential degeneracy in synthetic spectra produced by two significantly different temperature profiles. Each pair of a temperature profile and a volume emission rate profile can be combined into and represented by a histogram of total volume emission rate (photons $m^{-3}s^{-1}$) vs. neutral (rotational) temperature; the volume emission rate is separated into bins of temperature, so that the specific amount of volume emission at each temperature is known. This is the first step in generating a synthetic emission spectrum, the remaining process of which is described in Section 4.6. A degeneracy between profiles can arise from this process because the altitude at which each unit of emission corresponds to is lost; only the sum of the total emission at each temperature is retained. The result is that two different temperature

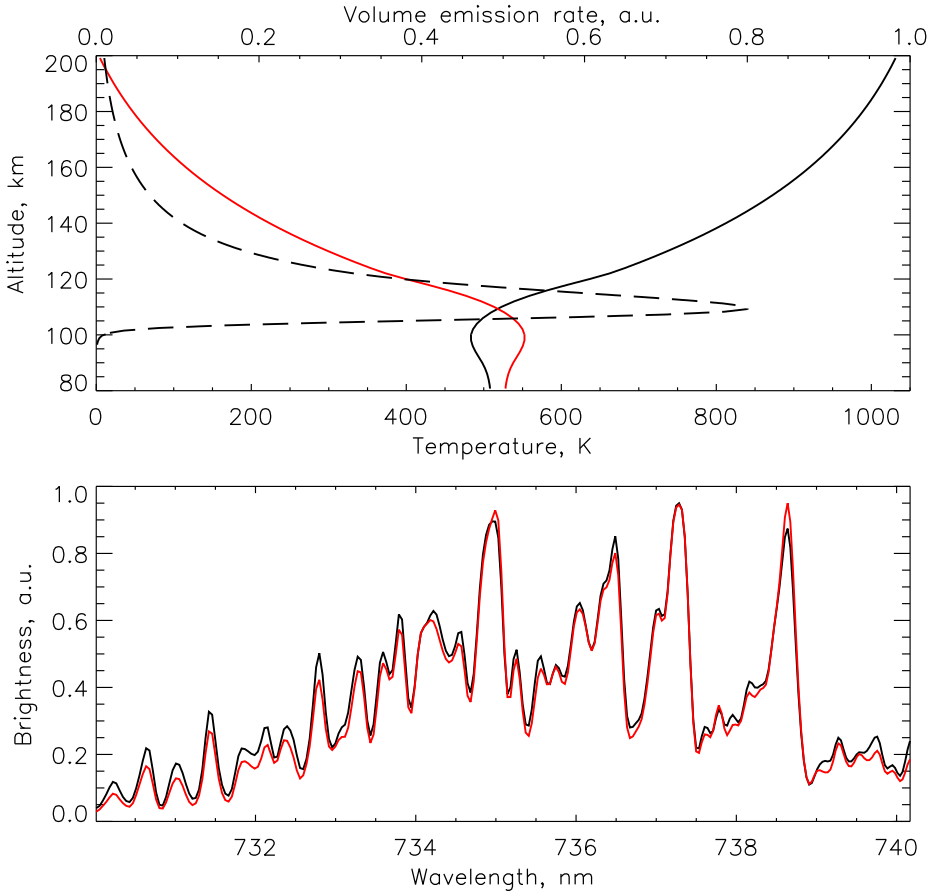


FIGURE 4.5: A schematic example of the degeneracy in two synthetic N_2 emission spectra generated by unique temperature profiles. The top panel shows an example of two temperature-altitude profiles (solid black and red lines) alongside an example N_2 volume emission rate profile (dashed black line). The bottom panel displays the resultant synthetic N_2 emission spectra for each of the profiles, displaying the similarity between them.

profiles (and a single N_2 volume emission rate profile) can produce identical volume emission-temperature histograms, and ultimately, synthetic N_2 emission spectra.

The problem is well represented by the extreme example displayed in Figure 4.5. For an approximately symmetrical (in altitude) N_2 volume emission rate profile, two profiles with shapes that are the temperature inverse of each other will produce very similar synthetic spectra. The top panel of Figure 4.5 displays two profiles that form an 'X' shape, the centre of which is located at the altitude of the peak volume emission rate. The first profile (black) would produce N_2 emission below the peak emission altitude at cooler temperatures, and emission from hotter temperatures above. The second profile (red) has the inverse structure, cooler temperatures above the peak emission, and hotter below. The temperature-volume emission histograms produced by these mirrored profiles are approximately identical - the total contribution of emission at each temperature is the same, despite the difference in the altitude of the temperatures and

emission. The bottom panel shows the similarity between the resulting normalised synthetic emission spectra for each profile. Whilst small differences are visible between the two, the expected difference in the goodness-of-fit between them would not represent the significant dichotomy between the two input profiles; these two profiles would be difficult to distinguish despite the red profile being exceptionally unrealistic.

The solution to this problem is to consider the large-scale structure of the thermospheric temperature profile (seen in the right-hand panel of Figure 3.6). In general, above the mesopause atmospheric temperatures are expected to increase with altitude. Therefore, in the 'X-degeneracy' case of two temperature profiles (as seen in Figure 4.5), the temperature profile that more closely matches this average behaviour is deemed to be more realistic. The temperature profile generation algorithm was adjusted to more strictly enforce this general temperature behaviour in altitude and thus reduce the chance of generating unrealistic profiles - effectively removing the negative temperature gradient 'branch' of the 'X' profile pair. The adjustments made were:

1. A stricter restraint on maintaining a positive gradient above the mesopause was introduced; the standard deviation of the normal distribution used to generate the new temperature gradient at each step was reduced - increasing the likelihood that the initial positive or negative gradient was maintained for longer into the walk. Furthermore, profiles with significantly cooler temperatures on average above the peak volume emission rate altitude than below it are discarded and regenerated.

2. An OH temperature condition was introduced. A general value of OH temperature and acceptable uncertainties can be supplied to the algorithm. In this case, temperature profiles that do not produce an expected OH temperature (as would be measured from the ground assuming an OH layer centred at 87 km with a width of 8 km [Baker and Stair Jr, 1988]) within the expected range are discarded and regenerated.

The acceptable range of OH temperatures for this version of the algorithm are generous and allow any temperature between 0 and 500 K. The result of these two additions to the algorithm was the production of more consistently realistic trial temperature profiles. Profiles which displayed extremely hot temperatures near the mesopause are discarded by sensible OH temperature constraints, and profiles with unrealistically cool temperatures at large altitudes are removed through the requirement for generally positive temperature gradients in altitude.

4.6 Synthetic N_2 Emission Spectra

Each randomly generated temperature profile was combined with an N_2 volume emission rate profile to generate a corresponding synthetic N_2 emission spectra. As mentioned in Section 4.5.2, the two profiles are integrated in altitude to produce a histogram

of temperature bins, where each bin contains the total amount of volume emission present at that temperature. The temperature bins have a width of 2 K, which was chosen because temperature changes smaller than 2 K have a negligible effect on the shape of the corresponding synthetic spectra. The contributions from each temperature bin are then combined into a single spectrum via a weighted sum. An individual synthetic spectrum, corresponding to a single rotational temperature, is created for the central temperature of each of the bins. The relative magnitude of each spectrum is scaled by the total volume emission contained in the temperature bin, and it is added to the final spectrum. Repeating this process for all temperature bins produces a synthetic spectrum that now contains relative contributions from each component spectra. The final spectrum is a representation of a column-integrated emission spectrum that would be expected to be observed by ground-based instrumentation.

This method has a number of consequences for further analysis, the most impactful of these is the already discussed 'X-degeneracy'. However, there are further considerations worth mentioning here. The most dominant shape in the resulting spectrum naturally comes from the temperature found at the altitude of peak emission. The temperature bin containing the largest amount of volume emission carries the highest weight in the integration process. Consequently, as the integration moves away from the altitude of peak emission, the spectral contributions from temperatures in these regions are appropriately reduced. At altitudes where the volume emission rate is negligible, the contributions from spectra become equivalently negligible. The natural result of this relationship is that the temperatures at and around the altitude of peak volume emission can be extremely well constrained by the fitting process, whereas temperatures at other altitudes become more uncertain with decreasing volume emission. A narrow volume emission rate profile allows for temperatures in an equally narrow altitude region to be accurately determined. Conversely, a wide volume emission rate profile can determine accurate temperatures over a wider altitude range, but they are not as well constrained as in the strongly peaked case.

4.7 Spectral Fitting

The pre-generated synthetic spectra are now used in a fitting process to determine which produces the closest match to the corresponding observed spectra from the HiTIES instrument. HiTIES observes auroral emissions at a time resolution of 0.5 seconds, but in order to ensure an appropriately low signal-to-noise ratio these half-second emission spectra are typically post-integrated in time. The length of the integration required to achieve an acceptable signal-to-noise ratio is inversely proportional to the brightness of the emission. During intervals of no auroral emission, such as during observations of the OH airglow, HiTIES spectra are usually integrated to a two minute resolution.

However, at times of intense auroral emission the total observed brightness can be orders of magnitude higher than what is typical for OH airglow, and the length of the post-integration HiTIES intervals can be significantly reduced.

During the observations presented in this thesis an adaptive integration interval was used; auroral brightnesses are highly variable over short time-scales and as such the integration interval was adjusted to ensure the best possible time resolution at any moment. Typically these intervals ranged from between two and five seconds long, depending on the intensity of the emission. The half-second resolution N_2 volume emission rate profiles output by the model are also integrated over the same intervals as the HiTIES observations. This ensures a temporal match between the observed spectra and the generated synthetic spectra during the fitting process.

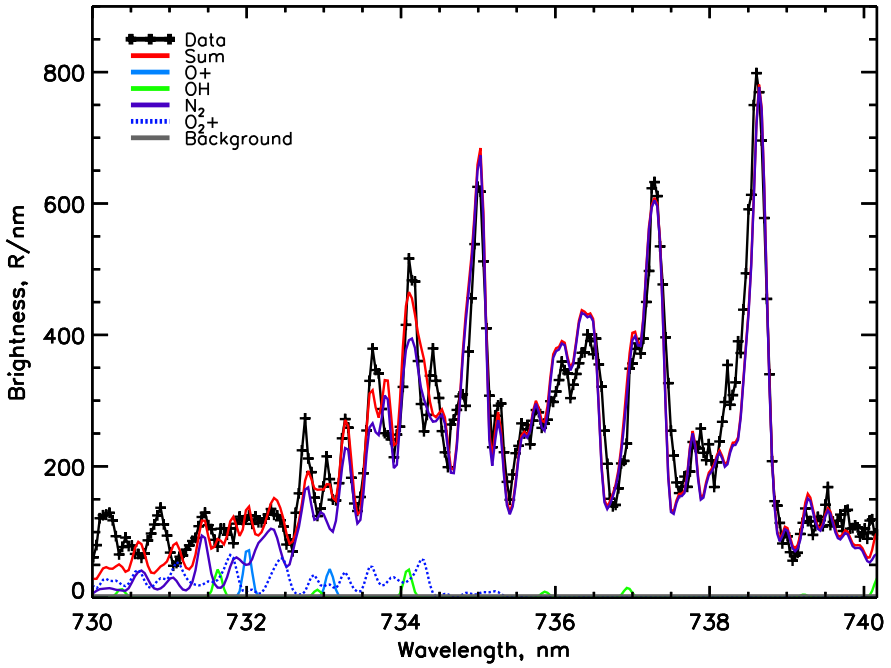


FIGURE 4.6: An example fit of a HiTIES emission spectrum during an interval of bright N_2 auroral emission. Solid black line and the solid red line indicate the observed spectrum the sum of the fit components, respectively. The remaining coloured lines represent the individual components of the fit: N_2 (purple), O+ (light blue), OH (green), O_2+ (dashed blue) and a constant background (grey).

The spectral fitting process is comprised of a least squares residual fitting routine which varies a number of free parameters until the best fitting values of each are found. Each fit consists of five components: the auroral N_2 1P(5-3) band emission, two auroral O^+ doublets, auroral O_2^+ (0-3) band emission, six OH(8-3) airglow emission lines and a constant background level. Since each synthetic N_2 spectrum has a unique fixed shape, determined by the combination of temperature and volume emission rate profiles (as described above), only its magnitude is allowed to vary as a free parameter in the fit. Other free parameters include the OH line intensities, the O^+ doublet intensities, a precipitable water vapour (PWV) value, and the O_2^+ peak intensity. The O_2^+ (0-3)

band emission is fit using synthetic emission spectra which are generated via the same method as for the N_2 spectra; the ionospheric model also produces O_2^+ volume emission rate profiles which are combined with the same temperature profile used for the N_2 emission. However, the intensity of the O_2^+ (0-3) band is usually negligible in comparison to the N_2 1P(5-3) band and as a result it does not have a significant impact on the fit. The auroral O^+ doublets, located at approximately 732.0 and 733.0 nm, and the six OH(8-3) airglow lines are fit by Gaussian profiles with a full-width-half-maximum corresponding to the instrument response function of HiTIES (0.12 nm). Initial guesses for the intensities of the OH(8-3) lines and the two auroral O^+ doublets are provided to the fitting routine based on the spectral intensities of single pixels at the relevant line wavelengths. The OH lines are fit as a series of ratios with respect to the intensity of the OH(8-3)P1(3) 'main line' at $\sim$734.0 nm. The value of the 'main line' is either left un-constrained or is weakly constrained by intensities observed before and after the auroral event. The fitted ratios for each of the remaining OH lines have adjustable upper and lower limits which are calculated theoretically to ensure a reasonable allowable OH temperature range (0–500 K). Finally, a value for the atmospheric PWV is calculated and returned using the method outlined in Chadney et al. [2017]. Both the PWV and background values can be constrained during the temperature profile fitting process using values obtained from quiet intervals before and after the event. This assumes that the background and PWV are constant over the short time interval of the event, which is reasonable given that the sky is clear with no clouds and the values obtained from before and after the event are very similar. An example fit, including all of these components is shown in Figure 4.6.

4.8 Iterating through many temperature profiles

A variety of different methods were trialled during the course of the research to determine an estimate of the best fitting temperature profile. The goals of these different methods were varied, but typically they were employed in an attempt to reduce wasted computational time and increase the final accuracy of the returned profile. Each method is discussed in detail in the following sections.

4.8.1 'Two-step' Method

Initially, a large (3000+) library of temperature profiles was generated using the first temperature profile algorithm (Section 4.5.1), the corresponding synthetic spectra were fit to an observed spectrum from a HiTIES integration interval. A measure of the goodness of fit is determined by comparing the observed emission spectra with the sum of the individual emission components produced in the fit. The fit error is calculated as the root mean square (RMS) of the residuals (SSR) between the observed spectra and the resulting fit. The RMS and SSR are given by:

$$RMS = \sqrt{\frac{SSR}{N}} \tag{4.2}$$

$$SSR = \sum_{i=1}^{N} (s_i - f(p_i))^2 \tag{4.3}$$

where N is the total number of pixels in the relevant HiTIES panel, i is the pixel number, s_i is the intensity of the observed spectrum at each pixel and $f(p_i)$ is fit-returned intensity at each pixel. The RMS value of the corresponding fit were assigned to each profile. However, it quickly became clear that a relatively small subset of the profiles were producing significantly lower RMS values than the rest and these profiles occupied a similar region of the temperature-altitude domain - particularly at the altitudes of the peak N_2 volume emission. It became clear that once a sufficiently large sample of profiles have been tested, the general region in the temperature-altitude domain that produced low RMS values became easy to locate. At this point, running the fit for a temperature profile well outside this region was using up computational time for no gain. In effect, a small number of trial profiles can quickly give an approximate idea of the general magnitude and shape of future well fitting temperature profiles. The method was therefore adjusted with an aim to save computational time and increase the accuracy of results.

The new method employed a 'two-step' fitting process. An initial 1000 temperature profiles are generated (still using algorithm 1 Section 4.5.1), converted to synthetic spectra, and fit to the observations. This number of temperature profiles was found to strike an appropriate balance between allowing for sufficient variety and coverage of the parameter space, whilst also reducing unnecessary computational expense. The RMS values for each of the 1000 synthetic spectra (and hence temperature profiles) are sorted from low to high, and the lowest 15.8% of the total population are selected. This populations constitutes the low RMS tail of values that are at least one standard deviation less than the mean. The temperature profiles that comprised this sample typically displayed a broad level of similarity. Therefore, a general region in the temperature-altitude domain can be defined such that any future temperature profiles that reside in this region can be assumed to produce well fitting synthetic spectra. Specifically, a range of allowed 'good' temperatures can be determined for each altitude in the emission region from this initial sample of 1000 profiles. The limits of this range are taken as the minimum and maximum temperatures covered by the profiles in the selected 'good' population. At this stage, the temperature profile generation algorithm (Section 4.5.1) was adjusted to include boundary conditions which restrict the minimum and maximum allowable temperatures at each altitude. A new, second library of 1000 temperature profiles are then generated under these boundary conditions and the fitting process is then run on the corresponding synthetic spectra, producing a final sample of profiles and RMS values. For this second step, the analysis is focused entirely on the region of interest, sampling it much more thoroughly than before and minimising

the computational time spent on temperature profiles that exhibit a large RMS. This method allowed for the temperatures in the region of intense volume emission to be rapidly and accurately determined.

4.8.2 OH Temperature Method

Whilst the N_2 1P band is always the dominant component in the fitted spectra, variation in the fitted OH(8-3) line intensities also proved to be a non-negligible source of uncertainty in the overall inversion process. The two main symptoms of a poor OH line fit are: extremely large or small OH 'main line' intensities and unrealistic calculated values of the OH temperature, which is obtained using the returned line intensities (see Section 2.3.3). These issues were primarily caused by the least squares fitting routine outputting unrealistic values for the un-constrained OH intensity of the main line - usually in an attempt to fit an N_2 spectral peak at the same wavelength and thus produce a lower RMS. Furthermore, this process would also impact the remaining OH line ratios, often forcing them to their minimum or maximum values to allow for the extreme main line intensity or in a further attempt to fit N_2 structure, thus impacting the calculated OH temperature. When this occurs, the OH temperature is either unable to be retrieved (no fit on the Boltzmann plot) or was unrealistically large or small. This problem has far reaching consequences for the accuracy of the temperature profile fit, an incorrect OH line component can artificially raise or lower the returned RMS value and misdiagnose a good temperature profile as bad, or vice-versa.

To tackle this issue, further improvements were made to the fitting process: the OH main line intensity and the secondary OH ratios were weakly constrained by values obtained from intervals before and after the aurora, and the method was adjusted to include the incorporation of a more reactive use of the temperature profile generation algorithm. Each of the generated profiles were checked for a consistent and reasonable OH temperature calculation. To facilitate this check we used a simple calculation to help determine an 'expected' OH temperature for a given temperature profile:

$$T_{OH}^{expected} = \int T(h) V_{OH}(h) dh \tag{4.4}$$

where h is altitude, $T(h)$ is the neutral temperature profile and $V_{OH}(h)$ is the OH volume emission rate profile. A generic OH volume emission rate profile is used, represented by a Gaussian emission layer with a central altitude of 87 km and a full width half maximum of 2 km. These values are appropriate for the OH (8–3) band observed by the selected HiTIES panel [Savigny et al., 2012]. The improved method now proceeded as follows:

1. Generate a random trial temperature profile, using algorithm 2 (Section 4.5.2), and convert it into a synthetic N_2 emission spectrum following the method described in Section 4.3 to Generating spectra.

2. Fit the synthetic N_2 emission spectrum to the observed spectrum from HiTIES (Section 4.7).

3. Calculate an OH temperature and uncertainty using the line intensities of the fitted OH(8-3) emission lines. If no linear fit is possible on the Boltzmann plot and this step fails, go to step 1, otherwise continue.

4. Calculate an expected OH temperature for the current temperature profile (Equation 4.4).

5. Check that the expected OH temperature is within the error of the OH temperature determined from the fit using that profile. If so, the expected and fitted OH temperatures are deemed to be consistent with each other and the temperature profile and RMS are saved. If not, the profile is not saved.

6. Repeat the process.

Including this check in the fitting process drastically reduced the chance of final temperature profile fits having unrealistic OH temperature values. The magnitude and ratio of the OH lines in the final fit must produce a realistic temperature that is comparable to the expected value from profile being trialled. In addition, the fitting process can no longer arbitrarily adjust the OH line intensities to produce the lowest error possible, which in turn reduces the chance of temperature profiles being assigned an incorrect RMS that is not representative of their true goodness of fit. This OH check not only increased the accuracy of the fitting method, but also improved the ability of future spectral measurements to resolve realistic OH line intensities during intervals of intense auroral emission.

A final improvement to the fitting process was also made at this stage of development. Despite the successful attempt to reduce extreme cases of temperature degeneracy, as outlined in Section 4.5.2, a further aspect of the ambiguity in temperature became apparent during the analysis. Due to the nature of the N_2 1P band structure and its variation with temperature, it is difficult to distinguish between contributions arising from a single temperature, or from those arising from the average of a number of temperatures.

The distinction between these two contributions is best explained using the simple example displayed in Figure 4.7. The shape of N_2 emission spectra changes relatively smoothly with respect to temperature (see top panel); whilst the overall shape changes drastically from cool to hot temperatures, the intensity change at any given wavelength is approximately constant with temperature. The degeneracy arises when considering the difference between a spectrum corresponding to a single rotational temperature (400 K) and a spectrum comprised of two equally weighted contributions, from two rotational temperatures spaced on either side of the first temperature (200 K and 600 K, for example). These two spectra show extremely similar overall structure, despite arising from two unique temperature regimes; the middle panel displays the small differences

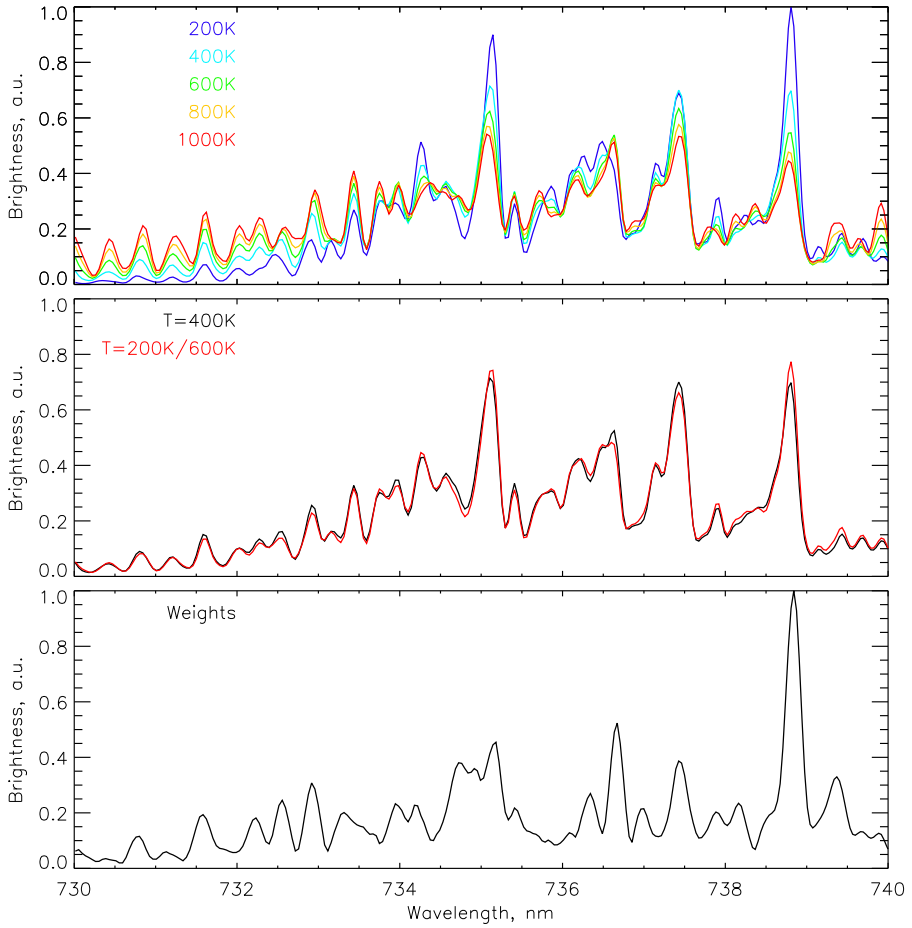


FIGURE 4.7: A schematic example of the degeneracy in two synthetic N_2 emission spectra generated by average N_2 emission temperatures. The top panel shows the spectrum expected for a range of temperatures. The middle panel shows the spectra expected from a temperature of 400 K and that expected from 50-50 contributions from two components with temperatures 200 and 600 K. The bottom panel shows the wavelength weighting scheme that was chosen to focus the fit on the regions where this degeneracy is reduced.

between the two. The simple case shown in Figure 4.7 is further complicated by the significantly more complex synthetic spectra that arise from a temperature profile, which are comprised of many separate temperature components. Two temperature profiles with distinct structures, but with a similar 'average' temperature across the emission region will have comparable emission spectra. In effect, this is the same problem as the 'X-degeneracy' discussed previously, but it cannot be reduced further by clever selection of temperature profiles as it arises from a fundamental and uncontrollable aspect of N_2 band emission.

Fortunately, there are wavelength regions within the HiTIES spectral range where the synthetic spectra show small departures from the generally constant relationship between intensity and temperature. These regions can be located by producing a number of plots such as the example shown in the middle panel of Figure 4.7: a large range

of temperatures, from 50 to 2200K, are iterated through and at each step a single temperature emission spectrum is compared to a spectrum produced by averaging the spectra from two equally spaced temperatures either side of the initial temperature. The relative differences between each spectrum are summed across the full range of temperatures. This difference spectrum represents the wavelength regions where the the deviation of the N_2 emission spectrum from a constant transition between temperatures is most extreme. The pixels corresponding to these wavelengths are given more weight in the fitting process to further increase its ability to remove potential degeneracies in temperature profiles by assigning small changes in the fitted intensities in these regions more importance.

4.8.3 Simulated Annealing Method

The final iteration of the inversion process utilises a simulated annealing algorithm (first outlined by Kirkpatrick et al. [1983]) to optimise the efficiency of the method and improve the accuracy of the final result. By considering the fitting process as an optimisation problem, a simulated annealing algorithm can be applied to determine an estimate of the global optimum solution in the large temperature-altitude search space. The fundamental process of simulated annealing is an iterative 'walk' through neighbouring states of a system, where the probability of accepting a move from one state to another is determined by a number of parameters: the system 'temperature', the 'energy' separation of the two states and an acceptance probability function. The acceptance probability function returns the likelihood of accepting a neighbouring state as an improvement to the solution, and is a function of the system's annealing temperature and the energy separation of the current and neighbouring state. Simulated annealing can be applied to the inversion process by defining a state of the system as an individual temperature profile and designating the energy of each state as the RMS goodness-of-fit error produced by the corresponding fit. The benefit of the simulated annealing algorithm is that it reduces the possibility of becoming trapped in a local error minimum. Early on in the process, when the annealing temperature is high, the acceptance probability function is designed to readily accept neighbouring solutions that return a larger error. This allows early iterations to effectively 'jump' out of local minima and sample a huge portion of the variable space quickly. For each iteration, the algorithm is set up such that the system's annealing temperature 'cools' slightly and consequently accepting higher error neighbours becomes less and less likely until eventually the algorithm will only accept direct improvements - cases where the error of the neighbouring state is strictly lower than the current state. The rate at which the system's temperature reduces in this way is known as the annealing schedule.

The algorithm's parameters require careful selection to ensure efficient and accurate results. Granville et al. [1994] proved the convergence of the simulated annealing algorithm for an increasingly extended annealing schedule. However, for many systems the

time required to achieve a significant probability of finding the optimum solution approaches unrealistic lengths. To ensure a reasonable total run-time, simulated annealing algorithms are set to have a fixed number of total possible iterations, and as such, it is crucial that these limited iterations are effectively utilised to obtain as accurate a final solution as possible. A balance between the annealing schedule and the neighbouring state selection process must be determined to ensure a wide enough parameter space is sufficiently sampled whilst the temperature is high and there are enough remaining iterations to fully explore the region of the global minimum before settling on a solution. If the cooling schedule is too rapid, such that the corresponding possibility of accepting neighbouring states with increased error diminishes too quickly, then the algorithm can still become trapped in a local minimum. In contrast, if the cooling schedule is too slow and the selection function too generous, then the final profile would not represent a significant minimum, but rather a random state within the system. The choice of the algorithm's parameters can impact its effectiveness and an efficient setup does not exist for all optimisation problems. One important requirement of the annealing schedule is that the annealing temperature should approach zero as the number of iterations reaches the maximum allowed, but the time (iteration) dependence and starting value of the annealing temperature have no conventional definitions. The probability acceptance function has a generally accepted format, and is defined below. Finally, the neighbouring state selection process is highly dependent on the individual problem in question, and the method used here is discussed in detail in the following sections.

A step by step definition of the simulated annealing process, applied to the optimisation problem of finding a best fitting altitude-temperature profile, is described by the following steps:

1. An initial temperature profile, or state s, and annealing temperature, T_a, are chosen to begin the algorithm (see Section 4.8.3.1). The initial profile is converted to a synthetic spectrum, fit to the observed spectrum, and an error, ϵ, is obtained.

2. The temperature profile (state) is perturbed slightly using a neighbour selection function (see Section 4.8.3.3).

3. The new profile represents a neighbouring state of the system, s_{new}, which is converted to a synthetic spectrum, fit to the observed spectrum and a new error, ϵ_{new}, is obtained.

4. At this stage, the previously discussed self-consistent OH temperature check between the temperature profile and fitted spectrum is performed (see Section 4.8.2). If the OH temperature check fails, then the process returns to step 2.

5. The current values of ϵ, ϵ_{new} and T_a are used as inputs into the acceptance probability function, defined as:

$$P(\epsilon, \epsilon_{new}, T_a) = \begin{cases} 1, & \text{if } \epsilon_{new} < \epsilon \\ \exp\Big[-\Big(\frac{\epsilon_{new}-\epsilon}{T_a}\Big)\Big], & \text{otherwise} \end{cases} \tag{4.5}$$

   The probability acceptance function, P, returns the probability of the new state being accepted. Once evaluated the result is compared to a randomly generated number, x, between 0 and 1, taken from a uniform distribution.

6. If $P(\epsilon, \epsilon_{new}, T_a) > x \in [0, 1]$ then the neighbouring state, or temperature profile, s_{new} is accepted and becomes the current state: $s = s_{new}$. Otherwise, the neighbour is not accepted and the algorithm remains at the current state without change.

7. After each iteration, regardless of if the neighbour was accepted or not, the annealing temperature, T_a, is reduced according to the user specified annealing schedule (see Section 4.8.3.2) and the process is repeated from step 2.

This loop runs until either the pre-determined maximum number of iterations is reached or a specified termination condition is met. Generally the maximum number of iterations was set to be 5000, which would take approximately 4-6 hours to complete, and the termination condition was set to trigger if a percentage of the maximum number of iterations occurs without a new state of the system being accepted. During the run, each of the trialled states of the system (temperature profiles) are saved and the steps at which a new solution was accepted by the algorithm are flagged. Due to the inherent random nature of the annealing algorithm, individual runs to fit an identical spectrum will produce slightly differing final results. In an attempt to minimise this effect the annealing algorithm is performed three times for each integration interval, and the results combined. This increases the chances of finding a global minimum even further, since the annealing is unlikely to find the same local minimum more than once.

4.8.3.1 Initial Conditions

The choice of initial annealing temperature should be appropriately large such that the algorithm makes multiple steps throughout the parameter space with little to no regard for the energy function. The system's annealing temperature can be thought to represent an approximate measure of the largest possible bad 'jump' or increase in error the algorithm can make. As stated, for all cases where the energy of the new state, ϵ_{new}, is greater than the energy of the current state, ϵ, the probability acceptance function evaluates as:

$$P(\epsilon_{new}, \epsilon, T_a) = \exp\Big[-\Big(\frac{\epsilon_{new}-\epsilon}{T_a}\Big)\Big] \tag{4.6}$$

which gives a $y = \exp(-x)$ relationship where x is the ratio of the change in system energy, $\Delta\epsilon$, to the annealing temperature, T_a. The effect of this ratio on the probability of accepting a neighbour can thus be easily calculated. A positive change in energy that is approximately equal to the annealing temperature of the system, gives a ratio $\frac{\Delta\epsilon}{T_a}$ equal to one and produces a probability of acceptance of $P = 0.367$; about a third of neighbours would be accepted. A ratio of approximately 0.01 ($T_a = 100\Delta\epsilon$) gives a probability of acceptance of $P = 0.99$, and so on. In practice, the range and magnitude of possible RMS values output by the spectral fit is strongly dependent on the total intensity of the observed spectrum; a 1% fit error on an observed spectrum with a peak intensity of 1000 R results in a 10x bigger RMS than the same 1% error on a spectrum with a peak intensity of 100 R. As a result, the initial annealing temperature was scaled to depend on the peak intensity of the spectrum being fit, to ensure that the annealing process does not spend disproportionate times at relatively high or low annealing temperatures.

The second initial condition is the choice of the first state or temperature profile for the simulation. As discussed, given sufficient iterations and a suitable annealing schedule, the simulated annealing algorithm would eventually sample a near complete set of states, and would converge to produce an accurate global optimum. This behaviour reduces the potential impact of a poor initial guess. However, in the interest of computational efficiency an extremely large number of iterations is not practical, and starting the simulation in an appropriate location ensures a more balanced investigation of the surrounding regions. Furthermore, the temperature-altitude parameter space in question is large, and the vast majority of it produces extremely high error results meaning that time or iterations spent sampling these regions are effectively wasted. It is useful then to ensure that the bulk of the iterations take place in the approximate region of low error, thus reducing both the total number of high error profiles being included in the final sample and the number of 'wasted' iterations in the high error regions. It was determined that the most appropriate starting profile, with the least amount of bias, would be the MSIS-E-90 neutral temperature profile for the event in question. Scaled versions of the MSIS profile informed by an initial single temperature N_2 fit (see Section 2.3.4) were trialled, but ultimately it was determined that this had little to no effect on the final result.

4.8.3.2 Annealing Schedule

The choice of annealing schedule, or the rate at which the annealing temperature decreases with each iteration, is the final input to the simulated annealing algorithm. The annealing schedule is chosen such that, during early iterations, the system moves slowly towards broad regions of low error, with annealing temperatures high enough that small scale features in the error function are ignored. Later in the process, the annealing temperature should restrict movement to favour 'good' neighbours more heavily, before finally enforcing strictly 'downhill' motion at the end of the algorithm.

As mentioned, the primary requirement is simply that the annealing temperature approaches zero towards the end of the designated number of iterations, and a number of possible schedules are thus applicable. An exponential decay schedule was settled on, with a decay constant chosen such that after the maximum number of iterations the annealing temperature would be equal to 1×10^{-5}, a value that was determined via trial and error during the verification processes (Section 4.10).

4.8.3.3 Neighbour Selection

The method for determining a neighbouring state is an extremely important aspect of the simulated annealing algorithm. There are two vital components for an efficient neighbour selection method. Firstly it is important that on average the energy of the neighbouring state is sufficiently close to that of the current state, and the possibility of making large jumps through the available states is rare. This is appropriate because a random move to a distant neighbour has a significantly higher chance of being a 'bad' move (i.e. towards a region of high error) than a 'good' move, simply because the majority of parameter space is assumed to produce large errors. In the temperature-altitude domain this was proven to be correct by previous fitting methods; there is a small region of temperatures, located around the N_2 volume emission rate peak, that produce the lowest errors. The second aspect is that the neighbour selection method must be able to provide a sufficiently short path from any state in the system to the region of the global optimum. That is to say, that no matter how far away the current state moves from the correct solution, neighbours that facilitate an appropriately quick return should always be possible. This way, even at low temperatures the algorithm can jump in large steps through the RMS domain, moving quickly from distant local minima to the approximate location of the the global minimum, simply by choosing a good neighbouring state. If this were not possible, at low annealing temperatures the algorithm could become trapped in local minima which have an effective RMS distance from the optimum solution that is larger than the maximum possible step in RMS to a neighbour. This also has the added benefit that early in the annealing schedule the possibility of large temperature jumps improves the algorithm's ability to quickly sample a wide region of the variable space. These two conditions are somewhat counter-intuitive, one favours small moves through the RMS domain, and the other requires the possibility of potentially large jumps to ensure the global minimum is always 'near by'. To reconcile these two requirements the neighbour selection process must strike a balance between both possibilities being allowed to occur often, but with the most common move favouring similar neighbours and thus a smaller step.

The final method determines a random neighbouring state by utilising a slightly modified version of the temperature profile generation algorithm discussed in Section 4.5.2. The current profile is combined with a second randomly generated profile using a 80-20 weighted average, where the original profile has the highest weight. This

effectively perturbs the initial profile by a variable amount depending on the magnitude and structure of the randomly generated profile. The chosen weighting scheme ensures that on average the perturbations in the profile are small and the energy of the neighbouring state is sufficiently close to the original. However, for its use in the simulated annealing algorithm the restrictions on the minimum and maximum temperatures of the random walk process are relaxed and a small percentage of the randomly generated profiles are permitted to exhibit extremely hot ($> 2200\ K$) or extremely cold ($< 0K$) temperatures. In these cases, despite the heavily weighted averaging of the two profiles, sufficiently large jumps in temperature from one neighbour to another are still rarely possible. Restrictions on the temperatures of the final averaged profile are still maintained; whilst the new random-walk profile can contain extreme temperatures the combination of the two must remain within the same restrictions as defined in Section 4.8.1. Fortunately, large jumps in the temperatures of the solution are likely to be rarely necessary in the latter stages of the algorithm as the system that is being solved has a strongly defined global minimum region. It is extremely unlikely that two distinctly different (separated by 100s of Kelvins) neutral temperatures in the altitude of peak volume emission give comparable RMS errors and so the existence of the aforementioned 'distant' local minima is equally unlikely. Furthermore, should a large jump out of the global minimum region be attempted late in the algorithm's process it would be discarded by the probability acceptance function's stipulation of a strictly decreasing error.

4.9 Determining the best neutral temperature profile

The last stage of the method is to analyse the distribution of the RMS errors and determine an estimate of the best fitting profile, as well as the uncertainties on the temperature at each altitude. The specific analysis varies depending on the version of the method that was employed during the research; analysis used for the research presented in Chapter 5 is discussed in that chapter. The remainder of the discussion in this chapter will focus on the analysis of the output produced by the simulated annealing algorithm, which is used in Chapter 6. However, much of the discussion here relates to the fundamental process of fitting an observed spectrum, and as such is relevant to all the research present in this work. It is important to note that the RMS value actually encapsulates the goodness of fit of all of the free parameters, as each of the emission components discussed in Section 4.7 contribute to the shape and magnitude of final fitted spectrum. For all spectra, especially during the dimmest intervals, the fitted intensities of the OH airglow emission, the O^+ doublets and the O_2^+ band emission have a non-negligible effect on the final value of the RMS for each profile. As mentioned, these parameters are appropriately constrained wherever possible to reduce the possibility and impact of extreme over-fitting. Fortunately, the majority of integration intervals are overwhelmingly dominated by the N_2 1P band emission, and as such the RMS value is

similarly dominated by that component of the fit. Furthermore, as was mentioned in Section 4.8.3.1, the calculated RMS values do not represent a universal goodness of fit metric - the RMS values are not standardised across all integration intervals. The magnitude of the RMS values depend on the total intensity of the observed spectrum, and a relative error of 1% on a bright emission spectra, produces a larger RMS value than the same relative error on a dim emission spectra. As such, the RMS values are unique to the interval in question, and the differences between them only help to distinguish between temperature profiles trialled on the same interval - a cross comparison of the fit RMS between different integration intervals and thus observed spectra is not possible.

With these factors considered, it is still appropriate to think of the RMS value as representative of the goodness of fit of each temperature profile. Temperature profiles with a relatively low RMS value have produced a spectral fit that was close to matching the observed spectrum. Each interval trials a sufficient number of unique temperature profiles to allow for a statistically robust sample size of RMS values to be analysed. Typically, when running a simulated annealing algorithm the final state of the system is accepted as the solution, however this method removes the possibility of calculating errors and it was determined that discarding the information contained in all the previous trialled temperature profiles, both accepted and unaccepted solutions, was unnecessary. The annealing algorithm can find good solutions at any point during its progression, and ultimately it was applied not with the goal of determining a single final temperature profile, but to efficiently and accurately sample the region of low error. As a result, for each of the integration intervals there are between 8,000-12,000 individual temperature profiles, each with an associated RMS value.

A crucial aspect of the following error analysis is the identification and understanding of the impact of the spectral noise present in the observed spectra on the output RMS values. Each spectrum has a variable noise component which is inherent to the observation and cannot be removed. The noise manifests as randomly distributed contributions to the spectrum at each pixel. Therefore, even if the perfect temperature profile is supplied to the least squares fitting process, and the remaining fit components are also correctly reproduced, the returned RMS value would be approximately equal to the RMS of the theoretical noise curve that would remain if all emission was removed. This 'noise RMS' therefore represents an approximate 'floor' on the output RMS, values below which are not possible to reliably achieve. Furthermore, the random fluctuations present in the noise can align with structures in the observed spectrum and thus affect the shape of various spectral features via either constructive addition or destructive subtraction. In these cases, the fitting routine will attempt to fit to these noise-influenced structures and can therefore can produce artificially raised or lowered RMS values for a particular trial profile. A bad profile could be misdiagnosed as a good fit, if the noise aligns with the spectral structure of its synthetic spectrum, and conversely a good profile could be miss-diagnosed as a bad fit if the opposite occurs. The result of these factors is that any temperatures profiles that deviate a small distance from the

correct solution, such that the differences between the resulting synthetic spectra is of the order of, or smaller than the noise, become essentially impossible to distinguish. It is then important to consider that the profile with the lowest absolute value of RMS is not necessarily the 'best' profile, since it could instead be the profile that the random noise aligns most beneficially with. Therefore, careful considerations must be made to produce a final estimate of the good profile, as past a certain threshold, differences in the RMS values are not an entirely accurate representation of the goodness of fit.

To account for this, a temperature profile weighting scheme must be devised to effectively combine a number of trial profiles and thus produce an accurate final estimate of the true profile for the spectrum in question. Each profile is assigned a weight, depending on its corresponding RMS error, and the profiles are averaged together using a weighted mean method to produce the final estimate. The weights must be designed such that the profile with the lowest RMS does not contribute an overwhelming component to the final result, and all profiles with RMS values at or around the noise floor provide a non-negligible contribution. This should mean that profiles which are near to the correct solution, and thus produce an error close to the RMS floor value, are given the highest weights and profiles that return errors far from this value are given much smaller weights. Ultimately, the 'benchmark' RMS error to which all other RMS vales are compared, is determined by taking the lowest RMS value from the interval in question and subtracting 1% of its value. This produces a standard value from which an RMS distance can be calculated for each profile and then converted into a weight by taking its square reciprocal. These weights are used to calculated a weighted mean and standard deviation of neutral temperatures at each altitude, in turn providing the final output of the inversion process.

Numerous weighting schemes of varying complexity were experimented with during the course of the research presented in this work. Significant difficulties presented themselves in the form of large variations in the distributions and magnitudes of the RMS errors for intervals of varying N_2 intensity and noise. A single weighting scheme needed to be determined that was applicable for all such scenarios. The final version was chosen during the method verification process (Section 4.10) and was believed to be the most robust and simple, requiring no unfounded assumptions on the expected distribution of errors. Contrary to the prior discussion, this method does assign the highest weight to the temperature profile with the lowest absolute RMS value. However, the RMS is fundamentally the only indicator of a profile's goodness-of-fit, and as a result is the only way to distinguish amongst profiles. This method ensures that an appropriate number of profiles have significant contributions to the final mean profile, and the lowest RMS profile does not carry too much weight.

4.10 Method Verification

The final stage in the development of the observational method was a verification process to investigate its accuracy. This verification test was achieved by fitting to artificial emission spectra which are generated by, and thus corresponds to, known or 'correct' temperature profiles. Comparisons between the true or verification temperature profile and the profile output by the inversion can thus inform on the validity of the method. The artificial emission spectra are created via an extension to the method described in Section 4.6. The first stage is to choose appropriate values for the N_2 peak intensity, O_2^+ peak intensity, O^+ doublet intensities, OH line intensities and the background intensity needed to produce a realistic verification spectra. The N_2, O_2^+, O^+, and background intensities are taken from an example fit during an interval of bright N_2 emission and are thus set to be fixed values. The OH line intensities are an important component of the fitting process and must be considered more carefully. To ensure a consistent verification spectra, the OH line intensities were determined by calculating the expected OH temperature from the verification profile (Section 4.4) and adjusting their values to ensure that the expected temperature is reproduced in the corresponding Boltzmann plot. This way, for a perfect fit to the verification spectra, the correct OH temperature will also be retrieved. The next stage is to choose an example N_2 volume emission rate profile to be combined with the verification temperature profile to determine the shape of the N_2 and O_2^+ band emissions. A volume emission profile produced by a moderately energetic (10 keV) period of Gaussian electron precipitation was chosen. This volume emission rate profile is representative of a typical interval in which there is significant N_2 volume emission. Finally, a Gaussian noise component is added to the spectrum to simulate the noise from the HiTIES instrument. The noise is randomly generated by sampling from a normal distribution, with a mean value of zero and a standard deviation of 10. This standard deviation value was determined to be approximately appropriate to match the magnitude of the real noise. The result of this process is a 'verification spectrum' which can then be processed by the method as if it was a normal observed spectrum.

Such a verification spectrum is produced for a selection of potential temperature profiles, chosen to help determine how accurately the method can reproduce large-scale temperature differences between two profiles, as well as resolve small-scale structures (local peaks) in the profiles themselves. Five verification profiles were created and tested with this in mind, these are:

- An example MSIS-E-90 profile from Svalbard.
- A 'flattened' version of the MSIS-E-90 profile, with cooler temperatures at high altitudes.
- A 'steepened' version of the MSIS-E-90 profile, with hotter temperatures at high altitudes.

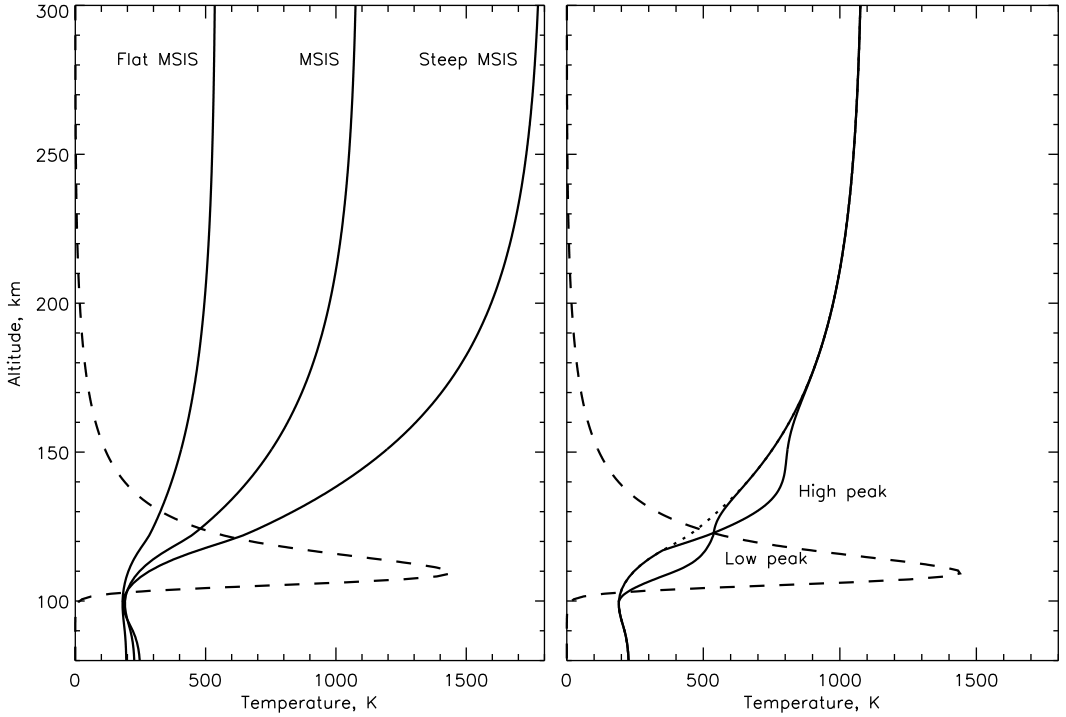


FIGURE 4.8: The five temperature profiles that were used in the verification process. Left hand panel: the MSIS-E-90 profile, the 'flat' MSIS-E-90 profile and the 'steep' MSIS-E-90 profile. Right hand panel: the two localised temperature peak profiles - a low peak altitude within the region of significant N_2 emission (dashed line) and a high altitude peak in a region of reduced N_2 emission.

- The MSIS-E-90 profile with a small localised peak in temperature centred at the altitude of peak N_2 volume emission.
- The MSIS-E-90 profile with a small localised peak in temperature centred at an altitude of approximately 50% the peak N_2 volume emission.

Figure 4.8 displays each of these profiles alongside the chosen N_2 volume emission rate profile, and the results from each the verification runs are displayed in Figure 4.9 to Figure 4.13. The accuracy of the outputs are individually discussed in the sections below. Each verification figure consists of four panels. The top left panel displays a selection of approximately ten percent of the trialled profiles used in the annealing process. Over-plotted on top of this selection is the verification profile that generated the synthetic spectrum, which is the temperature profile that we wish to recreate, and a scaled representation of the corresponding N_2 volume emission rate profile, in thick and thin black lines respectively. The central top panel displays the final result of the annealing process. The weighted mean output profile with one standard deviation of errors in temperature is plotted (orange line and grey shading), alongside the lowest RMS temperature profile (dashed red) and the 'closest' temperature profile (dashed blue, explained below). The N_2 volume emission rate profile and the verification profile

are again included (solid black lines) and the altitudes at which the N_2 volume emission is equal to 10% of the peak value are indicated by two horizontal black dashed lines. The MSIS-E-90 profile is also plotted as a dashed black curve, as a reference. The top right panel displays a scatter plot of 'RMS distance' against fitted RMS value. The RMS distance is calculated in a similar way as for the errors on the fit. The sum of the square residuals (SSR) in temperature is determined between the inversion profile and the true profile, but with the residuals at each altitude weighted with respect to the N_2 volume emission at that altitude. Profiles which are extremely similar to the true profile in the region of significant volume emission are therefore assigned a lower SSR value than those that are similar elsewhere. For simplicity when plotting, the SSR is then converted to a root mean squared value. The points in this panel therefore describe the relationship between the distance of each trialled profile from the desired profile and the corresponding RMS value obtained by the fitting process. The red and blue diamonds indicate the positions of the lowest RMS profile and the closest distance profile respectively. Finally, the bottom panel displays the fit RMS value for each iteration of the annealing process. Each plus symbol represents a single trialled temperature profile; orange pluses represent iterations in which the solution of the annealing process was updated (accepted by the probability acceptance function), and black pluses represent failed iterations (discarded by the probability function). Both states are saved in the final output. The solid blue curve indicates the system's annealing temperature and the green points and curve are the calculated values of the probability acceptance function (Equation 4.5), and a smoothed representation of its evolution in time. As mentioned in Section 4.8.3, the fitting process consists of three back to back annealing simulations which are reset to identical initial conditions each time. This can be seen in the periodicity of the bottom panel.

Each figure displays a number of interesting signatures, common to all outputs. The impact of the OH temperature consistency check (Section 4.8.2) is visible in the top left and top central panels. Since each trial profile must produce an expected OH temperature that is consistent with the value calculated from the fit (within errors), we are able to accurately constrain the temperatures at low altitudes ($<$100 km) despite the lack of N_2 emission in this region. Conversely, at high altitudes where the N_2 volume emission rate is significantly reduced, we are unable to effectively constrain the temperatures (see Section 4.6) and the trialled profiles display significant variation. As a result, it is a fairly consistent feature across all verification runs that the accuracy of the method is reduced at altitudes above the 10% maximum N_2 volume emission point (top dashed black line). Contributions from these altitudes are simply too small to have a significant effect on the resulting fit, and thus the method cannot distinguish between temperatures in those regions.

Reassuringly, the RMS distance vs fit RMS scatter plots (top right panel) display a positive correlation - generally, the annealing algorithm can reliably distinguish between profiles based on their distance from the true profile. However, each verification

run has a variable amount of non-conformity to the linear relationship, particularly in the lower left corner of the plot at low distance and RMS values. There appears to be a floor or minimum value that the RMS of the fit can take, and a range of distances from the true profile can produce an RMS at or near this value. As predicted, this value appears to be equal to approximately 10, which is the standard deviation of the random noise component which is added to the verification spectrum. Another feature observed in this panel, one that is present in all of the verification runs, is the fact that the closest profile (blue diamond) is never the same as the lowest RMS profile (red diamond).

Finally, the wide bottom panel displays the evolution of the annealing algorithm in time. Initially, when the annealing temperature (blue curve) is high, the probability acceptance function returns values close to 1 (green points and curve) and almost all neighbouring states of the system are accepted (orange pluses). As the algorithm progresses, the annealing temperature decreases and correspondingly a higher proportion of trialled neighbouring states are discarded. Interestingly, it appears that the algorithm can produce 'good' (low RMS) states at any point in the process; a fraction of early guesses are still producing RMS values near to the RMS floor. However, there is a slight negative trend in the minimum RMS value as the iterations increase in each loop, as is expected for a correctly working algorithm. This behaviour is most noticeable as a step in the RMS at the points where the algorithm is reset to its initial conditions. Regardless, it is clear from this panel that the annealing process is performing as required and its behaviour matches that which is expected from the theoretical algorithm. Furthermore, the user supplied parameters appear to be appropriate for the system in question.

4.10.1 MSIS-E-90

The MSIS-E-90 verification run performs well in the regions where the N_2 volume emission rate has values equal to or above approximately 10% of its peak. The returned profile and errors (top central panel - orange curve and grey shading) are very close to and tightly constrained around the inputted profile (black) at altitudes between 80 and 160 km, particularly so in the altitudes around the peak volume emission. However, at altitudes above this point the returned profile deviates from the true profile to increasingly cooler temperatures. The lowest RMS profile (dashed red curve) does not show as extreme behaviour at higher altitudes, but still favours cold temperatures. The shape of the output profile can be better understood with the context provided by the top left panel. The variability in the trialled profiles increases significantly at approximately the same altitude as the location at which the output profile begins to deviate from the MSIS-E-90 profile. As mentioned, this is a result of the reducing volume emission rate meaning that temperatures in these regions do not contribute significantly to the fit, and thus the determined RMS. It is likely then that once the N_2 volume emission is significantly diminished, the output temperatures approach the mean value of all the

possible profiles that also pass through the well constrained temperatures at lower altitudes. Some profiles, such as the lowest RMS profile, show increased temperatures in this region, but the majority do not and this affects the mean temperatures at each altitude, and thus the output mean profile.

This verification run suggests that a stronger RMS weighting scheme may produce a better final profile, since the lowest RMS profile (red dashed line) is significantly closer to the solution than the output profile (orange line). However this phenomena is likely the result of random chance, not only as it is not recreated in the other verification runs, but also because a particular aspect of this verification run provides further evidence to the contrary. As mentioned in Section 4.8.3.1, it was decided that the annealing process should use the MSIS-E-90 temperature profile as the initial state. Therefore for this particular verification case, the exact solution is provided to the algorithm at the start of every loop. Interestingly, as a result of the spectral noise and the other contributing factors discussed in Section 4.9, the RMS value corresponding to this exact solution is not the lowest returned RMS, further reinforcing the fact that the returned RMS values should not be given too strong a weighting, as they do not truly represent a perfect indicator of distance from the correct profile.

4.10.2 Flat MSIS-E-90

The flattened MSIS-E-90 verification run performs extremely well at all altitudes, only beginning to seriously deviate from the true profile at altitudes above 200 km, where it is still within one standard deviation of error. The lowest RMS temperature profile (dashed red) is much warmer than the true profile at altitudes above 140 km, but the inclusion of the weighting scheme discussed in Section 4.9 has ensured that the output profile is not significantly impacted in these regions.

The increased accuracy of this run for altitudes above the 10% volume emission cut-off point, particularly in comparison to the MSIS-E-90 run, is of note. The top left panel again provides useful context for this result; whilst there is still significant spread and variation in profiles at high altitudes, there appears to be a denser population of profiles at cooler temperatures. This is probably a similar affect as is observed in the MSIS-E-90 run, namely that the output profile in these regions defaults towards the mean of all possible profiles that pass through the correct temperature region at lower altitudes. In other words, once the correct temperatures and gradients have been determined in the volume emission region, the remaining shape of each profile can only evolve as allowed by the temperature profile generation algorithm. In this case this evolution appears to be a more appropriate estimate at high altitudes than in the MSIS-E-90 case, which favoured an incorrect cooler evolution.

4.10.3 Steep MSIS-E-90

The steep MSIS-E-90 run also performs well at all altitudes where the N_2 volume emission rate is more than 10% of its peak value. There is a deviation from the true profile at approximately 100 km, where the method has had difficulty reproducing the extremely sharp 'cut-back' in temperature present at the mesopause. This effect is possibly because the true profile is an extreme case - it is unusually steep - and as such the fraction of the total generated profiles that exhibit the correct temperature values and gradients is reduced in comparison to the less extreme profiles. As a result, the lowest RMS profile (red-dashed) has had a significant impact on the output profile and there are not enough similarly good profiles to effectively smooth out its more extreme random structures. Regardless, the large scale structure of the steep MSIS-E-90 profile has been effectively reproduced.

4.10.4 Local Scale Temperature Peaks

The remaining two verification runs were chosen as an attempt to push the method to its limit. Once it was clear that it could accurately reproduce large scale variations in the temperatures of the input profiles, much smaller and more localised heating signatures were trialled. The magnitude of the change in temperature is smaller, and it is located over a narrow altitude range, and not the entire profile. To test this scenario two slightly skewed Gaussian curves were added to the MSIS-E-90 profile, one with a peak altitude of $\sim$120 km, co-located with the altitude of peak volume emission, and one with a peak altitude of $\sim$140 km at approximately half the peak volume emission. The results are displayed in Figure 4.12 and Figure 4.13, respectively. These small heating structures are perhaps not realistic examples of heating that might occur in the thermosphere as a result of auroral electrodynamics, but they are a useful tool to examine the sensitivity of the observational method.

The low altitude peak verification run performs moderately well. The addition of the local enhancement at 120 km has resulted in a much steeper output profile in the region of significant volume emission than in the similar MSIS-E-90 case. The difference in the two cases is clearly visible in the temperature separation between the output curve (orange) and the MSIS-E-90 curve (dashed black) which was well fit at these altitudes in the MSIS-E-90 case. The method has responded to the increased temperature gradient as a result of the localised temperature enhancement. However, it has struggled to reproduce the precise structure of the heating, specifically it has been unable to retrieve the back turn in temperature at $\sim$130 km, instead favouring to pass through the middle of the curve. This choice to essentially ignore the shape of the curve and produce an output through the middle is a consequence of the temperature degeneracies discussed in Section 4.5 and Section 4.7. The two spectra that are generated for example by a temperature profile that follows this curve exactly, and one that passes through the central or average temperature of it, are extremely similar. Combining this problem with the

fact that the weighted average of many profiles will tend to approach the smoother case, has resulted in the method favouring the latter solution. However, as mentioned, the portion of the structure located in the volume emission region has been adequately reproduced and the difference between the MSIS-E-90 verification (Figure 4.9) run and this one is clear. As a result, a comparable heating signature in the real data would be identifiable.

The high altitude peak verification run is understandably less accurate. Since the bulk of the localised heating has now been moved to altitudes at which the N_2 volume emission is diminished, the method's ability to accurately resolve temperature structures here is similarly reduced. The final profile has once again resolved a shape that passes through the central temperatures of the structure, and it does not significantly react to its curved shape. The final output is slightly different with respect to the solution produced by the standard MSIS-E-90 profile, but since the majority of the differences between the two profiles is at altitudes outside the volume emission region, the magnitude of the difference is small.

4.10.5 Conclusions

Reassuringly, the method appears to perform exceptionally well at altitudes where there is significant N_2 volume emission. It is able to effectively and accurately determine large scale changes in the temperature profile and easily distinguish between profiles exhibiting a range of temperatures and gradients. Furthermore, localised temperature enhancements are detectable using the method temperature, so long as they occur in regions where there are appropriate levels of N_2 volume emission.

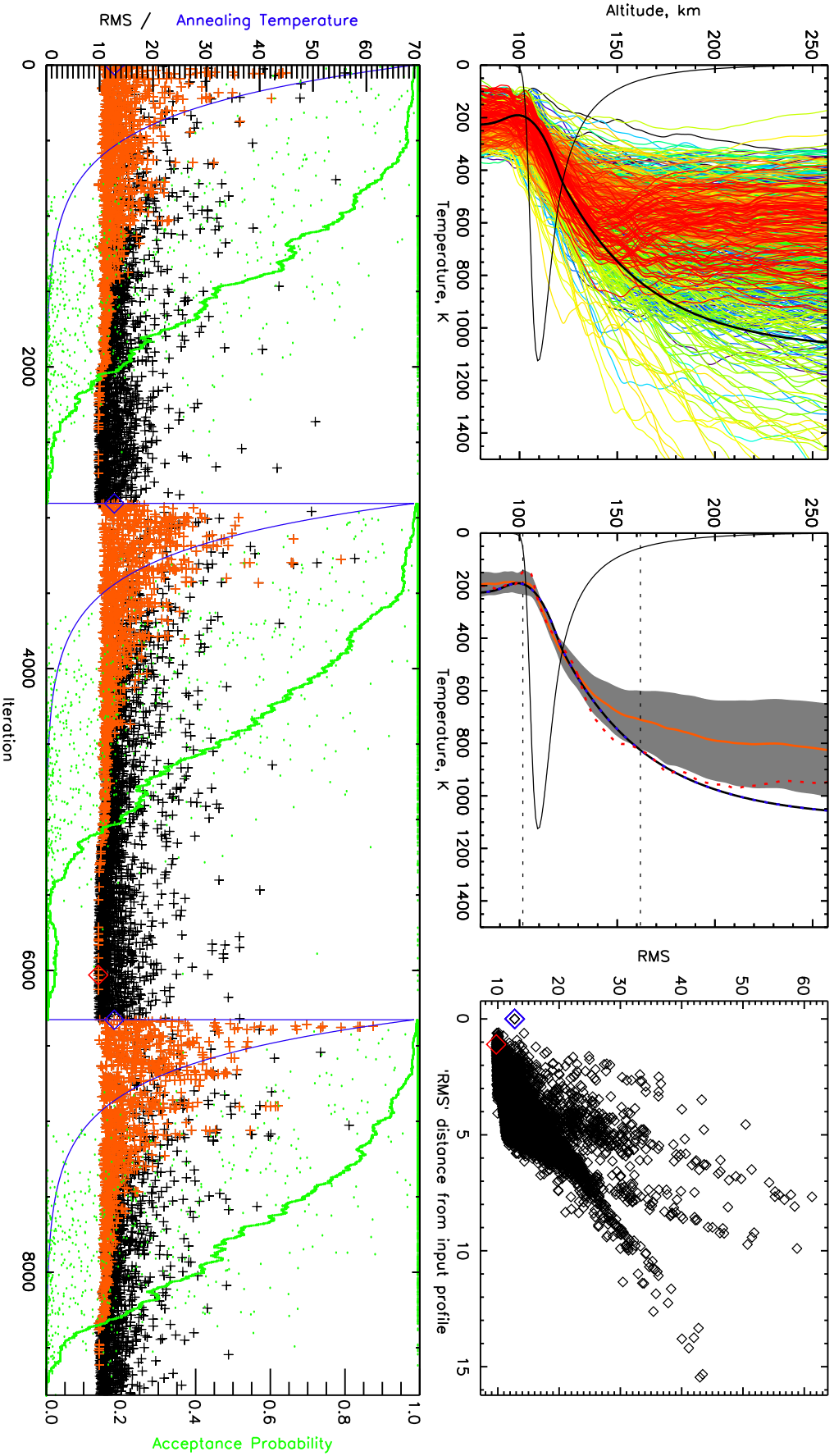


FIGURE 4.9: MSIS profile verification. The top left panel shows the true profile (thick black line), a selection of trialled profiles (multi coloured solid lines) and the N_2 volume emission rate profile (thin black line). The top central panel shows the resulting fitted profile (solid orange line), one standard deviation in temperature errors (grey shaded region), the true profile (thick black line), the N_2 volume emission rate profile (thin black line) and indications of the altitudes at which the N_2 volume emission is reduced to 10% of its peak value (horizontal dashed black lines). The top right panel shows a scatter plot of RMS temperature distance of each of the profiles from the true profile against the output RMS of the corresponding spectral fit. Blue and Red diamonds represent the closest profile (in temperature) and best fitting (lowest RMS) profiles respectively. The bottom panel shows the results of the simulated annealing algorithm plotted as fit RMS vs iteration. Crosses represent the fitted RMS of each iteration, orange crosses are iterations in which the solution was updated. The annealing temperature (solid blue line) and probability acceptance function (green line and points) are over-plotted to show their evolution as a function of iteration.

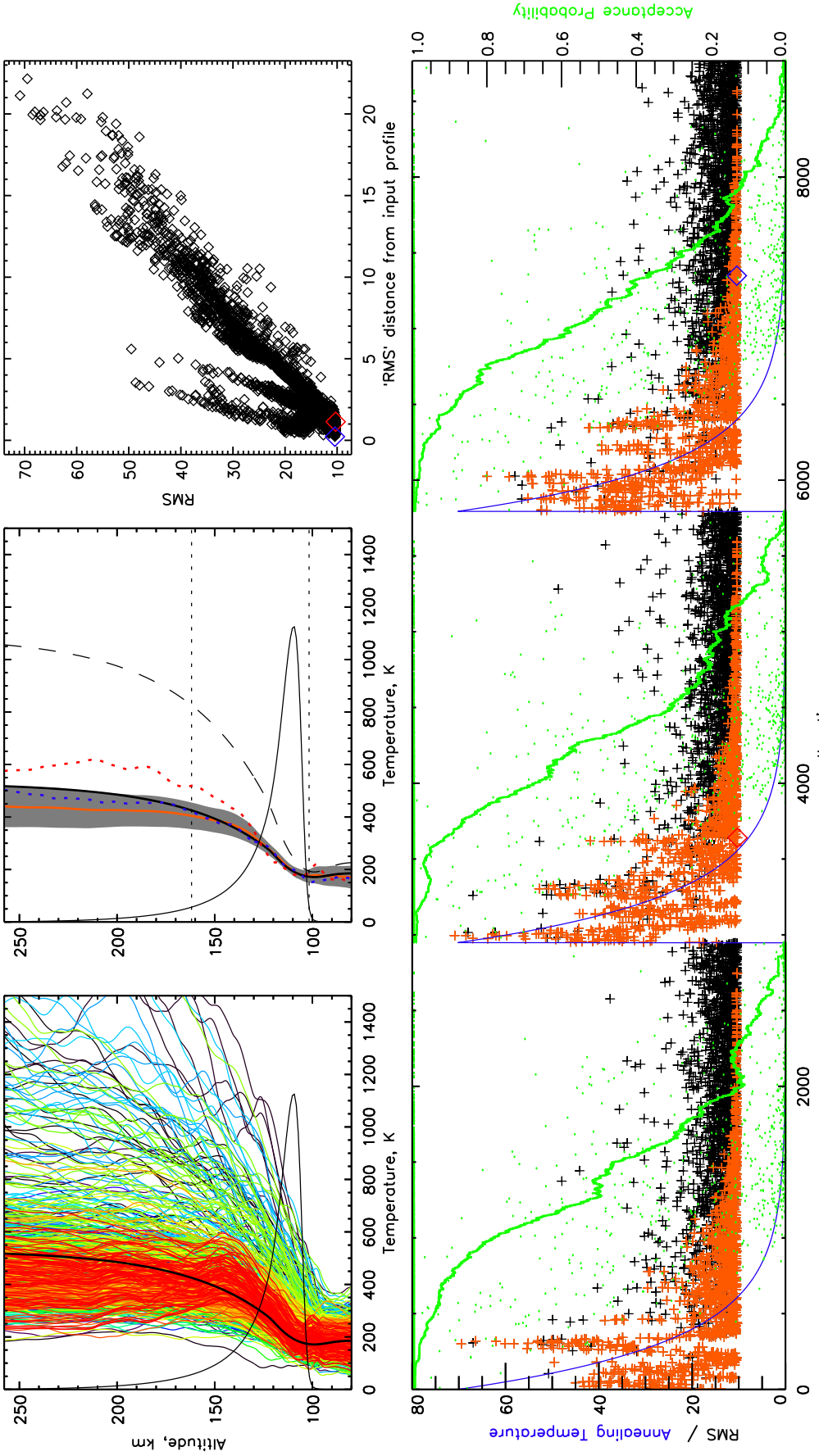


FIGURE 4.10: 'Flat' MSIS profile verification.

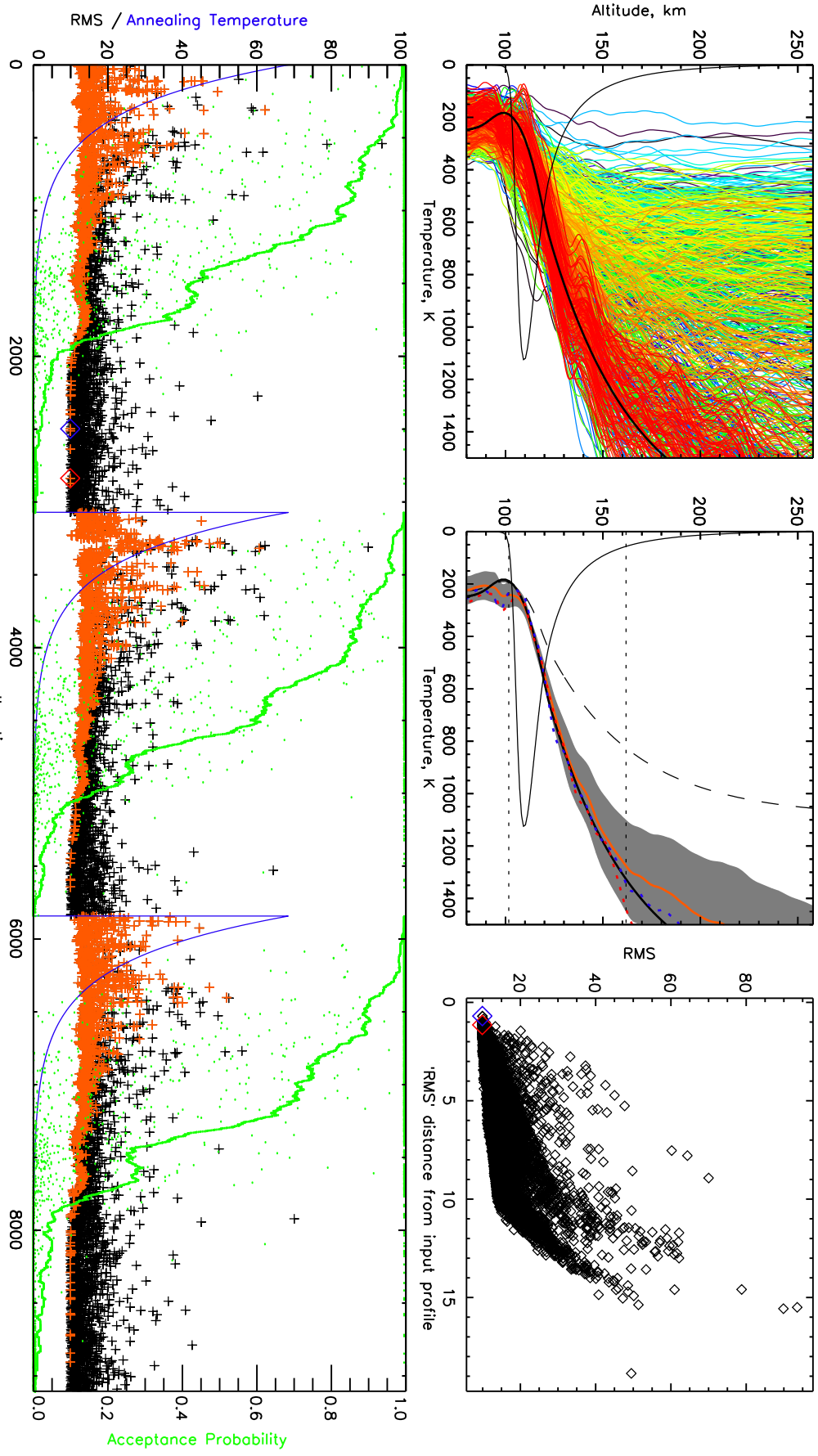


FIGURE 4.11: 'Steep' MSIS profile verification.

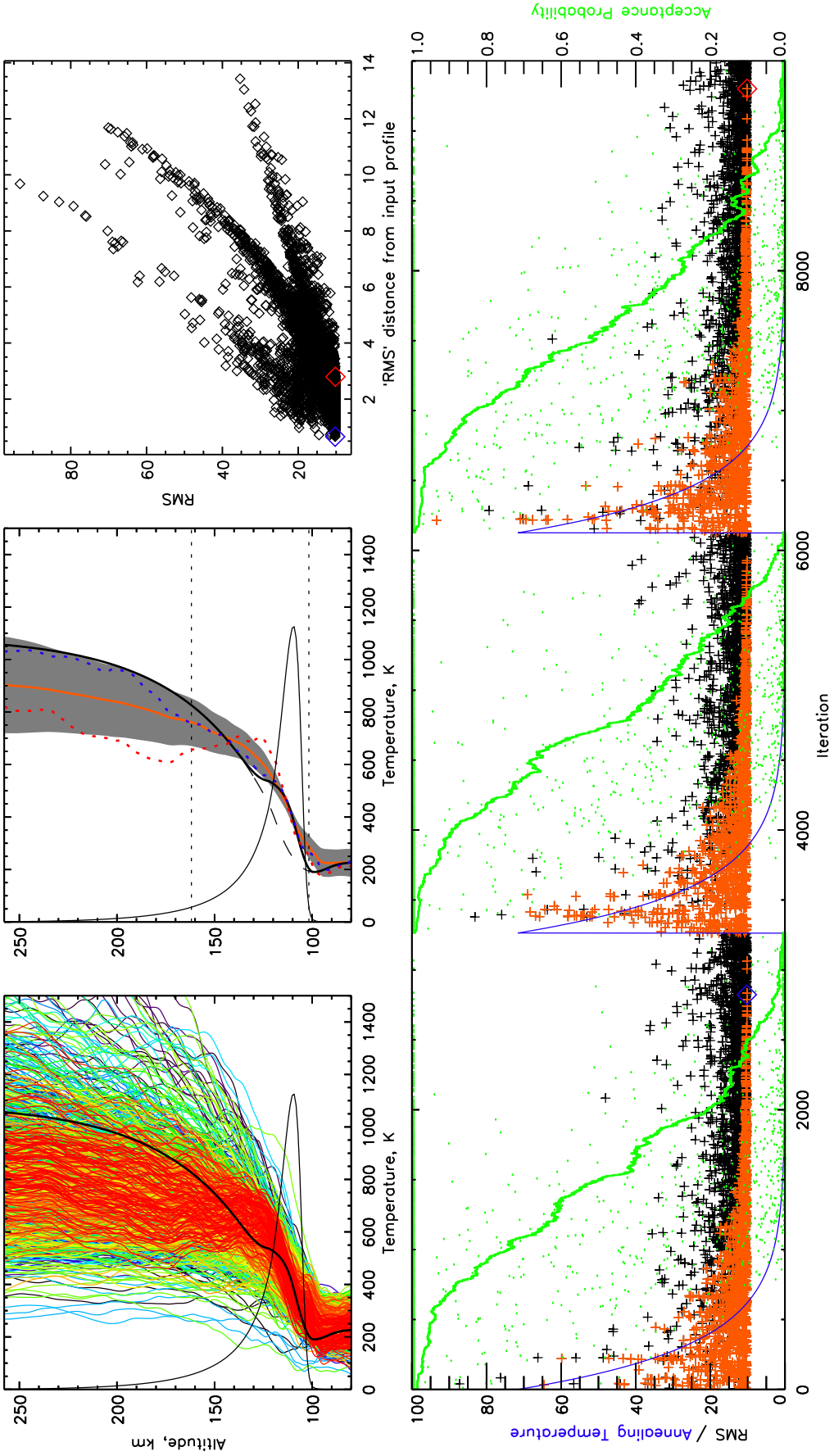


FIGURE 4.12: Low-altitude temperature peak profile verification.

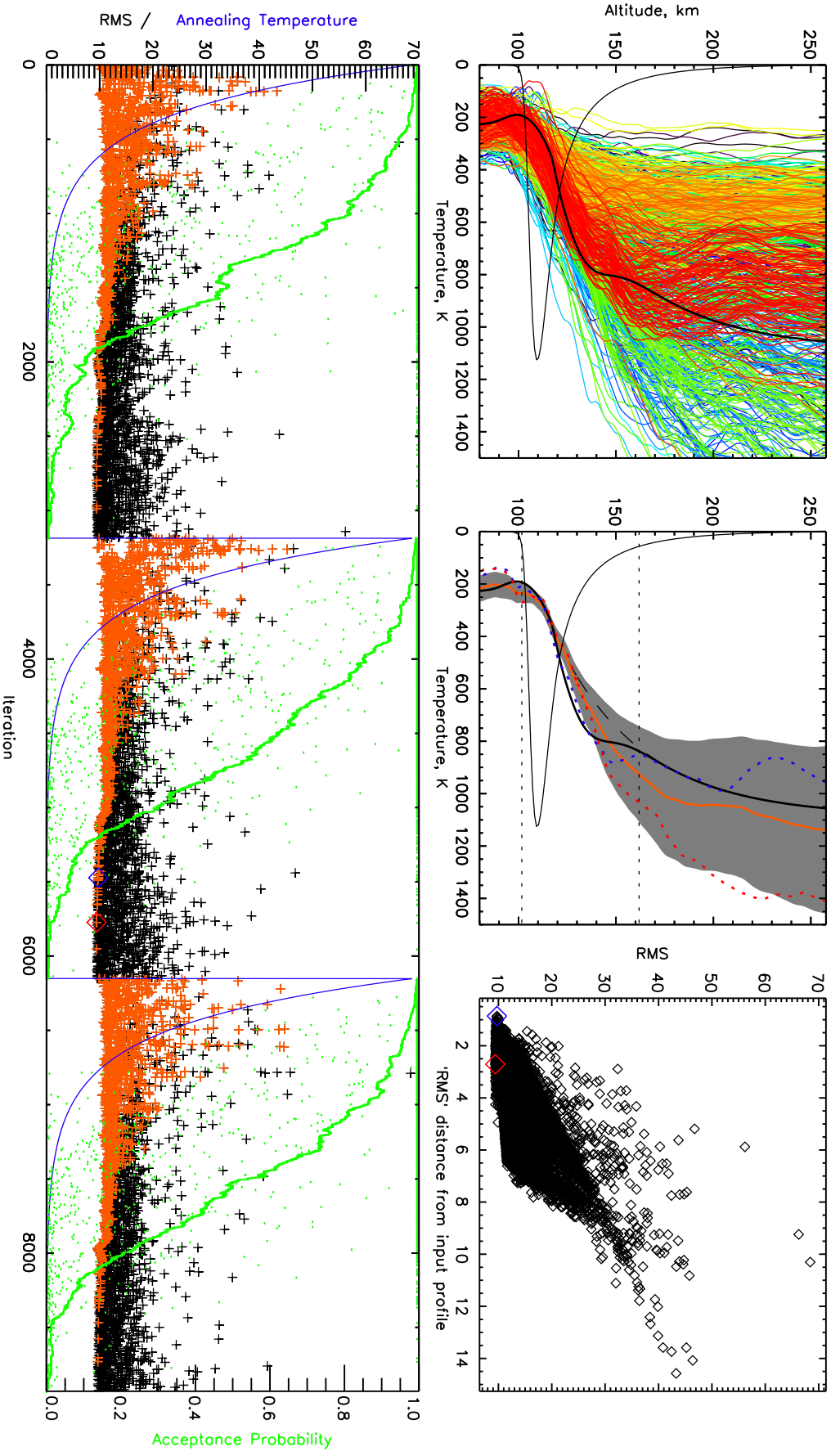


FIGURE 4.13: High-altitude temperature peak profile verification.

Chapter 5

Joule Heating

The research presented here was published as: DJ Price, DK Whiter, JM Chadney, and BS Lanchester. High-resolution optical observations of neutral heating associated with the electrodynamics of an auroral arc. *Journal of Geophysical Research: Space Physics*, 124 (11):9577–9591, 2019. The data from this study is publicly available and hosted by the University of Southampton (DOI: 10.5258/SOTON/D1056).

The analysis from Price et al. [2019] has been expanded to include further considerations on potential sources of the observed heating signatures as well as a more complete discussion on Joule heating theory. The study was entirely carried out by the author and outlines the first application of the method developed in Chapter 4.

5.1 Introduction

Understanding the closely coupled magnetosphere-ionosphere-thermosphere (M-I-T) system is a crucial aspect of developing an accurate and applicable understanding of the energy transfer processes and storage budget associated with the Earth's constant solar interactions. During active periods as much as 10^{12} W of power is transferred across the dayside magnetopause, via reconnection processes, and redistributed throughout the M-I-T system [Rodger et al., 2001]. Efforts to quantify the sinks of the incoming power often describe three main components: Joule heating of the high latitude thermosphere, particle precipitation, and energy storage in the ring current. However, it is commonly reported that Joule heating has the most significant contribution of these processes [Østgaard et al., 2002]. Joule heating is frictional heating of the upper atmosphere that results from collisions between the ion and neutral populations when the two populations have different net velocities. This typically occurs in the presence of an electric field, in which the ions feel the effect of the field and the neutrals do not - decoupling their motion. Joule heating and other ionosphere-thermosphere (I-T) coupling mechanisms have significant and global consequences for the atmosphere, and as a result the inclusion of Joule heating and particle precipitation in atmospheric models is of crucial importance. A brief summary of a number of methods used to estimate Joule heating was given in Section 2.2.4, which outlined the assumptions that are typically

needed to estimate the the relevant ionospheric electric field patterns and conductivities. The vast majority of studies resort to computation using the mean of averaged convection and precipitation patterns, dictated by the spatial and temporal resolution of the observation techniques available. Methods that must rely on assumptions or simplifications, such as those discussed previously, carry inherent uncertainties in the returned estimates of the Joule heating. Of particular note, it was Codrescu et al. [1995] who first pointed out the importance of accounting for the extreme variability of the high-latitude E-field when calculating estimates of Joule heating; neglection of potential E-field variability on sub-grid spatial and temporal scales can lead to a considerable underestimation of the total Joule heating rate [Rodger et al., 2001]. Since then, many studies have attempted to account for this effect, but computation of the average of the square field requires knowledge of the statistical characteristics of E-field variability.

Typical values of the background convective electric field at auroral latitudes are a few mVm^{-1}, however some observations have described enhancements of up to 100 mVm^{-1} in the region of discrete auroral features [Aikio et al., 1993]. These large E-fields are often also accompanied by enhanced ion or electron temperatures [Williams et al., 1990]. Lanchester et al. [1998] report electric field measurements during the passage of an arc system through an EISCAT radar beam at 3 second resolution. They observed a peak value of electric field strength of 600 mVm^{-1} approximately 10 km from, and oriented towards, the main arc structure. Their conclusions place emphasis on the importance of high resolution optical observations, alongside radar measurements of the electric field, to interpret the variability and response to temporal changes in the arc unambiguously. It has thus become increasingly clear that considerable improvements in the parametrisation of the small-scale variability of not only E-fields, but also current systems and particle precipitation events, is required to model the atmospheric response accurately [Cosgrove et al., 2011, Deng and Ridley, 2007, Rodger et al., 2001, Yiğit and Ridley, 2011, Zhu et al., 2018]. Intervals of rapid energy deposition via the intense field-aligned currents and electric fields associated with small-scale auroral structures are likely to be some of the most extreme cases contributing to the variability of ionospheric electric fields. The observations of aurora such as this, using the method discussed in Chapter 4, was the primary goal of the research conducted in this chapter.

5.2 Observations

The subject of the observations presented in this work is an extremely bright auroral arc, observed near Longyearbyen, Svalbard (78.15° N, 16.03° E) on the evening of the 27th of January, 2017, during a fieldwork campaign. At this time, Svalbard was located in the near-midnight sector of the polar cap - just eastward of the convection reversal region. All-sky images taken from the Kjell Henriksen Observatory (KHO) show a structured, east-west aligned auroral arc which spans the sky just south of the magnetic zenith. At 20:50:40UT the bright and dynamic arc begins expanding northward and its

poleward edge passes into the magnetic zenith. In the following ninety seconds a large fold develops in the centre of the all-sky field of view, moving the central section of the arc structure northward until it completely passes over the magnetic zenith.

During this time interval, the auroral arc was observed in the magnetic zenith by both the ASK and HiTIES instruments. The large semi-circular fold seen in the arc from the all-sky field of view is not apparent in ASK; instead, the aurora appears as a nearly straight east-west aligned band of emissions that progressed steadily northward until it fills the entire ASK field of view.

This study utilises emissions present in ASK 1 (N_2) and ASK 3 (O). HiTIES observations are taken from the north-south aligned integration slit, co-aligned with both the centre of the ASK image and the magnetic zenith, and with an angular size of 8° along the slit and 0.05° across it.

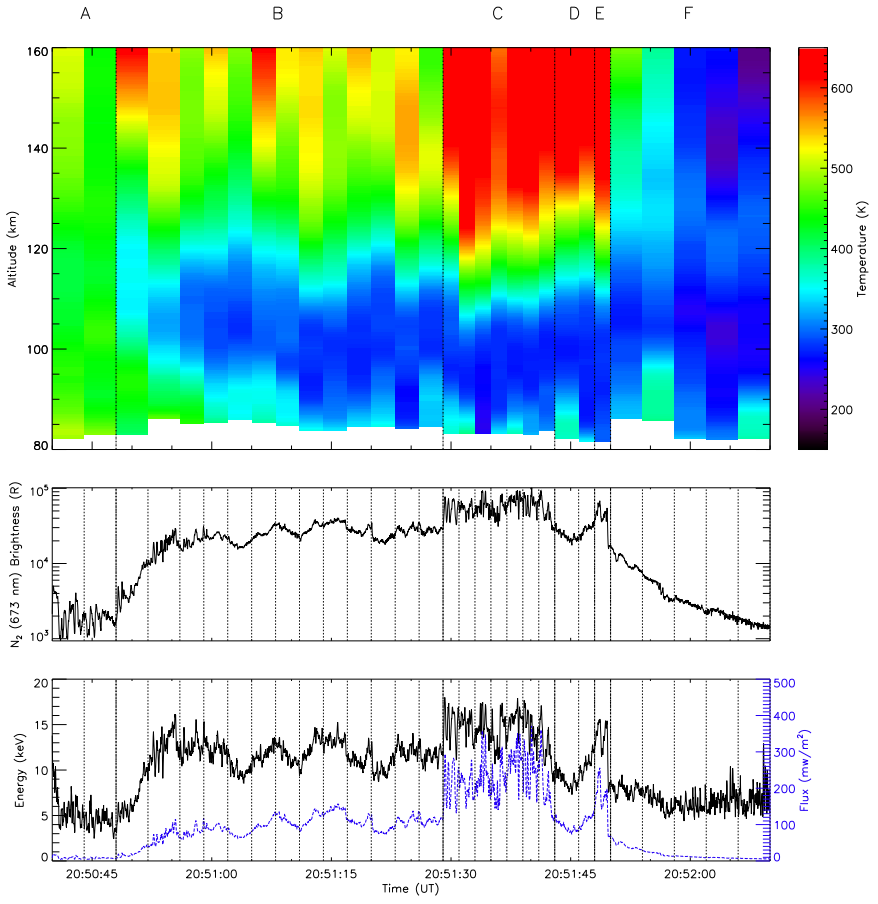


FIGURE 5.1: Top panel: mean best fit temperature profiles. Each vertical stripe corresponds to the HiTIES integration interval used to generate the observed spectrum (between 2–4 seconds). Bottom panel(s): the calibrated N_2 (673 nm) brightness (R) and measured particle precipitation energy and flux, respectively, as a function of time during the event. The grey vertical dashed lines correspond to the same integration intervals as in the top panel and the bold black vertical dashed lines to the time intervals labelled above the plot and discussed in section 4.

The emission brightness of the auroral arc varies significantly over the 90 second interval for which it is observable in the HiTIES integration slit. HiTIES spectra can therefore be integrated for a variable amount of time, depending on the emission brightness, before an acceptable signal-to-noise ratio is reached. The consequence of this is displayed in Figure 5.1. The central panel shows the ASK1 N_2 (673 nm) brightness as a function of time. When the brightness of the N_2 emission is low, then HiTIES observations must be integrated over a longer interval of time. Conversely, when the emission brightness is high, the opposite is true, and the time resolution of the HiTIES observations improve. The resulting intervals are shown by the vertical sectioning present in all 3 panels of Figure 5.1. In addition, the N_2 volume emission rate profiles from the ionospheric model (at 0.5 s resolution) are also integrated to line up with these intervals to ensure a temporal match between data sets. For this event, the result of these intervals is a time series of 29 individual observed emission spectra, each with a corresponding N_2 volume emission rate profile derived from the ionospheric model. The spectra and volume emission rate profiles thus cover the entire interval of time in which the arc is present in the magnetic zenith - documenting the full motion of the emission through the magnetic zenith. Once the intervals have been obtained, the next step to is retrieve a best-fitting temperature profile for each of them and thus produce a time-series of profiles from which the variation in the neutral temperature can be examined.

5.3 Methods

The method used to retrieve a time series of neutral temperature altitude profiles is outlined in Chapter 4 - specifically, the version of analysis used was the 'Two step method' which was discussed in Section 4.8.1 and is briefly repeated here.

The fitting process is separated into two steps. First, 1000 temperature profiles covering a large range of potential temperatures (50–2200 K) and altitudes (80–200 km) are generated using the method described in Section 4.5.1. This number of initial temperature profiles was found to strike an appropriate balance between sufficient variety in candidates whilst also reducing unnecessary computational expense. Each of the temperature profiles in this set are converted to the corresponding temperature-volume emission histogram, used in the spectral fitting process (Section 4.7), and assigned a root mean square value representative of its goodness-of-fit. An example fit is displayed in Figure 4.6. The errors for each profile are sorted from low to high, and the lowest 15.8% of the total population are selected. This value represents the population of temperature profiles that have RMS values that are smaller than one standard deviation less than the mean. From this selection, a new range of allowable 'good' temperatures for each altitude are defined by examining the minimum and maximum temperatures covered by the profiles in the selected low-error population. Using these newly determined boundary conditions, a second set of 1000 temperature profiles are

generated. This new selection of profiles provides a more rigorous and complete sampling of the low-error region of the temperature-altitude domain, whilst also avoiding running the fit and wasting time on temperature profiles that will likely produce high errors. Finally, the fitting process is run a second time using the synthetic spectra produced by the new temperature profiles and a final data set of profiles and RMS values are obtained. The RMS values corresponding to each profile are sorted once again, and the same selection of the best 15.8% is made. The final population of well fitting temperature profiles is used to calculate an estimate of a mean temperature and uncertainty at each altitude, giving the final temperature profile. The fitting process for each observed spectrum takes approximately 3 hours to complete. The analysis is repeated for each integration interval (or time step) of the auroral event and a time-series of neutral temperature profiles is produced.

5.4 Results and Discussion

Figure 5.1 (top) shows the resulting neutral temperature profiles, each of which corresponds to a unique HiTIES integration interval taken during the passage of the arc through the magnetic zenith. The event can be divided into 6 segments of time, separated by the vertical black dot-dashed lines, they are: A) an interval prior to the arrival of the optical arc in the HiTIES field of view, thus observing the region immediately poleward of the bright emission; B) an interval covering the passage of the leading edge of the arc across the HiTIES field of view, structurally the aurora in this interval consists of dynamic curls propagating along the edge of the arc and surrounded by a moderately bright diffuse background, the N_2 brightness fluctuates around a relatively constant value of approximately 20 kR; C) an interval of considerably enhanced brightness due to the formation of an extremely narrow, bright, and dynamic auroral arc embedded in the larger-scale auroral structure; D) a short interval in which the narrow, bright arc moves equatorward of the zenith, thus briefly reducing the N_2 brightness in the HiTIES field of view to previous levels; E) a second short interval, during which the narrow arc moves poleward again, entering the zenith for a brief moment and resulting in a intense spike of N_2 emission, before continuing further poleward and leaving the HiTIES field of view; F) an interval of diffuse aurora on the equatorward edge of the arc that gradually reduces in brightness to approach background levels as the large scale structure of the auroral arc completes its passage over the zenith.

Figure 5.2 displays a selection of characteristic temperature profiles, and an example of the associated auroral structure seen in ASK, for each of the 6 time intervals defined above. The ASK images on the left-hand side show a single frame snapshot taken during each interval, with the position of the HiTIES slit corresponding to the O^+ panel marked (2° by 0.05°). The panels on the right show an example temperature profile retrieved during each interval, as well as the corresponding N_2 volume emission rate profile. Panel A shows the first measured temperature profile, taken during the time

interval before the bright optical arc moves into the magnetic zenith. The HiTIES integration slit (curved white line) is observing the region directly poleward of the oncoming arc. This temperature profile displays hot (>400 K) neutral temperatures across the entire range of altitudes, even as low as 80 km in altitude. Panel B shows the temperature profile from when the highly dynamic poleward edge of the arc is in the process of passing through the HiTIES slit. At this time, the temperature profile is cooler at the lower altitudes ($<\sim$140 km) than that observed in Panel A, but interestingly also shows a localised increase or 'bump' in temperature in the 90–100 km altitude region. Neutral temperatures in this low altitude region ($<\sim$100 km) fluctuate throughout the event, but those observed during interval B are generally hotter than temperatures seen at the same altitudes, at later times - when the arc's leading edge has completely passed over the HiTIES field of view. The increased temperatures are most significant in the 10 seconds following the initial appearance of the arc in the zenith, as seen by the green region in Figure 5.1, located at altitudes below 100 km at the start of section B. After the first three integration intervals in section B, the temperatures at low altitudes cool and remain at lower temperatures for the remainder of the event. Panel C is a snapshot from the interval of most intense emission, when a narrow and bright arc filament develops within the larger arc structure and moves into the HiTIES field of view. This is approximately 40 seconds after the image shown in panel A and the temperatures in the altitude range between 90–100 km have decreased by approximately 150 K. Panel D shows the brief interval of time when the intense arc has moved equatorward of the HiTIES integration slit. This panel displays a similar shaped profile to panel C, with slightly enhanced temperatures at the lowest altitudes. There is still significant emission within the HiTIES slit during these intervals, which was positioned well within the large scale optical arc structure, but the ASK image appears darker in this region due to scaling effects caused by the bright filament. Panel E is taken when the narrow arc is moving poleward, back through the HiTIES integration slit, before leaving the field of view. This panel also shows a similar profile to that seen in panel C. Finally, panel F displays a temperature profile from equatorward edge of the arc. At this point the aurora has become less dynamic and has lost much of its small-scale structure, instead appearing as diffuse emission that slowly fades as the arc continues to move poleward. The corresponding temperature profile displays cooler temperatures than those observed at the leading edge of the arc over the majority of the altitude range.

It is important to note here that a meaningful consideration of the shape of the N_2 volume emission profile for each of the corresponding neutral temperature profiles is required to ensure the accurate analysis of the presented results. As mentioned in Section 4.6, the altitude regions in which the N_2 volume emission is most significant (with respect to the rest of the profile) are the same regions in which the process is most sensitive to neutral temperatures. The contributions to the synthetic spectra from neutral temperatures in regions of low emission are relatively smaller than contributions from the peak of the emission. The consequence of this relationship is that the uncertainty in

retrieved temperatures increases as the volume emission rate decreases. Thus, temperatures at altitudes where the volume emission rate is low are likely to be less reliable. This effect is most consequential inside the auroral arc, where the N_2 volume emission profile is near Gaussian and thus is centered on a small altitude range at which the mono-energetic electron precipitation spectrum deposits the majority of the energy. Figure 5.1 and Figure 5.2 both display the signature of this effect: In Figure 5.1, the hot (>600 K) temperatures at altitudes above 120 km in sections C, D and E occur when the N_2 volume emission is at its narrowest, causing the temperatures outside of this region to be nearly unconstrained; differences in the shape of the profiles in these regions have very little impact on the returned error and likely approximate the mean shape of profiles generated by the random walk process (Section 4.5.1). In Figure 5.2, panels C,D and E show the same effect more clearly. However, the reverse effect can be seen at the edges of the arc (e.g. Figure 5.2, Panels A and B). At these times, the N_2 volume emission rate profile exhibits a wider, less Gaussian shape, and has a larger extent in altitude. Consequently, the range over which the temperatures will have a strong effect on the shape of the synthetic spectra is also increased and they can be constrained more effectively at higher altitudes.

It is also noteworthy that temperatures in the volume emission regions of the auroral arc do not show significant changes in response to the varying N_2 emission brightness. After 20:51:00UT, the neutral temperature between approximately 90 and 120 km remains near the 300 K mark (blue shaded area in Figure 5.1), including the interval of time during which the equatorward edge of the arc is passing over (section F). The only interval of time in which temperatures below 120 km significantly deviate from this approximately similar structure is in and around the leading, poleward edge of the arc. This suggests that neutral temperature variations and any potentially responsible heating processes are not directly correlated with emission brightness, as expected, but are instead the result of other features of auroral emission - namely the electrodynamical structure associated with their formation.

5.4.1 Joule Heating on the Poleward edge

Figure 5.1 displays a strongly asymmetrical structure in the determined neutral temperature profiles for the time intervals before and after the passage of the auroral arc. On the advancing, poleward edge of the arc (interval A), temperatures are observed to be as high as 400 K at altitudes below 120 km. Conversely, the same altitudes exhibit cooler temperatures of approximately 300 K on the retreating, equatorward edge of the arc (interval F). The clear difference in neutral temperatures at these times suggests the presence of an equally asymmetrical neutral heating process, likely associated with the underlying electrodynamics of the auroral arc.

Auroral arcs form in response to the precipitation of magnetospheric electrons into

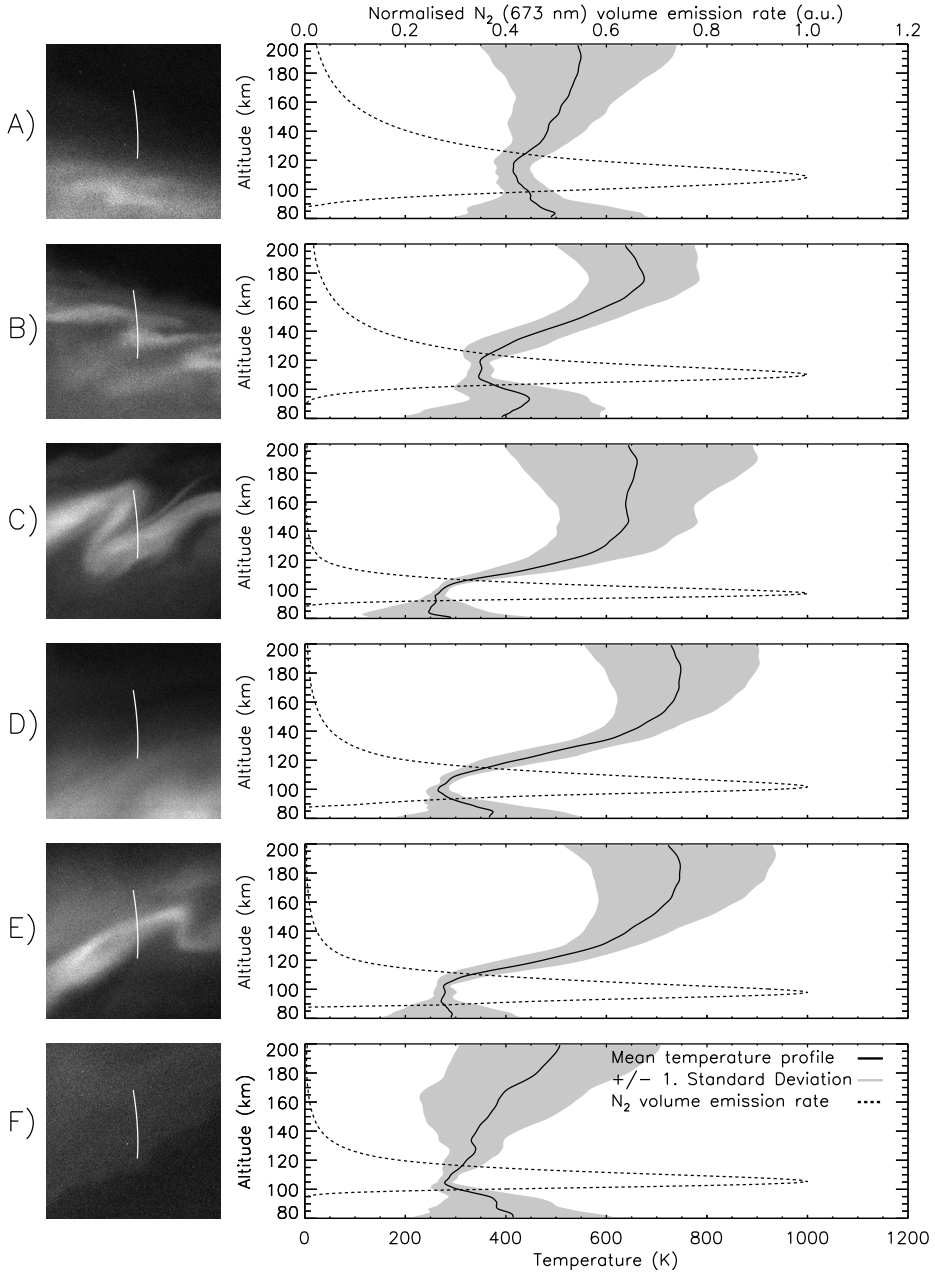


FIGURE 5.2: Temperature profiles associated with the time intervals outlined in Section 4. Each profile has been selected to show the dominant features present during those times. The images on the left show a single frame snapshot of the form of the optical arc, as observed by ASK during those intervals.

the thermosphere. The flow of these electrons constitutes an upward field aligned current (FAC), which is formed into a sheet like structure along the length of the arc, resulting in a net current flowing out of the thermosphere. Thus, in order to ensure current continuity in the ionosphere, a current closure mechanism must develop in response. The resulting current closure mechanism plays a large role in the final structure of the electrodynamical system that develops around auroral arcs. Marklund [1984] identified two primary mechanisms for closing current in the thermosphere: polarisation electric fields and Birkeland currents (field aligned currents). Polarisation fields are formed

when the precipitating electrons create a region of dramatically increased conductivity within the arc, in this region the background convection electric field drives a charge separation process between the free electrons and the ionised atmosphere. Eventually the polarisation field reaches an equal magnitude to the convection field, but in the opposite direction, and the superposition of these two electric fields results in a zero net field within the arc. Arcs where the majority of current closure is provided by the polarisation field are thus dubbed 'polarisation arcs' in the Marklund [1984] classification scheme. Birkeland currents are field-aligned currents produced by the motion of charged particles, primarily electrons, along the Earth's magnetic field lines. If the polarisation effect is not significant enough to ensure total current continuity, then the upward FAC sheet that is associated with the optical aurora is linked to and closed with a downward FAC sheet via meridional Pedersen currents. The downward FAC sheet is carried by upward moving cold ionospheric electrons. Arcs that require a significant parallel or field-aligned current component to ensure current continuity are thus classified as 'Birkeland arcs' in the Marklund [1984] scheme. The final classification, 'combination arcs' is a transient case wherein both mechanisms provide close to equally significant contributions. More recently, case studies of auroral arcs have shifted attention to observations of enhanced 'arc-associated' electric fields. These electric fields are significantly larger than the background convection field and are located adjacent to the optical arc and perpendicularly orientated (e.g. north or south pointing, for a standard east-west aligned arc). A point of particular interest to this study is that these enhanced electric fields are generally found to be asymmetrically distributed in space depending on the geomagnetic sector they are observed in; they are found on the equatorward edge of an arc in the evening sector, and on the poleward edge in the morning sector. This configuration is believed to be a consequence of the direction of the background convection field [Aikio et al., 1993, Marklund et al., 1982, Opgenoorth et al., 1990, Timofeev et al., 1987]. The conclusion of a case study of an auroral arc by Aikio et al. [1993] was that the enhanced electric field is produced in order to ensure sufficient Pedersen current to close the loop between the two FAC sheets. In this regime, the outflow of the cold ionospheric electrons associated with the downward FAC sheet creates a region of low conductivity on the appropriate edge of the arc [Doe et al., 1995], leading to a sharp 'step like' transition in the ionospheric conductivity. To overcome this steep conductivity gradient and maintain the required current continuity a large localised enhancement of the electric field is required.

The asymmetrical nature of the arc-associated electric field is a result of the orientation of the background convection field; to close the loop the Pedersen current must transport electrons away from the arc and therefore the **E** field must in turn be directed towards the arc (see Figure 2.5). The enhanced arc-associated electric field will naturally form on the edge of the arc where the direction of polar convection field aligns with the required direction of **E** field for the Pedersen current to flow; simply put, the

field manifests on the side of the arc where the background electric field is directed towards the arc. As a result, the orientation of the current closure system is dependent on whether the arc is located in the evening sector (northward convection field) or morning sector (southward convection field), with equatorward and poleward connecting currents found in each respectively [Timofeev et al., 1987]. For the auroral arc observed in this study, SuperDARN (Super Dual Auroral Radar Network) polar convection plots indicate an extremely contracted polar cap such that Svalbard was located in the morning sector of the auroral oval and thus the convection field was measured to be orientated equatorward. As such, the meridional Pedersen closing current also flows southward and the arc associated electric field is expected to be located on the arc's poleward edge. This conclusion is further supported by the EISCAT power profiles displayed in Figure 5.3 (panel 1) which cover an extended interval of time surrounding the event; intervals of enhanced electron density indicate the presence of the electron precipitation associated with the optical arc. Prior to the arc entering the radar beam, in the time interval between 20:50:30 and 20:50:40 UT, a localised decrease in electron density can be observed. This decrease in electron density corresponds to the closing (downward) FAC current sheet, carried by the cold ionospheric electron population. The complimentary evidence from both ESR and SuperDARN measurements suggest that the arc in question ensures current continuity in the ionosphere via a strong FAC system, in which case, due to its location in the convection field, the enhanced arc-associated electric field should be located on its poleward (leading) edge.

It is proposed then that the observed increase in neutral temperature, seen poleward of the arc between 80 and 160 km, is a result of the transfer of energy from ionospheric ions to neutrals via a frictional or Joule heating effect, in turn driven by the presence of a localised arc-associated electric field. As discussed in Section 2.2.4, the conventional expression for the rate of local Joule heating is given by:

$$Q_j(h) = \sigma_p(h)[\mathbf{E}_\perp + \mathbf{u}(h) \times \mathbf{B}]^2 \tag{5.1}$$

where $\sigma_p(h)$ is the Pedersen conductivity, $\mathbf{E}_\perp$ is the perpendicular electric field, $\mathbf{u}(h)$ is the ionospheric neutral wind velocity, $\mathbf{B}$ is the ambient geomagnetic field and h is the altitude. Since the ionospheric magnetic field can be assumed to be approximately constant ($\sim 5 \times 10^4$ nT), the main contributing factors to the heating rate are Pedersen conductivity ($\sigma_p(h)$), the electric field strength ($\mathbf{E}_\perp$) and the ionospheric neutral wind speed ($\mathbf{u}(h)$). Ions in the upper E and F-region ionosphere move according to the $\mathbf{E} \times \mathbf{B}$ drift, and thus substituting $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, we can re-write Equation 5.1 in terms of the difference in velocity between ions and neutrals as:

$$Q_j(h) = \sigma_p(h)\Big\{\big[\mathbf{u}(h) - \mathbf{v}(h)\big] \times \mathbf{B}\Big\}^2 \tag{5.2}$$

where $\mathbf{v}(h)$ is the ionospheric ion velocity. For Birkeland classification arcs, a large

and localised arc-associated electric field drives the necessary horizontal Pedersen currents in the ionosphere to close the forming current loop. These Pedersen currents are primarily carried by ionospheric ions moving in the **E** direction, whereas the neutrals are unaffected. Therefore, the Pedersen currents constitute a difference in velocity between the ion and neutral populations, increasing the $\left\{ \left[\mathbf{u}(h) - \mathbf{v}(h) \right] \times \mathbf{B} \right\}^2$ term in Equation 5.2 and producing a region where the Joule heating rate, $Q_j(h)$, is equally enhanced. Panels 2, 3 and 4 of Figure 5.3 display a comparison between the measured N_2 neutral temperatures and the ESR derived electron and ion temperatures, respectively. The electron and ion temperatures (panels 3 & 4) show a clear region of enhanced temperature, between 20:50:10 and 20:50:40 UT that is present at altitudes as low as 100 km. The increased ion and electron temperatures only appear on the poleward edge of the arc, and the same signature is not present on the equatorward edge, which passes through the radar beam at approximately 20:52:10 UT. This feature is in agreement with our results and is likely the lower limit of a large altitude region over which the Joule heating is present.

5.4.2 Arc-Associated Neutral Heating

Another feature of interest in the retrieved neutral temperature profiles is the regional temperature enhancements ($\sim$ 50 K) observed over a narrow altitude range between approximately 90 and 110 km. These smaller-scale temperature increases are seen during various integration intervals, but are only present during the total interval in which the optical arc is directly observable in the HiTIES field of view, rather than the interval before its appearance when the Joule heating is present. As such, from this point on wards, it will be referred to as 'Arc-associated neutral heating'. The observed neutral temperatures in this altitude region, during the passage of the arc, are highest during the window of time just after the bright edge of the arc has moved into the HiTIES slit. This can be seen in the top panel of Figure 5.1 as the green shaded region starting from the third integration interval, between 20:50:48–20:50:59 UT and centered on approximately 95 km. It is also visible in the temperature profile in panel B of Figure 5.2: the temperature shows a steep positive increase or 'bump' at altitudes below 100 km. In addition, in panels 3 & 4 of Figure 5.3, a signature of increased ion and electron temperatures can be seen at extremely low altitudes (80–95 km) during the same time interval. The highest magnitude ion and electron temperatures in this altitude region are seen between 85–90 km for the 5 second interval from 20:51:00 to 20:51:05 UT. Unfortunately, the N_2 neutral temperatures at this time do not display a similar signature; N_2 volume emission drops off significantly below 90 km, becoming negligible at altitudes below 85 km (signified by the white shaded region) and consequently accurate and reliable neutral temperatures are difficult to retrieve in this range.

However, in the fifteen seconds preceding this, the enhanced ion and electron temperatures are present at slightly higher altitudes, where there is more N_2 volume emission (lending more reliability to the method) and the aforementioned increase in the

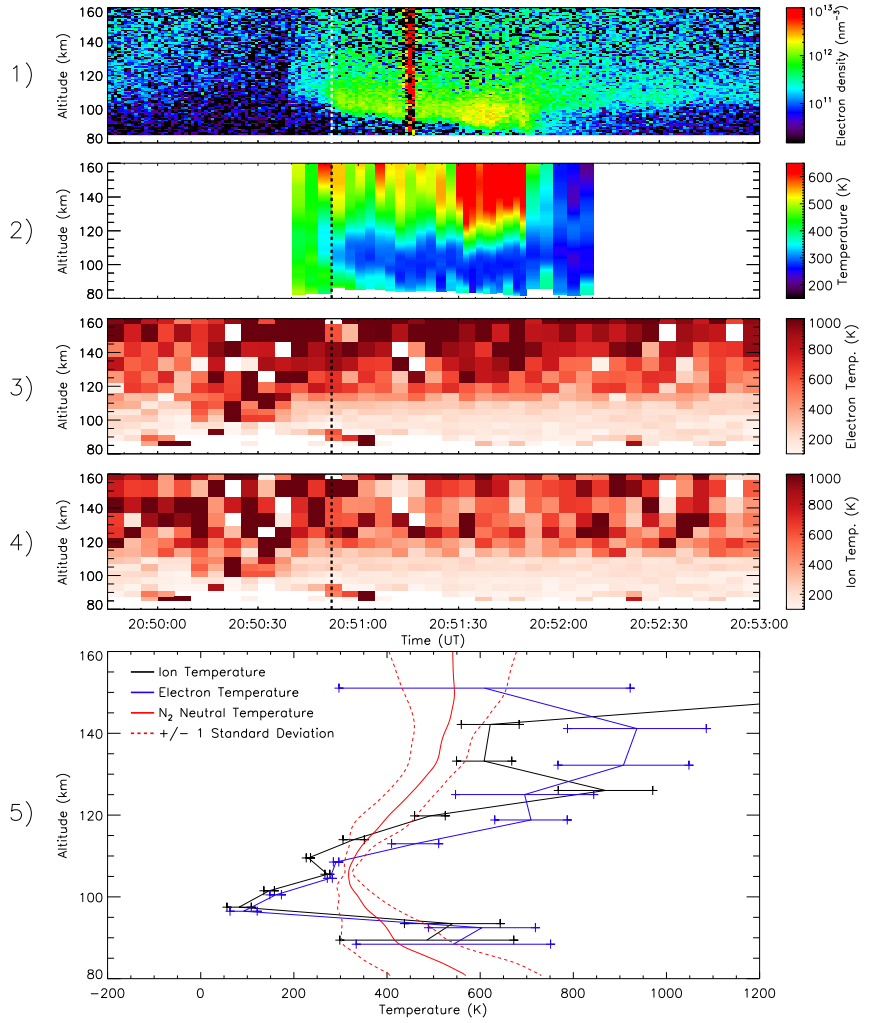


FIGURE 5.3: EISCAT Svalbard radar derived plasma parameters from 20:50:00 - 20:53:00 UT. A reduction in electron density immediately prior to the arrival of the arc in the radar beam, at 20:50:40 UT, is present in panel1. Panel 2 displays the measured N_2 neutral temperature profiles for the duration of the arc. Panels 3 & 4 show an asymmetric increase in electron and ion temperature, observed only on the poleward edge of the arc, as well as a smaller heating signature at low altitudes (90–100 km) associated with the arcs first entry into the ESR radar beam. The vertical dashed lines in the top four panels mark the time at which the electron and ion temperature profiles are plotted in panel 5. Ion an electron temperature profiles in panel 5 are displaced 0.5 km upwards and downwards, respectively, to assist readability.

N_2 neutral temperature is clearly observed. Individual ion and electron temperature profiles are plotted in the bottom panel of Figure 5.3 for the time stamp that is indicated with a vertical dashed line in the top four panels. The ESR derived electron and ion temperature profiles for this time show a clear spike in temperature below 95 km - a change of temperature that is significant in comparison to the errors associated with the measurements. This indicates that although this signature is near the floor of the ESR signal range, it is likely real. The corresponding N_2 neutral temperature profile, with a similar temperature enhancement at low altitudes, is displayed in red.

Further brief intervals of enhanced N_2 neutral temperatures between 80–100 km are observed in the top panel of Figure 5.1 for the remainder of the event. These observed temperature changes are not as significant in magnitude, do not typically persist for more than one HiTIES integration interval and are derived from a region of low volume emission, and hence large uncertainty. In addition, they display no clear correlation with the ESR ion and electron temperatures. As a result of this, for the remainder of this discussion we will focus solely on the enhanced temperatures observed in the leading edge of the arc.

The physical mechanism behind the arc-associated neutral heating is not as clear or well documented as the Joule heating discussed in the previous section. The majority of studies prior to this one have not reported a feature of this nature, particularly since it is seen in the neutral temperature. The signature occurs well within the E-region ionosphere where ion-neutral collision rates strongly couple the motion of both species and inhibit the formation of significant ion Pedersen currents. There are two potential heating mechanisms, related to the electrodynamics of a Birkeland arc, that could be responsible for the observed neutral heating signature - each of which is discussed in the following sections.

5.4.2.1 Ohmic Dissipative Heating from Field Aligned Currents

Zhu et al. [2001] presented results from a two-dimensional three-fluid model of the ionosphere-magnetosphere system. Along with a companion paper [Lanchester et al., 2001], they conducted an extensive study into the various sources of ionospheric ion and electron heating in order to explain optical and radar observations of a filamentary aurora event. Of particular interest is their treatment of electron heating sources, which they separate into two regimes: particle precipitation and Ohmic heating. They found that to account for the observed changes in electron temperature, a large contribution from FAC-associated Ohmic heating is needed. In the Zhu et al. [2001] model, the Ohmic heating effects are observed in a narrow current sheet that is located at sharp precipitation boundaries, such as those observed in our event and others [Arnoldy, 1974]. The Ohmic heating term is proportional to $\eta \mathbf{j}$, where η is the resistivity and $\mathbf{j}$ is the current density. The model shows an approximately constant current density and an exponential increase in resistivity with decreasing altitude; as a result, it would be expected that the maximum Ohmic heating rate is at the bottom of the FAC sheet [Zhu et al., 2001, Figure 7b.]. Furthermore, since the electron-neutral collision frequency drastically increases with decreasing altitude, the electron cooling via neutral collisions also increases, and the characteristic timescale to achieve a steady state between the two is short. Whilst the neutral temperature would likely not respond significantly to the electron temperature fluctuations, we would expect that given a strong enough field-aligned current, the corresponding signature in neutral temperatures, if any, would be located in this region.

The primary counterpoint to this theory is the extremely localised nature of the observed temperature enhancements. Whilst the combination of the exponentially decreasing (with altitude) atmospheric resistivity profile and relatively constant field-aligned current sheet could produce a small 'sweet-spot' of altitude where heating would be most efficient (where the product of the resistivity and field-aligned current density is at its maximum) it could potentially also produce a more gradual heating signature, across a larger range of altitudes as the resistivity changes. However, the true shapes of the resistivity and field-aligned current density profiles are not known, and they are unlikely to be uniform or smoothly varying. The field-aligned current density likely decreases somewhat with decreasing altitude, as the FAC closes through horizontal currents, and the resistivity profile is highly dependent on local atmospheric conditions. It is possible the resistivity 'spikes' just below the enhanced electron densities, leading to a sharp boundary in the resistance and a consequent rapid increase of the Ohmic heating rate, but further work would be needed to determine the feasibility and effect of such a structure.

5.4.2.2 Electron Pedersen Current Heating

The second potential source for the enhanced neutral temperatures observed within the arc is the 'Pedersen current carried by electrons' which was described by Buchert et al. [2008]. Buchert et al. [2008] present ESR observations of the cusp ionosphere which show considerably enhanced electron temperatures, to values as high as 4000 K, at similar altitudes (<120 km) to the signatures seen in Figure 5.1. The increased electron temperatures last for a significantly longer interval of time that those presented here (approximately 2.5 hours) and are observed at 6.5 minute resolution - the time taken to cycle over the three pointing directions used in the chosen radar experiment. However, the electron temperatures fluctuate greatly during the observation window, and the highest temperatures generally only persist for a single time step. Furthermore, coincident observations of the ion temperatures indicate the presence of a similar frictional or Joule heating effect at altitudes above 120 km.

Buchert et al. [2008] attribute the ion and electron heating signatures to the presence of a perpendicular **E**-field, which is derived from observed line-of-sight velocities along the three pointing directions. The resulting component shows an increase in the average electric field strength from approximately 0 to 100 mV/m during the same interval. They determined that the ion heating is driven by the usual Pedersen current, which arises from collisions between ions and neutrals, as was discussed in Section 5.4.1, whilst the electron heating is the result of "irregularities and plasma turbulence" which exist in the presence of high electric fields [Buchert et al., 2006]. They parameterised this effect as an anomalous ion-electron collision frequency which affects the mean electron motion analogously to standard Pedersen conductivity. Thus, the effect can also be thought of as an anomalous Pedersen conductivity. The key points are that the resulting Pedersen current is primarily carried by electrons, not the ions,

and the peak altitude of the heating rate is at around 100-115 km rather than 125-130 km as is more traditionally expected from classic Pedersen currents. These warm electrons transfer heat to the neutral gas via inelastic collisions with the neutral molecules that are abundant at the lower altitudes. As a result of the anomalous Pedersen conductivity, the distribution of atmospheric heating via field-aligned currents is shifted towards lower heights where enhanced electric fields and the associated irregularities are located.

The Buchert et al. [2008] result is of particular interest to this study as it presents an extremely similar set of circumstances: an auroral event with strong field-aligned currents closing through the ionosphere in the presence of an enhanced electric field, whilst also reproducing ion and electron temperature observations that display similar structures in altitude. The core difference between the two studies is the temporal and spatial resolutions of the observations; the results presented here are at a significantly smaller scale, particularly in time, and at slightly lower altitudes. Furthermore the Buchert et al. [2008] observations are taken from the cusp of the ionosphere, a region of open magnetic field lines where structured aurora like that which is presented in this study is unlikely. However, the mechanism proposed requires only the presence of large electric fields, something which is already invoked to explain frictional or Joule heating effects, and thus offers a potential secondary explanation for the observed low altitude electron and neutral temperature enhancements seen within the event.

5.4.2.3 Associated Auroral Dynamics

In an effort to further investigate the arc-associated heating signatures, the ASK video of the corresponding auroral dynamics can be examined. The ESR ion and electron temperature data (Figure 5.3, panels 3 & 4) show three distinct time intervals that contain enhanced temperatures ($>\sim$ 500 K) at the expected altitude range: X) between 20:49:55 and 20:50:10 UT; Y) between 20:50:45 and 20:51:05 UT; and Z) between 20:52:50 and 20:52:55 UT. Only the second of these intervals (Y) occurred during a interval of significant N_2 emission, thus allowing for the corresponding N_2 neutral temperatures to be determined, and a corresponding signature in the neutral temperatures was observed.

The ASK video for interval Y displays extremely dynamic auroral features, primarily consisting of strong shear flows and a system of multiple bright curls propagating along the poleward arc boundary, from west to east. Interval X occurs during a interval of negligible auroral emission and therefore has no associated structure. Interval Z occurs approximately 10 seconds after the end of N_2 neutral temperature observations at a time when the N_2 brightness is too low for any appreciable results. However, the ASK video for this time shows an extremely faint set of curls moving slowly through the magnetic zenith.

The presence of auroral curls in the ASK field of view during two of the three highlighted time intervals (Y & Z) in which there was observable N_2 emission, suggests that

the source of the arc-associated heating is potentially linked with the formation mechanism responsible for the curls. Auroral curls are a commonly observed auroral feature and Hallinan and Davis [1970] provided the initial outline for what is a generally well accepted theory for their formation: the Kelvin-Helmholtz instability. The K-H instability grows from a region of shear velocity in the ionospheric plasma which is located on either side of a sheet of negative charge [Hallinan, 1976, Vogt et al., 1999]. At the boundary between the two oppositely directed plasma flows, the instability forms and grows into a system of spiral like structures known as curls. Interestingly, the field-aligned currents demonstrated in the Zhu et al. [2001] model also develop as a consquence of shear flow in the ionospheric plasma, and a sheet of field-aligned current is a required component for both of the heating mechanisms presented in the above section, since both arise from closure currents through the ionosphere. Thus, if the arc-associated heating is seen to be co-located with auroral curls, these observations lend further evidence to both the Kelvin-Helmholtz theory for the formation of auroral curls and potentially provide a concrete link to further associate their formation with intense field-aligned currents. Unfortunately, the ASK observations alone are not sufficient to distinguish between the two proposed mechanisms - further analysis on more curl systems is required.

5.5 Conclusions

This work provides direct observational evidence of the large scale electrodynamical system responsible for the formation of an auroral arc, including a strongly enhanced perpendicular electric field adjacent to the optical emission, which arises due to current closure requirements in the auroral atmosphere. A schematic of the inferred electrodynamical structure for an auroral arc is presented in Figure 5.4; ionospheric currents are indicated by yellow arrows, and ionospheric electric fields by blue arrows. The auroral arc (green shaded area) is embedded in a region where the polar convection electric field ($\overline{\mathbf{E}}_{conv}$) is oriented southward. This enhances the arc-associated electric field ($\overline{\mathbf{E}}_{arc}$) on the poleward edge of the arc (left hand side) and reduces it on the equatorward edge of the arc (right hand side). The Pedersen (J_p) and downward field-aligned currents that close the ionospheric current loop form on the side of the arc that exhibits the enhanced electric field (poleward edge). The Joule heating affect discussed in Section 5.4.1 occurs in the region of enhanced arc-associated electric field on the poleward edge of the arc between the two regions of oppositely oriented field-aligned currents. The strong electric field in this region drives the connecting Pedersen currents and the Joule heating via collisions between ions and neutrals.The Arc-associated neutral heating (Section 5.4.2) is believed to occur in two similar but distinct locations depending on the proposed mechanism: the Ohmic heating mechanism proposed in Section 5.4.2.1 arises from a resistive or dissipate heating effect along the intense field aligned current structures, the most significant portion of which is present at the low altitudes where

the atmospheric resistivity is increased; the electron Pedersen current heating mechanism proposed in Section 5.4.2.2 is also a result of the enhanced electric field poleward of the arc, which produces an electron Pedersen current, heating the ionosphere via electron neutral collisions.

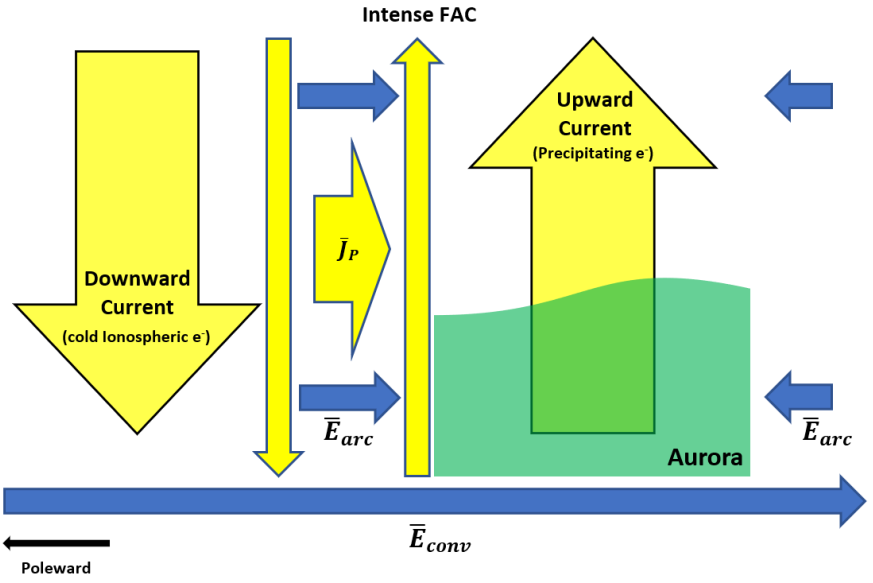


FIGURE 5.4: A schematic overview of the proposed electrodynamical structure of an auroral arc - the perspective is looking along the arc from West to East, with North to the left of the diagram. Field aligned currents are indicated by yellow arrows, and ionospheric electric fields by blue arrows. The auroral arc is represented by the green shaded area. Regions of upward and downward field-aligned current are connected through the ionosphere by meridonial Pedersen currents (J_p), which are driven by an enhanced arc-associated electric field.

The thermosphere is an important energy sink for magnetosphere-ionospheric interactions and observations of electrodynamic process at high resolution are extremely important for informing and quantifying the total magnetospheric-ionospheric energy budget. The neutral temperature has shown a rapid and significant response to the presence of electric currents on time scales of the order of seconds and over spatial scales smaller than 10 km. First results from a new observational technique provide high resolution measurements of the evolution of the atmospheric neutral temperature profile during a high energy auroral event over Svalbard. The process involves simultaneous observations from two optical instruments, Auroral Structure and Kinetics (ASK) and the High Throughput Imaging Echelle Spectrograph (HiTIES), as well as electron density measurements from the EISCAT Svalbard Radar (ESR). By fitting a library of synthetic N_2 emission spectra to observed spectra taken from HiTIES, an estimate of the shape of the neutral temperature profile over a series of time steps was produced for the duration of the auroral event. The key results of this study are summarised below.

Optical observations begin at 20:50:40 UT, just after a decrease in ESR electron density (Figure 5.3, panel 1) that is associated with a downward current of upward flowing cold

ionospheric electrons (Section 5.4.1). For the first 8 seconds (20:50:40–20:50:48 UT) of observation, the neutral temperature profile shows enhanced temperatures at altitudes as low as 80 km, generated by ion-neutral frictional or Joule heating which results from a strong arc-perpendicular electric field and consequent meridional Pedersen currents. These Pedersen currents act as a current closure mechanism linking the downward current with the upward current associated with the optical arc.

As the arc moves in to the field of view over the next 8 seconds (20:50:48–20:50:56 UT) the associated temperature profile rapidly evolves. Neutral temperatures between 100 and 130 km decrease by approximately 100 K and a small increase of approximately 50 K is seen between 90 and 100 km, coincident with simultaneous ESR ion and electron temperature enhancements. Neutral temperatures during this interval may be increased due to an electrodynamical structure associated with the presence of a bright system of auroral curls on the leading edge of the arc, suggesting that the heating is potentially associated with the specific formation mechanism responsible for the curls themselves. Further investigations into this phenomenon are needed before final conclusions can be confidently drawn (see Chapter 6).

The arc then remains in the field of view for approximately 60 seconds following its first entry, and the corresponding temperature profiles show relatively small variation with time. Neutral heating at the same altitudes as previous signatures (90–100 km) is present, potentially caused by filamentary current systems embedded in the arc itself, although no accompanying ESR ion or electron temperatures variations are observed. When the arc exits the the field of view, the temperature profiles measured on its equatorward edge transition to a significantly cooler regime than those observed on the poleward edge, suggesting an asymmetrical arc-associated electric field.

5.6 Future Work

The work presented here was a reassuring first application of the in-development observational technique discussed in Section 4. Results were able to confirm the expected heating signature associated with the electrodynamic structure of a simple auroral arc, the orientation of which aligned with independent observations of increased ion and electron temperatures taken from ESR, and with the convection field from SuperDARN measurements. In addition, it highlighted a potential avenue of further research via the retrieval of smaller-scale neutral heating signatures at altitudes not typically associated with the usual Joule heating mechanism. The mechanism behind these heating signatures was unclear, but a promising correlation was observed with the presence of auroral curls in the magnetic zenith at coincident time intervals. As a result, it was determined the next auroral event chosen for analysis should ideally contain a well behaved system of curls that would allow for a more rigorous investigation of potential neutral heating processes.

Chapter 6

Curl Associated Neutral Heating

This Chapter contains a discussion of the research relating to neutral heating processes that are associated with regions of intense field-aligned currents embedded within auroral forms. The work presented submitted to JGR as 'Neutral heating by strong field-aligned currents in auroral curls.'

The data from this study is publicly available and hosted by the University of Southampton (DOI: 10.5258/SOTON/D1965).

6.1 Introduction

The results presented in Chapter 5 provided clear empirical evidence of enhanced neutral temperatures in the vicinity of an auroral arc. The observed temperature enhancements are driven by neutral heating processes that are associated with electrodynamical structure in the ionosphere responsible for the arc's formation. The primary result was a region of enhanced temperatures (relative to the remainder of the event) that was located in the area directly adjacent to the poleward edge of the auroral arc. This signature was identified as a region of Joule heating, driven by an asymmetrically structured arc-associated electric field which develops alongside the arc in response to current continuity considerations in the ionosphere. EISCAT Svalbard Radar measurements displayed a similar structure in the simultaneously observed ion and electron temperatures. This result aligned well with contemporary theories for arc-formation mechanisms and provided observations of the effect of the predicted electric field structures on thermospheric temperatures at unprecedented resolution.

A second result of interest was also discussed in Chapter 5; a smaller and more localised heating signature was observed at low altitudes within the arc, during time intervals in which ASK images displayed the presence of auroral curls of varying brightness. Hereafter known as the 'curl-associated heating signature'. The important distinction between the two temperature signatures is that the curl-associated heating signature was localised over a much smaller height extent, confined to lower altitudes (between $\sim$80 and 100 km), and was observed to exist only for brief intervals of time

throughout the passage of the arc. The most notable and consistent increase in temperature of this nature was present during the first eight seconds of the arc's direct appearance in the field of view. The location of these enhanced temperatures in the context of the entire auroral arc was of particular interest as it is generally accepted that ionospheric electric fields are significantly diminished in the highly ionised region of the thermosphere that is generated by precipitating electrons. Consequently, the potential heating mechanism responsible for the observed signatures is unlikely to be the result of an electric field, and thus cannot be attributed to a traditional Joule or frictional heating effect. From the results presented in Chapter 5, it was difficult to reliably determine the heating mechanism responsible for the observed temperature structures, and two competing theories were proposed: dissipative heating via strong-field aligned currents and electron Pedersen heating.

As mentioned, the curl-associated heating signature was observed during intervals containing auroral curls. Auroral curls are some of the most commonly observed features present in the dynamic small-scale ($<$10 km) aurora seen by the ASK instrument. Auroral curls are typically observed as a quasi-periodic chain or system of counter-clockwise rotating spiral features, with a separation between curls or 'wavelength' of less than 10 km. Curls are short lived and fast-moving structures; individual curls are often only visible for a number of seconds before either dissipating or moving out of the field of view. The initial theory for the formation mechanism of auroral curls was established by Hallinan and Davis [1970], who concluded from optical observations that they develop from the Kelvin-Helmholtz instability. They postulated that the K-H instability grows from a region of strong shear velocity in the ionospheric plasma, located on either side of a vertical sheet of negative charge. Over a brief interval, the instability then develops into a short-lived system of approximately evenly spaced counter-clockwise rotating curls. Vogt et al. [1999] expanded on this theory by using a newly developed technique to measure the shear velocity and vorticity of a long-lived curl system and test the K-H dispersion relation determined by Hallinan and Davis [1970]. They found a discrepancy between the two results and suggested that the auroral acceleration region or other transient processes should be considered in theories for the formation of auroral curls. This possibility was also considered in Hallinan [1976], who reference a theory developed by Swift [1976] which described an electrostatic shock model for auroral formation that results in V-shaped equipotential lines (with respect to the vertical magnetic field). Crucially, the Swift [1976] model stipulates that there would be a field-aligned sheet of negative charge at the centre of the structure, as required for curl formation, which is supported by an upward field-aligned current sheet. Thus auroral curls are not only associated with negative charge sheets, but also potentially with field-aligned currents. The plasma shear flow component of curl formation, driven by the $\mathbf{E} \times \mathbf{B}$ drifts on either side of the sheet of negative charge, is easily verified observationally, and counter-streaming flows are often observed to develop into auroral curls. The sheet of negative charge itself, and the associated upward field-aligned

current structure, is much more difficult to verify observationally. Incoherent scatter radars struggle to resolve charge structures that are narrower than the beam width. However, the observations presented in Chapter 5 provided a first glimpse into an observable consequence of these field-aligned currents in the neutral temperatures, in turn opening the door for the verification of this aspect of the curl formation theories.

Narrow sub-arc scale field aligned currents are believed to arise as part of the larger current structure that results from the precipitation of magnetospheric electrons. The general picture for a 'Birkeland' classification arc (see Marklund [1984]) is two oppositely directed regions of field aligned current: an upward current carried by precipitating magnetospheric electrons and a downward current carried by up-flowing cold ionospheric electrons. The two field-aligned current regions are oriented parallel to the arc structure and are connected meridonally through the ionosphere by horizontal Pedersen currents, completing the required current loop. This 'large-scale' picture or overview of an arc is complicated by the likelihood of filamentary current structures present on sub-arc scales. The upward and downward current structures are better described as being regions where the total or net flow of the currents are orientated in a specific direction, and they are actually comprised of multiple smaller current structures of varying intensities and direction [Lühr et al., 2015, Stasiewicz and Potemra, 1998]. Therefore the field-aligned current density within the arc is expected to be non-uniform, with regions of intense upward current existing alongside regions of relatively reduced and even downward current. The variations in neutral temperatures that are observed within the arc in Chapter 5 are believed to be a response to these structures and as such provide further evidence to support the delamination of the large scale current structure.

The present work makes a direct correlation between auroral curls and filamentary field-aligned currents. This is achieved by applying a greatly improved version of the analysis technique applied in Chapter 5. The new auroral event is from a different MLT and has a different morphology from that in Chapter 5.

6.2 Methods

Significant improvements have been made to the method that was used in Chapter 5. Primarily, as was discussed at length in Chapter 4, the fundamental analysis method was re-imagined as a optimisation problem and a simulated annealing algorithm was developed and employed to estimate the solution. The purpose of this update was to ensure that the inversion technique could accurately identify localised heating. The analysis still relies on the fitting of synthetic spectra to observed spectra, with each synthetic spectrum corresponding to a trial neutral temperature profile, but sources of error have been minimised and the new algorithm approaches the problem in a more robust and efficient manner.

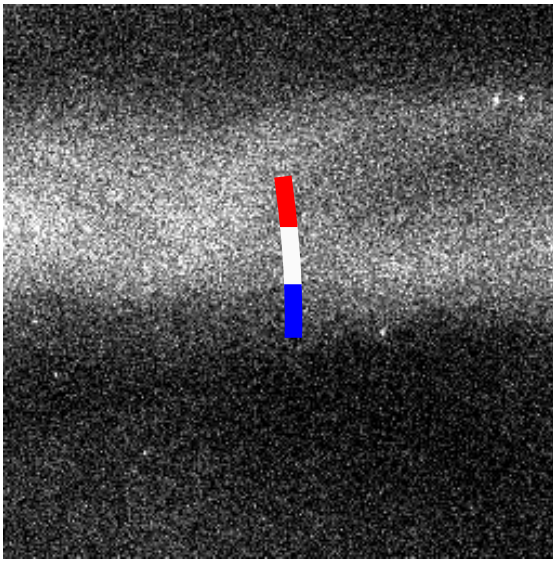


FIGURE 6.1: An example frame from ASK, with the sectioned HiTIES observation slit included.

Further improvements were made to the resolution of the results by dividing the HiTIES integration slit into three sections, allowing the temporal and spatial development of the auroral structure to be analysed more thoroughly. The slit was partitioned lengthwise (in the North-South direction) and each of the sections are exposed to a different region of auroral emission - shown by the red, white and blue sections of the line on the ASK frame in Figure 6.1. Each section thus produces an independent spectrum, with properties that correspond to the emission in that region only. Dividing the HiTIES integration slit in this way means that the full observational method described in Chapter 4 was separately applied to each section for each particular integration interval. The characteristic energy and flux of precipitation were calculated from the ASK pixels that are co-located within each section of the slit, then used in three separate model runs to produce three sets of N_2 volume emission rate profiles. The model run from the central section of the slit is still verified by the ESR derived electron power profiles, but the flanking top and bottom sections are no longer spatially aligned with the radar beam and thus direct comparisons with ESR are less reliable. However, comparison with the central section still ensures realistic volume emission rate profiles across the length of the slit. Finally, the full annealing algorithm is individually applied to each of the slit sections and three completely independent temperature profiles are produced for every HiITES integration interval. With three neutral temperature profiles, neutral temperature variations in both space and time can be decoupled from each other and spatial structures in the neutral temperature can be resolved more accurately. Previously, if the slit straddled a boundary in the aurora, the resulting spectrum was the result of the combined auroral emissions originating from both sides of the boundary, which often contained significantly different emission. This affect can produce ambiguous spectra that are comprised of a mixture of emission from both regions, resulting in the fitting

process outputting a temperature profile that favours an average solution between the two regimes. Splitting the slit and carefully choosing appropriate integration intervals mitigates this problem and allows for temperature changes across structured auroral forms and boundaries to be investigated more accurately.

6.3 Observations

The auroral observations in this study were made near Longyearbyen, Svalbard (78.15° N, 16.03° E) at approximately 06 UT (9 MLT) on the 1st of February, 2017. At this time, Svalbard is located between the noon and dawn sectors of the Polar Cap. A bright and variable auroral arc structure formed at 06:06:00 UT in the top of the ASK image. The arc moved southward across the local magnetic zenith, filling the ASK field of view and remaining as such for approximately one minute before it moved northward again. The entire event took place over two minutes. The arc initially appeared as a bright and narrow structure, with plasma flowing eastward along its length. The arc then widened and an interval of intense counter-streaming flows developed. The resulting region of maximum shear flow between the two plasma populations was co-located with the approximate location of the magnetic zenith. As the event progressed further, the counter-streaming flows transitioned into a stable and tightly wrapped system of bright auroral curls, which moved steadily in an west-to-east direction along the length of the auroral arc. Finally the curl system broke up and the arc faded and evolved into a region of diffuse and relatively dim emission, before it moved further northward and out of the field of view completely.

The passage of the arc was simultaneously observed by the ASK and HiTIES instruments (Chapter 3) as well as the ESR beam - providing the variety of observations that are needed to facilitate the observational technique described in Chapter 4. As a result, estimates of the neutral temperature-height profile were produced for a number of HiTIES integration intervals and for the three spatially adjacent sections of the HiTIES slit. These observations allowed for a detailed investigation of the temperature response of the neutral atmosphere to the presence of the varied and dynamic small-scale auroral structures discussed above.

6.4 Results and Discussion

The two minute auroral event is divided into a total of 47 integration intervals, each of which produces three temperatures profiles, resulting in a full set of 141 neutral temperature profiles. The length of each of the integration intervals varies between two and five seconds, and is dependent on the emission intensity present at that point in the event. Individual intervals of interest are examined in Section 6.4.1 and Section 6.4.2 and an overview of the large scale time-dependent structure of the temperature profiles is given in Section 6.5.

Auroral structures on the scales observed by the ASK instrument (<10 km) are exceptionally variable and more often than not display significant and dynamic structural changes on time scales significantly shorter than seconds. Since the integration intervals cannot be reduced further without an unacceptable signal-to-noise ratio in the spectra, some intervals suffer from a similar smoothing effect as the spatial problem that was discussed in the previous section. If the auroral structures change significantly within a particular interval, then the resulting spectrum corresponds to a combination of emissions from potentially distinct structures, thus producing a temperature profile that is an average of the different auroral forms. As a result, intervals during which the auroral structure was relatively stable, or at least displayed the same type of auroral structure through the full duration, are favoured in the analysis and more confidence can be placed in the results. Generally, profiles from the same time intervals exhibited similar large scale structures across all three sections of the slit, with the error range of the temperatures having significant overlap. This is to be expected during intervals in which the scale size of the auroral structure is larger than the HiTIES integration slit, or during intervals where the temporal variations of the aurora are so rapid that significant blurring along the slit could occur (as discussed above). At these times, the fact that the three independently determined temperature profiles showed such similarities is a reassuring result. Furthermore, it means that intervals of interest are easily identifiable as times when the returned temperature profiles show significant differences along the slit. These temperature differences correspond to real and significant structures in the neutral temperature.

6.4.1 Joule Heating on the Equatorward Edge

Temperature profiles from three chosen integration intervals are displayed in Figure 6.2. From left to right, each column contains: a representative ASK frame that highlights the primary auroral structure present during that interval; a combination plot containing all three temperature profiles, each derived from a different section of the HiTIES slit; and the individual temperature profiles from each section, displayed with errors of one standard deviation on the temperature. The grey shaded region in the second column indicates the primary altitude region of interest, where the most significant temperature differences are observed. Select panels also contain a plot of the error range of one of the other profiles from that interval, to highlight observed differences between the profiles. The colour of each profile corresponds to the section of the HiTIES slit it was retrieved from, as indicated in the ASK images, with the exception of the central (white) section which is plotted in black. The transition from a solid coloured line to a dotted coloured line in the 'Top', 'Middle' and 'Bottom' columns represents the altitudes above and below which the N_2 volume emission rate is less than 10% of the peak value.

The first two intervals (rows A and B) occurred near the start of the event when the aurora was a bright and filamentary arc. Interval B follows on directly from interval

A, and the two track the motion of the arc as it moves southward through the HiTIES slit. Interval C occurs much later in the event and displays similar auroral structure; namely a strong boundary between a region of relatively enhanced emission intensity in the north and one of significantly reduced emission intensity in the south. All three intervals (A, B & C) were integrated over a five second duration, during which the aurora is approximately stable. There is sufficient emission present in the southern, dimmer sections to produce acceptable spectra. All three of these intervals are individually discussed in the following sections.

6.4.1.1 Interval A

A narrow auroral arc is present in the northern portion of the ASK field of view. The topmost section of the HiTIES slit (red) samples this arc directly, whilst the remaining two sections (white & blue) sample the region directly equatorward of the arc. The top profile exhibits generally cooler temperatures than the middle and bottom profiles.

6.4.1.2 Interval B

The same narrow auroral arc has partially brightened and widened, moving southward in the ASK field of view. During this interval, which immediately succeeds interval A, the top and middle sections of the HiTIES slit (red & white) are now sampling the region of bright emission, whilst the bottom section (blue) remains to the southward of the border. In this interval, the top and middle profiles have approximately similar shapes and the bottom profile exhibits a generally hotter structure. A comparison of the first column of the temperature-altitude panels in Figure 6.2 (combination of all three profiles) during interval A and B shows how the warmer black profile (central section) in interval A now displays cooler temperatures similar to the red profile (top section) in interval B. This reduction in the observed temperatures from the central section of the slit corresponds to the passage of the boundary of the emission region as it progresses along the length of the HiTIES slit. During both intervals, the temperatures directly adjacent to the arc are enhanced with respect to those observed within the arc itself.

6.4.1.3 Interval C

A more diffuse auroral structure is present during this interval, which is taken much later in the progression of the auroral event. Nonetheless, a similar boundary in the emission intensity to that which is observed in intervals A & C is present. At this time, the top and middle sections of the HiTIES slit are observing the region of enhanced emission, whilst the bottom section is once again sampling the region directly equatorward. A very similar structure in the temperature profiles to that in interval B is observed; the bottom profile (blue) exhibits generally warmer temperatures than the middle and top profiles.

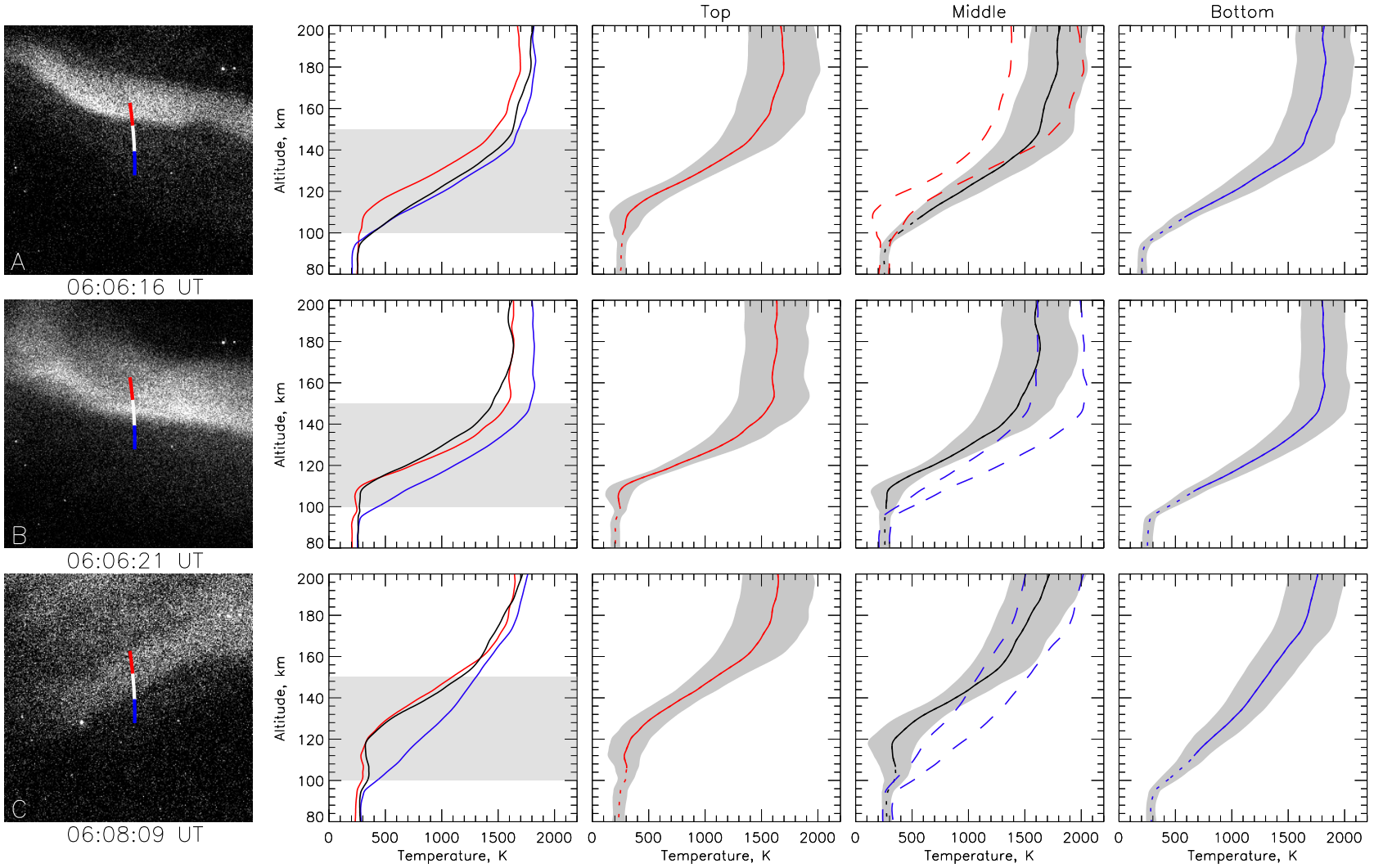


FIGURE 6.2: Joule heating on the arc's equatorward edge. Each row indicates a separate integration interval in HiTIES. A representative frame from ASK is included to indicate the dominant auroral structure present during each interval. Solid coloured lines display the mean temperature profile from each integration interval and each section of the HiTIES integration slit. The dotted segments of these profiles represent altitudes at which the N_2 volume emission rate is below 10% of its peak value during that interval. Light grey shading in the second column indicate the altitude region of interest, where the most significant temperature differences are observed. Dark grey shaded regions in the remaining columns display the error range in temperature, which is equal to one standard deviation of the averaged profiles. Long dashed lines indicate the error envelope of a temperature profile from the same integration interval (row) but from a different section of the slit (column).

Comparison of the error ranges for each of the returned profiles (grey shaded regions) allows for a determination of the significance of the observed temperature differences between profiles. If there is little to no overlap between error envelopes, it is reasonable to say that the temperature differences between the two profiles are likely to be meaningful and real. Conversely, if there is significant overlap, the profiles cannot be reliably distinguished from each other and no substantial conclusions can be drawn. At some altitudes, overlap between the temperature profiles indicates the reliability of the method. For example, there are significant similarities between all profiles at altitudes below 100 km. As discussed in Chapter 4, the neutral temperature in this region is constrained by an OH temperature consistency check that is applied to every trialled profile during the analysis. Each of the independent temperature profile estimates, across not only the different sections of the slit but also for different integration intervals entirely, produced comparable neutral temperatures at OH airglow altitudes. In fact, all 147 profiles demonstrated a temperature of approximately 250 K in this altitude region. This means that the best fitting temperature profiles across the entire event utilised similar OH line intensity components in their respective spectral fits, resulting in a consistent OH temperature and thus constraining the temperatures in that region effectively. This result is expected as the low altitude of the OH layer means that it is unlikely that auroral processes will significantly impact the chemistry there. Another consistent feature in all of the plotted intervals is that for altitudes above approximately 150 km, each of the temperature profiles also display similarities, with considerable overlap in the error envelopes in this region for every interval. Certainly, at these altitudes the profiles cannot be reliably distinguished from one another, and any observed temperature differences between the mean profiles are well within the respective error limits. As was discussed in Section 4.10, the expanding errors in this region are a consequence of the diminishing N_2 volume emission rates at higher altitudes restricting the accuracy of the method.

The remaining altitude region, between approximately 100 and 150 km, is where the majority of the N_2 volume emission occurs and is accordingly where the method is most accurate. In the middle column of Figure 6.2, the error limits of relevant profiles have been over-plotted alongside the profile from the middle section of the slit. Red dashed lines represent the errors on the profile from the top of the slit, and blue dashed lines the errors on the profile from the bottom of the slit. These were chosen to display the significance of the difference in temperatures between the relevant profiles. For all three intervals (A, B, & C), there are clear and significant differences in the neutral temperatures in the profiles from inside the auroral arc, to the profiles from the region directly adjacent to and equatorward of the auroral arc. This is particularly clear in the progression from interval A to interval B; as the filamentary arc structure moves southward through the HiTIES integration slit, the temperature structure follows it. In interval A, two completely independent outputted temperature profiles, both adjacent to the arc (black and blue), display an extremely similar shape, with enhanced

temperatures with respect to the profile in the arc (red). The maximum difference in temperatures between the two regimes is approximately 300 K at 110 km. In interval B, the same arc moves southward and the middle section of the slit (white) is now also directly observing it. The corresponding temperature profile (black) cools, displaying similar temperatures to the profile derived from top (red) section of the slit. Now only the bottom section (blue) of the slit is observing the region adjacent to the arc, and it maintains its relatively enhanced temperatures - with a similar maximum of about 350 K at approximately 110 km. The significance of the difference in temperature is even more apparent during this interval. Finally, interval C displays a comparable difference in temperatures between the profiles inside of and adjacent to the arc, with a maximum difference of approximately 350 K at a slightly higher altitude of 120 km.

This clear and consistent structure in the neutral temperatures observed across the equatorward boundary of the auroral arcs suggests a similar underlying structure in the process responsible for the heating. A comparable region of enhanced neutral temperatures was observed adjacent to an arc structure during the event discussed in Chapter 5. These enhanced temperatures were attributed to a Joule heating effect that arises from a strong arc-associated electric field. The present result aligns with this theory. A full discussion of the Joule heating mechanism is given in Chapter 5, but to aid discussion the fundamentals will be repeated here. Local scale auroral Joule heating arises from a strong arc-associated electric field which forms in the ionosphere to ensure current continuity by facilitating the flow of meridional Pedersen currents. These Pedersen currents connect the two regions of upward (auroral) and downward (return) field aligned current, closing the loop. The background polar convection electric field is significantly enhanced in this region in order to overcome a large step like gradient in conductivity across the edge of the arc. This new arc-associated electric field is generally only observed on one side of the auroral arc. This asymmetry is believed to relate to the orientation of the convection field, with the arc-associated field developing on the edge of the arc at which the direction of the convection field aligns with the required current flow and thus contributes to its magnitude. Thus, the resulting Joule heating is similarly asymmetrical and is observed on either the equatorward or poleward edge of the arc - as was seen in Chapter 5.

Supporting observations from SuperDARN and ESR were not as useful for the analysis of this arc as they were in Section 5. A reliable determination of the orientation of the convection field was not possible for this event. SuperDARN convection plots covering the interval of this arc's passage are inconclusive: there are no data points in the vicinity of Svalbard and the archipelago appears to be roughly in or near to the magnetic cusp. Both of these factors make the determination of the direction and magnitude of the polar convection field unreliable. Furthermore, ion and electron temperatures from ESR did not prove useful for comparison with the neutral temperature. Generally speaking the arc in question is not as well-behaved as that studied in Chapter 5, which was a stable east-west aligned arc that moved steadily through the magnetic zenith. The current

event is an extremely variable mixture of various dynamics, including filamentary arcs, curls, and large patches of diffuse emission. As a result of this behaviour, the relatively large radar beam and five second minimum integration times which are inherent to the ESR experiment appear to have difficulties resolving the short lived and extremely localised temperature structures described here.

However, the consistent and clear neutral temperature structures displayed by the relevant profiles on the equatorward edge of auroral structures provide clear evidence of enhanced temperatures in these regions, which is suggestive of a heating mechanism. Furthermore, the evidence presented here aligns with the same Joule heating process that was discussed in Section 5.

6.4.2 Curl Associated Heating

The primary auroral feature of interest in the present event is a bright eastward-moving system of auroral curls that develops in the magnetic zenith following an interval of counter-streaming flow. The curl system in question is short lived and is observed by HiTIES for a total of four integration intervals. During those intervals its structure remains relatively stable and various sections of the HiTIES integration slit are able to consistently observe it. Observations during the passage of the curl are presented in Figure 6.3, in the same format as Figure 6.2 but with some additions. The coloured arrows pointing from one panel to another represent the over-plotting of a particular profile's errors onto another panel. The base of the arrow indicates the panel that the errors are taken from, and the head of the arrow points towards the panel that they are added to. For example, all three of the errors from each profile in interval W have been over-plotted onto the corresponding profiles in interval X, as indicated by the downward pointing arrows. All four intervals (W – Z) are two seconds long and are plotted in chronological order. The auroral dynamics present in each are described below:

W: The interval immediately prior to the entry of the auroral curl, which can be seen to the right (west) of the HiTIES slit in the ASK image. The HiTIES slit is observing a region of relatively weak emission with respect to the curl. The top (red) section is aligned with the leading edge of the eastward travelling curl system, the bottom (blue) section observes a dim arc-like structure that moves west-ward and passes below the coming curl. The central (white) section spends much of the interval observing a dimmer patch of aurora between these two structures.

X: The main passage of the most stable and bright auroral curl. The tightly wound curl structure seen in the ASK panel progresses steadily through the centre of the HiTIES slit; all three sections observe the curl, with the centre of the rotational structure moving through the central (white) section of the HiTIES slit.

Y: The curl system progresses eastward and moves slightly northward into the upper regions of the ASK field of view. The stable circular curl from interval X

moves out of the field of view and is trailed by a system of more elongated curls which are connected by highly folded and narrow filamentary arc-like structure. The curls maintain a clearly visible rotational component in the ASK video. The bottom (blue) section of the HiTIES slit is no longer observing the curl system, and is instead sampling the diffuse emission region that is directly equatorward.

Z: The curl system continues to move eastward and northward. The top section (red) of the HiTIES slit observes the filamentary connecting structures for the first half of the interval, before the structure moves further northward. The middle and bottom sections are in the diffuse and dimmer region to the south for the entirety of the interval.

The curl system described here is a classic example of the evolution described in Section 6.1; immediately prior to its formation there is a sustained patch of counter-streaming or shear flow in the aurora, which fills the ASK field of view. The curl system appears to be embedded in the middle of a larger mostly diffuse auroral arc. The dimmer sections visible to the south of the curl structures in the ASK frames in Figure 6.3 are in fact regions of diffuse but quite bright emission. However, they appear dark due to image scaling effects intended to help highlight the spiral structures in the curl.

Once again, each of the temperature profiles displayed in Figure 6.3 show consistent temperatures below ∼100 km, a result of the OH temperature verification process. However, the main feature of interest is the clear increase in the neutral temperatures during intervals in which the auroral curl is in the HiTIES field of view. Intervals W, X and Y display this mostly clearly. During interval W, immediately preceding the entry of the curl into the HiTIES slit, all three profiles show approximately the same shape and there is significant overlap of the errors corresponding to each profile. In the following interval (X), all three sections of the HiTIES slit are exposed to a bright and stable curl, and again they all show similar shapes in both the mean profiles and the error envelopes. The error envelopes from each section of the slit in interval W are overplotted onto the corresponding panels in interval X. It is clear that between intervals W and X there is a significant enhancement in neutral temperature, sometimes as large as ∼400 K, in all three sections of the HiTIES slit at altitudes between 120 and 160 km. Furthermore, the neutral temperature gradients observed in interval X (between 100 and 160 km) are the steepest seen during intervals taken from the middle of the auroral arc. In the next interval (Y), the curl system has progressed slightly northward and is no longer observed by the bottom section of the HiTIES slit. The corresponding temperature profile shows a statistically significant decrease in neutral temperature at the same altitudes as before (∼100–160 km), with respect to both the profiles from the previous interval (X) and the profiles from the remaining sections of the slit in the same interval. The profiles from the top and middle sections, which observe the curl system continuously from interval X to Y, do not change significantly. In interval Z, generally reduced neutral temperatures are observed; the curl system has moved rapidly northward and

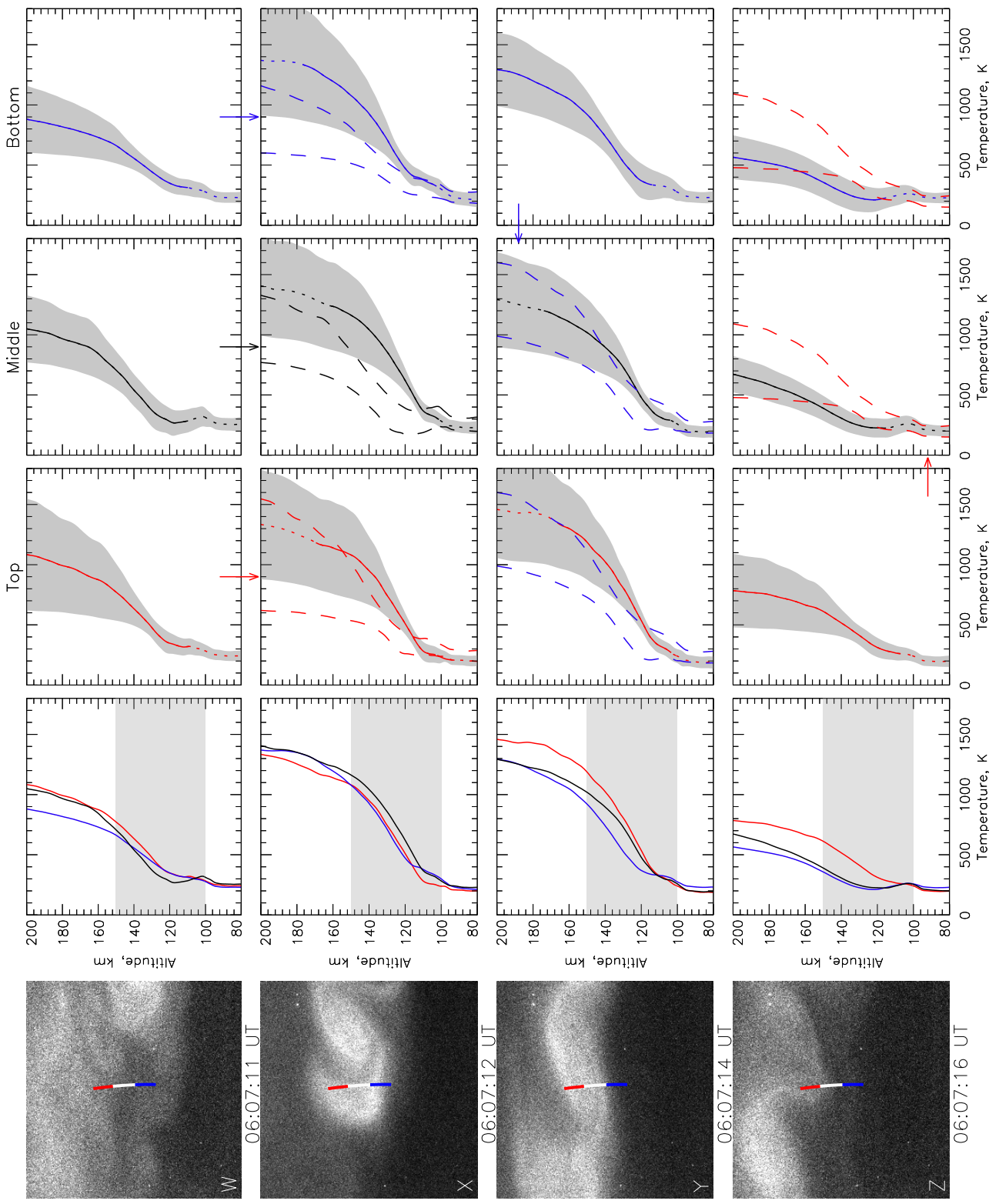


FIGURE 6.3: Curl associated neutral heating. Coloured arrows indicate the source of the dashed error envelopes that are over-plotted onto certain panels. The base of the arrow indicates the panel the errors are taken from, and the head points towards the panel they are added to.

the top section (red) of the slit only observes the bright filamentary curl structure for half of the integration interval. The bottom two sections (white & blue) spend the entire interval in the diffuse auroral region to the south. The profile derived from the region of the curls is significantly warmer than those from the adjacent southward sections, but not as warm as the intervals prior. This is likely the result of an averaging effect wherein half the interval is comprised of observations from the warmer curl structure, and the other half from the cooler diffuse region, resulting in a combination of the profiles from the two different regimes being produced. The profiles following the curls passage are some of the coolest retrieved during the entire event - in particular, the Z-middle and Z-bottom profiles are extremely cold with respect to the remainder of the intervals. Potential reasons for this feature are discussed in Section 6.4.2.1.

The temperature structures displayed in Figure 6.3 are clearly different to those discussed in Section 6.4.1; the enhanced neutral temperatures are now observed in the region of intense emission that is associated with the curls, and temperatures derived from the relatively dimmer adjacent regions are cooler. This is effectively the reverse of the previous result, which showed enhanced temperatures on the outside of sharp boundaries in the aurora. The difference between the two observations is in the location of the intervals in the context of the larger arc: those presented in Section 6.4.1 are from intervals at the beginning and end of the arc's passage and thus sample the aurora on its southernmost edge. The observations presented in this section are of a curl system that is embedded in the middle of a larger, diffuse region of auroral emission. This distinction has important consequences for understanding the associated electrodynamics. The significantly enhanced arc-associated electric field that drives the observed Joule heating is produced in response to a large boundary in ionospheric conductivity at the edge of the precipitation region. Inside the arc itself, there are no such boundaries in the conductivity and the ionosphere is highly ionised, inhibiting the formation of large electric fields. Therefore, a new heating mechanism that is compatible with auroral curl formation is required to explain the observed structures in the neutral temperature.

It is important to discuss the difference between the curl-associated heating signatures observed in this Chapter and those presented in Chapter 5. The primary discrepancy between the two events is the altitude extent of the observed heating signatures. The curl-associated heating signature in Chapter 5 was seen over a narrow altitude range below 100 km, whereas in this study, the altitude of the temperature enhancement is much higher and wider, between about 120 and 150 km. This result gives cause to revisit the two potential mechanisms that were theorised to be responsible for the curl-associated heating in Chapter 5: electron Pedersen current heating and Ohmic dissipative heating. The electron Pedersen current mechanism is outlined in detail in Buchert et al. [2008], but the most important aspect of the theory is the requirement for a large electric field - a situation which has been ruled out in this case due to the belief that there is a lack of conductivity gradients in the vicinity of the observed curls. The result in Chapter 5 did not have this contradiction, since the most significant curl-associated

heating signature was also seen at the leading edge of the arc, where the required electric field is expected. Furthermore, as mentioned above, the curl-associated heating signature in this event displays a considerably different structure in altitude - distancing it further from an electron Pedersen heating mechanism. Since electric fields are not applicable, the present result suggests that the enhanced neutral temperatures seen in Figure 6.4.2 are due to the second theoretical heating mechanism: an Ohmic or resistive heating effect that is driven by narrow sheets of field-aligned current. The main drawback of this theory in Chapter 5 was the thin altitude extent of the heating signature, where it was perhaps more realistic to expect a gradually increasing heating rate as the atmospheric resistivity increases with decreasing altitude. Reassuringly, this is now the general structure of the temperature enhancements associated with the auroral curls that are observed in this event. The most significant temperature differences are at $\sim$120 km, and the differences are gradually reduced with increasing altitude and hence decreasing resistivity. It seems logical to claim that the signature in Chapter 5 was likely the result of electron Pedersen current heating and the curl-associated heating observed here is the result of resistive heating via strong field-aligned currents. The presence of auroral curls during both instances of heating is of interest, particularly since there is no clear evidence of temperature enhancements between 100 and 160 km during auroral curls in the first event. Improvements to the observational method likely made the difference in ensuring the curl-associated heating could be resolved more accurately; the curl heating appears to be very localised to the curls themselves, and the single slit observations in Chapter 5 may have struggled to resolve it.

To summarise, the evidence suggests that the observed curl-associated heating is a resistive heating effect that is driven by strong and narrow field-aligned current sheets embedded within the larger scale auroral structure. Current sheet structures with horizontal scales of the order of the ASK field of view ($\sim$10 km), have often been observed in the auroral region by satellites [Lühr et al., 2015, Stasiewicz and Potemra, 1998]. A statistical study by Rother et al. [2007] reported a significant number of small-scale ($<$5 km transverse width) auroral field-aligned currents with current densities ranging between 250 μAm^{-2} and 1000 μAm^{-2}. The highly dynamic and structured auroral emissions observed in the ASK instrument are therefore expected to have equally dynamic FAC systems, which vary on similar temporal and spatial scales. Lanchester et al. [2001] presented evidence of strong field-aligned currents in filamentary aurora by invoking a heating effect they deemed as 'Ohmic heating'. During observations of a filamentary auroral arc, the corresponding EISCAT radar measured a large increase in the electron temperature immediately adjacent to enhanced E-region electron densities. Using a 1-D transport, ion chemistry and electron energy model they conclude that electron precipitation heating is unable to reproduce the observed electron temperatures, without also modifying the electron density profiles. Therefore, an additional source of electron heating is required, i.e Ohmic heating, the resistive heating process produced by Coulomb collisions between the thermal electrons in a strong field-aligned

current. The Ohmic heating term is described as ηj^2, where η is the ionospheric resistivity and j is the field-aligned current. Including this term in the electron energy equation, and choosing a large, but not unreasonable peak field-aligned current magnitude of $400 \mu A m^{-2}$, reproduced the observed electron temperature increase exceptionally well. Zhu et al. [2001] extended the analysis of the Lanchester et. al. event to a 2-D three-fluid simulation of the ionosphere-magnetosphere system. This simulation contained an extensive consideration of a variety of ionospheric heat sources and sinks. They demonstrate that a perturbation in the boundary conditions at the top of the ionosphere, in the form of a shear in the magnetic field and plasma velocity, propagate downwards in altitude as a pair of Alfvén waves. The Alfvén waves produce two field-aligned current layers of near constant density, with opposite directions (upwards and downwards), which are then closed by Pedersen currents in the E-region. These Alfvén waves arrive at the lower boundary (100 km) in approximately 0.6 s, meaning that these effects can propagate into the ionosphere on timescales comparable to the resolution of our observations.

The Zhu et al. [2001] model also includes an electron heating process via Ohmic dissipation, implemented in the same form as the Lanchester model (ηj^2), which too is associated with regions of field-aligned current. The modelled field-aligned current density is near constant with altitude, such that the Ohmic heating rate is dictated almost exclusively by the atmospheric resistivity, increasing exponentially with decreasing altitude. Similarly, the electron cooling rate, via collisions with the neutral population, is also inversely proportional to the altitude. As a result, temperature equilibrium between the electron and neutral populations is rapidly obtained below 160 km altitude. The combination of these relationships results in the greatest capacity for a net increase in the neutral temperature to occur at the base of field-aligned current structures. Hence our measured increases in neutral temperatures between 100 and 160 km, observed during intervals containing auroral curls, provide direct evidence of the presence of short-lived and intense field-aligned currents in the aurora. Furthermore, these observations agree with the prevailing theory of curl formation, that they are caused by the Kelvin-Helmholtz instability in regions of ionospheric plasma shear, on either side of a field-aligned current sheet.

6.4.2.1 Cold Profiles

The extremely cold profiles derived from the Z-middle and Z-bottom intervals are difficult to explain. The change in temperature from interval Y to interval Z at altitudes above 120 km can be as large as $\sim$ 700 K in a two second interval, which is unrealistic without some source of extreme cooling. It is particularly interesting that these cool intervals only occur directly after the curl system, and are absent from the remainder of the observations. However, no coherent explanation for what could have possibly caused such large temperature changes has been found. It is feasible that that low altitude and localised heating associated with the curls results in an ion outflow response

at altitudes above the energy deposition, in turn driving the rapid cooling of the neutral population via expansion into the resulting low density cavity.

Some speculation on the reliability of the method is worth considering here as well. During the cool intervals, the emission spectra are integrated for a total of two seconds. This is an acceptable duration for bright emission, where signal to noise ratio remains adequately high. However, a common feature of the two cool temperature profiles is that they are derived from a region of relatively weak and diffuse emission south of the curl system. The short integration interval, chosen to best evaluate the rapidly evolving auroral structure, and the relatively reduced emission in these sections results in a lower than normal signal-to-noise ratio. A high noise component makes it difficult for the method to determine the difference between a good profile and an average profile, since the noise washes out the potential RMS difference between the two fits. The result is that, whilst the method can still accurately determine the neutral temperatures at the altitude of peak N_2 volume emission rate, its ability to distinguish temperatures outside of this region is diminished further than usual. Counter-intuitively, this can be observed by the tighter than usual error envelopes, which is likely a similar feature to that discussed in Section 4.10.2. Namely, that for unusually cool neutral temperatures at low altitudes, many of the possible permutations of the temperature profile generation algorithm favour similar structures at high altitude, artificially pinching the error range. The cool temperatures at 100 to 130 km are likely real, but the values above that become increasingly unreliable.

6.5 Overview

Finally, an overview of all 147 of the derived temperature profiles is presented in Figure 6.4. The top three panels display 2-dimensional plots of the neutral temperature profiles that were derived from each section of the HiTIES slit, as a function of time. The timestamps of the intervals discussed in Section 6.4.1 and Section 6.4.2 are indicated by the vertical dashed lines and letter labels in the top and bottom panels. The bottom panel displays a temperature slice from each panel at an altitude of 130 km, indicated by a horizontal dashed line in the 3rd panel. The white shaded areas in the top three panels represent altitudes at which the N_2 volume emission rate is less than 10% of the peak during that interval and the determined temperatures are deemed unreliable.

A clear general trend is observable in the top three 2-D temperature panels: there is a cooler interval of neutral temperatures in the 40 seconds between 06:06:35 UT and 06:07:15UT. This interval of time corresponds to times when the emission fills the ASK field of view, during which the HiTIES slit is sampling aurora in the centre of the arc structure, separated spatially from its equatorward edge. This interval consists of a region of steadily increasing counter-streaming flows which eventually develop into the curl system discussed in Section 6.4.2: intervals W–Z. The warmer profiles that

appear before this interval (06:06:05–06:06:35 UT) are from the intervals covering the southward motion of the arc's equatorward edge through the HiTIES slit. The most significant Joule heating signatures are observed during this time. The second warm interval between 06:07:15 and 06:08:10 UT, which follows the break-up of the curl system, corresponds to a generally more diffuse region of the auroral arc as it begins to move gradually northward and out of the field of view.

The time evolution of the neutral temperature at 130 km can be seen in the bottom panel of Figure 6.4. The warmest temperatures at this altitude are as hot as 1400 K and are found at the beginning of the event, during the Joule heating intervals discussed in Section 6.4.1. This peak is followed by a relatively steady decrease in temperature as the arc moves southward, until reaching a consistent low of approximately 600 K during the interval of counter-streaming flows and curls. The ionisation in this region is approximately two orders of magnitude higher than the background level, and equatorward boundary of the arc is far to the south of the HiTIES slit. Thus the formation of electric fields is inhibited by charge separation processes and destructive polarisation electric fields. As a result, it seems likely that the cooler neutral temperatures present during this time are closer to the background temperatures of the ionosphere without the interference of enhanced electric fields. The only exception to this regime is during the passage of the curl system, where a clear peak in the neutral temperature is visible between ∼06:07:05 and 06:07:12 UT in Section 6.4.2. Finally, as the arc moves northward again the temperature increases slightly to a maximum of about 1000 K, but does not reach the same temperatures as at the start of the event.

The increased temperatures seen at the beginning of the event with respect to those towards the end are of interest. The temperatures in each are derived from approximately the same region of the auroral arc: its equatorward edge. However, the optical structure of the aurora and the sharpness of its equatorward edge in the early intervals is much more pronounced than in the later intervals. The ionosphere at this time (early intervals) is in a quiescent state, having not had any prior auroral activity, and as such has what can be assumed to be a close to background level of ionisation and conductivity. The arrival of the precipitating magnetospheric electrons that herald the beginning of the event leads to a rapid increase in ionisation over a localised region of the ionosphere. This produces the extremely sharp 'step like' gradient in conductivity that results in the strongly enhanced arc-associated electric field and the corresponding Joule heating. However, later in the interval when the arc is moving northward again, having occupied the region around the magnetic zenith, the ionosphere cannot be thought to be in a quiescent state and it still contains ionisation that has yet to fully recombine to pre-auroral levels. This latent ionisation increases the background conductivity of the ionosphere and thus reduces the gradient in conductivity at the edge of the arc. The resulting electric field that drives the closing Pedersen currents is therefore no longer as large as in the earlier intervals, and the Joule heating rate is similarly reduced. This is potentially the reason for the discrepancy in neutral temperatures before

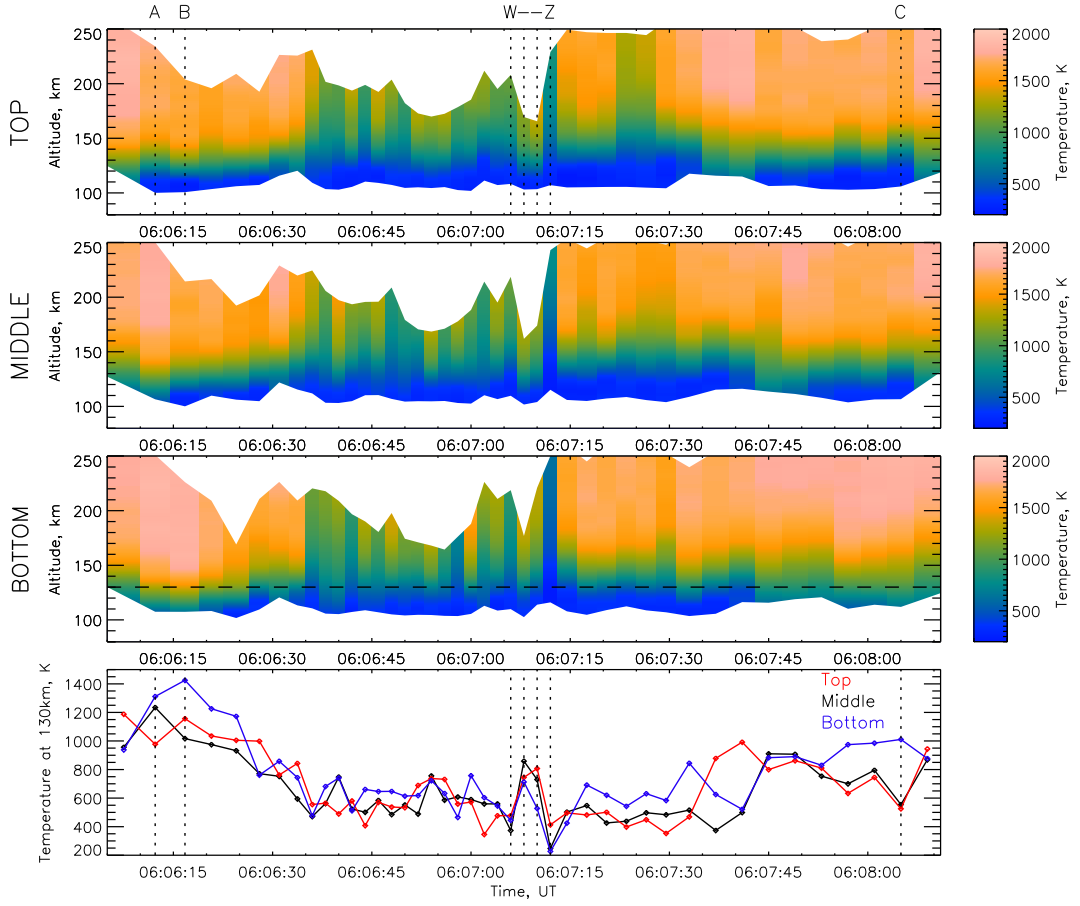


FIGURE 6.4: Full neutral temperature data set for the event in question. Top three panels display 2-dimensional plots of the neutral temperature as a function of altitude and time. Vertical coloured stripes indicate each of the HiTIES integration intervals. Shaded white sections represent altitudes at which the corresponding N_2 volume emission rate is less than 10% its maximum value. The timestamps of the intervals discussed in the text are indicated by labels above the first panel and by the vertical dashed lines in the first and fourth panels. The bottom panel displays a line plot of the neutral temperatures in each section of the slit at 130 km in altitude as a function of time. This altitude is indicated by the horizontal dashed line in the third panel.

and after the arc's passage and is a reminder that many of the theoretical discussions regarding the formation mechanisms of auroral arcs assume ideal background conditions with arcs forming in a contextual vacuum. Most aurora does not manifest as perfectly east-west aligned quiet arcs and usually occurs during intervals of enhanced magnetic activity wherein the ionosphere is already in a perturbed state.

6.6 Conclusion

This study is the first to present evidence of a neutral heating process associated with field aligned currents and auroral curls. The identification and analysis of the observed heating signatures provides clear confirmation of the presence of intense filamentary

field-aligned currents, embedded within large-scale auroral structure. The observations display significant but short-lived enhancements in neutral temperatures during intervals that contain auroral curls. Improvements to the resolution of the inversion technique allowed for a more thorough investigation of the competing heating mechanisms presented in Chapter 5. As a result of the new observations, the heating signatures from both events were distinguished and assigned the appropriate mechanism.

A second result is observed in the neutral temperature profiles at times when the HiTIES slit is sampling the equatorward region directly adjacent to filamentary structure in the optical aurora. During these intervals, the corresponding temperature profiles display a significant increase in the neutral temperature across a range of altitudes, which is identified to be the result of a Joule heating process driven by an arc-associated electric field. Arc-associated electric fields, such as the one inferred by these observations are necessary for current closure mechanisms in the ionosphere, and are an important component of a complete understanding of auroral formation and structure.

Quantifying the variability and magnitude of small-scale auroral electrodynamics, as well as their impact on atmospheric properties such as temperature, is essential in building an applicable understanding of the underlying formation mechanisms. Furthermore, the characterisation of the temporal and spatial scales of these events provides crucial insight into the variability of coupling between the magnetosphere and the thermosphere, which informs more accurate high resolution theoretical models. Satellite measurements of narrow field-aligned currents must be interpreted as spatial changes because of the speed with which they pass through the aurora. The advantage of ground-based measurements such as those described here, is that the spatial and temporal changes can be separated in the region surrounding the magnetic zenith. The combination of spectrographic measurements over a slit that can be integrated in sections, the very high temporal and spatial resolution of the ASK cameras, and ionospheric modelling using the key information within the wavelength dependent emissions, provides a powerful data set to determine very small-scale changes in the auroral ionosphere.

Chapter 7

N_2 Contamination Removal

7.1 Introduction

In addition to its application as a primary analysis tool, the inversion technique developed in Chapter 4 can also be used to facilitate and support other research goals. For example, during intervals of high energy precipitation, the resulting N_2 emission can often overwhelm and hinder observations of other atmospheric emissions, both auroral (e.g. O_2^+) and otherwise (e.g. OH). These other emissions prove useful in other areas of research [Chadney et al., 2017, Whiter et al., 2014] and the contaminating N_2 can reduce the effectiveness or even completely inhibit applications that require them. Fortunately, the developed inversion process has the side-effect of producing an extremely accurate estimate of the N_2 spectral shape and intensity. This component can thus be subtracted from the observations, removing the contamination and allowing for more detailed analysis of the remaining emissions.

With the author's help, the inversion technique was applied to observations presented by Reidy et al. [2020]. By removing the N_2 band emission component from the observed HiTIES spectra, the author was able to isolate the H-α emission line and effectively resolve its shape. The results of Reidy et al. [2020] and the author's contribution to their work are summarised in this chapter.

7.2 Observations

Reidy et al. [2020] presented multi-scale observations of two auroral events containing a 'polar cap arc' (PCA) over Svalbard. Polar cap arcs are auroral arcs which occur in the typically quiet polar cap region - i.e. aurora is observed to be inside (at higher latitudes than) the auroral oval. The exact mechanism responsible for the appearance of polar cap arcs is debated. However, previous case studies have determined that they can occur on either open or closed magnetospheric fields lines, with a different mechanism proposed for each case. The PCAs discussed by Reidy et al. [2020] were observed by a number of instruments, including both HiTIES and ASK. By investigating differences in each arc's observable characteristics they were able to produce compelling evidence

that one of the arcs occurred on closed field lines, and the other on open field lines. Of particular note to this thesis is the work that was done comparing the HiTIES emission spectra for both events.

7.3 Background and Method

At the time of both events HiTIES was fitted with a mosaic which contained the "H-α" panel, which selects auroral emission at wavelengths between 649 and 663 nm. This panel is so named as it contains the H-α spectral line, an emission that originates from the Balmer series and has a wavelength in air of 656.28 nm. H-α airglow emission is observable in the upper atmosphere during quiescent periods, and it does not explicitly require auroral mechanisms to be produced. However, this panel was specifically designed to investigate 'proton aurora' events - intervals in which magnetospheric protons precipitate into the high-latitude atmosphere, as opposed to the more commonly observed electron precipitation. Precipitating protons undergo charge-exchange interactions in the atmosphere, which results in a population of neutral Hydrogen atoms with varying states of excitation. The excited Hydrogen atoms relax and emit observable Balmer series emissions, primarily Hα, Hβ and Hγ. The H-α emission has a relatively low brightness compared to other aurora-associated emissions, and does not produce strong visible signatures. The distinction between auroral and non-auroral H-α emission is determined by the presence of Doppler-shift in the narrow emission line; the precipitating protons (and resulting Hydrogen atoms) maintain a component of their downward velocity following each of the charge-exchange interactions, and as such, the emissions appear blue shifted in ground based observations. Non-auroral H-α emission should not exhibit this feature, as the emitting Hydrogen atoms exhibit random thermal motion and thus have no net velocity. Therefore, a blue tail or wing on the H-α emission line is indicative of precipitating protons or proton aurora, and isolating the H-α emission in order to check for this feature can allow for easy detection of precipitating protons.

The method developed in Chapter 4 is easily adaptable to different HiTIES mosaic panels (and the corresponding wavelength ranges). Energy and flux values were calculated from ASK, as per the method described in Section 4.3, and the Southampton ionospheric model was used to determine appropriate estimates of the N_2 volume emission rate profile, for both events. Unfortunately, the Reidy et al. [2020] observations did not coincide with an appropriate ESR experiment, so the model verification stage was omitted. Synthetic N_2 spectra were generated for the H-α panel and the fitting routine was slightly adjusted to account for the different emissions present in the new wavelength range. The fit contained free parameters for the N_2 1P band emission, a number of OH(6-1) lines and a N^+ emission line at 658.1 nm. The fitting process did not include an H-α component and thus no attempt to fit H-α emission was made. Fortunately, the H-α emission is in a wavelength region of reduced (but non-zero) N_2 emission between

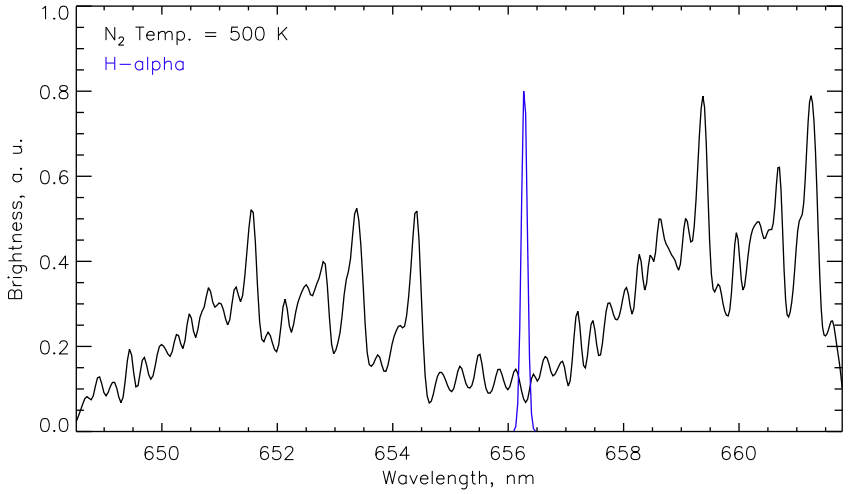


FIGURE 7.1: Example synthetic spectra for the H-α HiTIES mosaic panel. The solid black line is a synthetic N_2 spectrum corresponding to a temperature of 500 K. The blue emission line is an example non-Doppler-shifted H-α emission line. The brightness of both emissions is arbitrarily scaled for clarity.

two band structures (see Figure 7.1), and the fitting routine is not significantly biased by its presence.

The Two-step method (Section 4.8.1) was applied to the HiTIES spectra from both of the PCA events. In this case, the neutral temperature profile is not a crucial aspect of the analysis, and the primary goal is to determine an accurate estimate of the N_2 emission. To this end, the fitted spectrum corresponding to the temperature profile with the lowest RMS value was chosen and the N_2 component was retrieved. The contaminant N_2 emission was then removed via subtraction from the total fit, leaving the now uncontaminated H-α emission line. The Doppler-shift (or lack thereof) of the line was then determined, without concerns of its shape being affected by nearby emission.

7.4 Results

Two auroral events are discussed in Reidy et al. [2020], and for the remainder of this section they will be referred to as 'Event 1' and 'Event2'. Each event includes simultaneous observations from ASK, HiTIES, and from instrumentation aboard four of the Defence Meteorological Satellite Program (DMSP) spacecraft, including: the Special Sensor Ultra-violet Spectrographic Imager (SSUSI), which provides hemisphere-scale UV imaging of the polar cap; and SSJ/5 particle spectrometers, which return in-situ spectrograms of ion and electron precipitation characteristics, along the satellites trajectory. Event 1 displayed a number of traits that are typically associated with aurora on closed field lines, namely: larger values of precipitation energy and flux, with respect to Event 2, (as derived from ASK - see Section 4.3), coincident DMSP proton flux measurements at times when the satellite's footprint was passing overhead, and conjugate

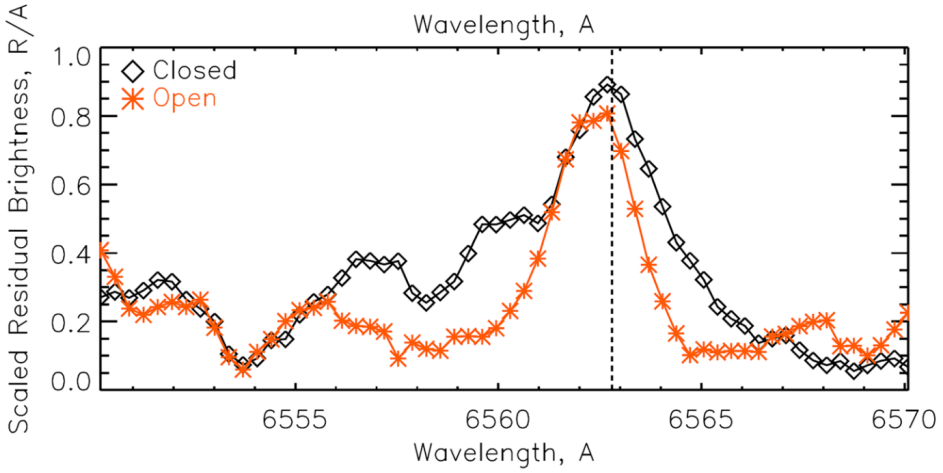


FIGURE 7.2: Residual spectra in the 6550 and 6570 Å (655 – 657 nm) wavelength region, surrounding the H-α emission at 6562.8 Å (dashed line), as measured by the H-α panel of the HiTIES instrument for the two auroral events in question. Black diamonds and red asterisks correspond to 'Event 1' and 'Event 2', each identified as occurring on closed and open magnetic field lines, respectively. An enhanced blue-shifted 'wing' on the H-α emission line is observable during the closed event. Figure from Reidy et al. [2020].

observations of PCA structures occurring in both hemisphere - as expected from closed field lines. Conversely, Event 2 did not display many of these characteristics, instead aligning with the expected properties of open field line aurora by having comparably smaller ASK-derived energy and flux values, electron only precipitation in the DMSP spectrographs and no conjugate structure in both hemispheres.

The last piece of evidence distinguishing between the type of the magnetic field line associated with each event can be seen in Figure 7.2, which shows the normalised residual emissions measured by HiTIES from a small wavelength region surrounding the H-α line. In the residual corresponding to Event 2 (red line), a clear Gaussian shaped emission line can be seen centred on the wavelength of the H-α emission (dashed line). This symmetrical signature is evidence of random thermal motion, likely associated with quiescent H-α day-glow [e.g. Weller et al., 1971]. In contrast, the residual from Event 1 (black) shows a strongly skewed shape, with relatively enhanced emissions in a blue-shifted (lower wavelengths) wing. The shape of this residual agrees with the predicted blue shifting of the H-α emission as a result of precipitating protons on closed field lines. Thus, the work here corroborates the evidence from the remaining observations and suggests that Event 1 and Event 2 occurred on two different types of magnetic field line.

7.5 Conclusions

The example presented here is just one potential avenue of analysis that is made available by the N_2 fitting process developed in Chapter 4. Spectral observations during

times of intense N_2 aurora no longer need to be discarded when observing other emissions such as OH airglow. The impact of the contaminant emission can be significantly reduced and the resulting gaps in the data can be appropriately filled with accurate measurements of the emissions of interest.

The data from this study is publicly available and hosted by the University of Southampton (DOI: 10.5258/SOTON/D1270).

Chapter 8

Conclusions

8.1 Research Summary

The emission spectra inversion method developed and outlined in Chapter 4 has produced measurements of thermospheric neutral temperature profiles during intervals of auroral emission at unprecedented temporal and spatial scale. The observations presented in Chapter 5 and Chapter 6 are reliable ground-based measurements of neutral temperatures at auroral altitudes. The detailed analysis made possible by these observations has allowed for entirely novel investigations into local scale auroral process, which in turn have facilitated the confirmation and development of existing theories regarding the electrodynamic structure of auroral arcs. The work presented in this thesis highlights the benefits of combining observations from a variety of instruments, with ionospheric and spectral modelling to produce a result that is more than the sum of their parts. Optical observations and analysis techniques developed and utilised by previous and current researchers have been united to produce a new and extremely promising technique that can be utilised extensively in the future - both as a principle analysis tool and also in a supportive role for a vast variety of related studies. Quantifying the impact of auroral process on sub-grid scales has the capacity to vastly improve the accuracy of the auroral component in atmospheric models. Additional studies to estimate the magnitude and extent of electrodynamic features, such as the arc-associated electric field, can provide concrete quantitative values of the statistical spread of polar ionospheric electric fields and help to improve estimates of the global Joule heating rate. In addition, increased understanding of the distribution of filamentary field-aligned currents and the specific auroral features associated with their presence can also help to develop a complete picture of magnetosphere-ionosphere coupling processes and auroral acceleration mechanisms. There are a range of potential avenues for future work; improvements to the inversion method are discussed in Section 8.2.1, further observational and research goals in Section 8.2.2 and the science chapters presented in this thesis are summarised below.

Chapter 5 presented results concerning the effect of an intense and localised Joule heating process on thermospheric neutral temperatures in the vicinity of an auroral

arc. Observations agreed with the prevailing theory for the formation of a 'Birkeland' classification arc [Marklund et al., 1982], and confirmed the electrodynamic structure associated with such an arc on fine scales. Furthermore, supporting observations suggested the presence of a secondary heating process, distinct from electric field driven Joule heating, which was embedded within the auroral arc itself. The observational technique showed significant promise and potential to uncover new results regarding auroral processes on scales that have previously been unobtainable.

Chapter 6 represented the next step in the continued application of the inversion method. Significant updates provided considerable increases to its accuracy and reliability, whilst also improving the spatial resolution of observations even further. The resulting case study of a dynamic auroral event provided further evidence to support the results presented in Chapter 5 as well as the confirmation of a heating process caused by Ohmic dissipation of field-aligned currents. The dissipative heating was observed to occur during the passage of an auroral curl, linking them to the presence of intense field-aligned currents.

Chapter 7 discussed a further application of the inversion method. The accurate spectral fitting of N_2 band emission made possible by the method can be used to remove contaminate emission from intervals where it is not desired. This provides the opportunity to obtain significantly more accurate observations of other emissions of interest, particularly during times of extremely active aurora. An example application of this nature, published in Reidy et al. [2020], is discussed.

8.2 Future Work

This section provides some suggestions for avenues of potential future work. Suggestions can be separated into two main categories: further development and fine-tuning of the observational technique, and the continued application of the technique to auroral observations. Each are discussed in the following subsections.

8.2.1 Method Development

The inversion method at the centre of this research was continuously developed throughout the course of the studentship, and two different iterations have been presented here. The final form of its development is the simulated annealing algorithm discussed in Section 4.8.3. However, this is simply its status at the end of the research period and is consequently subject to the time constraints inherent to a PhD project. Suggestions for potential improvements regarding the method are discussed here.

8.2.1.1 Simulated Annealing Algorithm

The simulated annealing algorithm proved to be extremely effective in increasing both the accuracy and the efficiency of the method. However, the appropriate initial conditions and algorithm parameters were determined via extensive trial and error and likely do not represent an optimum final state. A more complete set of verification runs, covering a wider range of potential combinations between temperature and volume emission rate profiles, would facilitate a more informed selection of the initial parameters and more appropriate analysis of the output. Furthermore, optimisation algorithms are a comprehensive topic in probabilistic analysis and there are considerable options in the literature for adapting and improving the version used in this thesis. It is possible that more applicable algorithms that are better suited for the specific optimisation problem presented here exist, but a full exploration of possibilities was unfortunately beyond the time scope of this thesis. Adjustments to the currently employed simulated annealing algorithm, or an adaptation to a potentially more appropriate algorithm can only further improve the accuracy and efficiency of the method, and in turn the results.

8.2.1.2 Trial Temperature Profile Algorithm

One of the most important and influential aspects of the inversion method was the approach used to create a random trial temperature profile. The impact of this choice is difficult to define qualitatively. Naturally, the variation in different trial profile possibilities affects the final result, as any average final profiles will retain features inherent to the individual profiles which comprise it. Whilst the annealing algorithm circumvents this somewhat, by using a cumulative average rather than a single profile, the shape of the final outputted profile is fundamentally influenced by the range of possible contributions made available by the temperature profile generation algorithm. As a result, the influence of this choice cannot be understated.

Section 4.10 of Chapter 4 provides essential context for outlining potential improvements to the temperature profile generation algorithm. The results presented therein made it extremely clear that the primary drawback of the method was the reduced accuracy at altitudes of relatively decreased N_2 volume emission. This is of course a fundamental characteristic of the method, given its reliance on N_2 emission spectra, but one that the impact of which can potentially be reduced further. As per the discussion in Section 4.10, the further the method moves in altitude from the peak volume emission region, the more impactful the temperature profile generation algorithm becomes. At extremely low volume emission rates, the outputted neutral temperatures simply approach the average temperature of all possible profiles, that exhibit the correct behaviour at the well constrained altitudes; the temperatures at the extreme altitudes are a function of how much the profiles generated by the algorithm are allowed to vary in temperature over that distance. This is not an unreasonable behaviour, as

long as the maximum allowed temperature gradients are based on values with realistic atmospheric considerations. Naturally this would not allow for the determination of anomalous heating signatures at low volume emission, since it would be preferred that in the uncertain regions the profiles approached temperatures representing an average or quiescent state. The current temperature profile algorithm makes a relatively successful attempt to achieve this, but improvements are possible. Fitted temperatures at the altitude of peak volume emission rate could restrict potentially well-fitting temperatures at other altitudes if the maximum allowed temperature gradients can't match the desired temperature change. For example, if the true profile has a temperature of 300 K at 120 km (the peak of emission) and 500 K at 130 km (at half the peak emission), but the trial profiles cannot easily replicate this temperature change, an underestimation of the temperature at 130 km is possible; the inversion process will favour getting the temperature correct at the peak volume emission altitude. Conversely, for extremely large allowable gradients, there is the possibility of over-fitting the observed spectrum with unrealistic temperature profiles. These factors require careful consideration and can have a serious impact on the final result if chosen incorrectly. Further research, verification runs and analysis would help to inform and adjust the temperature profile generation algorithm to be as appropriate and unbiased as possible.

Furthermore, the author believes that a significant improvement in accuracy is possible at altitudes in the transition region between the peak and negligible emission via improvements to the trial profiles. As can be seen in Figure 4.12 and Figure 4.13, the method is unable to replicate profiles that exhibit sharp local changes in temperature gradient. As mentioned, this is primarily due to the temperature degeneracies in the synthetic emission spectra and noise in the observed emission spectra. However, the impact of these factors could potentially be lessened further via intelligent changes to the temperature profile generation process. If a good estimate of correct 'scale' of the temperature features is known, the generation method could be adjusted to only allow variability on a similar scale. The potential upside of this adjustment would be that any structures which do not match the correct shape are effectively a scale distance away in temperature, which would ideally manifest itself as a significant difference in RMS. This change could ultimately facilitate an improved ability to distinguish between good and average trial profiles. The primary difficulty with an adjustment such as this is the lack of knowledge regarding the expected scales of temperature structures in the thermosphere; a limitation which can be reduced via repeated applications of the current method to help develop an understanding of the magnitude and extent of the heating processes of interest.

8.2.1.3 Automation

Aside from the development of the inversion technique, much of the research presented in this thesis has been in the form of case studies of individual auroral events. The developed analysis technique lends itself to such applications due to its expensive observational requirements; data from several independent instruments is required and the manual tweaking of ionospheric model parameters is preferred to ensure as accurate an output as possible. Nonetheless, more extensive observations and analysis would help to build an understanding of the range of electrodynamical features associated with different auroral events.

The primary bottleneck limiting the number of events the method can be applied to is the availability of concurrent EISCAT observations. However, other than providing a useful contextual picture, the EISCAT observations are only used as a model verification tool, and are not a fundamentally required aspect. If a certain level of inaccuracy can be accepted in the ion chemistry model results, then far more events are made available for analysis. The ASK and HiTIES instruments have 100s of hours of historic measurements to which the inversion technique could be applied; as long as an applicable mosaic filter is in use on HiTIES. Serious work is needed to achieve this eventual goal, and many aspects of the method would likely need to be adjusted to accommodate more automation. Primarily, the modelling stage would need to be adapted to use general case outputs from a variety of energies and fluxes, as well as a number of different combinations of Gaussian and Maxwellian input spectra. The model verification portion of the method would likely need to be skipped entirely due to the lack of the EISCAT power profiles, in turn making the manual tweaking of each model run impossible. In addition, a process for automatically determining appropriate HiTIES integration intervals depending on the N_2 emission intensity, and potentially the auroral structure in ASK, would need to be developed. However, with these improvements the inversion technique could theoretically be run on historic or live data whenever the appropriate auroral conditions are met. Automation in this manner would also allow for large statistical studies of multiple auroral events in different conditions to be undertaken (see Section 8.2.2.2).

8.2.2 Scientific Potential

Joule heating alone is believed to drive a variety of atmospheric dynamics, from ion up-welling [Clemmons et al., 2008, Prölss, 2011] to gravity waves Williams et al. [1988], which in turn impact thermospheric densities and compositions. Some studies even report travelling atmospheric disturbances, generated by Joule heating events, that propagate far outside of the polar regions [Buonsanto et al., 1999]. Clearly the impact of polar energy deposition is far reaching and the inclusion of Joule heating and particle precipitation processes in atmospheric models is of crucial importance. Reliable observations of these events are required to facilitate the desired sub-scale modelling

improvements - a task that the inversion method developed in this thesis is uniquely equipped for.

8.2.2.1 Further Case Studies

With the continued development and improvements made to the method between the research presented in Chapter 5 and Chapter 6, an obvious place to start would be to re-apply the current version of the method on the event presented in Chapter 5. This could prove beneficial as the auroral event in that case is generally more well behaved; it exhibits a stable east-west aligned equatorward boundary which progresses through the entire ASK field of view at a consistent velocity. This makes it an ideal event for studying the Joule heating process in more detail, particularly with the increased spatial resolution offered by the three sections of the HiTIES slit. Arc-associated electric fields have a magnitude which is expected to be non-uniform in space, with larger enhancements closer to the optical boundary in the arc. By taking three measurements spatially, as well as more carefully tracking the approach and motion of the equatorward boundary through the HiTIES slit, an estimate of the horizontal and vertical neutral temperature gradient is possible. Quantitative analysis on the electrodynamical system could be achieved by combining these neutral temperature measurements with energy balance equations and current continuity considerations. In addition to the arc in Chapter 5, further case studies of any auroral events that display specific structures or behaviours of interest would only improve the general understanding of the thermospheric temperature response and the associated electrodynamic structures.

Supporting observations of relevant atmospheric properties would also be of interest to investigate local responses beyond neutral temperature variations. It is well understood that large-scale Joule heating causes ion up-welling, but there are limited observations on the scales reported here. Contextual satellite observations of high-altitude densities, Fabry-Perot derived line of site velocities, or EISCAT ion velocities could provide invaluable insight into ion and neutral dynamics in response to the neutral heating. The difficulty is in finding supporting observations that operate on similar scales to the neutral temperature measurements - a fortuitous and relevant satellite passing likely provides the best case scenario.

8.2.2.2 Conditional Analysis

Increasing the data coverage of different auroral events via the repeated application of the inversion method (see Section 8.2.1.3) would improve contextual understanding of the thermospheric response. Auroral electrodynamics are believed to be heavily dependent on the background ionospheric conditions (see Section 5.4.1 in Chapter 5). There are numerous arc-classification schemes which separate arcs into distinct categories based on their expected electrodynamic properties. The method presented here could be utilised to investigate the legitimacy of such classification systems and contribute to improving and refining our understanding of auroral electrodynamics.Observations

during different magnetic conditions, local times, convection field orientations and even latitudes could illuminate much by providing powerful statistics on a number of parameters of interest. Statistics on the magnitude of the neutral temperature response, the structure of the horizontal temperature gradients adjacent to arcs (and the inferred electric fields and field-aligned currents), the primary altitudes at which the energy is deposited are all invaluable inputs for the future development of global atmospheric models.

A significant difficulty would be in correctly identifying events in the ASK instrument that display enough N_2 emission and auroral structure to be worth investigating. Machine learning and image recognition software could potentially be employed to flag events of interest in the ASK images, but fortunately there is already a large and increasing database of ASK event classifications thanks to the Aurora Zoo project. Citizen scientists have been given access to large intervals of ASK data, via the Aurora Zoo web page, and have already contributed useful scientific outputs [Whiter et al., 2021]. Using Aurora Zoo, ASK images have been accurately and efficiently sorted into pre-determined categories - including features such as single arcs, auroral curls and counter-streaming flows. This database provides the perfect companion to an expanded application of the inversion method and would no doubt unveil invaluable research outputs.

8.2.3 EISCAT 3D

The soon to be employed EISCAT 3D system promises to bring significant advancements to a long list of auroral and atmospheric sub-fields of research. EISCAT 3D will be able to produce volumetric measurements of ion velocities, ion and electron temperatures, and electron densities. A 3-D grid of these parameters can be used to fill an ASK field of view with multiple simultaneous measurements, as opposed to the current single radar beam. Better spatial coverage of electron density measurements will allow for the verification of the ionospheric model in the sections of the HiTIES slit that are currently not aligned with the radar beam, increasing confidence in temperature measurements outside the magnetic zenith. Furthermore, the EISCAT-3D ion velocity vectors will significantly increase the resolution of electric field measurements in the vertical and horizontal dimensions, which when combined with neutral temperatures will undoubtedly lead to enormous leaps in auroral thermosphere-ionosphere research.

References

AT Aikio, HJ Opgenoorth, MAL Persson, and KU Kaila. Ground-based measurements of an arc-associated electric field. *Journal of atmospheric and Terrestrial physics*, 55(4-5): 797–808, 1993.

Callum Anderson, MJ Kosch, MJ Nicolls, and Mark Conde. Ion–neutral coupling in earth's thermosphere, estimated from concurrent radar and optical observations above alaska. *Journal of Atmospheric and Solar-Terrestrial Physics*, 105:313–324, 2013.

RL Arnoldy. Auroral particle precipitation and birkeland currents. *Reviews of Geophysics*, 12(2):217–231, 1974.

Mina Ashrafi. Ask: Auroral structure and kinetics in action. *Astronomy & Geophysics*, 48(4):4–35, 2007.

Doran J Baker and AT Stair Jr. Rocket measurements of the altitude distributions of the hydroxyl airglow. *Physica scripta*, 37(4):611, 1988.

David R Bates and Marcel Nicolet. The photochemistry of atmospheric water vapor. *Journal of Geophysical Research*, 55(3):301–327, 1950.

W Baumjohann and R Nakamura. Magnetospheric contributions to the terrestrial magnetic field. 2007.

Rolf Boström. A model of the auroral electrojets. *Journal of Geophysical Research*, 69(23): 4983–4999, 1964.

SC Buchert, T Hagfors, and JF McKenzie. Effect of electrojet irregularities on dc current flow. *Journal of Geophysical Research: Space Physics*, 111(A2), 2006.

SC Buchert, T Tsuda, R Fujii, and S Nozawa. The pedersen current carried by electrons: a non-linear response of the ionosphere to magnetospheric forcing. 26(9):2837–2844, 2008.

MJ Buonsanto, SA Gonzalez, X Pi, JM Ruohoniemi, MP Sulzer, WE Swartz, JP Thayer, and DN Yuan. Radar chain study of the may, 1995 storm. *Journal of atmospheric and solar-terrestrial physics*, 61(3-4):233–248, 1999.

Joshua M Chadney, Daniel K Whiter, and Betty S Lanchester. Effect of water vapour absorption on hydroxyl temperatures measured from svalbard. 35:481–491, 2017.

S Chakrabarti, D Pallamraju, J Baumgardner, and J Vaillancourt. Hities: A high throughput imaging echelle spectrogragh for ground-based visible airglow and auroral studies. *Journal of Geophysical Research: Space Physics*, 106(A12):30337–30348, 2001.

Francis K Chun, Delores J Knipp, Matthew G McHarg, Gang Lu, Barbara A Emery, Susanne Vennerstrøm, and Oleg A Troshichev. Polar cap index as a proxy for hemispheric joule heating. *Geophysical Research Letters*, 26(8):1101–1104, 1999.

Francis K Chun, Delores J Knipp, Matthew G McHarg, James R Lacey, Gang Lu, and Barbara A Emery. Joule heating patterns as a function of polar cap index. *Journal of Geophysical Research: Space Physics*, 107(A7), 2002.

JH Clemmons, JH Hecht, DR Salem, and DJ Strickland. Thermospheric density in the earth's magnetic cusp as observed by the streak mission. *Geophysical Research Letters*, 35(24), 2008.

MV Codrescu, TJ Fuller-Rowell, and JC Foster. On the importance of e-field variability for joule heating in the high-latitude thermosphere. *Geophysical Research Letters*, 22 (17):2393–2396, 1995.

Russell Cosgrove, Mary McCready, Roland Tsunoda, and Anja Stromme. The bias on the joule heating estimate: Small-scale variability versus resolved-scale model uncertainty and the correlation of electric field and conductance. *Journal of Geophysical Research: Space Physics*, 116(A9), 2011.

Hanna Dahlgren, Nickolay Ivchenko, J Sullivan, BS Lanchester, Göran Marklund, and D Whiter. Morphology and dynamics of aurora at fine scale: first results for the ask instrument. 26(5):1041–1048, 2008.

Yue Deng and Aaron J Ridley. Possible reasons for underestimating joule heating in global models: E field variability, spatial resolution, and vertical velocity. *Journal of Geophysical Research: Space Physics*, 112(A9), 2007.

RA Doe, JF Vickrey, and M Mendillo. Electrodynamic model for the formation of auroral ionospheric cavities. *Journal of Geophysical Research: Space Physics*, 100(A6):9683–9696, 1995.

JC Foster, J Mo Holt, RG Musgrove, and DS Evans. Ionospheric convection associated with discrete levels of particle precipitation. *Geophysical Research Letters*, 13(7):656–659, 1986.

Ryoichi Fujii, Satonori Nozawa, Stephan C Buchert, and Asgeir Brekke. Statistical characteristics of electromagnetic energy transfer between the magnetosphere, the ionosphere, and the thermosphere. *Journal of Geophysical Research: Space Physics*, 104(A2): 2357–2365, 1999.

Vincent Granville, Mirko Krivánek, and J-P Rasson. Simulated annealing: A proof of convergence. *IEEE transactions on pattern analysis and machine intelligence*, 16(6):652–656, 1994.

Thomas J Hallinan and T Neil Davis. Small-scale auroral arc distortions. *Planetary and Space Science*, 18(12):1735–1744, 1970.

TJ Hallinan. Auroral spirals, 2. theory. *Journal of Geophysical Research*, 81(22):3959–3965, 1976.

De-Sheng Han, Tong Xu, Yaqi Jin, K Oksavik, Xiang-Cai Chen, Jian-Jun Liu, Qinghe Zhang, Lisa Baddeley, and Katie Herlingshaw. Observational evidence for throat aurora being associated with magnetopause reconnection. *Geophysical Research Letters*, 2019.

Alan E Hedin. Extension of the msis thermosphere model into the middle and lower atmosphere. *Journal of Geophysical Research: Space Physics*, 96(A2):1159–1172, 1991.

K Henriksen, L Veseth, CS Deehr, and RW Smith. Neutral temperatures and emission height changes in an e-region aurora. *Planetary and space science*, 35(10):1317–1321, 1987.

Gerhard Herzberg. Molecular spectra and molecular structure. vol. 1: Spectra of diatomic molecules. *New York: Van Nostrand Reinhold*, 1950.

Takesi Iijima and Thomas A Potemra. The amplitude distribution of field-aligned currents at northern high latitudes observed by triad. *Journal of Geophysical Research*, 81 (13):2165–2174, 1976.

Takesi Iijima and Thomas A Potemra. Large-scale characteristics of field-aligned currents associated with substorms. *Journal of Geophysical Research: Space Physics*, 83(A2): 599–615, 1978.

O Jokiaho, BS Lanchester, N Ivchenko, GJ Daniell, LCH Miller, and D Lummerzheim. Rotational temperature of n 2+(0, 2) ions from spectrographic measurements used to infer the energy of precipitation in different auroral forms and compared with radar measurements. 26(4):853–866, 2008.

Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

RA Koehler, MM Shepherd, GG Shepherd, and KV Paulson. Rotational temperature variations in pulsating auroras. *Canadian Journal of Physics*, 59(8):1143–1149, 1981.

Mike J Kosch and E Nielsen. Coherent radar estimates of average high-latitude ionospheric joule heating. *Journal of Geophysical Research: Space Physics*, 100(A7):12201–12215, 1995.

MJ Kosch, I Yiu, Callum Anderson, Takuo Tsuda, Yasunobu Ogawa, Satonori Nozawa, Anasuya Aruliah, Victoria Howells, LJ Baddeley, IW McCrea, et al. Mesoscale observations of joule heating near an auroral arc and ion-neutral collision frequency in the polar cap e region. *Journal of Geophysical Research: Space Physics*, 116(A5), 2011.

J Kurihara, S Oyama, S Nozawa, TT Tsuda, R Fujii, Y Ogawa, H Miyaoka, N Iwagami, T Abe, K-I Oyama, et al. Temperature enhancements and vertical winds in the lower thermosphere associated with auroral heating during the delta campaign. *Journal of Geophysical Research: Space Physics*, 114(A12), 2009.

Junichi Kurihara, Takumi Abe, Koh-Ichiro Oyama, Eoghan Griffin, Mike Kosch, Anasuya Aruliah, Kirsti Kauristie, Yasunobu Ogawa, Sayaka Komada, and Naomoto Iwagami. Observations of the lower thermospheric neutral temperature and density in the delta campaign. *Earth, planets and space*, 58(9):1123–1130, 2006.

BS Lanchester, K Kaila, and IW McCrea. Relationship between large horizontal electric fields and auroral arc elements. *Journal of Geophysical Research: Space Physics*, 101(A3): 5075–5084, 1996.

BS Lanchester, MH Rees, KJF Sedgemore, JR Palmer, HU Frey, and KU Kaila. Ionospheric response to variable electric fields in small-scale auroral structures. 16(10): 1343–1354, 1998.

BS Lanchester, MH Rees, D Lummerzheim, A Otto, KJF Sedgemore-Schulthess, H Zhu, and IW McCrea. Ohmic heating as evidence for strong field-aligned currents in filamentary aurora. *Journal of Geophysical Research: Space Physics*, 106(A2):1785–1794, 2001.

Markku S Lehtinen and Ingemar Häggström. A new modulation principle for incoherent scatter measurements. *Radio Science*, 22(4):625–634, 1987.

H Lühr and J Warnecke. Fine structure of field-aligned current sheets deduced from spacecraft and ground-based observations: Initial freja results. *Geophysical research letters*, 21(17):1883–1886, 1994.

Hermann Lühr, Jaeheung Park, Jesper W Gjerloev, Jan Rauberg, Ingo Michaelis, Jose MG Merayo, and Peter Brauer. Field-aligned currents' scale analysis performed with the Swarm constellation. *Geophysical Research Letters*, 42(1):1–8, 2015.

Dirk Lummerzheim. *Electron transport and optical emissions in the aurora*. PhD thesis, 1987.

Sawako Maeda, Satonori Nozawa, Yasunobu Ogawa, and Hitoshi Fujiwara. Comparative study of the high-latitude e region ion and neutral temperatures in the polar cap and the auroral region derived from the eiscat radar observations. *Journal of Geophysical Research: Space Physics*, 110(A8), 2005.

Octav Marghitu. Auroral arc electrodynamics: Review and outlook. *Auroral Phenomenology and Magnetospheric Processes: Earth And Other Planets*, 197:143–158, 2012.

Göran Marklund. Auroral arc classification scheme based on the observed arc-associated electric field pattern. *Planetary and space science*, 32(2):193–211, 1984.

Göran Marklund, I Sandahl, and H Opgenoorth. A study of the dynamics of a discrete auroral arc. *Planetary and Space Science*, 30(2):179–197, 1982.

IAB Meinel. Oh emission bands in the spectrum of the night sky. *The Astrophysical Journal*, 111:555, 1950.

Frederick H Mies. Calculated vibrational transition probabilities of oh (x2π). *Journal of Molecular Spectroscopy*, 53(2):150–188, 1974.

S Nozawa, Y Ogawa, A Brekke, T Tsuda, Hall CM, H Miyaoka, J Kurihara, T Abe, and R Fujii. Eiscat observational results during the delta campaign. *Earth, planets and space*, 58(9):1183–1191, 2006.

HJ Opgenoorth, I Hägström, PJS Williams, and GOL Jones. Regions of strongly enhanced perpendicular electric fields adjacent to auroral arcs. *Journal of Atmospheric and Terrestrial Physics*, 52(6-8):449–458, 1990.

N Østgaard, G Germany, J Stadsnes, and RR Vondrak. Energy analysis of substorms based on remote sensing techniques, solar wind measurements, and geomagnetic indices. *Journal of Geophysical Research: Space Physics*, 107(A9), 2002.

J. R. Palmer. *Plasma density variations in aurora*. PhD thesis, University of Southampton, University Road, Southampton, 7 1995.

N Partamies, M Syrjäsuo, E Donovan, M Connors, D Charrois, D Knudsen, and Z Kryzanowsky. Observations of the auroral width spectrum at kilometre-scale size. *Annales Geophysicae (09927689)*, 28(3), 2010.

JM Picone, AE Hedin, D Pj Drob, and AC Aikin. Nrlmsise-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research: Space Physics*, 107(A12):SIA–15, 2002.

DJ Price, DK Whiter, JM Chadney, and BS Lanchester. High-resolution optical observations of neutral heating associated with the electrodynamics of an auroral arc. *Journal of Geophysical Research: Space Physics*, 124(11):9577–9591, 2019.

Gerd W Prölss. Density perturbations in the upper atmosphere caused by the dissipation of solar wind energy. *Surveys in Geophysics*, 32(2):101–195, 2011.

JA Reidy, RC Fear, DK Whiter, BS Lanchester, AJ Kavanagh, David John Price, Joshua M Chadney, Y Zhang, and LJ Paxton. Multiscale observation of two polar cap arcs occurring on different magnetic field topologies. *Journal of Geophysical Research: Space Physics*, 125(8):e2019JA027611, 2020.

AD Richmond. Assimilative mapping of ionospheric electrodynamics. *Advances in Space Research*, 12(6):59–68, 1992.

AD Richmond, Y Kamide, S-I Akasofu, D Alcaydé, M Blanc, O De la Beaujardière, DS Evans, JC Foster, E Friis-Christensen, JM Holt, et al. Global measures of ionospheric electrodynamic activity inferred from combined incoherent scatter radar and ground magnetometer observations. *Journal of Geophysical Research: Space Physics*, 95 (A2):1061–1071, 1990.

AJ Ridley, Y Deng, and G Toth. The global ionosphere–thermosphere model. *Journal of Atmospheric and Solar-Terrestrial Physics*, 68(8):839–864, 2006.

AS Rodger, GD Wells, RJ Moffett, and GJ Bailey. The variability of joule heating, and its effects on the ionosphere and thermosphere. 19(7):773–781, 2001.

Martin Rother, K Schlegel, and Hermann Lühr. CHAMP observation of intense kilometer-scale field-aligned currents, evidence for an ionospheric Alfvén resonator. 25(7):1603–1615, 2007.

Ingrid Sandahl, Tima Sergienko, and Urban Brändström. Fine structure of optical aurora. *Journal of Atmospheric and Solar-Terrestrial Physics*, 70(18):2275–2292, 2008.

C von Savigny, IC McDade, K-U Eichmann, and JP Burrows. On the dependence of the oh* meinel emission altitude on vibrational level: Sciamachy observations and model simulations. *Atmospheric Chemistry and Physics*, 12(18):8813–8828, 2012.

JR Sharber, RA Frahm, R Link, G Crowley, JD Winningham, EE Gaines, RW Nightingale, DL Chenette, BJ Anderson, and CA Gurgiolo. Uars particle environment monitor observations during the november 1993 storm: Auroral morphology, spectral characterization, and energy deposition. *Journal of Geophysical Research: Space Physics*, 103(A11):26307–26322, 1998.

F Sigernes, N Shumilov, CS Deehr, KP Nielsen, T Svenøe, and O Havnes. Hydroxyl rotational temperature record from the auroral station in adventdalen, svalbard (78 n, 15 e). *Journal of Geophysical Research: Space Physics*, 108(A9), 2003.

GG Sivjee. Airglow hydroxyl emissions. *Planetary and space science*, 40(2-3):235–242, 1992.

K Stasiewicz and T Potemra. Multiscale current structures observed by freja. *Journal of Geophysical Research: Space Physics*, 103(A3):4315–4325, 1998.

Daniel W Swift. An equipotential model for auroral arcs, 2. numerical solutions. *Journal of Geophysical Research*, 81(22):3935–3943, 1976.

EE Timofeev, MK Vallinkoski, TV Kozelova, AG Yahnin, and RJ Pellinen. Systematics of arc-associated electric-fields and currents as inferred from radar backscatter measurements. *Journal of Geophysics*, 61(2):122–137, 1987.

S Tuttle, B Gustavsson, and B Lanchester. Temporal and spatial evolution of auroral electron energy spectra in a region surrounding the magnetic zenith. *Journal of Geophysical Research: Space Physics*, 119(3):2318–2327, 2014.

J Vogt, HU Frey, G Haerendel, H Höfner, and JL Semeter. Shear velocity profiles associated with auroral curls. *Journal of Geophysical Research: Space Physics*, 104(A8): 17277–17288, 1999.

DR Weimer. Improved ionospheric electrodynamic models and application to calculating joule heating rates. *Journal of Geophysical Research: Space Physics*, 110(A5), 2005.

CS Weller, RR Meier, and BA Tinsley. Simultaneous measurements of the hydrogen airglow emissions of lyman alpha, lyman beta, and balmer alpha. *Journal of Geophysical Research*, 76(31):7734–7744, 1971.

Daniel K Whiter, Hanna Dahlgren, Betty S Lanchester, Joshua Dreyer, Noora Partamies, Nickolay Ivchenko, Marco Zaccaria Di Fraia, Rosie Oliver, Amanda Serpell-Stevens, Tiffany Shaw-Diaz, et al. Fine scale dynamics of fragmented aurora-like emission. *Annales Geophysicae Discussions*, pages 1–20, 2021.

DK Whiter, BS Lanchester, B Gustavsson, NIB Jallo, O Jokiaho, Nickolay Ivchenko, and Hanna Dahlgren. Relative brightness of the o+ (2 d-2 p) doublets in low-energy aurorae. *The Astrophysical Journal*, 797(1):64, 2014.

PJS Williams, G Crowley, K Schlegel, TS Virdi, I McCrea, G Watkins, N Wade, JK Hargreaves, T Lachlan-Cope, H Muller, et al. The generation and propagation of atmospheric gravity waves observed during the worldwide atmospheric gravity-wave study (wags). *Journal of Atmospheric and Terrestrial Physics*, 50(4-5):323–338, 1988.

PJS Williams, GOL Jones, B Jones, H Opgenoorth, and I Hägström. High-resolution measurements of magnetospheric electric fields. *Journal of Atmospheric and Terrestrial Physics*, 52(6-8):439–448, 1990.

J Wu, DJ Knudsen, DM Gillies, EF Donovan, and JK Burchill. Swarm observation of field-aligned currents associated with multiple auroral arc systems. *Journal of Geophysical Research: Space Physics*, 122(10), 2017.

Erdal Yiğit and Aaron J Ridley. Effects of high-latitude thermosphere heating at various scale sizes simulated by a nonhydrostatic global thermosphere–ionosphere model. *Journal of atmospheric and solar-terrestrial physics*, 73(5-6):592–600, 2011.

H Zhu, A Otto, D Lummerzheim, MH Rees, and BS Lanchester. Ionosphere-magnetosphere simulation of small-scale structure and dynamics. *Journal of Geophysical Research: Space Physics*, 106(A2):1795–1806, 2001.

Qingyu Zhu, Yue Deng, Arthur Richmond, and Astrid Maute. Small-scale and mesoscale variabilities in the electric field and particle precipitation and their impacts on joule heating. *Journal of Geophysical Research: Space Physics*, 123(11):9862–9872, 2018.